IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 6, December 2026, pp. 5581~5591
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i6.pp5581-5591

A differentiable surrogate optimization framework for adaptive thresholds in multi-metric face verification

Anan Nugroho[1], Ahmad Zein Al Wafi[1], Muhammad Iqbal[1], Mona Subagja[2]
[1]Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Negeri Semarang, Semarang, Indonesia
[2]Directorate of Information System and Public Relations, Universitas Negeri Semarang, Semarang, Indonesia

Article Info

Article history:

Received Nov 5, 2025
Revised Aug 7, 2026
Accepted Aug 31, 2026

Keywords:

Deep learning
Face recognition
FaceNet
Loss function
Machine learning

ABSTRACT

Face verification systems conventionally rely on fixed, pre-defined thresholds to classify pairs of facial embeddings, a practice that lacks robustness under varying real-world conditions. However, learning adaptive thresholds has been hindered by the non-differentiable nature of the comparison operation, which prevents end-to-end optimization through gradient-based methods. This paper presents a novel framework, termed differentiable surrogate optimization (DSO), to address this limitation. The proposed approach introduces a continuous, margin-based surrogate loss function that acts as a differentiable proxy for the hard decision boundary, thereby enabling a lightweight neural network to learn and dynamically predict context-aware thresholds for multiple similarity metrics, including Euclidean distance, Cosine distance, and Euclidean norm (L2-norm), based on FaceNet embeddings. Experimental results demonstrate that the proposed system effectively learns to produce a broad distribution of context-aware thresholds, whose means align with regions of high-class ambiguity. The framework achieves an area under the curve (AUC) exceeding 0.94 across all metrics, confirming its discriminative capability. These results validate DSO as a feasible and interpretable solution for developing self-adjusting face verification systems.

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

Corresponding Author:

Anan Nugroho
Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Negeri Semarang
Sekaran, Gunungpati, Semarang, Central Java, Indonesia
Email: anannugroho@mail.unnes.ac.id

1. INTRODUCTION

Facial recognition technology provides a rapid, accurate, and contactless alternative to conventional identification or attendance systems, which are often prone to inaccuracies and security vulnerabilities [1]–[3]. These systems enhance security by mitigating impersonation risk [4] and exhibit strong robustness across challenging real-world conditions [5]–[7]. Their adoption has further accelerated due to the increasing demand for hygienic, touchless interfaces [8], making them a pivotal component of modern identity management and resource allocation infrastructures [9]. Among deep learning-based approaches, FaceNet is a state-of-the-art model [10] that has inspired a wide range of subsequent research [11]–[13]. By mapping facial images into a compact Euclidean embedding space, where distance directly corresponds to visual similarity [14], FaceNet enables efficient and accurate verification [15]. Through end-to-end feature learning, it produces a discriminative and robust representation that continues to serve as a benchmark model for face verification research [16].

Following the extraction of high-dimensional embeddings using models such as FaceNet, the verification pipeline culminates in a decisive binary classification stage that determines whether two images

represent the same identity. This stage typically compresses the comparison of complex feature vectors into a single interpretable scalar value through well-established similarity metrics. The predominant methodologies employ metrics that include the Euclidean distance, which effectively measures geometric similarity across multi-dimensional embeddings [17]; the cosine distance, which quantifies the angular deviation between feature vectors [18]; and the Euclidean norm (L2-norm) of the direct embedding difference. Once a similarity score is computed, it is compared against a pre-defined, invariant decision boundary [19]. Such cut-off points are commonly determined through offline calibration using validation datasets and balanced according to the receiver operating characteristic (ROC) curve [20]. In multi-metric systems, each metric is calibrated independently, producing disjoint threshold constants that collectively dictate the final verification decision.

Despite their widespread adoption, these static methodologies exhibit inherent limitations that significantly constrain robustness in practical deployments. A static threshold forms a rigid decision boundary, incapable of adapting to variations in acquisition conditions such as illumination, camera quality, or facial pose [21]. Consequently, embeddings obtained from low-quality or poorly lit images are often evaluated under the same fixed standard as those from ideal conditions, resulting in unstable performance [22]. This inflexibility leads to a brittle system with degraded performance in uncontrolled environments. While fully learned classification models can introduce adaptability, they often function as opaque "black-box" systems with limited transparency and interpretability. The core obstacle to maintaining this dichotomy between brittle-yet-interpretable fixed decision limits and adaptive-yet-opaque classifiers lies in the non-differentiable nature of the thresholding operation (e.g., score ≤ threshold). This discontinuity creates a gradient barrier that precludes the verification boundary from being directly optimized as learnable parameters within an end-to-end framework.

Addressing these limitations, this study introduces differentiable surrogate optimization (DSO), a novel framework that enables the end-to-end learning of dynamic decision boundaries in face verification systems. By formulating a continuous, margin-based surrogate loss function that approximates the discrete thresholding process, DSO effectively circumvents the gradient barrier and allows these boundaries to be optimized through standard gradient-based learning. The approach employs a lightweight neural network to dynamically predict context-aware operating points for multiple metrics, including Euclidean distance, cosine similarity, and L2-norm, using FaceNet embeddings as inputs. The resulting system is both adaptive and capable of adjusting its decision boundary to input variability—and interpretable, since the final verification remains a transparent comparison between a metric score and its corresponding learned cut-off points. The study contributes a new optimization framework for learning explicit verification decision points, an interpretable dynamic mechanism that offers a transparent alternative to opaque models, and a validated multi-metric implementation demonstrating superior discriminative performance and training accuracy.

2. METHOD

2.1. System architecture overview

The proposed methodology is instantiated in a system designed to perform adaptive multi-metric face verification, the architecture of which is illustrated in Figure 1. The framework operates on a pair of FaceNet feature embeddings, an anchor embedding (e_a), and an observed embedding (e_o), to produce a set of three independent binary verification decisions. The system is comprised of two parallel processing streams that converge at a final decision function.

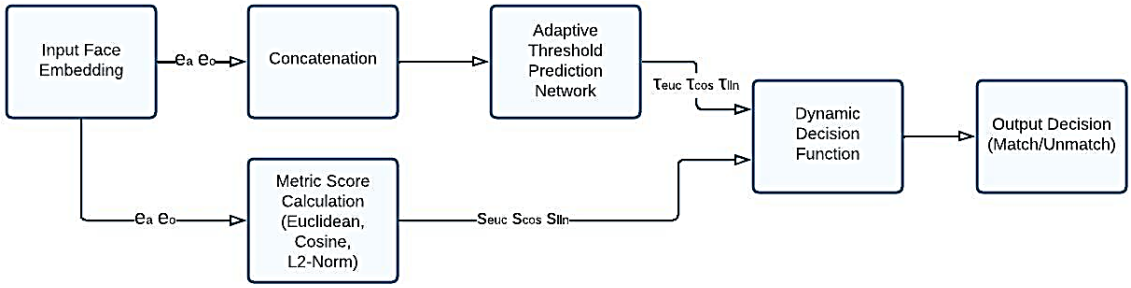

Figure 1. System architecture of the proposed adaptive multi-metric verification framework

The first stream is dedicated to the dynamic prediction of decision thresholds. The input embeddings, e_a and e_o, are initially concatenated to form a comprehensive feature vector that captures the

combined characteristics of the pair. This vector is fed into the adaptive threshold prediction network, a multi-layer perceptron (MLP) specifically trained to regress a 3-dimensional threshold vector, τ, where each element corresponds to a specific distance metric threshold. The second stream calculates the metric scores. The same input embeddings, e_a and e_o, are processed by the metric score calculation module to compute the scalar distance scores for each of the three predefined metrics: Euclidean distance (s_{euc}), cosine distance (s_{cos}), and the L2-norm (s_{lin}). These scores are consolidated into a score vector (s). Finally, the outputs of both streams (the predicted threshold vector (τ) and the calculated s) are passed to the dynamic decision function. This function performs an element-wise comparison between the corresponding elements of the two vectors, yielding the final multi-metric verification outcome.

2.2. Dataset and feature representation

The dataset employed in this study was curated from operational attendance records logged at the Digital Center Building, Universitas Negeri Semarang (UNNES), over a period from 17 April 2024 to 12 March 2025. The data acquisition process involved subjects presenting a unique personal identifier, at which point a corresponding facial image was captured and associated with that ID. This procedure ensures a high-fidelity link between each image and its ground truth identity. To evaluate the system's performance on naturalistic inputs, no form of data augmentation was applied to the source images. The resulting corpus contains a variable number of images for each subject, reflecting total authentic attendance sampling. The specific distribution of image samples per subject is detailed in Figure 2.

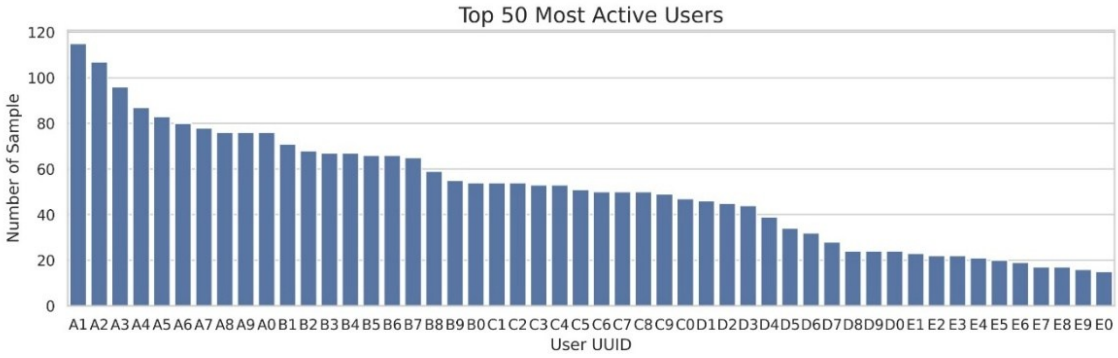

Figure 2. Diagram of image distribution per user

For the training and evaluation of the verification framework, a labeled set of image pairs was systematically generated. Positive pairs were formed using an intra-class sampling strategy, where two distinct images of the same individual were selected. Conversely, negative pairs were generated via an inter-class strategy, selecting images from two different individuals. This initial sampling process yielded a significant class imbalance, comprising approximately 150,000 positive pairs and 12 million negative pairs. To mitigate the risk of model bias towards the majority (negative) class, a random downsampling procedure was applied. This resulted in a balanced training corpus with a 1:1 ratio of positive to negative pairs, ensuring equitable representation of both classes during model optimization.

Feature extraction was performed using a pre-trained FaceNet model, which functions as a fixed feature extractor to convert each facial image into a high-dimensional vector representation. This process yields an anchor embedding ($e_a \in \mathbb{R}^{512}$) and an observed embedding ($e_o \in \mathbb{R}^{512}$) for each image pair. These two vectors are subsequently concatenated to form the final input vector for the adaptive threshold prediction network. This concatenated vector, embedding_overall, is thus defined as a real-valued vector in $\mathbb{R}^{1024}$, providing a comprehensive representation that encapsulates the features of both faces in the pair.

2.3. Adaptive threshold prediction network

The core of the self-adjusting mechanism is the adaptive threshold prediction network, a lightweight neural network designed to dynamically regress the decision thresholds for each verification pair. Unlike a static system where thresholds are fixed hyperparameters, this network learns a mapping from the combined features of an input pair to a set of optimal, context-specific thresholds. The network's function is to analyze the holistic characteristics of the concatenated embedding, embedding_overall, thereby inferring the quality

and complexity of the comparison task at hand. It then outputs a threshold vector tailored to that specific instance, enabling the system to apply a more lenient or stringent decision boundary as warranted.

The MLP architecture is designed to progressively reduce the dimensionality of the input feature vector while extracting the most salient information for the threshold prediction task. To mitigate overfitting and improve generalization, Dropout is employed as a regularization technique. The complete architectural specifications, including layer types, output dimensions, and activation functions, are detailed in Table 1.

Table 1. Detailed MLP model architecture

Layer name	Layer type	Shape	Activation function
Input layer	-	(Batch size, 1024)	-
Hidden layer 1	Fully connected (linear)	(Batch size, 512)	ReLU
Regularization 1	Dropout	(Batch size, 512)	-
Hidden layer 2	Fully connected (linear)	(Batch size, 256)	ReLU
Regularization 2	Dropout	(Batch size, 256)	-
Output layer	Fully connected (linear)	(Batch size, 3)	ReLU

The network accepts the concatenated feature vector, embedding_overall $\in \mathbb{R}^{1024}$, as its sole input. This vector provides a comprehensive representation of both the anchor and observed embeddings simultaneously. The network's output is a 3-dimensional real-valued vector, $\tau \in \mathbb{R}^3$, where each element corresponds to a predicted threshold for one of the system's verification metrics. The output vector is formally defined as $\tau = [\tau_{euc}, \tau_{cos}, \tau_{lin}]$, representing the dynamically determined thresholds for the Euclidean distance, cosine distance, and L2-norm, respectively.

2.4. Multi-metric verification framework

The multi-metric verification framework operates in parallel to the threshold prediction network and is responsible for quantifying the similarity between the embedding pair and executing the final decision. This process involves two primary stages: metric score calculation and the dynamic decision function. First, a set of three distinct distance metrics is computed from the e_a and e_o. The Euclidean distance on L2-norm vectors, s_{euc} is defined as (1).

$$s_{euc} = \left\| \frac{e_a}{\|e_a\|_2} - \frac{e_o}{\|e_o\|_2} \right\|_2 \quad (1)$$

The cosine distance, s_{cos}, which measures the angular disparity, is given by (2).

$$s_{cos} = 1 - \frac{e_a \cdot e_o}{\|e_a\|_2 \, \|e_o\|_2} \quad (2)$$

Finally, the direct L2-norm of the embedding difference, s_{lin}, is calculated as (3).

$$s_{lin} = \|e_a - e_o\|_2 \quad (3)$$

These three scores are consolidated into a score vector, $s = [s_{euc}, s_{cos}, s_{lin}]$.

In the second stage, the dynamic decision function receives both the calculated score vector s and the predicted threshold vector $\tau = [\tau_{euc}, \tau_{cos}, \tau_{lin}]$ from the adaptive threshold prediction network. A verification decision for each metric i is produced by an element-wise comparison, formally defined as (4).

$$Decision_i = (s_i \leq \tau_i) \quad (4)$$

This function yields a final Boolean vector, $[Decision_{euc}, Decision_{cos}, Decision_{lin}]$, representing the context-aware verification outcome for each metric.

2.5. Differentiable surrogate optimization

The central challenge in learning a dynamic verification threshold, τ, is that the decision function defined as a hard binary comparison ($s \leq \tau$) is non-differentiable. An objective function based on this comparison would be discontinuous, exhibiting a zero gradient almost everywhere. This property, known as a "gradient barrier," precludes the use of standard gradient-based optimization techniques like backpropagation to learn the parameters that produce τ. To circumvent this fundamental limitation, this study introduces a DSO framework.

2.5.1. The differentiable surrogate loss function

This study reparametrizes the non-differentiable decision objective by introducing a continuous and differentiable surrogate loss function. This employs a margin-based hinge loss for maximizing the separation between classes. First, the binary ground truth label, $\acute{y}_{true} \in \{0,1\}$, is transformed to $\acute{y} \in \{-1,+1\}$ using $\acute{y} = 2y_{true} - 1$. For a single metric i, the loss, L_i, is formulated to penalize predictions that do not satisfy the decision boundary with a sufficient margin, m_i, which is formally defined in (5).

$$L_i(\tau_i, s_i, \acute{y}_i) = \max(0, m_i - \acute{y}_i(\tau_i - s_i)) \quad (5)$$

In (5), the term $\acute{y}_i(\tau_i - s_i)$ represents the signed distance from the decision boundary, scaled by the correctness of the classification. A loss is incurred if the value is less than the specified margin m_i. Because the max operator is differentiable almost everywhere, gradients can be computed with respect to τ_i.

2.5.2. Combined multi-metric loss formulation

To train the single adaptive threshold prediction network, the individual loss components from each of the three metrics must be consolidated into a single scalar objective function. The total loss (L_{total}) is therefore defined as a weighted sum of the individual surrogate losses in (6).

$$L_{total} = \sum_{i \in \{euc,cos,lin\}} w_i \cdot L_i \quad (6)$$

Here, w_i represents the weight hyperparameter assigned to each metric's loss component, allowing for control over each metric's relative contribution to the total gradient during training. This composite loss function, L_{total}, is differentiable with respect to each predicted threshold τ_i, and consequently, with respect to all weights of the prediction network. This property renders the entire system amenable to end-to-end training via standard backpropagation algorithms.

2.6. Experimental setup

All experiments were conducted using the Python programming language (version 3.12.3) and the PyTorch deep learning framework (version 2.7.0). To ensure the deterministic behavior and reproducibility of our results, including data sampling, network weight initialization, and other stochastic processes, a global random seed of 42 was consistently applied. The computational procedures were executed on a workstation equipped with a 32-core Intel Core i9-13900K processor, supported by 32 GB of system RAM.

The adaptive threshold prediction network was trained for 200 epochs using the adaptive moment estimation (Adam) optimizer with an initial learning rate set to $1 \times 10 - 3$. A batch size of 128 was employed for all training iterations. The weights for the composite loss function (w_i in 6) were set uniformly ($w_{euc} = w_{cos} = w_{lin} = 1.0$) to provide an equitable contribution from each metric during optimization. The margin hyperparameters (m_i in 5) were empirically selected to align with the distinct scale of each metric's score distribution, with values set to $m_{euc} = 0.1, m_{cos} = 0.1$, and $m_{lin} = 0.5$.

3. RESULTS AND DISCUSSION

3.1. System performance evaluation

The performance of the proposed framework was evaluated throughout the training process and on the final test set. Figure 3 illustrates the accuracy history for both training and test data across 200 epochs for each of the three metric channels: Euclidean (E), cosine (C), and Linalg (L). The learning curves reveal a crucial insight into the optimization process: each metric exhibits a unique convergence timeline. The L metric demonstrates the most rapid learning, achieving high accuracy within the initial epochs. The E metric follows a more gradual but steady learning trajectory, requiring a greater number of epochs to stabilize. Most distinctively, the C metric displays a prolonged initial phase of low performance before undergoing a sharp, rapid convergence around the 75-epoch mark. This divergence in learning rates indicates that the unified optimization framework enables independent convergence paths tailored to the specific complexity and characteristics of each metric.

The final discriminative capability of each metric on the unseen test set is quantified by the ROC curves presented in Figure 4, where Figure 4(a) shows the E metric, Figure 4(b) shows the C metric, and Figure 4(c) shows the L metric. The area under the curve (AUC) for all three metrics is notably high, with both the E and C metrics achieving an AUC of 0.949, and the L metric achieving a marginally lower AUC of 0.944. These results confirm the excellent discriminative power of the fully trained system, as all three channels are highly capable of distinguishing between positive and negative pairs. The proximity of the curves to the top-left corner further substantiates their strong performance, demonstrating a high true positive rate for a low false positive rate across a range of potential decision thresholds.

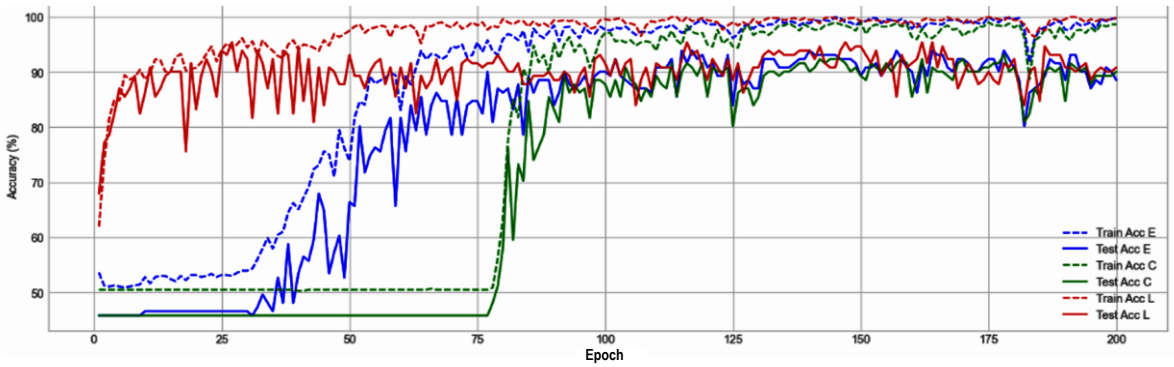

Figure 3. Accuracy history for training and test sets over 200 epochs for E, C, and L metrics

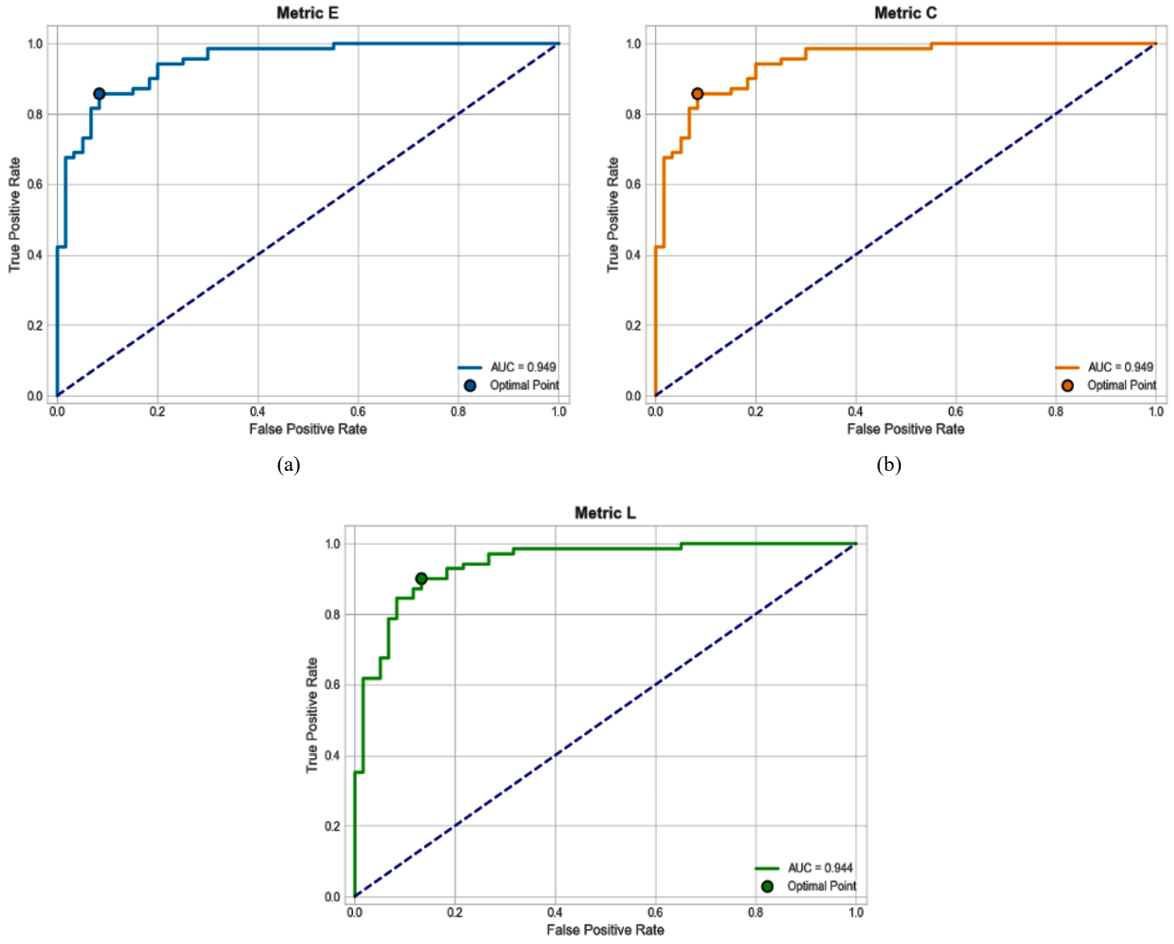

Figure 4. ROC curves for on the test set, with corresponding AUC values: (a) E metric, (b) C metric, and (c) L metric

3.2. Analysis of the learned dynamic verification threshold

An analysis of the system's learned behavior begins with an examination of the underlying data structure, followed by an evaluation of the model's output. Figure 5 presents the frequency distributions of the calculated metric scores (s) across the dataset, where Figure 5(a) shows the E metric, Figure 5(b) shows the C metric, and Figure 5(c) shows the L metric. A distinct bimodal characteristic is evident for all three metrics: s_{euc}, s_{cos}, and s_{lin}. These two modes correspond to the intra-class (match) and inter-class (no-match) populations, respectively.

Figure 6 illustrates the distributions of the corresponding operating threshold (τ) predicted by the adaptive network on the test set. This figure illustrating the dynamic nature of the learned decision boundary. The histograms confirm that the network produces a range of values for each metric: E metric (Figure 6(a)), C metric (Figure 6(b)), and L metric (Figure 6(c)), rather than collapsing to a single static constant. Each threshold distribution exhibits a notable positive skew, indicating a higher frequency of lower predicted threshold values. A key observation is the positional alignment between the mean of each predicted verification boundary distribution in Figure 6 and the valley region separating the two modes of its corresponding metric score distribution in Figure 5. This indicates that the model learned to center its decision boundary within the area of highest-class ambiguity. The observed variance and skew in the predicted threshold validate that the DSO framework successfully trained the network to implement an automatically adjusted decision function rather than a static one.

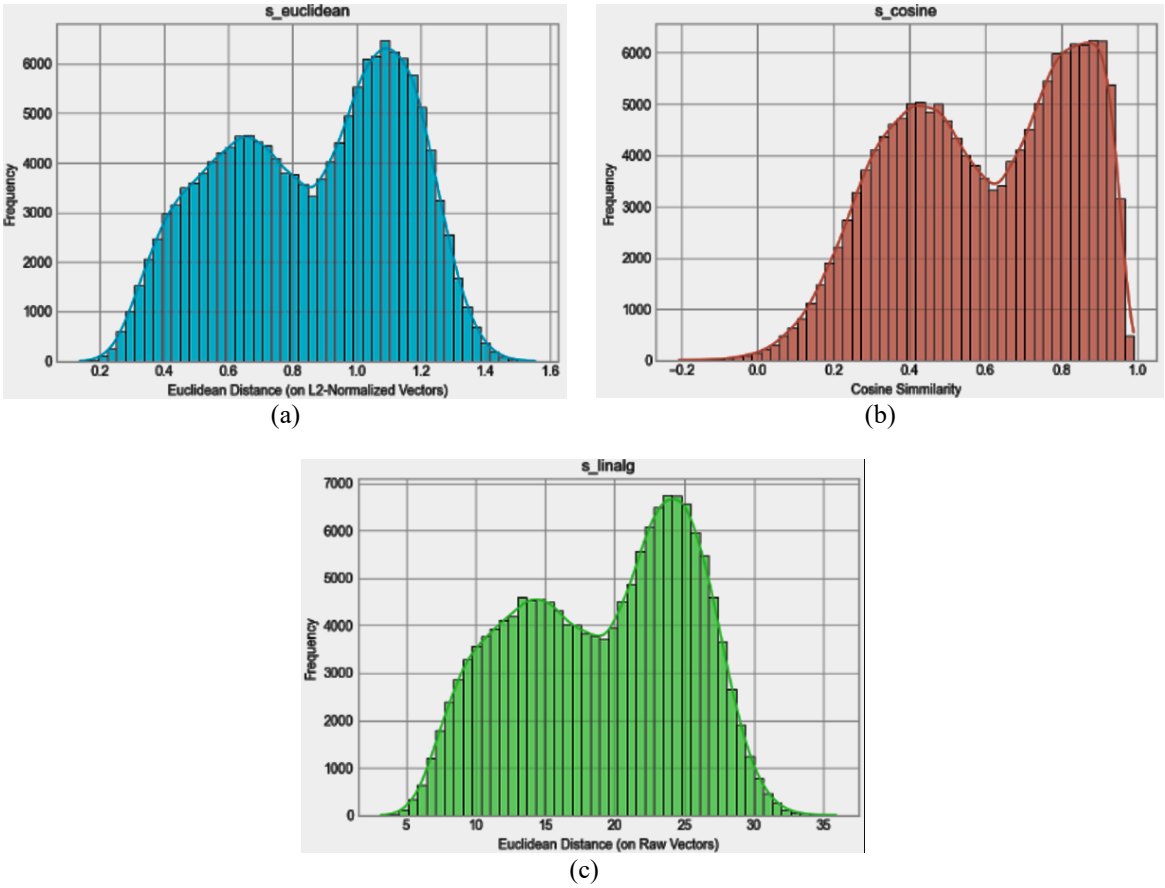

Figure 5. Bimodal frequency distributions of the calculated metric scores (s) across the dataset for (a) E metric, (b) C metric, and (c) L metric

3.3. Performance interpretation of surrogate optimization framework

The empirical results provide compelling evidence for the efficacy of the proposed framework, confirming that the system learns a truly self-adjusting decision-making policy. The bimodal nature of the calculated metric score distributions, presented in Figure 5, delineates the fundamental challenge of the verification task: the two peaks represent clear intra-class (match) and inter-class (no-match) populations, while the overlapping valley region signifies the most ambiguous and error-prone cases. An analysis of the predicted threshold in Figure 6 reveals that the network does not converge to a single static value. Instead, it produces a wide spectrum of decision boundaries for each metric, with the mean of each distribution strategically positioned within this critical valley of ambiguity. This demonstrates that the model has successfully learned to center its threshold in the region where adaptation is most crucial. Furthermore, the distinct positive skew observed in all three threshold distributions indicates the development of a sophisticated, risk-aware policy: the system defaults to a stringent baseline for clear-cut cases but has learned

to identify challenging instances for which it must dynamically predict a more lenient tolerance level to ensure correct classification.

The research addresses a long-standing challenge in machine learning known as the "gradient barrier," which has traditionally hindered models from handling hard, all-or-nothing decisions effectively. In the context of face verification, the proposed system represents a notable advancement over conventional fixed frameworks. It dynamically adapts its decision-making process for each individual face, enabling robust performance across the diverse range of facial variations encountered in real-world scenarios. This improvement is achieved by converting rigid binary comparisons into a more flexible, continuous learning mechanism, allowing end-to-end training with standard deep learning techniques. The findings demonstrate that decision-making rules need not be predefined manually; instead, they can be learned directly by the system. This approach effectively bridges the gap between the data-driven adaptability of neural networks and the interpretability of rule-based systems.

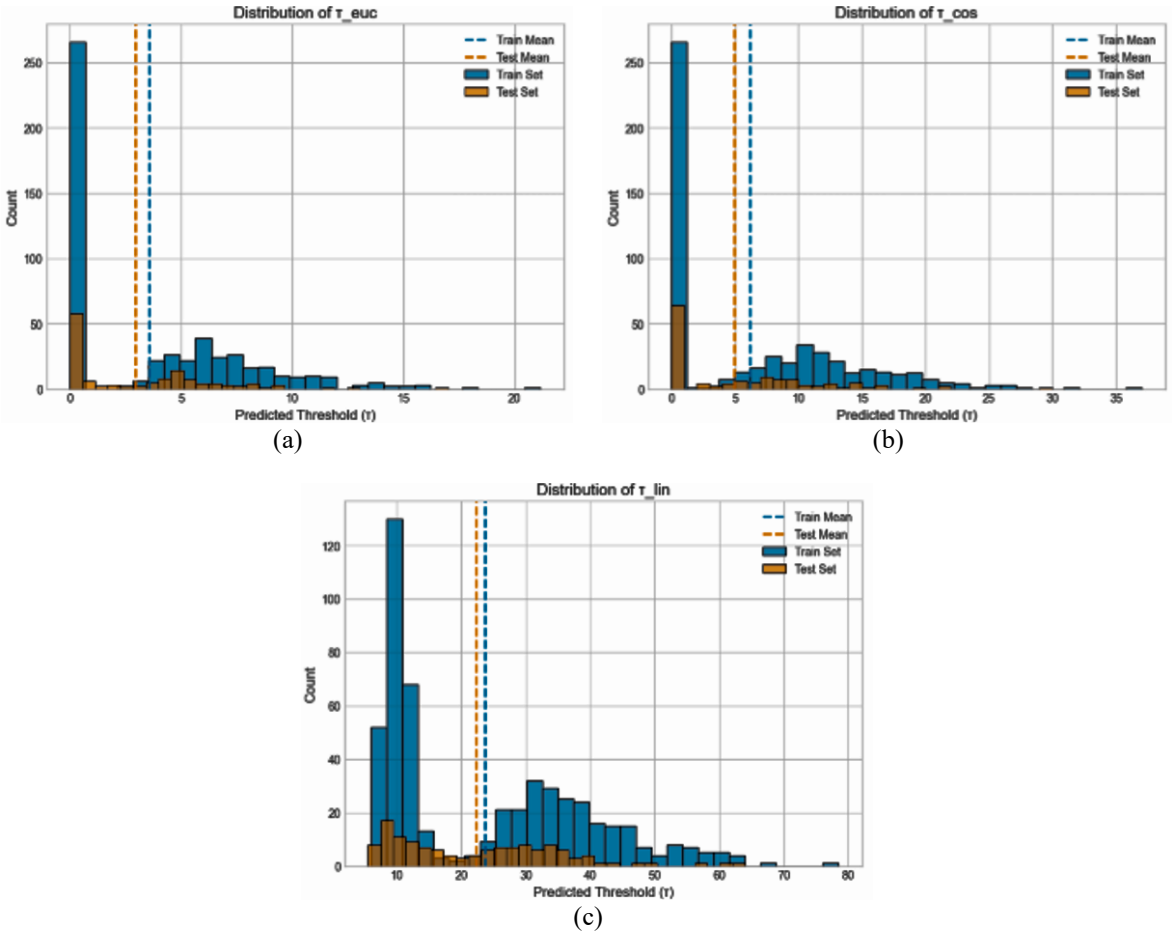

Figure 6. The distribution of dynamically predicted thresholds (τ) on the training and test sets for (a) E metric, (b) C metric, and (c) L metric

The implications of this framework extend far beyond the domain of face verification. The methodology provides a generalizable blueprint for a broad class of systems in science and engineering that rely on threshold-based decision rules. In healthcare, this hybrid approach fuses artificial intelligence (AI) predictions with established clinical guidelines, conditioned on a patient's profile, leading to diagnostic and treatment strategies that are both innovative and trustworthy [23]. In industrial internet of thing (IoT), it could enable predictive maintenance systems where failure detection thresholds for sensor data are dynamically adjusted based on a machine's operational state [24]. In financial modeling, it could be used to create more resilient credit scoring systems with automatically adjusted approval verification boundaries that respond to global financial dynamics and policy [25]. This work demonstrates a powerful

principle for building the next generation of intelligent systems that integrates explicit, interpretable rules directly into the learning process.

3.4. Limitations and future work

The scope of this study is primarily focused on establishing the viability of the DSO framework as a novel methodology. Consequently, the empirical validation is limited to a single feature extraction architecture, FaceNet, and a dataset sourced from a specific environment at Universitas Negeri Semarang due to resource constraints. While this provides a robust proof-of-concept, the results do not yet encompass a direct, quantitative performance benchmark against established baseline methods such as static thresholding or direct "black-box" classifiers. The current work demonstrates that the system learns an adaptive policy, but does not yet quantify the precise performance gains this adaptation provides over conventional approaches.

These limitations define clear and compelling directions for future research. This study intentionally focused on an in-house dataset to establish a controlled proof-of-concept. Future research could extend this study by exploring the framework's scalability across diverse public benchmark datasets (e.g., labeled faces in the wild (LFW) and visual geometry group face 2 (VGGFace2)), thereby assessing its generalizability beyond the current controlled setting. Furthermore, the generalizability of the framework should be tested by integrating it with other state-of-the-art feature extractors. Finally, a deeper investigation into the model's learned policy could be pursued by analyzing the correlation between the magnitude of the predicted thresholds and quantifiable image quality metrics, which would provide further insight into the decision-making process.

4. CONCLUSION

This paper introduces and validates a methodology, termed DSO, to address the challenge of dynamic decision boundaries learning in face verification systems where the decision function is inherently non-differentiable. By employing a continuous surrogate loss function, this study demonstrated that it is possible to train a neural network to dynamically regress explicit decision parameters via standard gradient-based optimization. The empirical results confirmed that the system learned to produce a distribution of self-adjusting decision limits, rather than a single static value, whose mean strategically aligns with the region of highest-class ambiguity in the data. These results demonstrate the viability of the proposed framework in creating verification systems that remain responsive to diverse inputs while preserving the structural interpretability characteristic of metric-and-threshold designs.

FUNDING INFORMATION

This work was supported by the Fundamental Research Scheme under the Ministry of Higher Education, Science, and Technology (Kemendiktisaintek) Grant No. 50.12.6/UN37/PPK.10/2024 administered by the Institute for Research and Community Service, Universitas Negeri Semarang.

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
Anan Nugroho				✓	✓	✓	✓		✓	✓		✓	✓	✓
Ahmad Zein Al Wafi	✓	✓	✓			✓			✓		✓			
Muhammad Iqbal			✓			✓	✓	✓	✓		✓			
Mona Subagja					✓		✓			✓		✓	✓	✓

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 author declares that there are no known conflicts of interest associated with this publication. There are no financial or personal relationships that could inappropriately influence or bias the content of this work.

INFORMED CONSENT

This section is not applicable, as this study did not involve human participants or personally identifiable information.

ETHICAL APPROVAL

This section is not applicable, as this study did not involve human or animal subjects and used only non-sensitive computational or public data.

DATA AVAILABILITY

The datasets generated and/or analyzed during the current study are available from the corresponding author, [AN], upon reasonable request. Initial inquiries and requests for the data can be directed to [MI] at xmod3905@students.unnes.ac.id.

REFERENCES

[1] E. M. Onyema, P. K. Shukla, S. Dalal, M. N. Mathur, M. Zakariah, and B. Tiwari, "Enhancement of patient facial recognition through deep learning algorithm: ConvNet," *Journal of Healthcare Engineering*, 2021, doi: 10.1155/2021/5196000.

[2] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. T.-Ahmed, "Past, present, and future of face recognition: a review," *Electronics*, vol. 9, no. 8, pp. 1–53, 2020, doi: 10.3390/electronics9081188.

[3] S. Negi, R. Chauhan, S. Tomar, S. Devliyal, and S. Gupta, "Portable fingerprint based attendance system with Raspberry Pi," in *Proceedings of the 2nd International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics, ICIITCEE 2024*, 2024, doi: 10.1109/IITCEE59897.2024.10467918.

[4] M. H.-de-Menendez, R. M.-Menendez, C. A. Escobar, and J. Arinez, "Biometric applications in education," *International Journal on Interactive Design and Manufacturing*, vol. 15, no. 2–3, pp. 365–380, 2021, doi: 10.1007/s12008-021-00760-6.

[5] P. Kaur, K. Krishan, S. K. Sharma, and T. Kanchan, "Facial-recognition algorithms: a literature review," *Medicine, Science and the Law*, vol. 60, no. 2, pp. 131–139, 2020, doi: 10.1177/0025802419893168.

[6] D. S. AbdELminaam, A. M. Almansori, M. Taha, and E. Badr, "A deep facial recognition system using computational intelligent algorithms," *Plos One*, vol. 15, 2020, doi: 10.1371/journal.pone.0242269.

[7] S. M. Bah and F. Ming, "An improved face recognition algorithm and its application in attendance management system," *Array*, vol. 5, 2020, doi: 10.1016/j.array.2019.100014.

[8] N. C. Basjaruddin, E. Rakhman, Y. Sudarsa, M. B. Z. Asyikin, and S. Permana, "Attendance system with face recognition, body temperature, and use of mask using multi-task cascaded convolutional neural network (MTCNN) method," *Green Intelligent Systems and Applications*, vol. 2, no. 2, pp. 71–83, 2022, doi: 10.53623/gisa.v2i2.109.

[9] N. Surantha and B. Sugijakko, "Lightweight face recognition-based portable attendance system with liveness detection," *Internet of Things*, vol. 25, 2024, doi: 10.1016/j.iot.2024.101089.

[10] A. Firmansyah, T. F. Kusumasari, and E. N. Alam, "Comparison of face recognition accuracy of ArcFace, FaceNet and FaceNet512 models on deepface framework," in *ICCoSITE 2023 - International Conference on Computer Science, Information Technology and Engineering: Digital Transformation Strategy in Facing the VUCA and TUNA Era*, 2023, pp. 535–539, doi: 10.1109/ICCoSITE57641.2023.10127799.

[11] M. Al Faqih, Hendrawan, E. Mulyana, and W. Hermawan, "On the performance of face recognition models for local faces dataset," in *Proceedings of 2023 17th International Conference on Telecommunication Systems, Services, and Applications, TSSA 2023*, 2023. doi: 10.1109/TSSA59948.2023.10366896.

[12] F. Cahyono, W. Wirawan, and R. F. Rachmadi, "Face recognition system using FaceNet algorithm for employee presence," in *4th International Conference on Vocational Education and Training, ICOVET 2020*, 2020, pp. 57–62, doi: 10.1109/ICOVET50258.2020.9229888.

[13] G. Bouallegue and R. Djemal, "EEG person identification using FaceNet, LSTM-RNN, and SVM," in *Proceedings of the 17th International Multi-Conference on Systems, Signals and Devices, SSD 2020*, 2020, pp. 22–28, doi: 10.1109/SSD49366.2020.9364129.

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: a unified embedding for face recognition and clustering," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.

[15] S. O. Olayemi, O. S. Olatunde, I. W. Oladimeji, and I. Folasade, "Enhancing authentication efficiency in computer-based examinations through advanced face recognition systems," *Journal of Digital Security and Forensics*, vol. 1, no. 1, 2024, doi: 10.29121/digisecforensics.v1.i1.2024.24.

[16] A. Alzu'bi, F. Albalas, T. Al-Hadhrami, L. B. Younis, and A. Bashayreh, "Masked face recognition using deep learning: a review," *Electronics*, vol. 10, no. 21, 2021, doi: 10.3390/electronics10212666.

[17] J. A. Mensah, J. K. Appati, E. K. A. Boateng, E. Ocran, and L. Asiedu, "FaceNet recognition algorithm subject to multiple constraints: assessment of the performance," *Scientific African*, vol. 23, 2024, doi: 10.1016/j.sciaf.2023.e02007.

[18] A. Nemavhola, C. Chibaya, and S. Viriri, "A systematic review of CNN architectures, databases, performance metrics, and applications in face recognition," *Information*, vol. 16, no. 2, 2025, doi: 10.3390/info16020107.

[19] A. Maafiri, O. Elharrouss, S. Rfifi, S. A. A.-Maadeed, and K. Chougdali, "DeepWTPCA-L1: a new deep face recognition model based on WTPCA-L1 norm features," *IEEE Access*, vol. 9, pp. 65091–65100, 2021, doi: 10.1109/ACCESS.2021.3076359.

[20] Z. Li, L. Xie, and G. Wang, "Deep learning features in facial identification and the likelihood ratio bound," *Forensic Science International*, vol. 344, 2023, doi: 10.1016/j.forsciint.2023.111576.

[21] J. D.-Tomillo, J. M. A.-Calero, and Q. Wang, "Dynamic-distance-based thresholding for UAV-based face verification algorithms," *Sensors*, vol. 23, no. 24, 2023, doi: 10.3390/s23249909.

[22] D. Akingbesote, Y. Zhan, R. Maskeliūnas, and R. Damaševičius, "Improving accuracy of face recognition in the era of mask-wearing: an evaluation of a Pareto-optimized FaceNet model with data preprocessing techniques," *Algorithms*, vol. 16, no. 6, 2023, doi: 10.3390/a16060292.
[23] S. M. Varnosfaderani and M. Forouzanfar, "The role of AI in hospitals and clinics: transforming healthcare in the 21st century," *Bioengineering*, vol. 11, no. 4, 2024, doi: 10.3390/bioengineering11040337.
[24] M. Pech, J. Vrchota, and J. Bednář, "Predictive maintenance and intelligent sensors in smart factory: review," *Sensors*, vol. 21, no. 4, pp. 1–39, 2021, doi: 10.3390/s21041470.
[25] J. Alvi, I. Arif, and K. Nizam, "Advancing financial resilience: a systematic review of default prediction models and future directions in credit risk management," *Heliyon*, vol. 10, no. 21, 2024, doi: 10.1016/j.heliyon.2024.e39770.

BIOGRAPHIES OF AUTHORS



Anan Nugroho received his B.Eng. degree in Physics Engineering, M.Eng. degree in Electrical and Information Engineering, and doctoral degree (Dr.) in Electrical Engineering, all from Universitas Gadjah Mada, Indonesia. He is currently an assistant professor with the Department of Electrical and Computer Engineering, Universitas Negeri Semarang, Indonesia. His research interests lie in intelligent vision and imaging (IVI), focusing on the development and application of algorithms, models, and systems that enable machines to interpret and understand visual information from the environment by adopting and extending human visual capabilities. He is a member of the Association for Computing Machinery (ACM) and currently serves as the Laboratory Coordinator of Artificial Intelligence, Robotics, and Internet of Things (AIRIoT) at Universitas Negeri Semarang (2023–present). He can be contacted at email: anannugroho@mail.unnes.ac.id.



Ahmad Zein Al Wafi holds a bachelor of engineering degree in Electrical Engineering from Universitas Negeri Semarang, Indonesia, in 2024. He is currently a freelance machine learning researcher. He was formerly involved in research and development through the UEESRG Lab and AIRIoT Lab. His research includes optimization, machine learning, system information, and intelligent robotics. In his professional work, he has already led projects in intelligent digitalization and robotics for smart tourism, building a containerized machine learning operations ecosystem platform, and becoming an external contributor at Dicoding Indonesia. He is deeply passionate about bridging the gap between research and real-world AI systems, emphasizing reproducibility, system reliability, and sustainable ML infrastructure. He can be contacted at email: ahmadzeinalwafi@outlook.com.



Muhammad Iqbal received his B.Eng. in Computer Engineering from Universitas Negeri Semarang, Indonesia (2025). His thesis on a "Prototype of facial biometric attendance edge device based on FaceNet algorithm" cemented his focus on optimized models. His primary research interests lie at the intersection of efficient deep learning, large-scale recognition systems, and facial anti-spoofing (presentation attack detection). He currently leads the implementation of these technologies as the president director of Wiyasa Rohmah Barokah Sejahtera (WIYORA), an AI solutions startup. He can be contacted at email: muhammad.iqbal@wiyoratechnology.com or xmod3905@proton.me.



Mona Subagja received his B.Ed. degree in Electrical Engineering Education from Universitas Negeri Semarang, Indonesia, in 2002, and his master of computer science (M.Comp.) degree from Universitas Dian Nuswantoro, Indonesia, in 2016. He is currently the director of the Directorate of Information System and Public Relations at Universitas Negeri Semarang, Indonesia. His research interests include information technology and computer networks, with a focus on the design, implementation, and management of information systems and network infrastructures in higher education environments. He can be contacted at email: mona@mail.unnes.ac.id.