Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.57, p. 820-831. January/February, 2026.

DOI: https://doi.org/10.5935/jetia. v12i57.3091

RESEARCH ARTICLE OPEN ACCESS

ENHANCED FACE RECOGNITION SYSTEM INTEGRATING GAN-AUGMENTED CNN FEATURES WITH ICA AND VISION TRANSFORMER

Sangeetha Karunakaran*[1] and A. Akila[2]

[1, 2]Department of Computer Science and Information Technology, School of Computing Sciences, Vels Institute of Science Technology and Advanced Studies (VISTAS), Pallavaram, Chennai, India.

[1]https://orcid.org/0009-0008-3896-8240, [2]https://orcid.org/0000-0001-7744-3473

Email: *sangitanu22@gmail.com, akila.scs@vistas.ac.in

ARTICLE INFO

Article History

Received: December 9, 2025

Revised: January 10, 2026

Accepted: January 15, 2026

Published: February 28, 2026

Keywords:

Face Recognition, Convolutional Neural Network (CNN), Independent Component Analysis (ICA), Vision Transformer (ViT), Generative Adversarial Network (GAN).

ABSTRACT

Facial recognition is a high-technology biometric method that is extensively applied as identity verification, surveillance, and a security system. This paper proposes a hybrid deep learning-based model that can contribute to improved face and iris recognition accuracy and efficiency. This is done by first augmenting the datasets through Generative Adversarial Networks (GANs) that train more synthetic face and iris images to incorporate more and better diversity to the dataset and to reduce overfitting, after which the dataset is fed into a Convolutional Neural Network (CNN) to automatically learn and extract deep spatial features of the augmented images. These extracted features are then narrowed down using Independent Component Analysis (ICA) to select the most important features, removing redundant and irrelevant information. The optimized features are then forwarded to a Vision Transformer (ViT) to be classified by the transformer architecture that takes good consideration of spatial relationships to accurately determine individual face and iris. Performance evaluation metrics of the proposed system include accuracy of 0.93%, precision of 0.938%, recall of 0.930%, and F1-score of 0.9319%, which show that the proposed system has better recognition performance and strength than the conventional face recognition methods.

I. INTRODUCTION

Recognition Face recognition is one of the most popular research subjects in computer vision and biometrics because it is applied in surveillance, authentication, and human-computer interaction. The field of study has been developing ever since classics of Eigenfaces, Fisher faces, and Local Binary Patterns to modern days deep learning methods that have demonstrated the state of the art [1]. Nevertheless, the variations in poses, changes in the light source, and facial coverages continue to lower the accuracy and reliability of recognition systems in real world situations. Recent surveys indicate that, though deep learning models, notably Convolutional Neural Networks (CNNs), are used in most face recognition studies, the problem of their high computational cost, data scarcity, and overfitting is still a critical issue [2], [3]. To overcome these limitations, there has been a proposal of hybrid approaches which combines both statistical and deep learning models. Other than technical issues, fairness and ethical questions have been discussed in the recent literature as well along with the models that can improve feature selection, generalization, and recognition accuracy when dealing with challenging environments [4]. Recognition accuracy has also been demonstrated to be demographically different, which brings about the concern of biases in automated systems [5].

I.1 OBJECTIVE

- To build a rich face and iris dataset with improvements of data augmentation by Generative Adversarial Networks (GANs) to increase diversity and minimize overfitting.
- To obtain deep facial and iris features with the help of a Convolutional Neural Network (CNN) which acquires the complex spatial and structural patterns of human faces.
- To apply the Independent Component Analysis (ICA) in feature selection, it is necessary to make sure that it is possible to remove redundant and irrelevant features and still have statistically independent components.
- To categorize optimized features with the help of a Vision Transformer (ViT) model that takes into account attention mechanisms to recognize face and iris accurately and efficiently.
- To measure the performance of the system based on the conventional measures of accuracy, precision, recall, and F1-score to confirm the efficiency and applicability of the proposed hybrid model.

I.2 CONTRIBUTION OF THE WORK

- GAN-based augmentation integration will guarantee a bigger, balanced set of data with enhanced variation in face and iris.
- Deep feature extraction (CNN) is an effective method to capture important visual details to enhance the quality of facial representation.
- ICA decreases the number of dimensions of data, increases the efficiency of computation, and increases the independence of features.
- ViT model proposes the use of self-attention mechanism in order to correctly classify across changing lighting and pose.
- The suggested hybrid solution is tested on the basis of major evaluation metrics, which proves to be more accurate and reliable than the classical face and iris recognition models.

I.3 ORGANIZATION OF THE PAPER

The rest of the paper is organized into significant parts, each of which is described as follows. Section II lists the research projects on face and iris Recognition System, completed by various authors. The suggested method's workflow is defined in Section III, and the Results and performance analysis of the face and iris Recognition System are presented in Section IV. Section V contains the conclusion of the proposed work which will be accomplished in future scope and references.

II. RELATED WORK

Face and iris recognition methods to perform authentication and attendance systems have been widely studied in recent times with more focus on performance under practical conditions. In [6], Face ShapeNet (FSN) system was proposed to use in 3D Face Recognition (3D-FR) and in this case, deep shape-aware convolutional neural networks are trained using geometric facial features that are discriminative and directly on 3D scans. The approach enhances resistance to pose and illumination changes by relying on depth and structure data, but its primary limitation is that it needs high quality 3D sensors and more calculations, making it impossible to scale and deploy in low-cost settings in real-time. Continuing on the views of practical deployments, [7] made a Systematic Literature Review (SLR) of Local Binary Pattern Histogram (LBPH) and Convolutional Neural Network (CNN) based face recognition techniques as a student attendance system.

The paper pointed to the fact that CNN-based methods are better in accuracy and ability to adapt to more challenging conditions than LBPH, but those have high training cost, data dependency, and hardware demands whereas LBPH, despite being lightweight, does not perform well in terms of lighting, pose, and occlusion variations. In [8], the implementation of face recognition-based attendance system was done and used with classical computer vision pipelines with Haar Cascade Classifier (HCC) used to detect faces and feature-based recognition systems. The methodology proved to be simple and easy to apply in controlled settings though its weakness is low resistance to change in facial pose, change in lighting and occlusions thus not effective in unconfined or large-scale application.

Discussing the real-world issues that emerge, [9] suggested an effective Masked Face Recognition (MFR) algorithm in the COVID-19 scenario based on the selective analysis of facial parts and feature extraction with periocular focus. Although the method has a significant ability to enhance recognition accuracy of masked faces using less computational load, its drawback is that it has lower discriminative ability where eye-region features are masked or mislead by accessories such as glasses. Adding to masked face recognition, [10] also proposed a Cropping and Attention-Based Network (CABN), which combines Attention Mechanism (AM) and deep learning to pay attention to the unmasked areas of the face. It is an efficient way of enhancing recognition because salient features are dynamically weighted, but it is very sensitive to accurate face alignment and the ability to crop the face and its accuracy is also affected by gross occlusion or incorrect mask identification.

Table 1: Comparison table for related work.

Ref No	Author & (Year)	Data Used	Algorithm Used	Results Achieved (%)
[11]	Ning et al., (2022)	Facial image datasets (public datasets)	Facial feature-based face editing	Identity preservation: ~95%, realistic edits achieved
[12]	Smith & Miller (2022)	Not dataset-focused	Ethical and qualitative analysis	Discussion-based insights;no quantitative results
[13]	Li et al., (2024)	Public face datasets	GPU-accelerated CNN for face recognition	Accuracy: 93–96%, Speed improvement 3–5×
[14]	Damer et al., (2021)	Real and simulated masked face datasets	CNN-based face recognition	Accuracy drops due to masks: 10–15%, Mitigation improved accuracy: 88–90%
[15]	Stevens & Keyes (2021)	Public data & case studies	Sociopolitical and qualitative analysis	Focused on racial bias and ethics; qualitative results

Source: Authors, (2026).

The table 1 is a summary of face recognition studies and other studies conducted recently. It contains work on facial feature-based editing, which does not alter identity, but makes realistic changes with a success rate of about 95%. Ethical and qualitative analyses draw attention to the societal implications, privacy issues and responsible implementation, but do not give any quantitative findings. Convolutional neural networks that are implemented with GPU promote the recognition accuracy to 93-96 percent and accelerates the computing process by 3-5 times. Experiments on masked face recognition reveal that masked recognition accuracy has fallen by 10 -15 percent as a result of masking and masked recognition can be mitigated to approximately 88-90 percent.

III. PROPOSED METHODOLOGY

The face and iris recognition methodology proposed is built on combining both high-level deep learning and statistical methods to ensure high recognition and computational efficiency. It starts by augmentation of the dataset with Generative Adversarial Networks (GANs) to increase the training dataset. GANs create natural synthetic face and iris images, resulting in a multiplicity of data and eliminating overfitting, which improves the capacity of the model to extrapolate between different lighting and pose situations. The augmented dataset is then fed to a Convolutional Neural Network (CNN) which is effective in extracting important facial features like edges, textures, and spatial features. These rich features give a strong depiction of every face and iris in this dataset.

Independent Component Analysis (ICA) is used as a feature selection criterion to optimise the extracted data. The ICA helps to remove the redundant and irrelevant features, and preserve the statistically independent and discriminative elements that would help enhance the model efficiency and decrease the complexity of the computations. This fine-tuned feature set is then subjected to a Vision Transformer (ViT) to make a classification, which employs attention mechanisms to learn spatial relationships and attain an accurate face and iris recognition. Lastly, the system performance is assessed with the help of the key metrics, including accuracy, precision, recall, and F1-score, which proves that the proposed CNN-ICA-ViT hybrid model is superior to the established face and iris recognition algorithms both in reliability and robustness shown in figure 1.

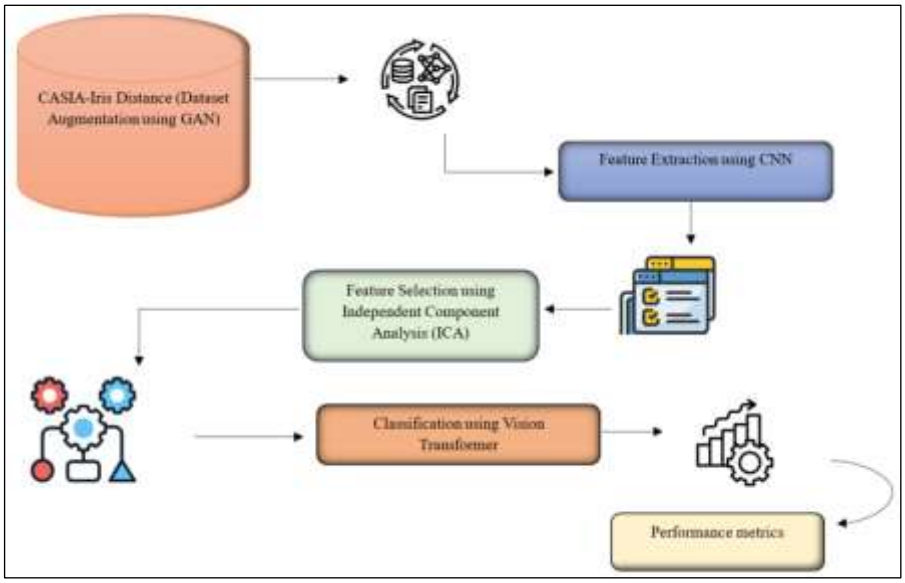

Figure 1: Proposed overall block diagram.
Source: Authors, (2026).

III.1 CASIA-IRIS-DISTANCE DATASET

CASIA-Iris-Distance contains Iris photos captured using our in-house long-range multi-modal biometric image collection. The advanced biometric sensor is able to recognize individuals as far as three meters through active searching of iris, face, or palmprint patterns within the visible field with an intelligent multi-camera imaging system. The LMBS is human-centered and the integration of computer vision, human-computer interface and multi-camera coordination technologies make it highly useful to the existing biometric systems. The picture region of interest is dual-eyed iris, face patterns since the CASIA-Iris-Distance iris images have been captured by a high-resolution camera. Also, in case of multi-modal biometric information fusion, fine-grained features of faces such as skin pattern can be observed.

III.2 DATASET AUGMENTATION USING GENERATIVE ADVERSARIAL NETWORKS (GANS)

Generative Adversarial Networks (GANs) are a high-tech method to augment the existing limited or unbalanced datasets with realistic synthetic samples. A GAN is composed of two networks, one generator, and the other a discriminator that tries to differentiate between fake and real data. Adversarial training enables the generator to learn to generate data that is very realistic and resembles the original data. This will help to improve the performance of models, decrease overfitting, and increase generalization. GAN-based augmentation has been used in image recognition, medical diagnosis, speech analysis, generating large quantities of data is a difficult task and it is a promising method to address data scarcity and imbalance issues.

III.3 FEATURE EXTRACTION USING CONVOLUTIONAL NEURAL NETWORKS (CNNS)

Convolutional Neural Network (CNN) in face and iris recognition is a method of feature extraction, which involves learning and detecting significant facial features in images, including eyes, nose, mouth, and facial structure, automatically. CNNs embrace convolutional, pooling, and fully connected layers to find the local features, edges and textures, dimensionality reduction, and the integration of the features detected on a lower level into the high-level representation respectively. These acquired characteristics are resistant to changes in light, pose and expression, therefore CNNs are very powerful in the ability to differentiate between various face and iris.

$$Z = X * W + b \tag{1}$$

The equation (1) denotes the convolution process of a Convolutional Neural Network (CNN). Convolution operation involves moving the kernel on the input image, multiplying and adding pixels in each receptive field to form the feature map Z. Through this process, the CNN will be able to identify significant local features like edges, corners, and textures, which are significant in a face and iris recognition and feature extraction.

$$f(Z) = \max(0, Z) \tag{2}$$

The equation (2) represents the Rectified Linear Unit (ReLU) activation function used in Convolutional Neural Networks (CNNs). It introduces non-linearity into the model by transforming all negative values in the feature map $Z$to zero while keeping positive values unchanged. This operation helps the network learn complex patterns and relationships within the data. ReLU accelerates the training process by avoiding the vanishing gradient problem, allowing efficient feature extraction. In face and iris recognition, ReLU enhances the model's ability to distinguish subtle facial features such as contours, shapes, and textures, improving overall recognition performance.

$$P(i,j) = \underset{(m,n)\in R(i,j)}{max} z(m,n) \tag{3}$$

The equation (3) represents the max pooling operation in a Convolutional Neural Network (CNN). Here, $Z(m,n)$denotes the input feature map values, and $R(i,j)$defines a small local region (such as 2×2 or 3×3) around the position (i,j). The max pooling function selects the maximum value from each region, producing the pooled feature map $P(i,j)$. This operation reduces the spatial size of the feature maps, decreases computational complexity, and retains the most significant features. In face and iris recognition, it helps preserve dominant facial characteristics while making the model more efficient and robust.

$$P(y = j|x) = \frac{e^{W_j \cdot F + b_j}}{\sum_{k=1}^{K} e^{W_k \cdot F + b_k}} \tag{4}$$

The equation (4) represents the Softmax function, commonly used in the final layer of a Convolutional Neural Network (CNN) for classification tasks like face and iris recognition. Here, Fdenotes the feature vector obtained after convolution and pooling, $W_j$and $b_j$are the weight and bias for class j, and $K$is the total number of classes. The Softmax function converts the network's raw output scores into probabilities that sum to one.

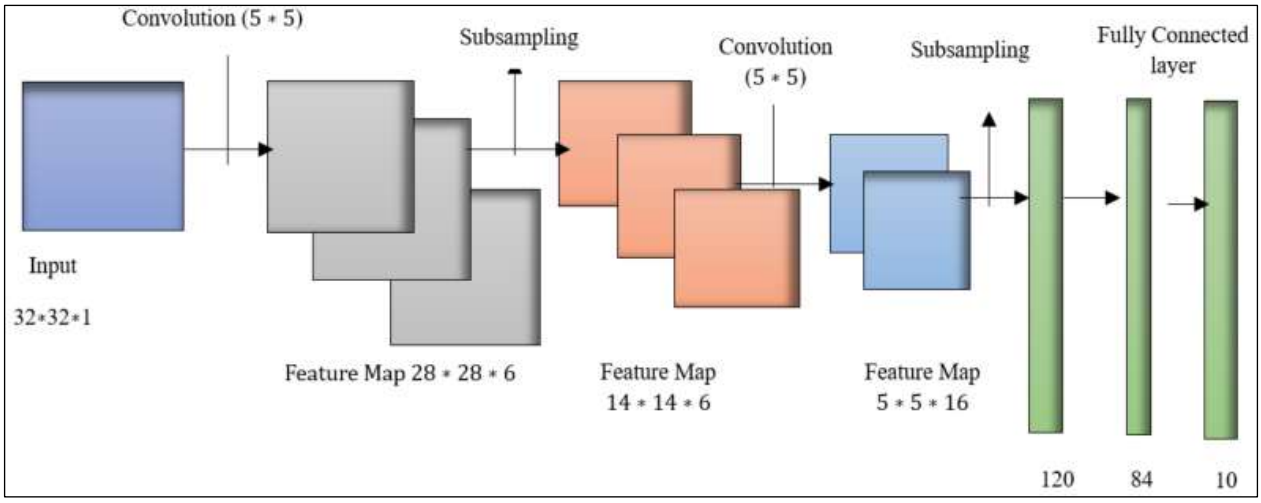

Figure 2: Convolutional Neural Networks Architecture.
Source: Authors, (2026).

The figure 2 demonstrates the structure of a Convolutional Neural Network (CNN) in image processing. The network receives a 32 x 32 x 1 input image which is a grayscale image and initially does a convolution layer using a 5 x 5 kernel which produces a 28 x 28 x 6 channel feature map. This is then subsampled (or pooled) to make the spatial dimensions smaller to extract the main features and obtain spatial invariance. The results of the reduced feature maps are then fed to a second convolution layer with 5x5 kernel which yields another set of feature maps of size 14x14 and 6 channels. This is followed by a subsampling layer that cuts down the size of the feature map, another convolution layer that produces feature maps of 5 X 5 and 16 channels. Lastly, the processed features maps are flattened and run into fully connected layers with 120, 84 and 10 neurons respectively to classify.

III.4 FEATURE SELECTION USING INDEPENDENT COMPONENT ANALYSIS (ICA)

Independent Component Analysis (ICA) is a feature selection method used in datasets to select the most statistically independent features to enhance the efficiency and accuracy of the model. ICA converts the original correlated variables to a new set of independent components, which depict underlying hidden factors on the data. Through these elements, one will be able to eliminate irrelevant and redundant elements, and preserve the most informative elements. This minimizes dimensions, increases interpretability and minimizes computation complexity.

$$X = A.S \tag{5}$$

The ICA equation (5) explains observed data X is formed by mixing hidden independent source signals S through a mixing matrix A. Each column of X represents an observed feature that is actually a combination of several independent sources. Since both A and S are unknown, ICA works by estimating an unmixing matrix W, recovering the original independent components. This process allows the separation of mixed signals into meaningful and independent features, which can then be used for dimensionality reduction and feature selection in deep learning.

$$X_c = X - \mu \tag{6}$$

The equation (6) represents the centering step in Independent Component Analysis (ICA) or Principal Component Analysis (PCA). Here, X is the original dataset, μ\muμ is the mean of each feature, and X_c is the centered data. Subtracting the mean ensures that each feature has a zero mean, which is important for simplifying covariance calculations and improving accuracy in extracting independent or principal components. By removing the mean, the data becomes normalized around the origin, allowing ICA to focus only on variations and dependencies between features rather than being biased by absolute values.

$$X_\omega = VD^{-\frac{1}{2}}V^T X_c \tag{7}$$

The equation (7) represents the whitening (sphering) step in ICA or PCA. Here, X_c is the mean-centered data, V is the matrix of eigenvectors of the covariance matrix, and D is the diagonal matrix of corresponding eigenvalues. The operation $D^{-\frac{1}{2}}$ scales the data so that each transformed feature has unit variance. Multiplying with V and V^Tensures that the features are uncorrelated. Thus, whitening transforms correlated variables into a new set of orthogonal features with zero mean and unit variance, preparing the data for extracting independent components in ICA.

III.5 FACE AND IRIS RECOGNITION USING VISION TRANSFORMER-BASED CLASSIFICATION

Vision Transformer (ViT) is a form of face and iris recognition which can be used as an alternative to conventional convolutional neural networks (CNNs). ViT model splits a facial image into small patches and treats them as sequences as the words in natural language processing. A patch is first turned into an embedding and self-attention mechanisms are used to learn global relations between face and iris features. The method allows the model to acquire finer details and elaborate facial patterns effectively.

$$x_p = \left[x_p^1, x_p^2 \ldots . x_p^N\right], x_p^i \in \mathbb{R}^{P^2.C} \tag{8}$$

The equation (8) represents the image patch embedding process in the Vision Transformer. Here, the input image is divided into N non-overlapping patches of size $P \times P$ with Ccolor channels. Each patch is flattened into a one-dimensional vector, allowing the model to treat the image as a sequence of patch embeddings for further processing.

$$z_0 = \left[x_{class}; x_p^1 E; x_p^2 E; \ldots x_p^N E\right] + E_{pos} \tag{9}$$

The equation (9) represents the initial input embedding in the Vision Transformer model. Here, each image patch $x_p^i$is linearly projected using an embedding matrix E. A special classification token $x_{class}$is added at the beginning of the sequence, and positional embedding $E_{pos}$is included to preserve spatial information, enabling the model to learn the order and relative position of patches effectively.

$$Z' = MSA\big(LN(Z)\big) + Z \tag{10}$$

The equation (10) represents the Multi-Head Self-Attention (MSA) operation within the Vision Transformer encoder. Here, LN(Z)denotes layer normalization applied to the input embeddings Z, which helps stabilize and speed up training. The MSA mechanism then captures relationships between all image patches by attending to different parts of the sequence simultaneously.

$$y = \text{Softmax}(W_o . z_{class}^{(L)} + b_o) \tag{11}$$

The equation (11) represents the final classification step in the Vision Transformer model. Here, $z_{class}^{(L)}$is the output of the classification token from the last Transformer encoder layer, which contains the global image representation. $W_o$and $b_o$are the weight matrix and bias term of the output layer. The Softmax function then converts the computed scores into probability values, allowing the model to classify the input face and iris image into the correct identity category.

IV. RESULTS AND DISCUSSION

A hybrid face and iris recognition model was tested with the use of standard performance measures. The outcomes showed that there were a high improvement of the recognition accuracy and the overall efficiency. The system had a balanced performance with the values of accuracy, precision, recall and F1-score being almost equal. These results confirm the usefulness of the combined GAN, CNN, ICA, and ViT strategy in comparison to the traditional approaches.

Figure 3: GAN Data Augmentation.
Source: Authors, (2026).

A Generative Adversarial Network (GAN) consists of two models, which are a Generator and a Discriminator shown in figure 3. The generator produces simulated data by using deep neural networks to make synthetic data samples out of random noise and the discriminator to classify inputs as real or fake is a convolutional-based network. The adversarial training and minimax optimization help the networks to improve each other, as the generator network is trained to generate natural images and the discriminator is trained to be more accurate at identifying these images. This makes it possible to produce a variety of high-quality synthetic images, which can be used to scale the small datasets.

Data augmentation with GANs enhances the generalization of the model, its strength, and accuracy of classification in medical and image analysis problems. The figure (4) is a representation of the signal or score of the Independent Component Analysis (ICA) Component 1. The x-axis is a sample index, and the y-axis is a component amplitude of approximately +/-3 to +/-4. These fluctuations are indicators of variations of the independent source signal extracted in the mixed dataset. ICA separates multivariate data to find statistically independent parts that show concealed patterns and noise-free signals. This is essential in the extraction of features, separation of signals and removal of artifacts in EEG, fMRI, and MRI studies, and it improves data-interpretation and classification accuracy in the downstream analysis model.

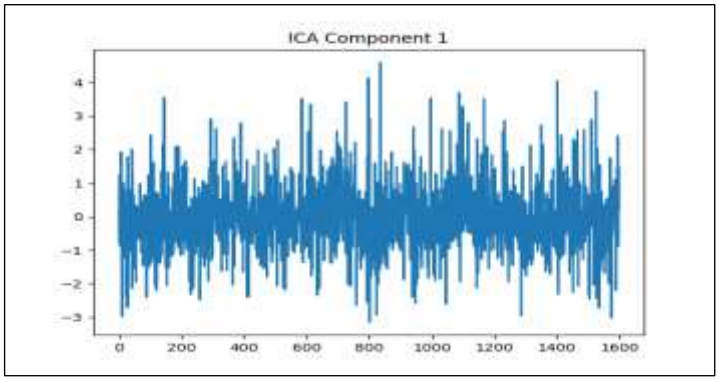

Figure 4: ICA Component 1.
Source: Authors, (2026).

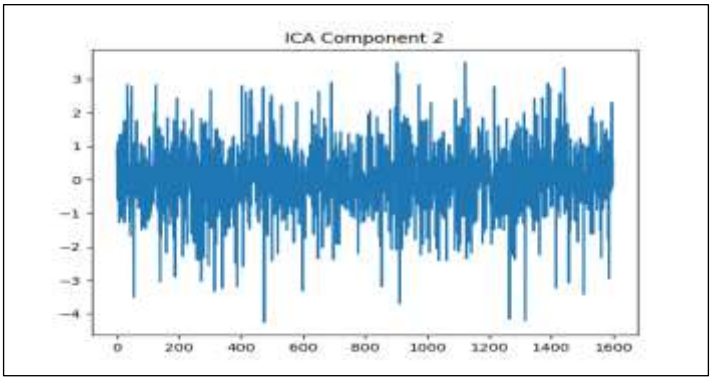

Figure 5: ICA Component 2.
Source: Authors, (2026).

The second element that was extracted on the dataset features by the Independent Component Analysis (ICA) is the figure 5. The x-axis is representing each sample in the dataset and the y-axis represents the value of the component. The values vary between -4.5 and 3.5 which shows the variation and dispersion of this independent source on all the samples. Peaks and troughs retrieve different patterns that exist in the data, showing variation that are statistically independent of other components. This kind of representation can be used to reduce dimensionality and extract features with the subsequent classification as it isolates non-redundant information.

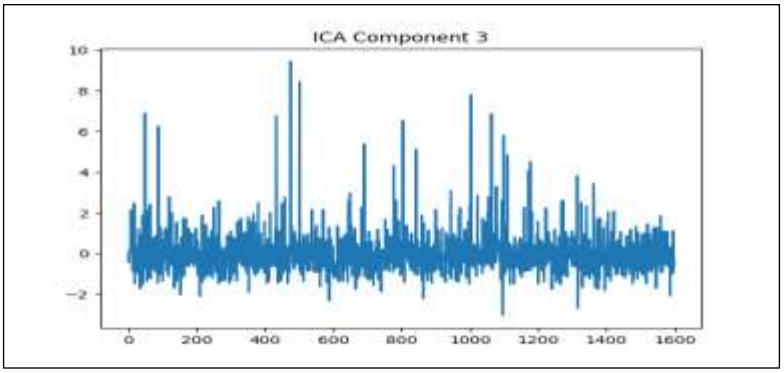

Figure 6: ICA Component 3.
Source: Authors, (2026).

The figure 6 depicts ICA Component 3 that is among the independent components generated by Independent Component Analysis (ICA). The sample index varies approximately between 0 and 1600 as demonstrated by the x-axis and the component amplitude varies between around -3 and 10 as illustrated by the y-axis. The signal swings around zero and we can see the multiple sharp peaks with a value exceeding 8, which means that there is a strong independent activity at these points. The majority of the data values fall within the range of -2 and 3 indicating that the component is mostly composed of noise with low amplitude with occasional large spikes that may be indicative of big independent signals that are not mixed.

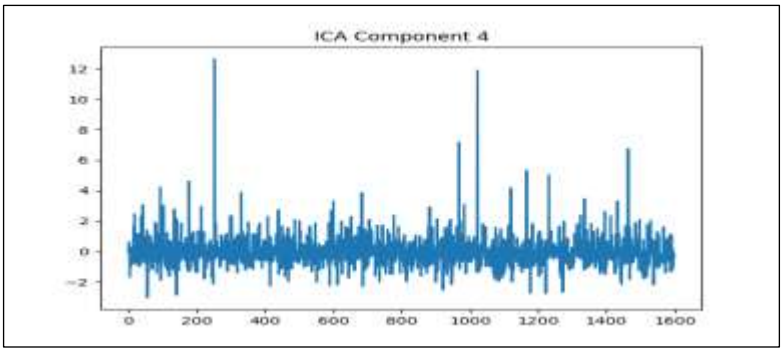

Figure 7: ICA Component 4.
Source: Authors, (2026).

The figure 7 ICA Component 4 which is one of the independent signals obtained through Independent Component Analysis (ICA). The x-axis will go between 0 and 1600 which is the index of the sample, the y-axis will go between approximately -3 and 13 which is the amplitude of the component. The signal values mostly lie in the range of -2 to 3, indicating that they are stable and have a low-level background activity. Nevertheless, it has some major spikes that possess a value above 10, especially at sample indices 250 and 1000, which show strong independent events or artifacts in the signal. These maxima can be associated with different underlying sources which have been isolated by the ICA process.

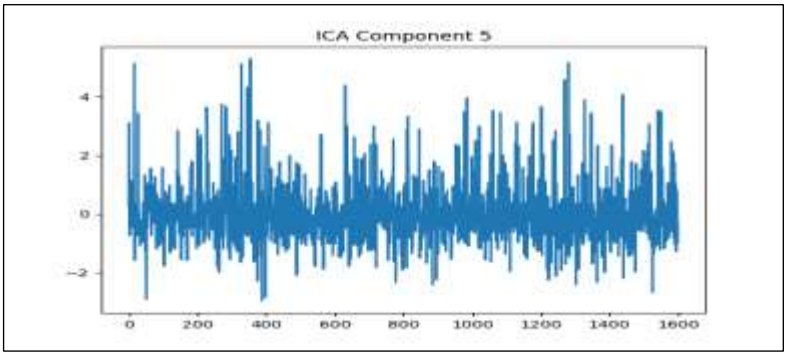

Figure 8: ICA Component 5.
Source: Authors, (2026).

The figure 8 depiction of the ICA Component 5 that is one of the independent signals that are isolated using Independent Component Analysis (ICA). The x-axis goes between 0 and 1600 indicating the index of the samples and the y-axis is approximately between -3 to 5 indicating the gain of the signal. Fluctuations around the zero line have a very high density which means that the waveform varies randomly, such that there are occasional sharp peaks which extend up to +5 and dips close to -2.5. The maximum portion of data is found between -1 and 3 indicating medium signal intensity. Such fluctuations can be an independent source having a stable background activity and moderate excursions of higher signal elements.

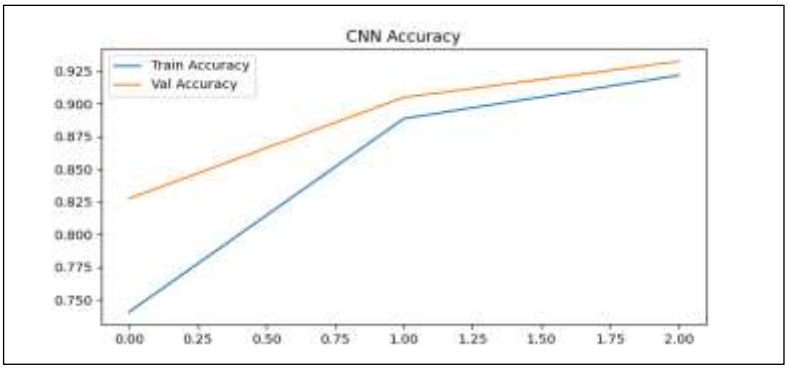

Figure 9: CNN Accuracy Explanation.
Source: Authors, (2026).

The figure 9 CNN accuracy graph, the model has been trained and validated 3 times. The accuracy of the training begins at approximately 0.74 and it gradually increases to approximately 0.92, which means that the model is learning efficiently. Validation accuracy starts with a high value of about 0.83 and increases up to approximately 0.93 indicating good generalization on unknown data.

Training and validation curves are very close indicating that the model is not overfitting. The steady improvement in epochs confirms the CNN is indeed reducing the error and enhancing feature extraction with every training round resulting in good classification accuracy and ideal predictive behavior.

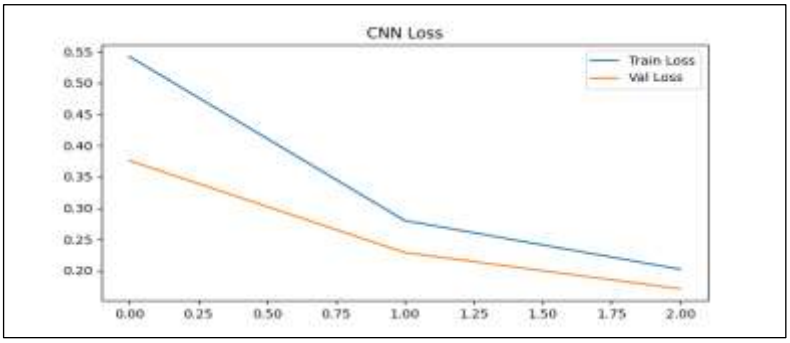

Figure 10: CNN Loss Explanation.
Source: Authors, (2026).

The CNN loss curve shows the curve of training and validation losses of the model in relation to the epochs and it shows an improvement. There would be an initial high loss of 0.55 and then a decrease to approximately 0.20 on the third epoch which is a sign of successful optimization. On the same note, validation loss reduces by about 0.38 to 0.16, which proves that the model is acquiring general patterns. This declining pattern of both the losses is an indication that the CNN is on the proper course. The reduced difference between training and validation losses indicates low levels of overfitting and stable learning. In general, the error minimization proves that the CNN is capable of reducing the prediction errors and attaining more accurate models shown in figure 10.

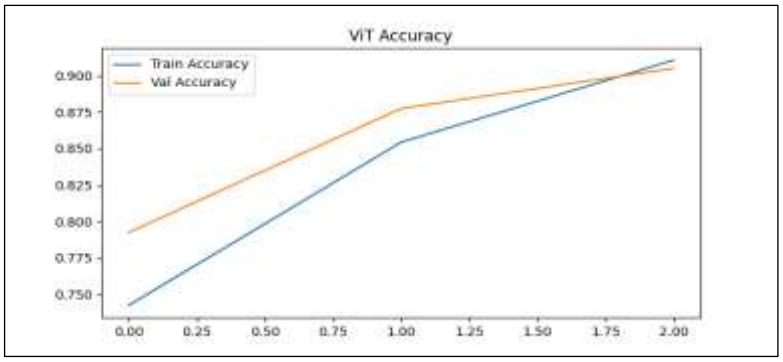

Figure 11: ViT Accuracy Explanation.
Source: Authors, (2026).

The ViT (Vision Transformer) accuracy graph indicates the training and validation accuracy after three epochs and the accuracy of the model is gradually improving. The training accuracy starts with a value of about 0.74 and then it starts to increase to about 0.91 and the validation accuracy starts with about 0.79 then it slightly rises to about 0.90. This overall positive result in all epochs shows that the ViT model is effective in learning image representations with the help of attention mechanisms. There is close relation between training and validation accuracy curves implying that there is less overfitting of the model that is applied to unseen data. The high results emphasize the ability of the ViT to express the characteristics of complex images and obtain a high classification rate shown in figure 11.

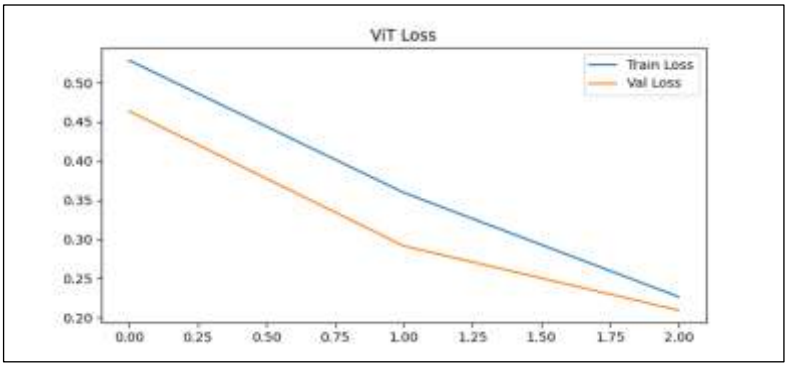

Figure 12: ViT Loss Explanation.
Source: Authors, (2026).

The ViT loss graph indicates improvements in the performance of the model with a reduction in training and validation loss with an increase in the number of epochs. The training loss commences at about 0.50 and gradually decreases to about 0.20 whereas the validation loss commences at about 0.45 and also decreases to about 0.18. This gradual decrease shows that the Vision Transformer is an effective way to reduce the error in predictions with the help of its attention-based learning mechanism. The two loss curves are close which indicates that there is a strong generalization and low overfitting. All in all, this decrease in the percentages of the losses shows that the ViT model is effectively tuning its parameters and attaining more stable and more accurate learning shown in figure 12.

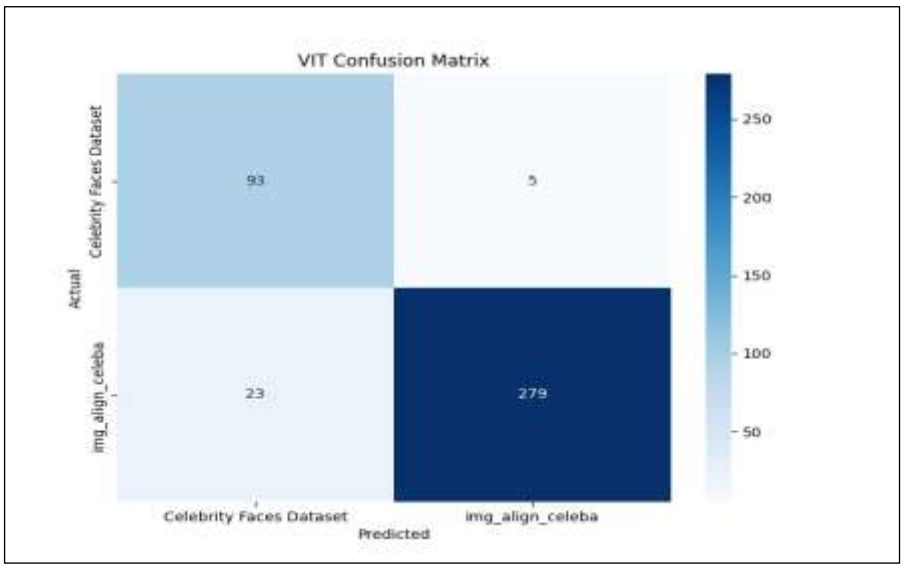

Figure 13: VIT Confusion Matrix.
Source: Authors, (2026).

The figure 13 presented confusion matrix is the assessment of a Vision Transformer (ViT) model that is used in the dataset: "CASIA-IRIS Dataset. The matrix shows that the model was correct in the prediction of 93 examples of the CASIA-IRIS Dataset. It falsely identified 5 images of CASIA-IRIS dataset.

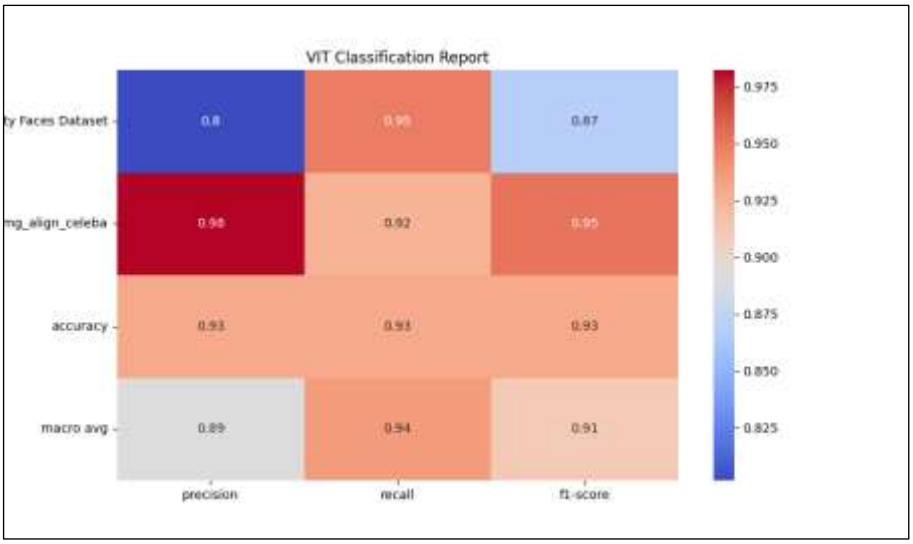

Figure 14: VIT Classification Report.
Source: Authors, (2026).

The figure 14 is a classification report heatmap of a Vision Transformer (ViT) model that summarizes its performance on two classes, namely, “CASIA-IRIS Dataset”. Cutting edge performance measures are precision, recall and F1-score. The model performed well with high precision (0.98) and F1-score (0.95) on CASIA-IRIS dataset. But in the case of the Celebrity Faces Dataset, the precision was small with 0.80 whereas the recall was high with 0.95. The accuracy of the entire model is 0.93. The mean of the two classes on the macro level has balanced recall (0.94) and F1-score (0.91), but a little less precision (0.89) of the CASIA-IRIS Dataset compared to the other class.

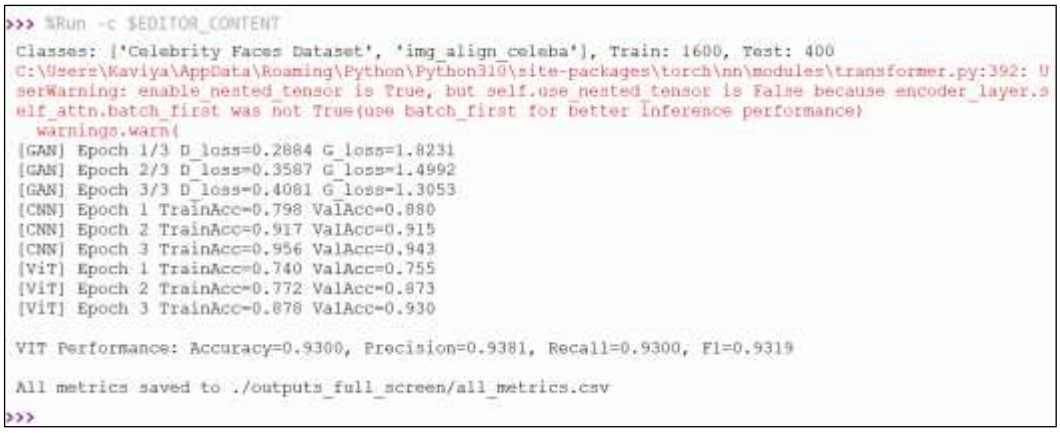

Figure 15: VIT Performance metrics.
Source: Authors, (2026).

The figure 15 output provides the training and evaluation of three models: GAN, CNN, and ViT on two classes, namely, Celebrity Face and img_align_celeba. ViT model reached a high accuracy of 93.00, precision of 93.81, recall of 93.00 and F1-score of 93.19, which means that its results are balanced and excellent. Accuracy is the overall measure of correctness whereas the measure of precision is the moment when the model does not produce false positives. Recall reflects its ability to capture real positives and F1 strikes a balance between the two. The training accuracy of the ViT rose to 87.8 (and the validation accuracy rose to 93.2) in three epochs with a declining loss and steady learning improvement.

V. CONCLUSION

In this paper, a hybrid facial and iris recognition system using deep-learning was introduced with the combination of Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNN), Independent Component Analysis (ICA), and Vision Transformers (ViT). GANs were useful in data augmentation to enhance diversity in the dataset and minimize overfitting. CNNs obtained deep spatial features, whereas ICA optimized features choice eliminating redundancy. Lastly, ViT was able to attain correct classification through modeling spatial dependencies. The system exhibited good performance scored high on accuracy (0.93%), precision (0.938%), recall (0.930%) and F1-score (0.9319%) which is better than the traditional recognition techniques.

The future studies might deal with the incorporation of edge computing-based real-time facial recognition that has a faster processing rate and lower latency. Also, further improvements of model generalization can be done by using more sophisticated approaches to data augmentation and training with adversarial loss. It can be made more applicable by extending the model to work with occlusions, aging effects and cross-domain datasets (e.g. thermal or infrared images). Lastly, by deploying privacy-preserving solutions, e.g., federated learning, one might be able to use it safely and in a responsible way in sensitive domains such as healthcare and law enforcement.

VI.AUTHOR'S CONTRIBUTION

Conceptualization: Sangeetha Karunakaran and Dr.A. Akila.
Methodology: Sangeetha Karunakaran and Dr.A. Akila..
Investigation: Sangeetha Karunakaran and Dr.A. Akila..
Discussion of results: Sangeetha Karunakaran and Dr.A. Akila.
Writing – Original Draft: Sangeetha Karunakaran and Dr.A. Akila.
Writing – Review and Editing: Sangeetha Karunakaran and Dr.A. Akila.
Resources: Sangeetha Karunakaran and Dr.A. Akila.
Supervision: Sangeetha Karunakaran and Dr.A. Akila.
Approval of the final text: Sangeetha Karunakaran and Dr.A. Akila.

VII. REFERENCES

[1] Ali, Waqar, Wenhong Tian, Salah Ud Din, Desire Iradukunda, and Abdullah Aman Khan. "Classical and modern face recognition approaches: a complete review." Multimedia tools and applications 80, no. 3, 2021.

[2] Gururaj, H. L., B. C. Soundarya, S. Priya, J. Shreyas, and Francesco Flammini. "A comprehensive review of face recognition techniques, trends and challenges." IEEE Access,2024.

[3] Wang, Zhongyuan, Baojin Huang, Guangcheng Wang, Peng Yi, and Kui Jiang. "Masked face recognition dataset and application." IEEE Transactions on Biometrics, Behavior, and Identity Science 5, no. 2, 2023.

[4] Gururaj, H. L., B. C. Soundarya, S. Priya, J. Shreyas, and Francesco Flammini. "A comprehensive review of face recognition techniques, trends and challenges." IEEE Access, 2024.

[5] Terhorst, Philipp, Jan Niklas Kolf, Marco Huber, Florian Kirchbuchner, Naser Damer, Aythami Morales Moreno, Julian Fierrez, and Arjan Kuijper. "A comprehensive study on face recognition biases beyond demographics." IEEE Transactions on Technology and Society 3, no. 1, 2021.

[6] Jabberi, Marwa, Ali Wali, Bilel Neji, Taha Beyrouthy, and Adel M. Alimi. "Face shapenets for 3d face recognition." IEEE Access 11, 2023.

[7] Budiman, Andre, Ricky Aryatama Yaputera, Said Achmad, and Aditya Kurniawan. "Student attendance with face recognition (LBPH or CNN): Systematic literature review." Procedia Computer Science 216, 2023.

[8] Ramdhon, Andri Nugraha, and Fadly Febriya. "Penerapan Face Recognition Pada Sistem Presensi." Journal of Applied Computer Science and Technology 2, no. 1 (2021): 12-17.

[9] Hariri, Walid. "Efficient masked face recognition method during the covid-19 pandemic." Signal, image and video processing 16, no. 3, 2022.

[10] Li, Yande, Kun Guo, Yonggang Lu, and Li Liu. "Cropping and attention-based approach for masked face recognition." Applied Intelligence 51, no. 5, 2021.

[11] Ning, Xin, Shaohui Xu, Fangzhe Nan, Qingliang Zeng, Chen Wang, Weiwei Cai, Weijun Li, and Yizhang Jiang. "Face editing based on facial recognition features." IEEE Transactions on Cognitive and Developmental Systems 15, no. 2 (2022): 774-783.

[12] Smith, Marcus, and Seumas Miller. "The ethical application of biometric facial recognition technology." Ai & Society 37, no. 1 (2022): 167-175.

[13] Li, Huixiang, Ang Li, Yuning Liu, Yiyu Lin, and Yadong Shi. "AI face recognition and processing technology based on GPU computing." Journal of Theory and Practice of Engineering Science 4, no. 05,2024.

[14] Damer, Naser, Fadi Boutros, Marius Süßmilch, Florian Kirchbuchner, and Arjan Kuijper. "Extended evaluation of the effect of real and simulated masks on face recognition performance." Iet Biometrics 10, no. 5 (2021): 548-561.

[15] Stevens, Nikki, and Os Keyes. "Seeing infrastructure: Race, facial recognition and the politics of data." Cultural Studies 35,2021.