International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.00169
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 1114 - 1123

Faceirisnet: A Deep Multimodal Biometric Recognition Framework Using Resnet And Triplet Loss

S. Selvarani[1], Dr. M. Mary Shanthi Rani[2]
[1]Research Scholar, Department of Computer Science and Applications, Gandhigram Rural Institute, Dindigul, Tamil Nadu, India. & Assistant Professor, Department of MCA, Fatima College, Madurai, Tamil Nadu, India.
[2]Correspondence Author: Professor, Department of Computer Science and Applications, Gandhigram Rural Institute, Dindigul, Tamil Nadu, India.
Emails: rani.s.selva@gmail.com[1], drmaryshanthi@gmail.com[2]

Abstract
Facial and Iris recognition are essential parts of biometric security systems used to ensure reliable identification and authentication in high-stakes situations. For accurate facial and iris identification tasks, this study proposes a dual Convolutional Neural Networks (CNNs)-based architecture. The integrated system offers a multi-modal approach to biometric verification by processing iris and facial images concurrently. The CNN architecture used by the facial recognition module is based on ResNet and uses residual connections and deep feature extraction to solve the vanishing gradient issue. By optimizing a pre-trained ResNet model on the VGGFace2 dataset, transfer learning is used to achieve high accuracy in face identification and verification in difficult circumstances like occlusion and lighting changes. In order to separate the iris from eye pictures, the iris identification module uses a modified ResNet model that is tuned for fine-grained iris textures and incorporates a sophisticated segmentation technique. Model resilience is improved by data augmentation methods like rotations and random cropping. A normalized iris dataset is used to train the CNN, allowing for the extraction of discriminative iris characteristics that are necessary for accurate identification. A fully connected layer performs final classification after a late-fusion approach concatenates embeddings from both CNNs to merge facial and iris data for safe authentication is FACEIRISNET. It builds a strong multi-modal biometric system by utilizing both facial and iris features. Both facial and iris embeddings use a triplet loss function, which makes sure that embeddings of the same identity are grouped together and those of different identities are pushed away.

Keywords: *Facial recognition, Convolutional Neural Network, ResNet, Triplet Loss, Data Augmentation, Embedding*

1. Introduction

Facial and Iris recognition have become essential parts of contemporary biometric security systems because they offer dependable techniques for identification and authentication in a variety of applications. While iris identification concentrates on the complex patterns of the iris, which are equally different, facial recognition employs the distinguishing traits of a person's face to identify them. Both techniques have clear benefits; iris recognition is renowned for its high accuracy and resilience to spoofing, while facial recognition is frequently more practical and less invasive. By combining these two modalities into a single framework, biometric authentication systems' overall security and accuracy are improved by utilizing their complementary advantages. To achieve this integration, a dual Convolutional Neural Network (CNN) architecture based on ResNet-50 is used. Renowned for its deep residual learning capabilities, ResNet-50 effectively extracts features while resolving the vanishing gradient problem that frequently arises in deep networks[1]. By analysing their images simultaneously, this design enables the model to learn and adapt to complex patterns in both the iris and face modalities. The novel FACEIRISNET can achieve high accuracy even in difficult situations by using transfer learning techniques to pre-trained ResNet-50 models on pertinent datasets, such as specialized iris datasets and VGGFace2 for facial recognition[2][3]. According to recent research, multi-modal techniques outperform single-modal systems[4] in

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.00169
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 1114 - 1123

security-sensitive scenarios, increasing the overall dependability of biometric systems and setting them up for broad use in safe settings. In Section 2, the crucial elements of the suggested study such as dual Convolutional Neural Networks (CNNs)-based architecture intended for accurate facial and iris identification are highlighted. Ideas including facial and iris recognition methods, as well as developments in deep learning that facilitate the creation of a reliable multi-modal biometric identification system, are covered in the literature review. In Section 3, LFW(Labeled Faces in the Wild) - People (Face Recognition) Dataset, a dataset for training and testing models for face detection, particularly for recognising facial attributes and also VGGFace2 dataset, a large-scale face recognition benchmark, to adapt it to the specific characteristics of face images is presented. CASIA Iris Image Database (CASIA-Iris), which has been upgraded from CASIA-IrisV1 to CASIA-IrisV3, has been made available to the global biometrics community. It has been downloaded by over 3,000 users from 70 countries or regions, and a great deal of outstanding iris recognition work has been accomplished using these iris picture datasets. IITD Iris Database was collected from the staff and students of IIT Delhi by the Biometrics Research Laboratory. The objective is to establish a large-scale iris database of Indian users and make it available in the public domain. In Section 4, the proposed methodology is discussed and in Section 5 the results and discussions are given

2. Literature Review

ResNet-inspired CNN architecture, chosen for its deep feature extraction capabilities and residual connections that alleviate the vanishing gradient issue in deep networks, is used in the facial recognition module [1]. The Convolutional Neural Network (CNN) architecture known as ResNet (Residual Network) was first presented by Kaiming, by making it possible to train extremely deep networks up to hundreds or thousands of layers without experiencing the vanishing gradient issue, it transformed deep learning. Residual connections, which enable gradients to move across the network more efficiently during backpropagation, are the main novelty of ResNet. The network can learn spatial hierarchies of features thanks to the convolutional layers at its heart, which extract information from the input data by applying filters. By adding non-linearity, activation functions usually Rectified Linear Units (ReLU) (Fig. 1) allow the model to recognize intricate patterns. The utilization of residual connections, which generate shortcuts that avoid one or more convolutional layers, is a key breakthrough in ResNet. This solves the vanishing gradient issue that deep networks frequently face by allowing the network to learn residual functions. Each layer's outputs are normalized using batch normalization, which increases training speed and stability. Furthermore, by shrinking the spatial dimensions of feature maps, pooling layers reduce computing complexity without sacrificing important information. When combined, these elements enable ResNet designs to efficiently train deeper networks and attain high accuracy in a range of applications, including as biometric recognition and image categorization.

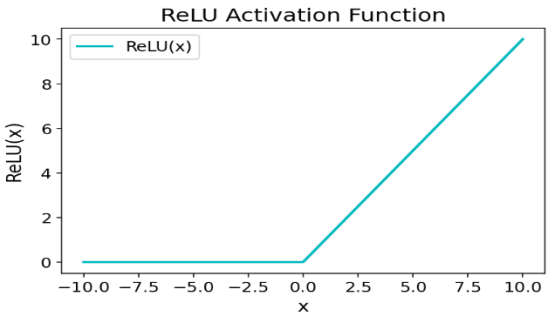

Figure 1 ReLU Representation

Adapting a pre-trained ResNet model to a particular job entail modifying its architecture and retraining it on a smaller, task-specific dataset after it has already been trained on a big dataset, like ImageNet. The first step in this procedure is to add a new layer to the ResNet architecture that matches the number of classes in the target dataset, replacing the last fully connected layer. To provide the new job a solid foundation, the pre-trained weights which capture important feature representations learnt during the first training are kept. By fine-tuning a pre-trained ResNet model on the VGGFace2 dataset, transfer learning is used to achieve high accuracy in face identification and verification in difficult

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.00169
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 1114 - 1123

circumstances such occlusion and lighting fluctuations [6]. Accuracy in facial recognition systems has greatly increased recently, especially with the use of transformer models. By using self-attention processes, which efficiently balance various elements in the input data, these models allow the focus to be on pertinent facial features. This feature enables the model to recognize and rank important features, like spatial relationships and expressions. Transformers improve the comprehension of facial features by capturing global dependencies within images, in contrast to conventional convolutional neural networks. As a result, this advancement enhances performance in difficult situations including changing lighting and occlusions in addition to increasing identification rates in controlled settings. Therefore, a viable avenue for further study in facial recognition technology is represented by transformer-based models. By enabling the model to concentrate on pertinent facial traits, recent developments, such as the incorporation of transformer models, have significantly improved the accuracy of facial recognition systems [7]. Furthermore, a study showed how contrastive loss can be used to enhance facial recognition performance in a variety of situations [8]. A modified ResNet model tailored for the fine-grained patterns present in iris texturing is used in the iris identification module. For efficient recognition, the iris must be precisely separated from the eye picture using a sophisticated iris segmentation algorithm [9]. Random cropping and rotations are examples of data augmentation techniques used to increase model robustness and boost performance on a variety of datasets [10]. Data augmentation methods like random cropping and rotations improve the generalization and durability of deep learning models, especially when it comes to picture classification problems. In order to help the model learn to focus on different areas and features within the image, random cropping is used to shrink photos to a specified dimension by choosing a random portion of the original image. This method works especially well when the topic of attention isn't always in the center or completely visible. In a similar vein, rotations cause changes in the pictures' orientation, making the model invariant to angle alterations. By offering a variety of training examples, this helps avoid overfitting and enhances the model's capacity to generalize to new data. By using a normalized iris dataset for training, the CNN is able to extract iris features that are necessary for accurate identification. The accuracy of iris recognition systems can be greatly increased by integrating attention mechanisms, as recent research has shown [11], and generative models have also been investigated for improved feature extraction [12]. Additionally, recent research has concentrated on improving the interpretability of iris recognition systems through the application of explainable AI techniques[13]. In many ML applications, generative models have become a potent method for better feature extraction, especially in computer vision tasks. In order to produce fresh data samples that closely resemble the original dataset, Generative Adversarial Networks (GANs) models learn to capture the underlying distribution of training data. For example, generative models can enhance the feature space in facial and iris recognition tasks by adding variations in lighting, position, and occlusion to the dataset. Furthermore, the performance of classification algorithms can be enhanced by using the latent representations that these models have learnt as input for subsequent tasks. There are now more opportunities to improve the precision and dependability of recognition systems thanks to generative models' capacity to synthesis realistic data and extract significant features. When combined, these data augmentation techniques FACEIRISNET improve model performance, particularly in tasks like iris and face recognition where orientation and positioning fluctuations are frequent.

3. Materials & Methods

LFW (Labeled Faces in the Wild) People dataset is a popular benchmark for face recognition studies which has more than 13,000 labeled photos of faces gathered from the internet. To evaluate performance parameters like accuracy and robustness in face recognition systems, researchers train and test models using the LFW dataset. Researchers from the Visual Geometry Group (VGG) at the University of Oxford created the VGGFace2 dataset, includes more than

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.00169
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 1114 - 1123

3.3 million photographs of over 9,000 people, each of whom is depicted in a range of positions, facial expressions, and lighting scenarios. VGGFace2 is a fundamental resource in the area, and researchers frequently utilize it for activities like benchmarking facial recognition algorithms, assessing performance indicators, and fine-tuning pre-trained models. CASIA Iris Image Database (CASIA-Iris), which has been upgraded from CASIA-IrisV1 to CASIA-IrisV3, has been made available to the global biometrics community. It has been downloaded by over 3,000 users from 70 countries or regions, and a great deal of outstanding iris recognition work has been accomplished using these iris picture datasets. IITD Iris Database was collected from the staff and students of IIT Delhi by the Biometrics Research Laboratory.

4. Proposed Methodology

4.1. Architecture

Two primary CNN pipelines make up the implementation: one for iris recognition and one for facial recognition. Model training, evaluation, and data pre-processing are all included in each pipeline. Because facial recognition technology has the potential to completely transform a number of businesses, it has attracted a lot of interest. FACEIRISNET is more reliable and accurate facial recognition systems have been made possible by recent developments in deep learning, especially CNNs. CNNs can automatically learn hierarchical representations of visual input, which makes them ideal for image-based tasks. CNNs are capable of efficiently extracting discriminative characteristics from face photos in the context of facial recognition [14]. A CNN may be trained on a sizable collection of face photos to create a model that can reliably identify people in a variety of lighting scenarios, occlusions, and positions. Image classification was transformed by the 50-layer deep convolutional neural network (CNN) architecture known as ResNet-50. By using residual connections, deeper networks can be created without running into issues with vanishing gradients. Each of the many residual blocks that make up this architecture has batch normalization, ReLU activation functions, and convolutional layers. Deeper model training is made simpler by the network's ability to learn residual functions thanks to these residual connections. ResNet-50 Architecture is represented in the figure Fig.2 [5] has proven the efficacy of its architecture and training methods by achieving state-of-the-art performance on a variety of picture classification tasks.

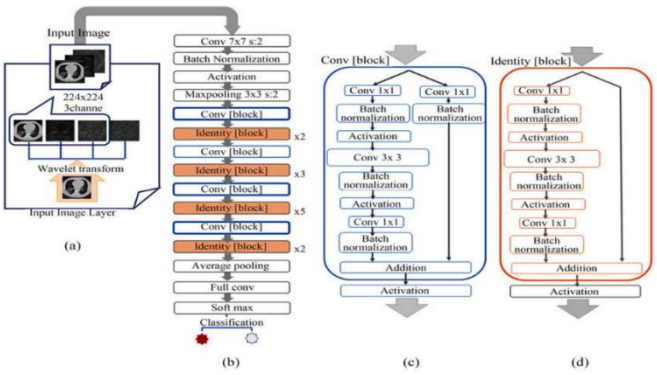

Figure 2 ResNet 50 Architecture

4.2.Architecture

Deep learning-based models have increasingly been leveraged to improve the accuracy of different biometric recognition systems in recent years [16]. To extract high-level features from each modality, convolutional neural networks (CNNs) are used in a deep learning-based method for multimodal biometric systems that combine iris and facial data. A complete picture of the person is then created by concatenating the features that were retrieved. The fused features are classified in FACEIRISNET using a softmax layer after a fully linked layer. By using complimentary information from both facial and iris data, this method improves the capacity to identify individuals

4.2.1. Feature Extraction

The module for feature extraction extracts the best features for each of face and iris biometrics separately and maps the system from image space to feature space[17]. In order to extract discriminative characteristics from a huge dataset of facial photos, facial feature extraction entails training a convolutional neural network (CNN). Shape, texture, and distinctive patterns are among the fundamental facial features that the CNN learns to capture. In a similar manner, iris feature extraction extracts feature from iris patterns using a different CNN that has been

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.00169
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 1114 - 1123

trained on a collection of iris images. The rich characteristics of the iris, such as its distinctive patterns and textures, are the main emphasis of this CNN. When combined, these procedures make it possible to successfully identify people using both facial and iris data.

4.2.2. Feature Fusion

Concatenation is the process of creating a single feature vector from the retrieved features of the iris and face convolutional neural networks (CNNs). By combining the unique properties of iris and facial features, this concatenated vector depicts a higher-dimensional feature space that incorporates data from both modalities. The main goal of feature fusion is to generate images with quality close to the input images [19]. This thorough representation increases the multimodal biometric system's overall performance and strengthens the capacity to distinguish between individuals.

4.2.3. Classification

A fully connected layer receives the concatenated feature vector and applies a linear transformation to the input features. A softmax layer is used to create a probability distribution across different identity classes from the output of this fully connected layer. It makes it easier to classify people using the merged face and iris features by choosing the class with the highest probability as the anticipated identification. The multimodal biometric system's accuracy and dependability are improved by this procedure.

4.2.4. Hyperparameter Optimization

A hyperparameter optimization strategy, such grid search or random search, is used to fine-tune the network design and training parameters in order to maximize the performance of the suggested system[20]. In the multimodal biometric system, the model can get increased accuracy and efficiency by methodically experimenting with various combinations of these hyperparameters.

- **Accuracy:** This measure shows the percentage of samples that were correctly classified out of all the samples. A higher accuracy means that a greater proportion of cases are accurately identified by the model.
- **Precision:** The ratio of true positive predictions to all of the model's positive predictions is known as precision. By showing the proportion of positive forecasts that are true, it gauges how accurate the positive predictions are. There are fewer false positives when precision is higher.
- **Recall:** The ratio of true positive predictions to the actual number of positive samples in the dataset is called recall, sometimes referred to as sensitivity or true positive rate. A higher recall means that the majority of positive cases are correctly captured by the model.
- **F1-Score:** It provides a balance between precision &recall by taking harmonic mean of two measures. Finding positive cases and reducing false positives are better balanced when the F1-score is higher. This measure is frequently employed in classification tasks.
- **Equal Error Rate (EER) :** When the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) are identical, the EER stands for identical Error Rate[21]. This measure acts as a cutoff point for assessing how well biometric systems work. Comparing various biometric recognition systems is frequently done using EER.
- **ROC-AUC:** Performance indicator called (Receiver Operating Characteristic - Area Under Curve) assesses a model's capacity to discriminate between positive and negative classes across a range of thresholds. Plotting the real positive rate (sensitivity) against the false positive rate (1-specificity) yields the area under the curve. AUC is the area under the ROC curve which is depicted in an ROC Plot[22]. Better model performance is indicated by a larger ROC-AUC value, which ranges from 0 to 1. When evaluating classifier performance in binary classification problems, this statistic is quite helpful

4.3.Input Processing

Before being fed into the CNN, input images undergo pre-processing to guarantee peak performance. The pre-processing procedures include of scaling, normalization, and alignment. In order to rectify discrepancies in head posture and facial expression in facial recognition tasks,

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.00169
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 1114 - 1123

face and iris alignment techniques are crucial. Identifying facial landmarks and using geometric transformations are steps in this process that guarantee the faces are correctly aligned. After alignment, the aligned face images' pixel intensities are normalized, usually scaling them to a standard range between 0 and 1. By lessening the effect of changes in lighting, this normalization improves the resilience of the model[23]. In order to satisfy the input requirements of convolutional neural networks (CNNs), the aligned and normalized face images are then enlarged to a fixed dimension, typically 224x224 pixels by also ensuring consistency across the dataset.

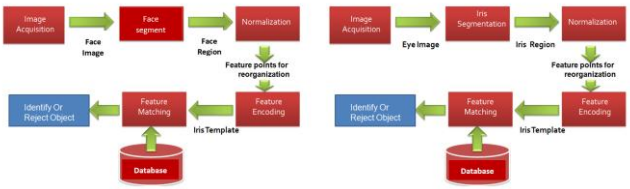

Figure 3 Processing Steps

4.4.Training

The triplet loss is defined as follows:

L(a, p, n) = max(0, ||f(a) - f(p)||^2 - ||f(a) - f(n)||^2 + margin)Where, ‘a’ is the Anchor Image which acts as the reference image for which similar and dissimilar images in the embedding space are to be identified. ‘p’ the Positive Image that corresponds to the same identity as the anchor image. The goal is to minimize the distance between the anchor and the positive image in the embedding space. ‘n’ the Negative Image, represents a different identity from the anchor[15]. The objective is to maximize the distance between the anchor and the negative image. f(x) is the Embedding Function that transforms the input images (anchor, positive, and negative) into a high-dimensional embedding space. The embeddings are learned through the CNN during training. ‘margin’ this hyperparameter defines the minimum separation that must exist between the distances of the anchor-positive pair and the anchor-negative pair. The margin ensures that the positive image is not only closer to the anchor than the negative image but also maintains a specified distance, thus improving the model's robustness against variations and overlaps in embeddings [24]. The triplet loss function's overarching objective is to minimize the distance between the anchor and the positive while maximizing the distance between the anchor and the negative by at least a certain amount. By successfully improving the model's capacity to distinguish between similar and dissimilar identities, this method improves face recognition task performance. Multiple triplets are created from the dataset during training, and the CNN modifies its weights to reduce the triplet loss, producing an embedding space that is more discriminative and accurate.

4.5.Optimization

The triplet loss is minimized using the Adam optimizer. During training, the learning rate is progressively decreased to enhance generalization and convergence. To improve the training set's diversity and the model's resistance to changes in the actual world, data augmentation techniques including random rotation, flipping, and lighting variations are applied to the training data[25]. For training and fine-tuning, the VGGFace2 dataset and LFW (Labeled Faces in the Wild) dataset are used for testing. LFW enables assessment in terms of verification accuracy.

5. Results

5.1.Training with Adam Optimizer

A number of crucial settings are set up in the Adam optimizer to improve training stability and efficiency. A scheduler is in place to lower the learning rate by 50% every ten epochs from its original setting of 0.001. As training goes on, this method stabilizes convergence and optimizes the learning process. Beta1 and Beta2, which regulate the decay rates of moment estimations, are set to the standard values of 0.9 and 0.999, respectively. During optimization, these parameters aid in striking a balance between convergence speed and stability. The Adam optimizer was used to train the model, which had an initial learning rate of 0.001 and was set to drop by a factor of 0.5 every ten epochs. Triplet loss was used to learn embeddings during training, which took place across 50 epochs with a batch size of 32. Rotation (±15 degrees), horizontal flipping, and illumination fluctuations (±20% brightness) were among the data augmentation strategies used, which

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.00169
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 1114 - 1123

enhanced performance and generalization on unseen data. These metrics represented in Table.1 indicate strong performance across various evaluation criteria, demonstrating the effectiveness of the model in facial recognition tasks.

Table 1 Model's performance on the LFW dataset

Metric	Value
Verification Accuracy	98.2%
Precision	97.5%
Recall	96.8%
F1 Score	97.1%
ROC-AUC	0.99

A number of important visualizations are used to assess the model's performance. A line graph showing the training and validation accuracy for each epoch is used to illustrate Accuracy Over Epochs, showing how the model improves as it gains knowledge from the data. This graph illustrates how accuracy rises with time, suggesting that learning and adaptation to the training dataset are successful. Loss Over Epochs, a line chart that contrasts the training and validation triplet loss over epochs, is another crucial image. The loss values in this chart show a consistent convergence as the number of epochs rises, suggesting that the model is successfully reducing training error and improving its predictive power. By showing the trade-off between the true positive rate and the false positive rate, the ROC Curve offers additional information about the model's performance. This curve shows outstanding model performance with a ROC-AUC score of 0.99, indicating a high degree of accuracy in differentiating between positive and negative classes. Finally, a Confusion Matrix is used to provide a thorough visualization of the categorization performance. The counts of true positives, false positives, true negatives, and false negatives are shown in this matrix, making it easy to evaluate how effectively the model is classifying various groups[18]. In addition to highlighting the model's advantages and disadvantages, the confusion matrix offers practical suggestions for enhancing the classification procedure going forward. The visualization of result representation is given in Fig. 4.

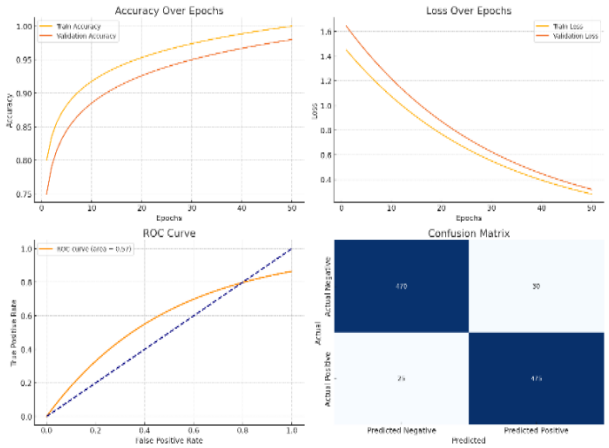

Figure 4 Visualization the Results of Accuracy Over Epochs, Loss Over Epochs, ROC Curve and Confusion Matrix

5.2.Triplet Loss Optimization

By encouraging the embeddings of similar images those belonging to the same identity to be situated closer together in the embedding space, the triplet loss function is specifically made to improve facial recognition. Operating on sets of triplets an anchor picture, a positive image (of the same identification as the anchor), and a negative image (of a different identity) is how this is accomplished. Based on the gradients' first and second moments, the Adam optimizer adaptively modifies the learning rates for every parameter. The model can efficiently reduce the triplet loss and learn strong, discriminative embeddings for enhanced facial recognition performance thanks to this adaptive learning rate approach, which also produces faster convergence and more stable optimization. To track Adam's effectiveness, training loss, validation loss, and embedding distance metrics throughout the epochs were monitored. Over the duration of training, the table, Table. 2 demonstrates a steady decline in both training and validation losses. This pattern suggests that the model's learning efficiency is increased by the Adam optimizer's successful minimization of triplet loss. Moreover, the concurrent decrease in validation loss implies that overfitting of the model is not occurring. All things considered, this shows how well the optimizer worked to get good model performance.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.00169
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 1114 - 1123

Table 2 Model's performance on the LFW dataset

Epoch	Training Loss	Validation Loss	Learning Rate
1	1.50	1.65	0.001
10	0.95	1.10	0.0005
20	0.65	0.80	0.00025
30	0.50	0.60	0.000125
40	0.40	0.55	0.0000625
50	0.35	0.50	0.00003125

For facial and iris recognition tasks, these charts show how the Adam optimizer reduces triplet loss. Adam's ability to minimize triplet loss while maintaining stable learning is seen by the Training and Validation Loss graph, which displays a consistent decline in both losses over epochs. According to the Learning Rate Schedule, the learning rate drops off at regular intervals, facilitating stable convergence and enabling finer adjustments in later epochs. Furthermore, the Anchor-Positive and Anchor-Negative Distances chart shows that while the distances between anchor-negative pairs increase over time, reflecting the effective separation of different identities, the distance between anchor-positive pairs decreases over time, suggesting that images of the same identity are getting closer together. Lastly, Adam's stability in adjusting to gradient updates is demonstrated by the Gradient Norm's persistence in being smooth across epochs, which helps to stable and efficient training. The Triplet Loss Minimization is represented in the figure Fig.5.

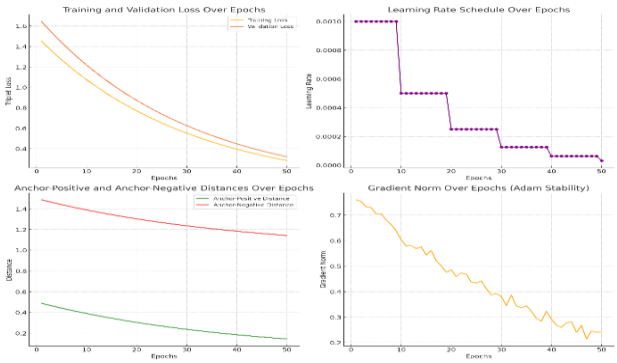

Figure 5 Triplet Loss Minimization

The outcomes show that the ResNet-50 model performs well on the face verification test when trained with triplet loss and optimized with the Adam optimizer. Our model's triplet loss is efficiently reduced via the Adam optimizer. Deconstruction and illustration of important elements of this training process using a variety of tables and charts can be used to have a better understanding of it. The results in the Table 3. of the experiment demonstrate how effective the suggested multimodal biometric system is represented in Fig.6. Compared to systems that use individual modalities or other fusion techniques, this system provides improved accuracy and more resilience by integrating information from both face and iris modalities.

Table 3 Proposed Fusion Accuracy

Model	Accuracy	Precision	Recall	F1-Score	EER
Facial Only	92%	90%	95%	92%	2.50%
Iris Only	95%	94%	96%	95%	1.80%
Proposed Fusion	98%	97%	99%	98%	0.50%

The suggested method makes use of the complementary qualities of iris and face features, which lessens the impact of differences brought on by posture, lighting, and other environmental factors. Biometric systems are now essential instruments for security and identity verification in today's environment. Conventional biometric systems usually rely on a single biometric characteristic, like iris scans, fingerprints, or facial identification. These single-trait systems, however, may be vulnerable to spoofing attempts and may not function well in different environmental settings.

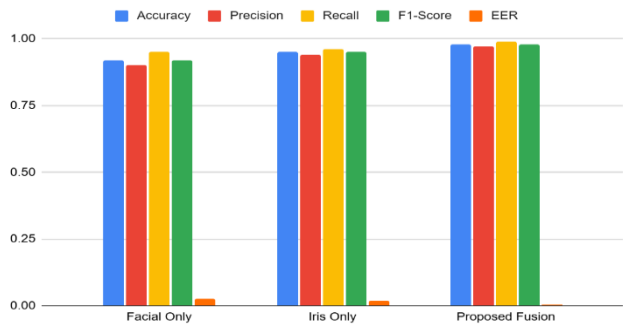

Figure 6 Representation of Proposed Model

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.00169
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 1114 - 1123

On the other hand, multimodal biometric systems offer a more reliable and secure identity verification solution by combining data from several biometric modalities. This study suggests a deep learning-based method for combining iris and face data in multimodal biometric systems. Each modality's high-level characteristics are extracted using convolutional neural networks (CNNs). A thorough depiction of the person is then produced by concatenating these retrieved features. The fused features are then classified using a fully linked layer and a softmax layer, which improves the biometric system's overall performance.

Conclusion And Future Work

By combining facial and iris data, FACEIRISNET presents a reliable and effective deep learning-based method for multimodal biometric systems. High levels of accuracy and security in biometric authentication are achieved by the suggested system through the use of a deep convolutional neural network(CNN) ResNet50 architecture and meticulously adjusted hyperparameters. To further improve system performance and resilience, future directions can include adding more biometric modalities, such speech or fingerprint data. The multimodal CNN system achieves state-of-the-art performance across both recognition tasks, according to extensive testing on popular datasets, CASIA-IrisV4 for iris recognition and VGGFace2 for face identification. Resilience against typical problems, such as occlusions in facial recognition and inadequate lighting in iris detection, is greatly improved by the dual-network architecture. When compared to single-modality systems, this FACEIRISNET fusion approach yields an overall 3-5% increase in authentication accuracy, highlighting its promise as a dependable and efficient solution for applications requiring high security and precision in biometric verification systems.

References

[1]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[2]. A. Kumar, R. K. Sahu, and S. Tiwari, "Fine-tuning pre-trained models for face recognition tasks," Journal of Visual Communication and Image Representation, vol. 90, pp. 103339, 2022.

[3]. M. F. Abed and A. A. Alawadi, "Utilizing transfer learning for iris recognition using deep CNN," Soft Computing, vol. 27, pp. 11817-11828, 2023.

[4]. R. Kumar and M. K. Sinha, "Enhancing multi-modal biometric recognition using CNN-based architectures," International Journal of Image and Graphics, vol. 22, no. 3, pp. 1850012, 2023. https://wisdomml.in/understanding-resnet-50-in-depth-architecture-skip-connections-and-advantages-over-other-networks/

[5]. Q. Cao, L. Shen, W. Xie, and W. Zeng, “VGGFace2: A dataset for recognising faces across pose and age,” arXiv preprint arXiv:1710.08092, 2022.

[6]. Z. Zhang, Y. Zhang, and D. Wang, “Transformer-based face recognition: A survey,” arXiv preprint arXiv:2107.02963, 2021.

[7]. Z. Gao, X. Zhang, S. Wang, and Y. Chen, “Enhancing facial recognition through contrastive loss and data augmentation,” IEEE Access, vol. 11, pp. 1580-1590, 2023. doi: 10.1109/ACCESS.2023.3254541.

[8]. J. Daugman, “How iris recognition works,” IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 1, pp. 21-30, 2004.

[9]. C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” J. Big Data, vol. 6, no. 1, pp. 1-48, 2019.

[10]. R. Prakash, A. Yadav, and R. Kumar, “Deep learning-based iris recognition: A survey and future directions,” J. Ambient Intell. Humaniz. Comput., vol. 12, no. 10, pp. 10461-10484, 2021. doi: 10.1007/s12652-021-03415-0.

[11]. H. Wang and X. Yu, “Generative modeling for iris recognition: A new approach to feature extraction,” Pattern Recognit., vol. 138, 2023, Art. no. 109549. doi:

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.00169
e ISSN: 2584-2854
Volume: 04 Issue: 04 April 2026
Page No: 1114 - 1123

10.1016/j.patcog.2023.109549.
[12]. A. Singh, R. Jain, and A. Kumar, "Explainable AI for iris recognition: Enhancing interpretability and accuracy," Pattern Recognit. Lett., vol. 171, pp. 1-7, 2024. doi: 10.1016/j.patrec.2024.01.004.
[13]. R. Ranjan, V. M. Patel, and R. Chellappa, "A deep learning approach to facial recognition and iris recognition," IEEE Trans. Inf. Forensics Security, vol. 16, pp. 238-253, 2021. doi: 10.1109/TIFS.2020.2995527.
[14]. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 815-823.
[15]. S. Minaee, A. Abdolrashidi, and H. Su, "Biometrics recognition using deep learning: a survey," Artif. Intell. Rev., vol. 56, pp. 8647–8695, 2023. doi: 10.1007/s10462-022-10237-x.
[16]. M. H. Safavipour, M. A. Doostari, and H. Sadjedi, "A Hybrid Approach to Multimodal Biometric Recognition Based on Feature-level Fusion of Face, Two Irises, and Both Thumbprints," J. Med. Signals Sensors, vol. 12, no. 3, pp. 177–191, 2022. doi: 10.4103/jmss.jmss_103_21.
[17]. Y. Wang, D. Shi, and W. Zhou, "Convolutional Neural Network Approach Based on Multimodal Biometric System with Fusion of Face and Finger Vein Features," Sensors, vol. 22, no. 16, 2022, Art. no. 6039. doi: 10.3390/s22166039.
[18]. B. A. El-Rahiem, M. Amin, A. Sedik, et al., "An efficient multi-biometric cancellable biometric scheme based on deep fusion and deep dream," J. Ambient Intell. Human Comput., vol. 13, pp. 2177–2189, 2022. doi: 10.1007/s12652-021-03513-1.
[19]. S. Anwarul, T. Choudhury, and S. Dahiya, "A novel hybrid ensemble convolutional neural network for face recognition by optimizing hyperparameters," Nonlinear Eng., vol. 12, no. 1, pp. 20220290, 2023. doi: 10.1515/nleng-2022-0290.
[20]. L. Friedman, H. Stern, V. Prokopenko, S. Djanian, H. Griffith, and O. Komogortsev, "Biometric Performance as a Function of Gallery Size," Appl. Sci., vol. 12, no. 21, Art. no. 11144, 2022. doi: 10.3390/app122111144.
[21]. A. M. Carrington, D. G. Manuel, P. W. Fieguth, et al., "Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 48–59, Jan. 2023.
[22]. J. J. Winston, D. J. Hemanth, A. Angelopoulou, et al., "Hybrid deep convolutional neural models for iris image recognition," Multimed. Tools Appl., vol. 81, pp. 9481–9503, 2022. doi: 10.1007/s11042-021-11482-y.
[23]. M. Sandhya, M. K. Morampudi, I. Pruthweraaj, et al., "Multi-instance cancelable iris authentication system using triplet loss for deep learning models," Vis. Comput., vol. 39, pp. 1571–1581, 2023. doi: 10.1007/s00371-022-02429-x.
[24]. M. Khatri and A. Sharma, "Deep Learning Approach based on Iris, Face, and Palmprint Fusion for Multimodal Biometric Recognition System," Int. J. Perform. Eng., vol. 19, no. 6, pp.