

# Hybrid CNN-ViT integration into Siamese networks for robust iris biometric verification

Samihah Abdul Latif, Khairul Azami Sidek, Aisha Hassan Abdalla Hashim

Department of Electrical and Computer Engineering, Kuliyyah of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

## Article Info

### Article history:

Received Aug 21, 2024  
Revised May 26, 2025  
Accepted Jul 5, 2025

### Keywords:

Convolutional neural network  
Convolutional neural network-  
vision transformer  
Hybrid  
Iris recognition  
Siamese network  
Vision transformer

## ABSTRACT

Iris recognition has emerged as a critical biometric verification method, valued for its high accuracy and resistance to forgery. However, traditional convolutional neural network (CNN)-based models, despite their strength in extracting local iris features, struggle to capture global dependencies, which limits their generalization across different datasets. Additionally, conventional classification-based approaches struggle to accurately verify new individuals with limited training data. Thus, this study proposed a hybrid CNN-vision transformer (CNN-ViT) model within a Siamese network to enhance one-shot learning capability by combining CNN's local feature extraction with vision transformers (ViT's) global attention. To evaluate its performance, the hybrid model was compared with VGG16 and ResNet under the same training conditions for 20 epochs. VGG16 and ResNet rely on pre-trained models, whereas the hybrid CNN-ViT model is specifically designed to achieve this task with an increment to 98.9% training accuracy, surpassing the TinySiamese model's benchmark accuracy. It also attained a recall of 75%, demonstrating strong sensitivity in correctly identifying positive matches. The hybrid model maintained an excellent balance between learning and generalization by employing the binary cross entropy (BCE) loss function. These findings contribute to the development of efficient iris recognition systems, paving the way for advanced biometric applications in financial transactions, border control and mobile security.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Khairul Azami Sidek  
Department of Electrical and Computer Engineering, Kuliyyah of Engineering  
International Islamic University Malaysia  
P.O. Box 10, 50728 Kuala Lumpur, Malaysia  
Email: azami@iiu.edu.my

## 1. INTRODUCTION

Iris biometric recognition is a sophisticated technology for identifying individuals based on the unique patterns in human irises [1], [2]. The full process which comprising iris image acquisition, segmentation, and feature extraction is crucial, as each phase significantly influences recognition accuracy [3], [4]. Recent advancements have greatly enhanced the performance of iris recognition systems in terms of accuracy, reliability, and usability. For instance, modern systems now operate effectively under challenging conditions such as low-light environments or when the subject is in motion, making them increasingly suitable for real-world applications [5].

This growing robustness has encouraged widespread adoption across various sectors like healthcare, retail, and transportation, driving the rapid expansion of the global market for iris recognition access control systems [6]. Furthermore, the integration of artificial intelligence (AI) and machine learning (ML) has played

a key role in improving the precision and adaptability of these systems [7]. Innovations such as multimodal biometrics, continuous authentication, and internet of things (IoT) integration have pushed iris recognition towards becoming a secure and scalable solution for identity verification [8].

In parallel with these developments, the rising computational power has fueled the emergence of ML techniques like one-shot learning, which are particularly beneficial for biometric recognition tasks where collecting large datasets is impractical [9], [10]. One-shot learning enables systems to accurately identify individuals from a single example, significantly enhancing efficiency and applicability in security contexts with limited training data [11], [12]. A prominent architecture supporting one-shot learning is the Siamese network, which has proven effective for biometric recognition including iris recognition under data-scarce conditions [13]. Siamese networks consist of two identical subnetworks that share parameters and are designed to process pairs of biometric samples. By extracting feature vectors and computing similarity scores based on the distance between these vectors, the model determines whether two inputs belong to the same individual [14].

To further improve feature representation in iris recognition, researchers have explored the integration of self-attention mechanisms into traditional convolutional neural networks (CNNs) [15], [16]. Inspired by human visual perception, these mechanisms help emphasize the most salient features in an image. When combined with the Siamese architecture, attention-augmented networks enhance the system's ability to focus on distinctive iris patterns, thus boosting recognition accuracy. This hybrid approach has shown promising results across various biometric modalities such as palmprint and facial recognition, underscoring its versatility and effectiveness [13], [17].

While CNNs have been widely adopted for iris biometric verification due to their powerful feature extraction capabilities, they still face notable limitations. Traditional CNN-based approaches often struggle to generalize across varying image conditions such as noise, illumination changes, occlusions, or off-angle iris images which are common in real-world environments. Moreover, CNNs can be heavily reliant on large-scale, labeled datasets to learn effective representations, which presents a challenge in biometric contexts where annotated iris datasets are often limited or imbalanced. These constraints reduce the scalability of CNN-based systems when deployed in uncontrolled or resource-constrained settings.

To address these challenges, recent studies have turned to Siamese networks, which are particularly suitable for scenarios with limited training data. The researcher [18], [19] explored the use of Siamese networks for face recognition, targeting the challenge of acquiring labelled data. These studies adopted a double-branch architecture and leveraged cross-entropy loss for binary classification, achieving results comparable to supervised baselines despite utilizing an unsupervised learning setup. Likewise, [20] compared two approaches for facial detection: one using CNNs combined with the k-nearest neighbours (KNN) classifier, and the other applying a Siamese network for similarity-based classification. The comparative results highlighted the flexibility and efficiency of the Siamese network in dealing with real-world variability and data limitations.

Beyond facial recognition, Siamese networks have also shown strong performance in broader domains such as image retrieval [21], one-shot learning for object recognition, medical imaging, and even robotics, proving effective in contexts where traditional models falter due to data scarcity. In visual object tracking, [14] illustrated how combining Siamese networks with discriminative correlation filters (DCF) improved tracking accuracy and robustness. Similarly, [22] proposed innovative pair generation techniques and feature fusion strategies that further enhance Siamese network performance in regression tasks. These studies collectively emphasize the versatility and adaptability of the Siamese architecture across tasks that require discriminative learning with limited data.

Despite the effectiveness, the performance of Siamese networks is still dependent on the quality of the extracted features. To this end, we propose a hybrid CNN-vision transformer (ViT) model integrated within the Siamese network framework to elevate feature extraction capabilities for iris biometric verification. While CNNs excel at capturing local texture and spatial hierarchies, ViTs add the power of global attention mechanisms that can represent long-range dependencies in the image, which is critical for detecting subtle iris patterns. By combining these two architectures, the hybrid CNN-ViT model captures both detailed local features and global structural patterns, leading to more robust and discriminative embeddings.

In summary, recent research confirms the effectiveness of Siamese networks across diverse domains [23], showing promising outcomes in limited-data environments. Building on this foundation, the key contribution of this study is the integration of a novel hybrid CNN-ViT model into a Siamese network architecture. This approach significantly enhances feature extraction and discriminative learning, thereby enhancing the system's performance in real-world iris biometric verification scenarios.

## 2. METHOD

This section will go over the Siamese network architecture, datasets, and performance matrix used in this iris recognition verification.

### 2.1. Overview of siamese networks architecture

The Siamese networks architecture uses two identical branches of the hybrid CNN-ViT model to process input iris image pairs illustrated in Figure 1, adopted and re-constructed from studies in [14], typically used for tasks like similarity learning and verification. The proposed model incorporates a hybrid CNN and ViT within the Siamese network architecture, designed to enhance iris biometric recognition. By integrating the hybrid model, the system combines the local feature extraction capabilities of CNN with the global attention mechanisms of ViT, addressing challenges such as occlusion, lighting variations and noisy data.

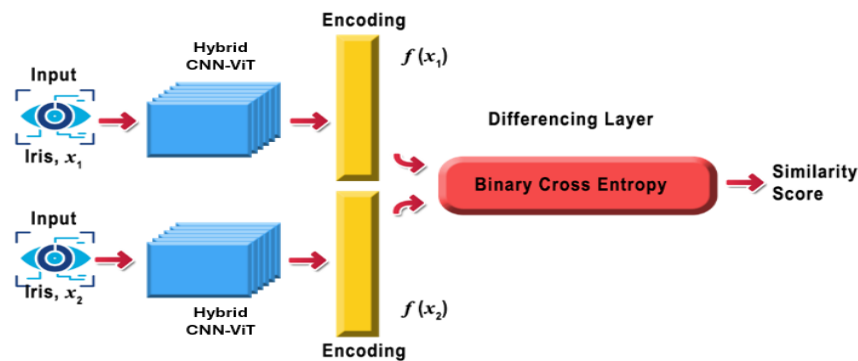


Figure 1. The Siamese network architecture

#### a. Input and preprocessing stage

The process begins with two input iris images,  $x_1$  and  $x_2$ , which are passed through identical hybrid CNN-ViT branches to extract feature embedding  $f(x)$ . Preprocessing includes applying Gaussian Blur to reduce noise, resizing the images to  $128 \times 128$  pixels to standardize their dimensions, and normalizing pixel values to scale them between 0 and 1. These steps ensure the network focuses on meaningful features rather than irrelevant noise or inconsistencies in image quality. Preprocessing is vital for minimizing the effect of lighting variations and ensuring uniformity, as any inconsistencies in this phase can propagate errors throughout the system.

#### b. Feature extraction with the hybrid CNN-ViT

The hybrid CNN-ViT model plays a central role in feature extraction, combining the strengths of CNN and ViT. CNN layers focus on capturing local features such as edges, textures and fine details, while the ViT layers use self-attention mechanisms to capture global relationships and contextual dependencies within the iris images. This dual approach allows the hybrid model to address challenges like occlusions and noisy data more effectively than traditional CNNs. The output of this phase is a high-dimensional embedding vector,  $f(x)$ , that represents the essential features of the input iris image in a robust and discriminative form.

#### c. Siamese architecture design

The Siamese network architecture ensures that both input images are processed identically using two identical hybrid CNN-ViT branches. This shared-weight design guarantees that the same transformations are applied to  $x_1$  and  $x_2$ , producing embeddings  $f(x_1)$  and  $f(x_2)$ . By processing the images through identical branches, the network avoids bias and ensures fair comparison of features. This architecture is key to enabling the network to learn relationships between the embeddings of genuine pairs (same person) and imposter pairs (different person) during training.

#### d. Differencing layer (Euclidean distance calculation)

After feature extraction, the embeddings  $f(x_1)$  and  $f(x_2)$  are passed through a differencing layer to calculate the similarity. The Euclidean distance, defined as in (1) is used as the similarity metric. A smaller distance indicates high similarity, suggesting the images are of the same iris, while larger distance implies dissimilarity. This differencing layer is fundamental to the Siamese network, as it quantifies the closeness of the two embeddings in the feature space, which directly correlates with the verification decision.

$$d(x_1, x_2) = \|f(x_1) - f(x_2)\|^2 \quad (1)$$

e. Similarity score computation

The calculated Euclidean distance is converted into a similarity score, which is compared against a predefined threshold,  $T$ , to classify the input pair. If the distance is below  $T$ , the pair is classified as "genuine," indicating that the two images belong to the same person. Conversely, if the distance exceeds  $T$ , the pair is classified as an "imposter." The threshold  $T$  is determined during the validation process and is critical to balancing the trade-off between false positives (FPs) and false negatives (FNs). This phase enables real-time decision-making based on the computed similarity score.

f. Loss function

During training, binary cross entropy (BCE) is the loss function, which measures the difference between the predicted similarity scores and the actual labels. BCE ensures that the network learns to produce embeddings where genuine pairs are closer together and imposter pairs are farther apart in the feature space. This training process enables the network to distinguish between similar and dissimilar iris images effectively.

## 2.2. Dataset

The Chinese Academy of Science Institute of Automation (CASIA) dataset is a cornerstone in biometric research, particularly for iris recognition. It offers a comprehensive collection of iris images that are essential for developing and evaluating advanced iris recognition algorithms [24]. This study utilizes the CASIA-IrisV1 dataset, a widely recognized subset of the CASIA database, to train and test a Siamese Network for iris recognition. The dataset's diversity and high-quality images make it an ideal choice for this research. CASIA-IrisV1 comprises 756 grayscale images of irises from 108 individuals, with each individual contributing seven different images of their iris. These images are stored in JPEG format, typically with a resolution of 320×280 pixels. The dataset was captured using a specially designed sensor under controlled indoor lighting conditions to ensure high-quality images suitable for biometric analysis. Each image is annotated with the subject ID and the eye (left or right), providing essential metadata for training and evaluating iris recognition algorithms. The training process involves creating pairs of images by using the creates pair function, where each pair consists of either matching (same subject) or non-matching (different subjects) irises. This pairing strategy helps the model learn to distinguish between unique iris patterns effectively. The model's performance is assessed using accuracy, equal error rate (EER) and recall.

## 2.3. Model evaluations

Evaluating the performance of an iris recognition model requires a thorough understanding of various metrics that provide insights into the model's effectiveness and reliability. In this study, the proposed model was assessed using accuracy, EER, recall and loss functions. Accuracy provides a general overview of how well the model distinguishes between similar and dissimilar pairs of iris images. Accuracy can be measured using in (2):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

where, true positive (TP) is the number of correctly identified similar pairs, true negative (TN) is the number of correctly identified dissimilar pairs, FP is the number of dissimilar pairs incorrectly identified as similar and FN is the number of similar pairs incorrectly identified as dissimilar. High accuracy implies that the model correctly identifies most iris pairs, though it does not differentiate between types of errors (FP and FN).

The EER provides a single value that summarizes the trade-off between security by minimizing false acceptances and usability by minimizing false rejections [25]. Typically, EER is calculated by plotting the receiver operating characteristic (ROC) curve, which depicts the trade-off between false acceptance rate (FAR) and false rejection rate (FRR) at various threshold values. In contrast, recall measures the model's ability to correctly identify all true matches. It is the ratio of true positives to the total number of actual matching pairs. High recall in iris recognition indicates that the model is effective at finding all true matches, thereby minimizing FNs.

In this study, we also explore how well the model's predictions align with expected outcomes by investigating the impact of different loss functions on the performance of a Siamese network for iris recognition. Loss refers to the difference between the predicted output of a model and the actual target value [26]. Specifically, we examine BCE, triplet loss, and contrastive loss, each of which has unique characteristics that influence how the network learns to distinguish between similar and dissimilar iris patterns. BCE is frequently used in binary classification tasks, where the objective is to classify pairs of inputs as either similar or dissimilar. The network generates a probability between 0 and 1, representing the

similarity between two inputs, and BCE is then used to calculate the loss between the predicted probability and the actual label (1 for similar and 0 for dissimilar).

Triplet and contrastive loss are used in metric learning tasks, where the goal is to learn a distance metric such that similar inputs are close together in the embedding space, and dissimilar inputs are far apart. A triplet consists of an anchor, a positive example (identical to the anchor) and a negative example (dissimilar to the anchor). Triplet loss is designed to ensure that the distance between the anchor and the positive is smaller than the distance between the anchor and the negative by a specified margin. Whereas contrastive loss calculates the distance between two embeddings. If the inputs are similar (same class), the loss function encourages the network to minimize the distance between the embeddings. If the inputs are dissimilar (different classes), the loss function encourages the network to maximize the distance between the embeddings, up to a certain margin.

Loss is a critical metric used during the training of ML models to guide the optimization process and improve the model's predictive performance. By minimizing the loss, the model's parameters are adjusted iteratively, leading to better alignment between the predicted outcomes and the actual targets. The choice of loss function, whether BCE, contrastive loss or triplet loss for fine-grained discrimination, directly impacts the model's ability to learn and generalize from data. Ultimately, the loss function's effectiveness determines the model's success in accurately capturing the underlying patterns in the data, thereby enhancing its utility in real-world applications.

### 3. RESULTS AND DISCUSSION

This section presents the experimental results of the hybrid CNN-ViT model implemented within a Siamese network for iris biometric verification. The discussion includes performance metrics, comparisons with existing methods, and an analysis of the model's strengths and limitations in addressing real-world challenges.

A systematic evaluation approach was adopted to compare the implementation of a Siamese network using three different models (hybrid CNN-ViT, VGG16, and ResNet). This evaluation involves training and testing each model under the same conditions to ensure a fair comparison of their performance. Specifically, a batch size of 16 was chosen for training, which refers to the number of training samples used in one iteration to update the model's parameters. This batch size balances memory efficiency and the stability of gradient updates. Additionally, each model was trained for 20 epochs, where an epoch refers to one complete pass through the entire training dataset. Training for 20 epochs ensures that the models have sufficient iterations to learn the underlying patterns in the data without overfitting.

Table 1 compares the performance metrics of three different neural networks across three performance parameters. The proposed hybrid CNN-ViT model achieves the highest accuracy at 98.9%, followed by VGG16 at 96% and ResNet V1 at 94.3%. VGG16 has the lowest EER at 0.0091, indicating the best balance between false acceptance and rejection rates. Furthermore, the hybrid CNN-ViT has a recall of 0.75, indicating it correctly identifies 75% of the positive instances. In contrast, VGG16 and ResNet V1 have a recall of 0.44 and 0.11, respectively indicating lower sensitivity in correctly identifying positive cases. VGG16 shows a balanced performance with low EER, moderate recall, and high accuracy at 96.0%. In comparison, the proposed model offers the best overall performance for iris recognition because it combines high accuracy with the best recall, ensuring that most true matches are detected. Although its EER is slightly higher than VGG16 model, the trade-off is worthwhile given its superior recall. This makes the proposed model the most balanced and reliable choice for iris recognition, particularly in applications where correctly identifying matches is crucial.

Table 1. Comparison of performance parameter

	Accuracy	EER	Recall
VGG16	96.0	0.0091	0.44
ResNet V1	94.3	0.0827	0.11
Hybrid CNN-ViT	<b>98.9</b>	0.0303	0.75

Table 2 provides a comparative study overview of the accuracy achieved by various biometric systems using Siamese network architecture. The lip-based biometric system, achieved a notably high accuracy of 98.24%, highlighting its effectiveness in distinguishing between individuals based on lip movement or patterns. In contrast, a TinySiamese modality for a biometric system that evaluates fingerprint, face, and gait recognition, produced various accuracies of 90.13%, 85.87%, and 98.39%, respectively, demonstrating the strengths and weaknesses of each modality. Face recognition systems, as cited in the study

[19], showed significant variability, with accuracies of 74.4%, indicating the challenges associated with facial recognition due to factors like lighting, pose, and facial expressions. Finally, the hybrid CNN-ViT for iris biometric in this study outperforms all the other approaches, achieving an accuracy of 98.9%, which suggests that the chosen layers were particularly well-suited to this system. These design choices likely contributed to its superior performance in biometric recognition, making it a highly effective and reliable approach.

Table 2. Performance analysis of Siamese network in different biometric

Types of biometric	Accuracy (%)
Lip-based [27]	98.24
Fingerprint	90.13
Face	85.87
Gait [28]	98.39
Face recognition [19]	74.4
Hybrid CNN-ViT	98.9

Furthermore, the study evaluated the impact of three distinct loss functions: BCE, contrastive and triplet loss. Figure 2 shows the BCE (green line) steady and consistent decrease over the epochs, with minimal fluctuations. The loss starts at 0.08 which is relatively high and quickly drops, stabilizing after a few epochs. This pattern suggests that the model using BCE loss is efficiently learning the distinctions between classes, leading to a fast convergence. The steady decline in loss indicates that the model is becoming more confident in its predictions as training progresses.

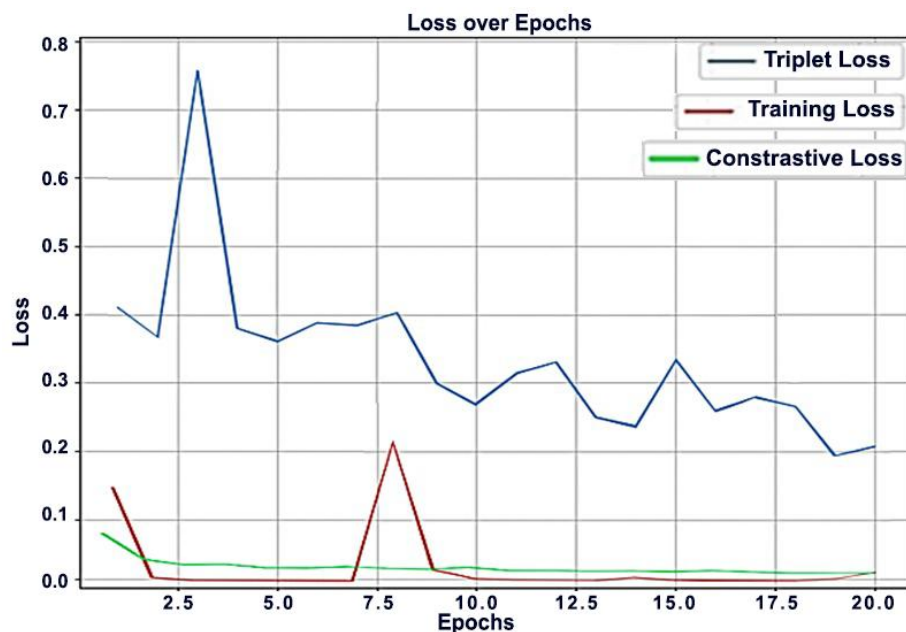


Figure 2. Graph of BCE loss, contrastive loss and triplet loss

The contrastive loss graph shown in the red line indicates more fluctuations compared to BCE. While the loss generally trends downward, there are significant spikes, particularly around the 7th and 8th epochs. The spikes in this graph suggest that the model encounters challenging pairs of iris images, causing a temporary increase in loss. However, the overall trends indicate that the model is still learning to differentiate between similar and dissimilar pairs, with the loss stabilizing towards the later epochs.

The triplet loss represented by the blue line is the most fluctuating among the three loss functions, with the loss starting at 0.4 and spiking to near 0.8 at the 3rd epochs then decreasing gradually, showing consistent peaks and troughs throughout the training process. The fluctuating triplet loss suggests that the model is continuously refining its decision boundaries, particularly for more challenging triplets. The gradual

decrease indicates that the model is slowly learning to enforce the margin between similar and dissimilar pairs, but the process is more complex and requires more time to stabilize.

Based on the analysis of the three loss functions, BCE appears to be the most suitable loss function for this specific task. The BCE loss demonstrates a steady and consistent decrease indicating that the model is efficiently learning the distinctions between different iris classes. The quick drop and early stabilization of the loss suggest that the model is confident in its predictions and converges faster compared to other loss functions. On the other hand, contrastive loss and Triplet loss exhibit more fluctuations and instability during training indicating that the models face challenges in consistently differentiating between similar and dissimilar pairs, which may lead to slower convergence and potential overfitting, especially in a complex dataset like the CASIA iris database. Therefore, given the stability and faster convergence of the BCE loss, it is the most appropriate choice for achieving reliable and efficient performance in this iris recognition system.

#### 4. CONCLUSION

The hybrid CNN-ViT model integrated within a Siamese network offers significant performance improvements over traditional CNN-based approaches for iris biometric recognition. While CNNs are effective at capturing local features, they often struggle with generalization and robustness, especially in the presence of limited data, lighting variations, occlusions, and noise. By combining CNNs with ViT, the hybrid model leverages both local feature extraction and global attention, enabling richer and more discriminative iris representations. Compared to three baseline CNN models within the same Siamese architecture, the hybrid CNN-ViT model consistently demonstrated superior generalization, maintaining high accuracy on validation data and avoiding overfitting. Its efficiency in one-shot learning scenarios makes it ideal for applications where labelled iris data is limited. Moreover, the model showed greater resilience under challenging real-world conditions, making it a reliable solution for robust biometric authentication.

Despite its strengths, the model does present some limitations. Notably, the increased complexity and longer inference time introduced by the ViT component may pose constraints for real-time or resource-limited edge applications. Additionally, the model's performance, while robust, still depends on careful preprocessing and balanced input pair generation, which may not always be feasible in uncontrolled environments. Future work will explore optimization techniques to reduce computational overhead, such as model pruning or knowledge distillation, to make the architecture more suitable for real-time deployment. Moreover, incorporating advanced regularization methods and dynamic data augmentation strategies could further improve generalization and reduce susceptibility to overfitting. Cross-dataset validation and testing across diverse demographics and sensor types would also help assess the model's scalability and fairness in broader deployments.

Importantly, the study demonstrates the practical robustness of this approach for mobile and edge deployment, where computational efficiency, memory constraints, and real-time performance are critical. By achieving high verification accuracy in low-data scenarios, the hybrid architecture supports the development of lightweight, secure, and reliable iris recognition solutions tailored for smartphones, embedded systems, and IoT-enabled security platforms. With targeted refinements such as model enhancement, pruning, or knowledge distillation, this hybrid CNN-ViT Siamese network presents a strong foundation for next-generation biometric authentication technologies that are not only accurate and robust but also scalable and deployable in real-world environments.

#### ACKNOWLEDGEMENTS

The authors would like to thank the Department of Electrical and Computer Engineering, Kuliyyah of Engineering, International Islamic University Malaysia and Ministry of Higher Education Malaysia for the continuous support.

#### FUNDING INFORMATION

The authors would like to thank the Ministry of Higher Education, Malaysia, for funding this research.

#### AUTHOR CONTRIBUTIONS STATEMENT

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



Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Samihah Abdul Latif	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Khairul Azami Sidek	✓	✓				✓		✓	✓	✓	✓	✓	✓	✓
Aisha Hassan Abdalla Hashim	✓	✓		✓		✓	✓			✓	✓	✓	✓	✓

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review &amp; Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

The dataset underlying this study's findings is publicly available in <https://www.kaggle.com/search?q=casia+iris+in%3Adatasets> (CASIA Iris-V1). This dataset includes all parameters used for model training and validation.

## REFERENCES




- [1] R. H. Farouk, H. Mohsen, and Y. M. A. El-Latif, "A proposed biometric technique for improving iris recognition," *International Journal of Computational Intelligence Systems*, vol. 15, no. 1, Sep. 2022, doi: 10.1007/s44196-022-00135-z.
- [2] G. Liu, W. Zhou, L. Tian, W. Liu, Y. Liu, and H. Xu, "An efficient and accurate iris recognition algorithm based on a novel condensed 2-ch deep convolutional neural network," *Sensors*, vol. 21, no. 11, Mei. 2021, doi: 10.3390/s21113721.
- [3] S. Lei, A. Shan, B. Liu, Y. Zhao, and W. Xiang, "Lightweight and efficient dual-path fusion network for iris segmentation," *Scientific Reports*, vol. 13, no. 1, Aug. 2023, doi: 10.1038/s41598-023-39743-w.
- [4] C. Wang, Y. Wang, B. Xu, Y. He, Z. Dong, and Z. Sun, "A lightweight multi-label segmentation network for mobile iris biometrics," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2020, pp. 1006–1010, doi: 10.1109/ICASSP40776.2020.9054353.
- [5] Y. Yin, S. He, R. Zhang, H. Chang, and J. Zhang, "Deep learning for iris recognition: a review," *Neural Computing and Applications*, vol. 37, no. 17, pp. 11125–11173, 2025, doi: 10.1007/s00521-025-11109-5.
- [6] L. Omelina, J. Goga, J. Pavlovicova, M. Oravec, and B. Jansen, "A survey of iris datasets," *Image and Vision Computing*, vol. 108, pp. 1–20, Apr. 2021, doi: 10.1016/j.imavis.2021.104109.
- [7] L. Nanni, G. Minchio, S. Brahnam, D. Sarraggiotto, and A. Lumini, "Closing the performance gap between siamese networks for dissimilarity image classification and convolutional neural networks," *Sensors*, vol. 21, no. 17, Aug. 2021, doi: 10.3390/s21175809.
- [8] L. Lin, Y. Zhao, J. Meng, and Q. Zhao, "A federated attention-based multimodal biometric recognition approach in IoT," *Sensors*, vol. 23, no. 13, Jun. 2023, doi: 10.3390/s23136006.
- [9] J. Mohr, F. Breidenbach, and J. Frochte, "An approach to one-shot identification with neural networks," *International Joint Conference on Computational Intelligence*, vol. 1, pp. 344–351, 2021, doi: 10.5220/0010684300003063.
- [10] N. I. A. Sabri and S. Setumin, "One-shot learning for facial sketch recognition using the Siamese convolutional neural network," in *ISCAIE 2021 - IEEE 11th Symposium on Computer Applications and Industrial Electronics*, Apr. 2021, pp. 307–312, doi: 10.1109/ISCAIE51753.2021.9431773.
- [11] N. Lee, S. Hong, and H. Kim, "Single-trace attack using one-shot learning with siamese network in non-profiled setting," *IEEE Access*, vol. 10, pp. 60778–60789, 2022, doi: 10.1109/ACCESS.2022.3180742.
- [12] L. Zhu, P. Xu, and C. Zhong, "Siamese network based on CNN for fingerprint recognition," in *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology, CEI 2021*, pp. 303–306, Sep 2021, doi: 10.1109/CEI52496.2021.9574487.
- [13] X. Chen and K. He, "Exploring simple Siamese representation learning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2021, pp. 15745–15753, doi: 10.1109/CVPR46437.2021.01549.
- [14] S. Javed, M. Danelljan, F. S. Khan, M. H. Khan, M. Felsberg, and J. Matas, "Visual object tracking with discriminative filters and siamese networks: a survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 6552–6574, 2023, doi: 10.1109/TPAMI.2022.3212594.
- [15] Y. Liu, G. Chang, G. Fu, Y. Wei, J. Lan, and J. Liu, "Self-attention based siamese neural network recognition model," in *Proceedings of the 34th Chinese Control and Decision Conference*, Aug. 2022, pp. 721–724, doi: 10.1109/CCDC55256.2022.10034228.
- [16] S. A. Latif, K. A. Sidek, and A. H. A. Hashim, "An efficient iris recognition technique using CNN and vision transformer," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 34, no. 2, pp. 235–245, Des. 2024, doi: 10.37934/araset.34.2.235245.
- [17] L. Utkin, M. Kovalev, and E. Kasimov, "An explanation method for Siamese neural networks," *Smart Innovation, Systems and Technologies*, vol. 220, pp. 219–230, 2021, doi: 10.1007/978-981-33-6632-9\_19.
- [18] E. Solomon, A. Woubie, and E. S. Emiru, "Deep learning based face recognition method using Siamese network," *arXiv*, 2023, doi: 10.48550/arXiv.2312.14001.






- [19] M. Meddad, C. Moujahdi, M. Mikram, and M. Rziza, "Convolutional Siamese neural network for few-shot multi-view face identification," *Signal, Image and Video Processing*, vol. 17, no. 6, pp. 3135–3144, 2023, doi: 10.1007/s11760-023-02535-w.
- [20] C. R. Kumar, N. Saranya, M. Priyadarshini, E. D. Gilchrist, and M. K. Rahman, "Face recognition using CNN and Siamese network," *Measurement: Sensors*, vol. 27, pp. 1–11, Jun. 2023, doi: 10.1016/j.measen.2023.100800.
- [21] K. L. Wiggers, A. S. Britto, L. Heutte, A. L. Koerich, and L. S. Oliveira, "Image retrieval and pattern spotting using Siamese neural network," in *Proceedings of the International Joint Conference on Neural Networks*, 2019, pp. 1–8, doi: 10.1109/IJCNN.2019.8852197.
- [22] Y. Zhang *et al.*, "Similarity-based pairing improves efficiency of Siamese neural networks for regression tasks and uncertainty quantification," *Journal of Cheminformatics*, vol. 15, no. 1, Aug. 2023, doi: 10.1186/s13321-023-00744-6.
- [23] W. Xiao and D. Wu, "An improved Siamese network model for handwritten signature verification," in *ICNSC 2021 - 18th IEEE International Conference on Networking, Sensing and Control: Industry 4.0 and AI*, Des. 2021, pp. 1–6, doi: 10.1109/ICNSC52481.2021.9702190.
- [24] M. M. Alrifae, M. M. Abdallah, and B. G. Al Okush, "A short survey of IRIS images databases," *The International Journal of Multimedia & Its Applications*, vol. 9, no. 2, pp. 1–14, Apr. 2017, doi: 10.5121/ijma.2017.9201.
- [25] S. Ayeswarya and K. J. Singh, "A comprehensive review on secure biometric-based continuous authentication and user profiling," *IEEE Access*, vol. 12, pp. 82996–83021, 2024, doi: 10.1109/ACCESS.2024.3411783.
- [26] J. Terven, D. M. C. Esparza, A. R. Pedraza, and E. A. C. Urbiola, "Loss Functions and Metrics in Deep Learning," *arXiv*, pp. 1–76, 2023, doi: 10.1007/s10462-025-11198-7.
- [27] A. Zakeri, H. Hassanpour, M. H. Khosravi, and A. M. Nourollah, "WhisperNetV2: SlowFast Siamese network for lip-based biometrics," *arXiv*, 2024, doi: 10.48550/arXiv.2407.08717.
- [28] I. Jarraya, T. M. Hamdani, H. Chabchoub, and A. M. Alimi, "TinySiamese network for biometric analysis," *arXiv*, 2023, doi: 10.48550/arXiv.2307.00578.

## BIOGRAPHIES OF AUTHORS






**Samihah Abdul Latif**    received a degree in Information Technology from University Tun Hussein Onn (UTHM), Malaysia in 2008. A year later, she received her Master's in Technical and Vocational Education from the same university. Currently, she is a Ph.D. candidate in the Department of Electrical and Computer Engineering, Kuliyyah of Engineering, International Islamic University Malaysia (IIUM). She can be contacted at email: samihahabdlatif@gmail.com.



**Khairul Azami Sidek**    a graduate of the International Islamic University Malaysia (IIUM) in Computer and Information Engineering (Hons), started his career as an assistant lecturer at the Department of Electrical and Computer Engineering, Kuliyyah of Engineering, IIUM in 2004. In 2007, he was appointed as a lecturer in the same department after completing his Master's degree in Communication and Computer Engineering from University Kebangsaan Malaysia. Later, in 2014, he was appointed as an Assistant Professor after finishing his Ph.D. studies in Computer Science at RMIT University, Melbourne, Australia. In August 2018, he was promoted to associate professor in the Department of Electrical and Computer Engineering. His area of interest is biometric recognition, pattern recognition, and biomedical signal processing. He can be contacted at email: azami@iium.edu.my.



**Aisha Hassan Abdalla Hashim**    received her Bachelor in Electronic Engineering (1990) from University of Gezira, M.Sc. in Computer Science (1996) from University of Khartoum and Ph.D. in Computer Engineering (2007) from International Islamic University (IIUM). She joined IIUM in 1997 and is currently a Professor at the Department of Electrical and Computer Engineering. She can be contacted at email: aisha@iium.edu.my.