

Multimodal Biometric Authentication Based On Finger Vein And Fingerprint Recognition Using Artificial Intelligence

Raghda M. Ibrahim^{1*}, Khamis A. Zidan² and Shereen S. Jumaa³

¹University of al-Iraqi, College of Engineering, Department of Computer Engineering, Iraq

²University of al-Iraqi, College of Engineering, Department of Computer Engineering, Iraq

³University of Technology, College of Artificial Intelligence Engineering, Department of Artificial Cognitive, Iraq

* Corresponding author's Email: Raghda M. Ibrahim

Abstract

As of now, the majority of multimodal biometric systems that utilize multiple modalities for biometric authentication utilize different approaches to train and tune deep learning models to obtain improved performance through training or fine-tuning the model to improve their performance. Many of these techniques can provide a high level of accuracy, but they also add high levels of computational complexity to the system, thus they may not work well in an environment that has limited training data available. The purpose of this study is to introduce a multimodal biometric authentication framework based on the use of pre-trained deep learning models to extract features from multiple modalities without using additional training or tuning to obtain the feature set. More specifically, the study uses VGG16, EfficientNetB0, and ResNet50 as feature extraction methods. In addition, an attention mechanism will be used to increase the effectiveness of the most informative features and will then be reduced by using Principal Component Analysis (PCA) while preserving 95% of the variance. K-Nearest Neighbors (KNN) will be used to classify the reduced features for finger vein recognition, while Random Forest will be used to classify the reduced features of fingerprint recognition. The experimental framework was developed using the Python programming language, and the results were 98.3% recognition accuracy for finger vein and 98.9% recognition accuracy for fingerprint. Finally, the fingerprint and finger vein modalities were fused at the decision level for highly accurate authentication performance. The proposed method shows that high accuracy of recognition can be obtained with low computational complexity using lightweight and interpretable techniques.

Keywords: Multimodal Biometric Authentication, Finger Vein, Fingerprint Recognition, Artificial Intelligence.

1. Introduction

The research community has been particularly interested in exploring the capability of multimodal authentication systems to enhance personal identification. By fusing together two or more different biometric traits, such as fingerprints and finger veins, with complementary characteristics of each biometric, multimodal systems can improve the overall reliability of recognition, thereby reducing the chance of an individual being falsely accepted or rejected. [1]

In contrast, unimodal biometric systems rely solely on one form of biometric trait and face various obstacles, including the effects of noise, intra-class variability, and being susceptible to spoofing attacks. As a result, there are limitations to the applicability of unimodal systems in real-world applications. [2]

The recent advances made in convolutional neural networks (CNNs) have permitted automatic extraction of discriminative features, and, therefore, improved performance levels in terms of recognition of biometric traits. Nevertheless, CNNs normally require a very large amount of labelled training data to either train or fine-tune the CNN model, which can be resource-intensive and is therefore often impractical for use in contexts where little training data is available. [3] [4] [5]

Numerous applications are characterized by small or limited dataset sizes, which results in traditional “deep-learning methods” being less effective with respect to overfitting and generating generalizable models. To test the “performance of systems.”

Under small dataset conditions, we used Virtual datasets to generate the “equivalent training data” in order to conduct preliminary tests [6][7][8][9].

Using a combination of fingerprint and finger vein modalities, this paper describes a “multimodal biometric system,” consisting of two image modalities (fingerprint and finger vein) derived from different sources (one pre-trained deep learning model for each modality). All feature extraction for this system has been done without any additional training on either of the two source images (i.e., no further training is required). The shirt features were compared using “similarity-based matching” techniques, whereas decision-based fusion techniques have been used to incorporate data from both sources. This approach will allow for accurate recognition, reduced processing costs, and interpretable results with very small datasets, while still providing sufficient information that can be utilized to deploy the system in real-world situations

2. Methodology

This work uses both a fingerprint and finger vein operator systems under one umbrella-type authentication process to help reinforce the security issuance of your fingerprint. The data for both operators will also be extracted using CNNs as an initial deep feature extraction, followed by enhancements and fusions of the features to produce greater discriminative power than either system

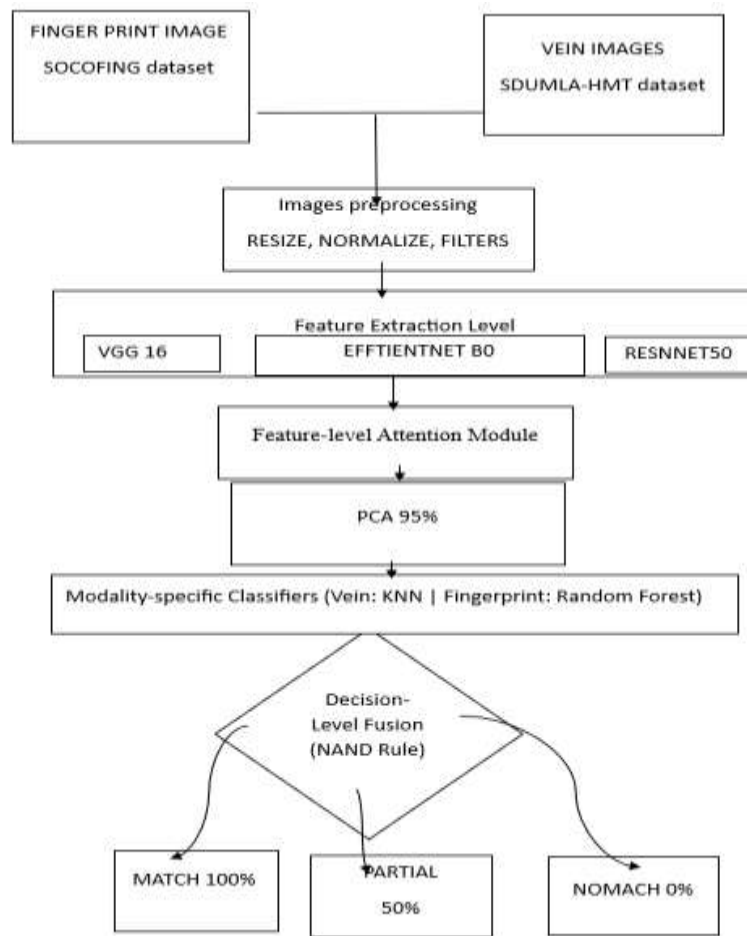


Fig. 1 Creative design methodology

2.1 image collection

The method involved in this study is a balanced dataset created from virtual pairing by the SOCOFing and SDUMLA-HMT databases using a one-to-one approach, which provides each of the 106 young people with a unique identifier to eliminate the challenges of having more than one multi-modal images for each user, which added up to a total of 1060 images for each of the modalities, was performed using an 80-to-20 split. More specifically, for feature extraction and construction of the reference gallery, 848 sample images were used, and then 212 sample images were used only in the final evaluation stage to determine the effectiveness of the system

2.2 Image processing

The pre-processing of images is vital to creating output data that is standardized for input as well as to enhancing the discriminating attributes associated with bio-signals (specifically:

fingerprints and finger veins). Once the output of the fingerprint image pre-processing pipeline had been acquired, the next steps for fingerprint images were to apply two forms of image pre- processing: normalizing and segmenting the images. Both of these pre-processing techniques serve to separate the parts of the image that contain the fingerprints (i.e. what one can consider to be the "foreground") from those parts of the image that do not contain any fingerprints (i.e. what one might refer to as the "background") [9]. The next step was then to improve the ridge-valley patterns within fingerprint images with the use of median filtering (removing noise) and Frangi filtering (improvement of image structures found at various scales which resemble blood vessels) [10]. The detection of the "minutiae" areas of the fingerprint images was done using the "difference of Gaussians" (DoG) approach, while the contrast of the fingerprint images had been equalised with the use of contrast limited adaptive histogram equalisation (CLAHE) to maintain a more stable local image contrast.[10].

To eliminate acquisition noise in the finger vein image acquisition process, Gaussian and median filters were applied, whereas Gamma correction and power law functions enabled for correction of the non-linear characteristics of the lighting used when capturing the images; thus, the background noise was reduced through the removal of reflections caused by glossy surfaces due to sunlight. Then, the vascular patterns of each finger vein image were enhanced through CLAHE so there would be contrast between the vascular pattern of the finger veins and surrounding soft tissue. Intelligent cropping algorithms were developed to automatically crop out unwanted areas of the finger vein image based on the shape of an individual's finger. Finally, both fingerprint and finger vein images were globally normalized to provide a common scale of measurement as input into an analysis using the VGG16-Attention deep learning model.[11].

2.3 Feature Extraction for Fingerprints and Finger Vein

Biometric image recognition relies on feature extraction to generate high-dimensional vector representations of identity attributes [12]. In this investigation, we employed three off-the-shelf pretrained Convolutional Neural Network (CNN) models VGG16, ResNet50, and EfficientNet to gather fingerprint and finger vein data. We did this by extracting features from these models without retraining them from scratch.

2.3.1 Fingerprint Feature Extraction

Using convolutional filters at multiple levels to analyses both the topography of ridge patterns beyond just the flat plane associated with fingerprints provides the capacity to calculate and classify the orientation of fingerprint ridges across the different layers in the model. The second set of parameters generated from the model is based on the relative positions and angles of the ridges to each other as well as on all other previously measured and recorded data on the fingerprints. A complete representation of minutiae is constructed as a result of the convolutional analysis of images of fingerprints through the different levels of the model. [12].

The filtering process using convolutional filters can essentially be represented mathematically through the following equation

$$G(x,y) = \sum_{i=-k}^k \sum_{j=-k}^k I(x-i,y-j) \cdot K(i,j) \quad (1)$$

[13]

- $I(x, y)$ represents the input fingerprint image.
- $K(i, j)$ is the filter kernel applied to the image.
- $G(x, y)$ is the resulting image after convolution.

- The convolution of an image enhances ridge patterns, highlighting fine fingerprint details that can be used for discrimination purposes.

ReLU activation is applied to retain meaningful ridge features:

$$f(z) = \max(0, z) \quad (2)$$

[13]. allows only positive activations to pass, helping the network focus on meaningful features

2.3.2 Extraction of Features from the Finger-Vein Modality

Researching the physical geometry of the vascular network is vital to exploring the finger vein area by relating how veins appear (generally beneath various levels of dermal tissue) and because they are not very distinct from the surface of the skin. The hierarchical arrangement in the CNN architecture takes advantage of this hierarchical organization to classify different sizes of vascular branches and the relationships among vascular branches.[15]

$$F_l = f(I * K_l \dots \dots) \quad (3)$$

[14] F_l represents the feature map at layer l . A convolution operation with K_l is performed to extract important features of a vein-patterned image, including branch width and orientation. The application of an activation function can highlight important features of the image and suppress unimportant features.

ReLU ($f(z) = \max(0, z)$) removes irrelevant signals [13]

The output is a Blood Vessel Map highlighting vein structure

2.3.3 Spatial Feature Representation

At block 5, the last convolutional layer is applied with regard to the two biometric modalities (fingerprint & finger-vein). Each has produced a 3-dimensional feature map by this point.:

$$F \in \mathbb{R}^{H \times W \times D} \quad (4)[15]$$

Each modality also has a spatial size ($H \times W$; two-dimensional) and associated depth (D ; number of feature channels). These outputs, together, create a representation of the spatial distribution or arrangement of biometric characteristics for each modality over the corresponding image

$$P_{m,n} = \max_{(i,j) \in R_{m,n}} F(i,j) \quad (5)[15]$$

The local pooling area in an image is represented by $R_{(m,n)}$. By performing this local pooling operation on every pixel in the image (i.e., a pixel's $R_{(m,n)}$ region), we gain robustness against small translations or rotations of a finger that may occur during acquisition.

3. Global Feature Aggregation (GAP)

Global Average Pooling (GAP) is a method to generate a compact, fixed-sized representation of the spatial feature maps (called a "fingerprint vector" and a "finger-vein vector") by reducing them to a single value

$$VC_c = \frac{H \times W}{\sum_{i=1}^H \sum_{j=1}^W F(i,j,c)} \quad (6)[16].$$

The value of pixel located (i,l) on channel c is denoted as $F(i, l, c)$. Thus, using this method will create vectors that are not dependent on their spatial location, and so the Fingerprint

and Finger-Vein Vectors will give more emphasis to the existence of the biometric patterns rather than their location.[16].

4. Attention Mechanisms

4.1. Fingerprint Modality (Dynamic Attention)

An intelligent gating system, known as a dynamic attention mechanism, improves features at three hierarchy levels: Channel Attention: Channels with the greatest informational importance, as identified by using Global Average Pooling and Global Max Pooling, can represent filters that effectively model fingerprint ridge patterns.

$$\begin{aligned} \mathbf{M}_c(\mathbf{F}) &= \sigma \left(\mathbf{W}_1 \right. \\ &\quad \left. + \mathbf{W}_1(\mathbf{W}_0(\mathbf{F}_{\max}^c)) \right) \end{aligned} \quad (8)[17].$$

Spatial Attention: The area of the fingerprint image that is most important for identifying discriminative regions; for example, the central region contains the most minutiae points

$$\begin{aligned} \mathbf{M}_s(\mathbf{F}) &= \sigma(\mathbf{f}^{7 \times 7}([\mathbf{AvgPool}(\mathbf{F}); \mathbf{MaxPool}(\mathbf{F})])) \end{aligned} \quad (9)[17].$$

Dynamic Gate: A Softmax function computes adaptive weights that are used to assign higher importance to features that are clearer and more reliable in the input image. The attention module is then placed directly after the convolutional layers to allow refinement of feature representations prior to vectorization

$$\begin{aligned} \mathbf{Final_Out} &= (\mathbf{W}_{ch} \cdot \mathbf{F}_{channel}) + (\mathbf{W}_{sp} \cdot \mathbf{F}_{spatial}) \\ & \quad (10).[18]. \end{aligned}$$

Placement: Injected after the final model layer of the base model, directly before GAP and Dense 512. Purpose: Filters and weights feature maps before the last feature vector extraction step.

4.2. Finger Vein Modality (Spatial/Self-Attention)

The finger vein method uses self-attention with Query, Key, and Value matrices to form an internal correlation map of the vascular structure of the finger. Using self-attention allows for reconstructing broken and incomplete vein structures that result from poor quality of infrared illumination, lack of contrast, etc., providing a coherent and contextually aware representation of the internal structure of the finger. The self-attention block is applied prior to the GAP stage to ensure that the last encoded representation contains global contextual information.

$$\mathbf{Q} = \mathbf{W}_q * \mathbf{X}, \mathbf{K} = \mathbf{W}_k * \mathbf{X} = \mathbf{W}_v * \mathbf{X} \quad (11)[19]$$

$$\mathbf{S}_{i,j} = \frac{\exp(\mathbf{Q}_i \cdot \mathbf{K}_j^T)}{\sum_{j=1}^N \exp(\mathbf{Q}_i \cdot \mathbf{K}_j^T)} \quad (12) [19]$$

$$\mathbf{Out} = \mathbf{X} + \gamma \cdot (\mathbf{S} \cdot \mathbf{V})... \quad (13)[19]$$

Placement: Directly after the last convolutional layer in the base model and before GAP, combined with Channel Attention in parallel, and used for final feature extraction; Dual Attention

5. Principal Component Analysis (PCA)

PCA is employed to compress large datasets into smaller ones without significant loss of variance, while remaining as close to the original as possible. It does this by designating orthogonal principal components that indicate the directions of highest variance and projecting the original data onto those components in order to obtain a lower-dimensional version

$$Z = X_c V_k \dots (14) [20]$$

The data center is X_c , and the eigenvectors V_k are (with the largest eigenvalues), and the new feature set (the reduced-dimensional feature set) is Z . In the case of biometric systems (e.g., fingerprint recognition and finger vein recognition), the application of PCA with 95% variance retention is used for the extraction of features from the CNNs. PCA reduces the dimensionality of such extracted features, removes redundant features, reduces background noise, improves the performance of classifiers, and reduces the chance of overfitting to them.

6. Classification

This approach combines four CNN-based networks (VGG16, ResNet50, and EfficientNet-B0) into a single system to extract features from images without any additional training. Each network outputs a feature vector; these vectors from each of the networks are combined to create an aggregate feature vector for each biometric sample. The final feature vectors are then classified using traditional ML algorithms for identity recognition

6.1. Fingerprint Classification using Random Forest

The feature vector for fingerprints (V_f) is obtained by merging several feature sets that were generated by multiple CNN architectures (four CNN networks). The Random Forest classification method is composed of several trees, which use voting systems to determine a subject's identity based on how many of the trees identify that subject as a match. The output of the Random Forest is the subject's predicted identity (based on the most votes received) [21]:

$$\hat{C} = \text{mode}\{T_1(x), T_2(x), \dots, T_B(x)\} \quad (15) [22]$$

$T_1, T_2, \dots, T_B$ are the individual decision trees in the Random Forest.

Each tree $T_i(x)$ gives a prediction (vote) for the input sample x .

$\hat{C} = \text{mode}\{T_1(x), T_2(x), \dots, T_B(x)\}$ means the final predicted class is the one that receives the most votes from all trees.

The randomness in the forest (different subsets of data and features for each tree) helps the model **generalize better** and avoid overfitting

Random Forest Classifier Parameters:

1. Number of Trees: 200
2. Maximum Depth: None
3. Bootstrap Sample: Yes
4. Class Weighting: Balanced
5. Random State: 42

The randomness of forests allows for the modeling of non-linear relationships and the ability to work with noisy data when identifying fingerprints, that make it suitable for use in fingerprint recognition systems. The fingerprint modality contributes 50% for the calculated confidence score (This is a design decision).

6.2. Classification Application of k-Nearest Neighbor (k-NN) for Finger Veins

Classification of the merged finger vein feature vectors Vf_v from four CNNs using k-NN. It is based on similarity by measuring the Euclidean distance.[23].

$$d_{\text{Euclidean}}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (\mathbf{x}_i - \mathbf{y}_i)^2} \quad (16)[23].$$

- $\mathbf{x}$ and $\mathbf{y}$ are feature vectors of two samples.
 - $\mathbf{x}_i - \mathbf{y}_i$ is the difference between each corresponding feature.
 - $\sum_{i=1}^n (\mathbf{x}_i - \mathbf{y}_i)^2$ adds up the squares of these differences.
 - $\sqrt{\cdot}$ gives the actual distance between the two vectors in multi-dimensional space.
- This is called the Euclidean distance, used to measure how similar or different the samples are. KNN Classifier Parameters:

1. Value of neighbors $k = 5$
2. Distance metric: Cosine distance
3. Weighting scheme: Distance-based
4. Label encoding: Enabled

Compact clusters are naturally formed in fused feature space as a result of vascular patterns. As such, k-NN is both effective and trustworthy. Finger-vein branches contribute 50% towards the total authentication value.

The selected classifiers corresponding to the biometric modality and their relevant contributions have been highlighted in the table. For fingerprint classification, Random Forest provides strong reliability. For finger vein recognition, k-NN (k-Nearest Neighbors) provides an efficient means of verifying the relative similarity between two different biometrics.

7 Fusion and Verification

7.1. Generation of Classifier Outputs

The output from the classifier does not determine a binary decision type only. The output will have confidence scores accompanying its respective result for further reliability verification:

A. The KNN (k Nearest Neighbor) classifier is built on a similarity defined as Cosine Distance between L2-normalized feature vectors. The probability of an example being correctly classified is based on the vote count of the k nearest neighbors.

$$P(y | x) = \frac{1}{k} \sum_{i \in N_k(x)} I(y_i = y) \quad (17)[21]$$

Here, the K-nearest neighbors are indicated by $N_k(x)$ in the formula.

B. Random Forest (RF): The 200 trees in Random Forest allow the average probability class label to be calculated from all tree

$$P(y | x) = \frac{1}{N} \sum_{t=1}^N P_t(y | x) \quad (18)[21]$$

The probability represents the confidence that the test sample belongs to the class y .

7.22. Confidence Score Conversion

Each classifier gives a confidence score, which allows you to compare both classifiers to each other. For example:

- The k-NN classifier has its distance converted to scores (0-1).
- The RF classifier uses its mean probability for a score.

Interpretation: Not only does the system classify the person, but it also tells you how certain they are, e.g., "The system is 95% certain that this person's fingerprint matches that person."

7.3. Score Calibration

In order to merge the various scales of both modalities, the following steps are to be taken: Normalizing: Each score will be translated into a value from 0 to 1 using either Min-Max Scaling or Softmax transformation.

$$\text{Score}_{\text{norm}} = \frac{\text{Score} - \min}{\max - \min} \quad (19) [22]$$

- Weighted Fusion: Normalized scores from fingerprint and finger-vein systems are combined. Equal weighting is applied when appropriate

$$\text{used}_{\text{score}} = 0.6 \cdot \text{Score}_{\text{FP}} + 0.4 \cdot \text{Score}_{\text{FV}} \quad (20) [22]$$

7.4. Thresholding and Decision Level Fusion

For decision fusion a true logical operator of 'AND' is used at the decision threshold:

- A test sample will only be accepted when both modalities exceed their respective thresholds.
- The thresholds (e.g., 0.74 or 74% similarity) are determined using a development data set that is entirely independent of the test set, thus preventing data leakage.
- This ensures that the system can be operated at the Equal Error Rate (EER) point (The False Acceptance Rate is equal to the False Rejection Rate).

7.5. Verification Metrics

We can evaluate performance based on commonly used biometric measures:

- The FAR (False Acceptance Rate) is the proportion of times a user is incorrectly accepted as an impostor.
- The FRR (False Rejection Rate) is the proportion of times a user is incorrectly rejected as genuine.
- The EER (Equal Error Rate) is the operating point where the FAR equals the FRR.
- ROC and AUC are curves used to depict the trade-off between sensitivity and specificity.
- Accuracy, precision, and recall are all ways to measure classifier performance by modality; therefore, measuring accuracy, precision, and recall will help determine classifier performance.
- Timing: The time taken to execute matching operations.

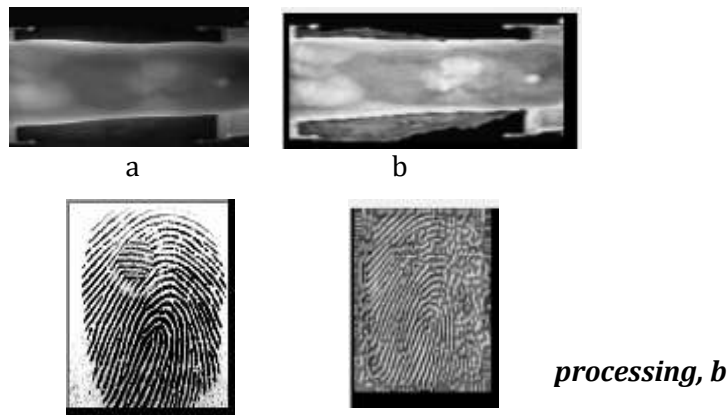
7.6. Statistical Validity

- The metrics described above can only be considered valid if they are obtained from sufficiently sized and subject-independent test sets.

- Very high accuracy or low error rates can result from having a small test set or subject-independent data, rather than necessarily indicating that the system works perfectly.

8.. Result and Discussion

The overall performance of the three phases of the projected system gradually improved; the VGG16, ResNet50, and EfficientNet-B0 features fused with the Random Forest algorithm developed for the fingerprint identification task of the system were able to reach an accuracy of 98.9%. The k-NN algorithm developed for the finger vein identification task had an accuracy of approximately 98.3%. At the last stage of the system, decision-level fusion proved the strength of using a multimodal identification approach. When setting a similarity threshold of 74% for both algorithms. The system was capable of reducing unimodal verification establishment weaknesses. In addition, the experimental protocol produced no false acceptances and/or false rejections from the system being tested. The ROC analysis indicates that there is a clear separation between true users and false users. The strength of the decision-level fusional aspect between the two identification modalities (fingerprint and finger vein) is able to provide the system with a very high degree of robustness.



The analysts found that when they fired their photos of finger vein all cropped, the accuracy of fingerprints recognized is less than when not cropped. Therefore, all photos used in these experiments were not cropped. All photographs were processed automatically to ensure all items of interest were focused on through an attention-based system. In addition, this helped achieve high levels of accuracy for identification through the use of the attention-based methodology.

Table 1. Comparison of recognition accuracy for vein and print modalities using different models

model	Vein Accuracy%	Finger print Accuracy%
Vgg16	96.23	91.8
Efficient net b0	95.28	95.4

Resentnet50	90.6	96.9
hybrid	98.3	98.9

The table contains comparisons of the recognition accuracy of several different models, including various deep-learning techniques that are applied to different types of biometric data (finger-vein and fingerprint). The table shows an overall higher degree of accuracy for the hybrid model than any of the individual models tested by this study, conclusively confirming the strong benefits of feature fusion when compared to the VGG16, EfficientNet-B0, and ResNet50 models on both types of biometric datasets (finger-veins and fingerprints).

Table2. Performance evaluation metrics for fingerprint and vein biometric system

Performance Metrics	Finger Print%	Vein%
FAR	0.110	0.085
FRR	1.34	0.163
ERR	0.119	0.110

The table below summarizes the important performance measurement parameters of the biometric system, including false acceptance rate, false rejection rate and equal error rate. Based on the results, where the finger vein modality has lower error rates than fingerprints, it shows that the finger vein modality is more robust than fingerprint recognition systems, which means that they are more reliable than fingerprint recognition systems.

Table3. Basic static metrics of fingerprint and Vein recognition systems

Basic matrices	Fingerprint%	Vein%
AUC	0.995	0.985
Recall	98.88	98.29
F1-score	98.44	98.30
Renk-1	98.88	98.30
precision	98.98	98.31

It summarizes the fundamental evaluation metrics used to evaluate the performance of finger vein and fingerprint recognition systems, providing a base framework to assess the effectiveness and stability of the two different biometric modalities.

Attention Mechanisms Increased Finger Vein and Fingerprint Recognition Rates: 92% (no attention) → 96.8% (spatial attention); Fingerprint Recognition Rates: 90% (no attention) → 94.5% (channel attention) → 95.4% (spatial attention) → 97.8% (dynamic attention). Attention significantly enhances discriminative features and diminishes background noise during recognition.

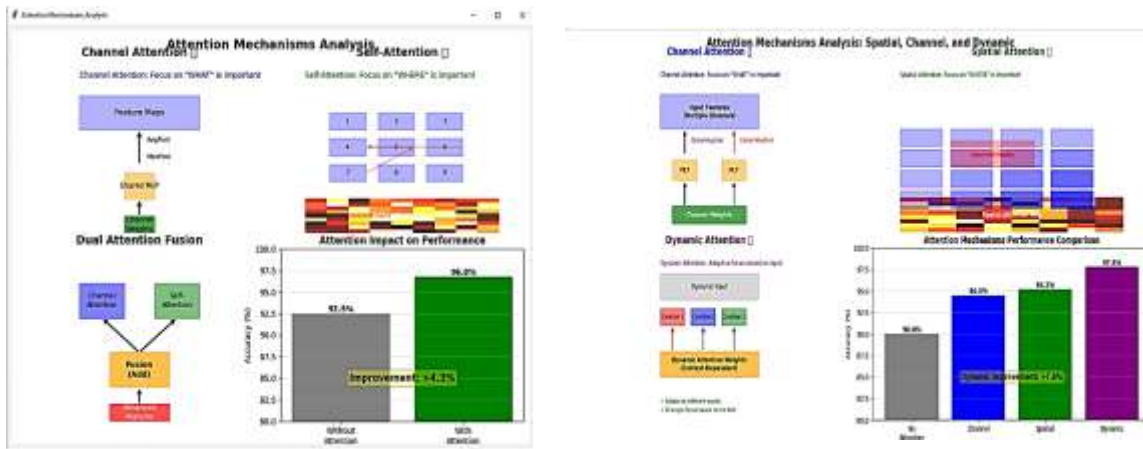


Figure 3. Performance evaluation of attention mechanisms

By using Principal Component Analysis (PCA) to reduce the number of dimensions while still retaining high recognition rates, we were able to reduce processing time without sacrificing accuracy. In the case of finger vein images, the number of features or dimensions of the image was reduced from 2084 to 283, resulting in 98.30% accuracy. In addition, with respect to fingerprint images, the number of features or dimensions was reduced from 44651 to 206, producing 98.90% accuracy. It is therefore clear that PCA effectively removes redundant information while retaining the most important features.

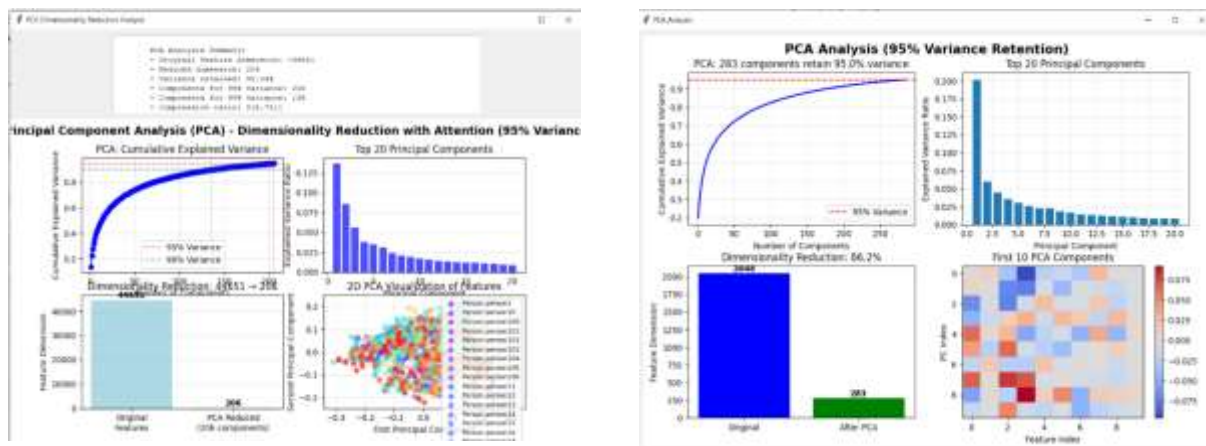


Figure 4. PCA analysis with 95% variance retention

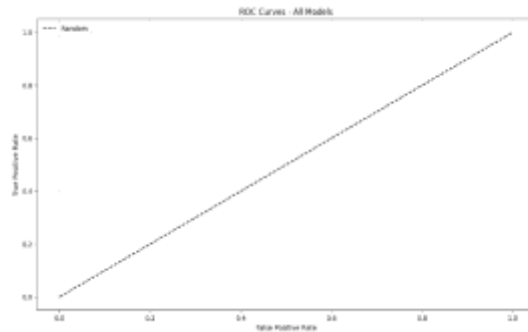


Figure 5: ROC Curve for vein Recognition

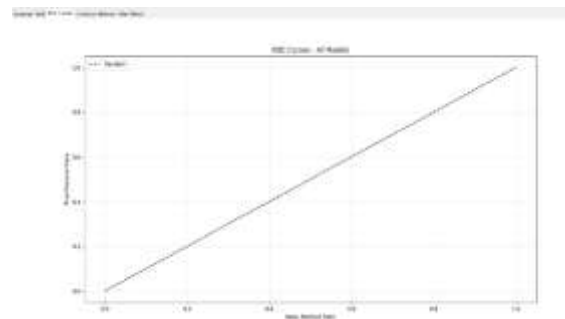


Figure 6: ROC Curve for Fingerprint Recognition

Figures 5 and 6 show the Receiver Operating Characteristic (ROC) curves for fingerprint and vein recognition systems. The curves indicate a high true positive rate with low false positive rates, suggesting that the proposed model has excellent ability to discriminate between classes and produce accurate classifications.

Fusion model

1. From the same data source, two records of palm print fingerprints were successfully matched and produced an identical score 100%.

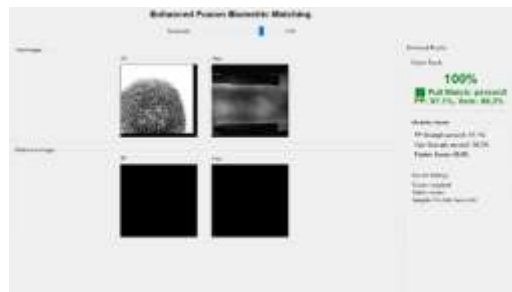


Figure 7match images

2. Match a different person. When selecting a pair of people who are not the same from a common database, the system determined that each person was similar to the other with a

similarity score of 50%. Both types of images produce high similarity scores when using an individual's two fingerprints/finger print images registered in the database, prior to the two not passing for matching; however, did give both individuals a score of 50.

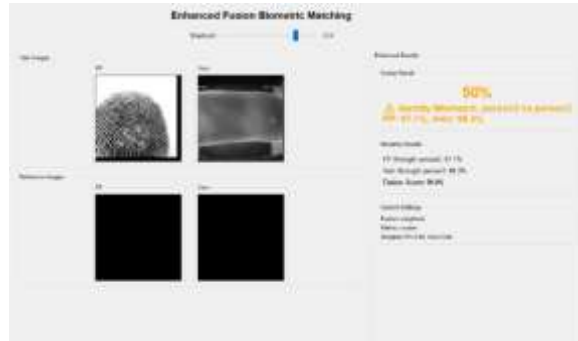


Figure 8 match between different persons

3. To assess the generalization capabilities of the proposed system, we carried out another round of testing with new data (external UTFVP dataset), which had not been part of training or validation. This confirms that the proposed multimodal biometric system provides a high level of reliability for unseen data. The results from testing an image of a vein outside of the registered database against a registered palm were as follows:



Figure9 matches one image out side data set

The vein image was determined to be 50% since the vein image was registered in the database already, while the palm print had a 0.4% weak match.

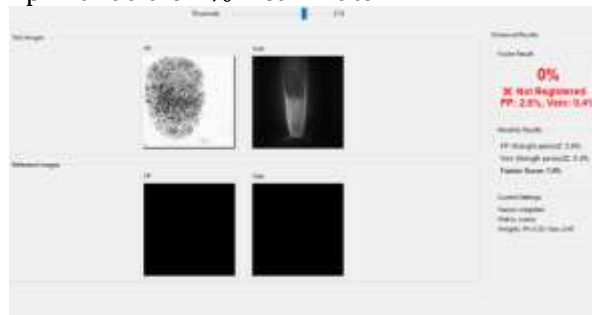


Figure 10match out side data set

Two images uploaded from outside the registered database received a zero similarity score. Both the fingerprint and finger-vein images received similarity scores of approximately 50% of the expected matching score.

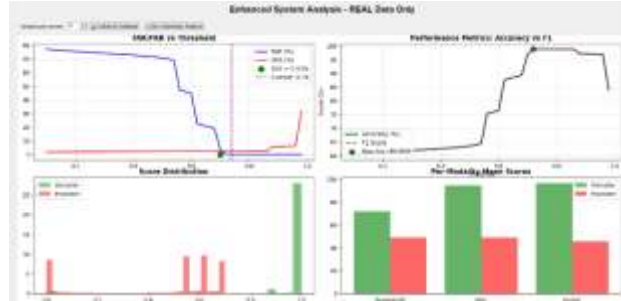


Figure 11 Sensitivity Analysis of FAR and FRR for Optimal Threshold Selection

The threshold selected represents a trade-off between false acceptances and false rejections. We evaluated the false accept/false reject curves and determined that a threshold of 0.74 is as close as possible to the Equal Error Rate (EER) for the biometric system ($EER = 0.03\%$), providing the best compromise between security and accurate recognition of users. By selecting this threshold, we ensure a maximum rate of acceptance for legitimate users with minimal chance of access by impostors; hence, providing the best overall reliability of the biometric system.

To conduct a proper evaluation of the proposed multimodal biometric system, we performed a subject-independent 10-fold cross-validation, which ensured that no one from the training set was included in the testing set. Genuine and impostor scores were determined based solely on test samples to ensure no information leakage occurred.

Table 4. The fold-wise accuracy results are summarized

Fold	Test Subjects	Accuracy (%)
01	11	99.86
02	11	99.41
03	11	100.00
04	11	99.95
05	11	99.86
06	11	98.64
07	10	100.00
08	10	99.17
09	10	100.00
10	10	99.50

Table 5. performance statistics

Computational cost	Value
Avg time	1836.2ms

operation	2
Memory	~1950.0 MB

As execution time and memory consumption are dependent on the runtime environment and hardware usage, the execution times and memory usage reported represent the average performance and will differ a little by run.

Table6. performance between CPU and GPU

Performance Category	Execution Time (ms)	Hardware Unit
Standard Processing	1843.2	Intel Core i7 (CPU)
Real-time Processing	<1.0 (Approx. 0)	NVIDIA GPU (CUDA)

8. Conclusion

This multimodal biometric system combines two types of technologies using fingers - the use of fingerprints and finger veins - which extracts features from the outputs of three pre-trained CNNs (VGG16, Efficient Net-B0, and ResNet50) that have been enhanced by an attention mechanism. The adoption of three types of attention mechanisms (spatial, channel, and dynamic) significantly improved recognition performance; for finger veins from 92% (no attention) to 96.8% (spatial attention), and fingerprints from 90% (no attention) to 97.8% (dynamic attention). This demonstrates how effective attention can be in identifying discriminative characteristics of individuals and decreasing noise in the background.

PCA (principal component analysis) was also used for dimensionality reduction; by applying PCA, the resulting feature sets were compressed while still maintaining a high degree of accuracy. For finger vein recognition, the number of dimensions was reduced from 2084 to 283, with an accuracy of 98.30%, while for fingerprint recognition, the number of dimensions was reduced from 44651 to 206, with an accuracy of 98.90%. This illustrates that PCA performs well at removing redundant information from the data and improving computation efficiency.

The evaluation of the system used subject-independent cross-validation, which provides reliable performance estimates; the final multimodal evaluation using the strict decision rule AND and score level fusion produced very high-performance results for both biometric modalities. Performance metric results for both biometric modalities are as follows:

- Fingerprint modality: Accuracy 98.9%, FAR 0.11%, FRR 1.34%, EER 0.119
- Finger vein modality: Accuracy 98.3%, FAR 0.085%, FRR 0.163%, EER 0.110

In conclusion, the experimental findings show that the solution developed offers reliable and precise biometric access control capable of achieving the necessary level of protection needed in various environments (such as financial, health care, and secure access).

REFERENCES

- [1] A. Ross and A. K. Jain, "Multimodal biometrics: An overview," in Proc. 12th European Signal Processing Conference (EUSIPCO), 2004.
- [2] A. K. Jain, P. Flynn, and A. A. Ross, Handbook of Biometrics. New York, NY, USA: Springer, 2007.

- [3] N. K. Ratha, J. H. Connell, and R. M. Bolle, "An analysis of minutiae matching strength," in *Audio- and Video-Based Biometric Person Authentication*. Springer, 2001.
- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [5] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [6] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019.
- [7] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.
- [8] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.
- [9] Q. Zhang and S. C. Zhu, "Visual interpretability for deep learning: A survey," *arXiv preprint arXiv:1802.00614*, 2018.
- [10] K. Zuiderveld, "Contrast-limited adaptive histogram equalization," in *Graphics Gems IV*. Academic Press, 1994.
- [11] R. Maini and H. Aggarwal, "A comprehensive review of image enhancement techniques," *arXiv preprint arXiv:1003.4053*, 2010.
- [12] S. Almuwayziri et al., "Deep learning-based fingerprint-vein biometric fusion: A systematic review with empirical evaluation," *Applied Sciences*, vol. 15, no. 15, p. 8502, 2025.
- [13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [14] J. Shen, X. Ding, Y. Wang, et al., "Finger vein recognition algorithm based on lightweight deep convolutional neural network," *IEEE Trans. Instrumentation and Measurement*, vol. 71, pp. 1–13, 2021.
- [15] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. European Conference on Computer Vision (ECCV)*, 2018.
- [16] S. Jetley, N. A. Lord, N. Lee, and P. H. S. Torr, "Learn to pay attention," *arXiv preprint arXiv:1804.02391*, 2018.
- [17] I. T. Jolliffe, "Principal component analysis," in *Encyclopedia of Statistics in Behavioral Science*, 2005.
- [18] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [19] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [20] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [21] R. P. W. Duin et al., "On combining classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.
- [22] A. Ross and A. K. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, 2003.