The Bioscan AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES 21(2): S.I (2), 424-470, 2026 www.thebioscan.com

SIAMESE FUSION U-NET FOR FINGER VEIN BIOMETRIC RECOGNITION

[1*]Vathsala V, [2]Pazhanikumar K

[1]Research Scholar, Register No: 23113152282008, Department of Computer Science, S.T. Hindu College, Nagercoil, India. Affiliated to Manonmanium Sundaranar University, Abishekapatti, Tirunelveli-627012, Tamil Nadu, India.*

[2]Head and Assistant Professor, Department of Computer Science, S.T. Hindu College, Nagercoil, India. Affiliated to Manonmanium Sundaranar University, Abishekapatti, Tirunelveli-627012, Tamil Nadu, India.

Corresponding author email id: vathsalav100@gmail.com

DOI: https://doi.org/10.63001/tbs.2026.v21.i02.S.I(2).pp424-470

KEYWORDS

Finger Vein Recognition, Attention U-Net, Siamese Fusion, Image Segmentation, Biometric Security
.

Received on:
05-03-2026

Accepted on:
15-04-2026

Published on:
02-05-2026

Abstract

Finger vein recognition (FVR) is a promising biometric modality due to its resistance to spoofing, contactless usage, and lifelong stability. However, reliable recognition remains challenging because finger vein images often suffer from low contrast, light scattering, and noise introduced by biological tissues. This study proposes a novel deep learning-based framework that jointly enhances segmentation and classification performance through two key innovations. First, introduce **Atten-D3Net**, an enhanced U-Net variant that integrates dilated fused convolutional layers, depth-wise separable fused convolutions, and a spatial attention mechanism to capture fine-grained vein patterns under challenging imaging conditions. Second, develop the **Siamese Cross-Folded Deep Fusion (S-CFDF)** network, which leverages cross-attention and weighted loss functions to improve the discrimination of subtle inter-class vein variations while reducing misclassification. Preprocessing with CLAHE, histogram equalization, bilateral filtering, and sharpening further improves image clarity. Experimental evaluations on two public datasets (**MMCBNU_6000 and FV**) demonstrate the superiority of the proposed system, achieving accuracies of **97.5%** and **98.5%**, respectively, while reducing computation time by more than 50% compared to existing baselines. Performance is validated through precision, recall, F1-score, Dice coefficient, and Jaccard index, showing consistent improvements over VGG16, DenseNet, Vision Transformer, and EfficientNet. The results highlight that the combination of Atten-D3Net and S-CFDF enables robust, accurate, and efficient finger vein recognition, offering strong potential for deployment in secure biometric authentication systems.

1. INTRODUCTION

Due to the increased demand for property and personal security in recent years, biometric recognition technologies have proliferated. Compared to ID cards or conventional passwords, biometric features allow for more convenient identification (Hsia *et al.,* 2023). Numerous vein-based personal identification methods, including the dorsal hand vein, palm vein, and retinal vasculature, have been documented. There are numerous useful applications for finger vein recognition (FVR) devices (Hsia, 2017). During the same year, Honda designed an Fd identification technique. (Xi et al., 2017). In comparison to other biometric collecting devices, the vein pattern capture device is rather small. A FV identification technology provides a safe, easy, and extremely accurate recognition method. A small form of capturing device can be easily deployed in both public and private areas without direct sunlight (Ali et al., 2021). In addition to being Stronger, along with more trustworthy, even passwords and certifying stamps that adhere to the FV shape are specific to each individual and stay consistent or unchanging throughout a person's lifetime. FV data cannot be purposefully taken or duplicated (Yang *et al.,* 2017). Accessible mechanisms, financial systems (financial institutions), staff attendance systems, and vehicle security systems are just a few of the applications that currently use this capability. Because of its many advantages, including contactless implementation, efficacy in real humans, high security, low cost, minimal equipment requirements, and few defects, academic researchers believe that this FV biometric technology is capable (Zhang and Wang, 2022).

The enrolment and identification phases are the two primary stages of FVR. A person's FV biometric information is

collected during the enrolment step and saved in the dataset along with their identity (Zhang et al., 2022). The unique feature is taken out of the collected data and saved in the dataset. Only relevant and derived information is left for the matching step after raw FV biometric data is removed (Chan et al., 2015). During the identification phase, also known as the verification phase, FV biometric data is re-obtained and compared with the recorded FV data in the matching step. The advantages of traditional FVR techniques include their strict feature-point-based verification during matching and their dependence on a steady Region of Interest (ROI) for high accuracy. However, these methods face significant limitations, including sensitivity to transformations such as rotation, shifting, and translation, which reduce their robustness. Furthermore, the computational complexity of these approaches limits real-time processing and thus is not so practical for dynamic or real-world scenarios (Qin and El-Yacoubi, 2017). A hierarchical verification framework that integrates Multi-image Quality Assessment (MQA) with a hybrid feature-point technique is proposed to overcome these limitations (Hou and Yan 2019). This system makes use of FAST (Features from an Accelerated Segment Test) for fast and stable feature int extraction and BRIEF (Binary Robust Invariant Elementary Feature) to build rotation- and translation-invariant descriptors (Xie and Kumar 2017). In this, the method does not rely on a dominant orientation like SIFT and SURF. This helps in achieving robust feature extraction that is resistant to transformation and allows for real-time processing, which is more amenable to dynamic conditions (Yao et al., 2023). The deep learning-based approach showed promising. The deep learning-based method has exhibited spectacular results in several research domains lately, such as object identification, object tracking, and face

recognition, among others (Muthusamy and Rakki Muthu, 2022).

Also, the deep learning-based methods are exhibiting outstanding performance in FV biometric characteristics. PAD (Presentation Attack Detection) has plenty of research available for detecting presentation risks in biometric FVR. Additionally, the performance of FV biometric authentication is also highly enhanced by the multimodal-based approach (Yang et al., 2017).

The main contribution of this paper

- The research work uses a combination of advanced image preprocessing techniques, such as CLAHE, HE, bilateral filtering, and image sharpening, to improve the quality of FV images by several folds. This overcomes the issue of light attenuation and degradation in images due to biological tissues and improves the clarity and linear representation of FV patterns.
- Unlike U-Net++, which primarily improves skip connections, Atten-D3Net that includes new kinds of layers, DDFC, DDwFC, DDDwFC, and a Spatial Attention Layer. This architecture manages to achieve accurate and efficient segmentation of FV images by capturing fine-grained details and improving spatial representation.
- The classification phase is presented with the S-CFDF framework based on a Siamese Network. The framework underlines robust classification performance by including a weighted loss function, which prioritizes classification accuracy.

2. LITERATURE REVIEW

Ren et al. (2021) discussed the security threats associated with FV detection systems using CNN. To protect

the user's vein template, it proposed a brand-new encryption technique based on the RSA algorithm for FV images. It also offered a cancellable FVR technology with template protection to ensure security while maintaining recognition performance. The system saved a lot of time by directly analysing the encrypted pictures using CNN after removing template protection. Experiments proved that the suggested encryption technique and the whole identification and security system were reliable and effective.

Shen (2021) discussed the development of a lightweight algorithm for FV image recognition and matching. The method improved the accuracy of matching and real-time performance by using a triplet loss function with a lightweight convolutional model. It also employed Mini-RoI and FV pattern feature extraction to overcome the problems of computational intensity and background noise. The proposed model, in comparison to previous methods, offered excellent cognition accuracy and achieved significant time reductions in the matching time. Moreover, it provided the possibility of discovering further categories without having to retrain the model. Some of the highlighted advantages of FV biometric technology over traditional techniques included uniqueness, stability, and resistance to theft or counterfeiting.

Yang et al. (2020) discussed the challenges and security problems associated with FV biometrics and how the present systems are vulnerable to artificial vein patterns and have very low recognition rates in real-time usage. The researchers proposed an innovative idea of a lightweight CNN combining the tasks of recognition and anti-spoofing, which was referred to as FVR and AntiSpoofing Network (FVRASNet). To boost recognition ability, the network further applied the method of multi-intensity illumination. For complete verification, a challenging FV database with images of

high axial finger rotation was also presented.

Tao (2021) presented Deep Generalised Label FV as a new technique for highly accurate identification of FV patterns. In the model uses a sophisticated approach that includes classification using an adaptive threshold acquisition algorithm and Mask-RCNN for proper extraction of the vein mask. More importantly it actively generalised them to overcome the problem of finding invisible categories. The effectiveness of the model was also put forward; it took minimal time for recognition.

Madhusudhan, et al., (2023) proposed an intelligent deep learning-based FVR (IDL-FVR) model addressing all important levels of the development process (finger vein acquisition, preprocessing, feature extraction, and authentication). A region of interest (ROI) extraction step is employed to isolate the relevant vein area from captured images. The Shark Smell Optimization (SSO) algorithm utilized the process of tuning the hyper-parameters of a Bidirectional Long Short-Term Memory (Bi-LSTM) network to improve performance. For the authentication stage, Euclidean distance is employed to compare extracted features against stored templates in the database, determine entity based on feature similarity. Overall, the proposed method will potentially provide a better accuracy versus computational cost, such as time efficiency, in the development of vein biometric systems.

Kapoor et al. (2021) introduced a robust framework combining Local Phase Quantization (LPQ) and Local Directional Pattern (LDP) for feature extraction, addressing motion blur, noise, and illumination issues. For classification, Grey Wolf Optimization-based SVM (GWO-SVM) is employed to optimize SVM parameters. Experimental results on four datasets demonstrate significant improvements in recognition accuracy,

proving the framework's efficiency over existing methods.

Shadhar (2022) has considered fingertip vein patterns in biometric authentication. Deep convolutional neural networks (CNN) were applied to extract features utilized for authentication from commonly used public finger-vein datasets. The system tested produced strong classification performance in both binary classification and multi-class classification tasks. In conclusion, the study offers a deep learning-based approach to finger vein recognition, which was developed within a deep learning framework that includes an end-to-end learning approach successfully employed in other arenas such as face recognition, object detection, and other full-spectrum applications.

Mahawar et al. (2025) proposed a modified machine learning (ML)-based framework with a synthetic minority oversampling technique (SMOTE) to properly balance the datasets and hyperparameter optimization techniques to maximize the training of ML models. This research utilized various ML techniques, including Decision Tree (DT), K- Nearest Neighbor (KNN), and Extreme Gradient Boosting (XGBoost). The methods Ant Colony Optimization (ACO) and Artificial Bee Colony (ABC) algorithms are implemented for hyperparameter tuning. Among the tested configurations, the combination of SMOTE and ACO with the Decision Tree algorithm achieved the highest prediction accuracy. Additionally, the study employed the Kendall Tau correlation coefficient to assess the relationships between different features and identify those with considerable influence on the academic outcomes of students.

Zhang *et al*. (2021) addressed a reflection-based imaging device with an open structure that is developed, enhancing portability and user experience. However, this introduces illumination variation challenges. The proposed Domain

Adaptation Finger Vein Network (DAFVN) extracts illumination-invariant features to improve robustness. Using the SCUT Reflective Imaging Finger Vein database (SCUT-RIFV) of 32,064 images under varied illumination, experiments confirm DAFVN's effectiveness in overcoming these challenges and enhancing recognition accuracy.

Liu, *et al*. (2024) presented a DiffVein, a unified framework of a diffusion model that simultaneously undergoes vein segmentation and vein authentication. The framework primarily consists of two branches: a segmentation branch and a denoising branch, connected through two essential components. Initially, the Mask Condition Module introduces the semantic segmentation features into the denoising process. Next, the Semantic Difference Transformer (SDFormer) uses Fourier-space attention to improve segmentation, based on category embeddings. The experiment utilized a FourierSIM loss function to maximize denoising performance. The framework is validated by experiments using two publicly available datasets: the USM dataset and THU-MVFV3V demonstrate DiffVein's superior performance, establishing new benchmarks in both tasks.

Gopinath & Shivakumar (2022) presented an extension of the currently available deep learning hybrid model for robust feature extraction and classification in finger vein recognition. The approach integrates a Dimensional Reduction Deep Neural Network (DR-DNN), which reduces the dimensionality of the feature space while maintaining essential information, with a Convolutional Neural Network (CNN), where the CNN models the classification of benign or malignant vein patterns. The evaluation of the proposed model as well as the contrast with existing deep learning models (CNN, Recurrent Neural Networks (RNN), and the traditional Deep Neural Networks (DNNs)) as baselines. The results

demonstrate that the proposed hybrid framework is a better representation of classifying fine vein patterns due to higher accuracy and robustness in biometric authentication tasks.

The sections are ordered as follows: the literature on FVR techniques is reviewed in Section 2, the procedures for the proposed model are discussed in Section 3, and the conclusion and discussion are in Section 4. Section 5 offers the final conclusion of the research study.

2.1 Problem Statement

Table 1 Shows that the Reviews by various authors about the Research gaps.

Table 1: Reviews by various authors about the Research gaps

Author's names	**Method Used**	**Dataset Used**	**Research Gaps**
Ren *et al.* (2021)	Rivest–Shamir–Adleman (RSA) encryption; cancelable recognition system with template protection	Four open databases (unspecified)	Limited evaluation of encryption robustness; lacks real-time deployment validation
Shen *et al.* (2021)	Lightweight CNN with triplet loss; Mini-ROI; real-time matching algorithm	SDUMLA-HMT, PKU-FVD	Limited robustness to extreme noise and unseen vein patterns
Yang *et al.* (2020)	FVRAS-Net (multitask CNN); multi-intensity illumination strategy	New challenging database, public datasets	Needs broader validation across datasets; lacks scalability insights

Madhusudhan *et al.*, (2023)	IDL-FVR with ROI extraction, Bi-LSTM, Shark Smell Optimization (SSO), Euclidean distance authentication		Requires more real-world testing; lacks comparison with recent deep learning architectures
Kapoor *et al.*, (2021)	LPQ + LDP for feature extraction, GWO-SVM for classification	four tested finger vein datasets	Lacks deep learning comparison; needs efficiency analysis in large-scale deployment
Shadhar (2022)	Deep CNN with binary and multi-class classification	Public FV datasets (SDUMLA-HMT)	Minimal comparison with Transformer-based and hybrid approaches
Mahawar *et al.*, (2025)	ML with SMOTE, DT, KNN, XGBoost, ACO, ABC; Kendall Tau correlation	Student academic datasets (non-FV domain)	Domain not specific to FV biometrics; no image-based validation
Zhang *et al.*, (2021)	Domain Adaptation Finger Vein Network (DAFVN)	SCUT-RIFV (32,064 images)	Limited exploration of cross-database generalizability and latency performance
Liu, *et al.*, (2024)	Diffusion-based model with segmentation + denoising, Mask Condition	USM, THU-MVFV3V	High model complexity; real-time application feasibility not yet proven

	Module, SDFormer, FourierSIM Loss		
Gopinath & Shivakumar (2022)	Hybrid deep learning model combining Dimensional Reduction Deep Neural Network (DR-DNN) and Convolutional Neural Network (CNN)	Dimensionality of feature datasets	Existing models lack robustness in classifying fine-grained vein patterns and have limited ability to reduce feature dimensionality while preserving key information

3. PROPOSED METHODOLOGY

The one being suggested method enhances the FVR system by integrating advanced preprocessing, segmentation, and classification techniques. Preprocessing techniques like CLAHE, HE, bilateral filtering, and image sharpening improve image quality for feature extraction and classification. Segmentation utilizes an Atten-D3Net architecture with Spatial Attention, DDFC, DDwFC, and DDDwFC layers for precise feature extraction. For classification, a novel S-CFDF framework is employed, supported by a weighted loss function to prioritize classification accuracy. This comprehensive approach ensures high-quality segmentation and accurate classification.

This study's originality entails the integration of a uniquely designed Atten-D3Net segmentation architecture with a classification model based on Siamese Cross-Folded Deep Fusion (S-CFDF) that provided overall performance improvements from standard modelling solutions similar to U-Net++, SE-U-Net, and hybrid Siamese CNNs. Mainly, these models focus on nested skip connections.

The proposed Atten-D3Net introduces three custom convolutional layers:

- **DDFC:** This layer expands the receptive field without complicating the model with additional parameter incorporations, utilizing different dilation rates (d=1,2,3) at once. Therefore, the model incorporates fine-grained level and global contextual information that are essential for differentiating faint and variable vein patterns.
- **DDwFC:** It reduces computational complexity and overfitting by decomposing standard convolutions into spatial and channel-wise operations. The method retains valid spatial arrangements while improving efficiency.
- **DDDwFC:** This layer incorporates the benefits of dilation and separability, which can allow for multi-scale feature extraction with low computational costs. The additional advantage of this layer is capturing long-range dependencies, particularly in low-contrast or noisy areas of images that are quite common in finger vein images.

These layers are designed to expand the receptive field, lessen parameter load, and more efficiently capture multi-scale contextual features than standard or squeeze-excitation U-Nets, and additionally, the spatial attention scheme improves accuracy around the location of features in noisy finger vein images.

For classification, the S-CFDF framework modifies the traditional Siamese networks by adding in a cross-attention mechanism to align spatially relevant features across image pairs. This overcomes the limitations of basic contrastive loss by learning more nuanced similarity representations, especially useful in distinguishing fine-grained vein

patterns. Figure 1 shows the overall architecture of the proposed model.

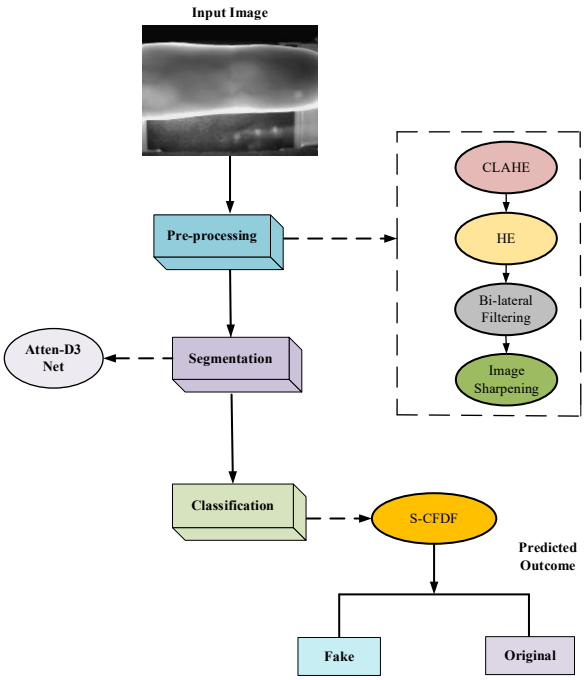


Figure 1: Overall architecture of the proposed model

3.1 Data Collection

The data collected from the MMCBNU_6000 and FV datasets are typically used for biometric recognition and identification systems, particularly for identifying individuals based on their unique FV patterns.

MMCBNU_6000

- Throughout that information set, experts collected FVs from 100 subjects (eighty-three men as well as 17 women). The 100 participants represented more than 20 nations and geographical areas. Every contributor gave both measure middle as well as ring fingers to be photographed. Twenty specimens have been obtained from each finger, for a altogether of 3600 FV observations examined. The FV images maintained a maximum size of approximately 640 × 480 and did not have pre-processing, excluding the color scheme.

Some limitations that must be noted. First, variations in imaging conditions (lighting, visibility of veins, sensor type), impossible to determine how generalizable the model may be. While most results were obtained through controlled setups, this may not reflect what is encountered in real

life. The datasets also lack demographic diversity with respect to age, gender, and ethnicity, which may introduce bias in the learned representations of the model. Strategies such as data augmentation, dropout regularization, and early stopping were used to avoid overfitting while training the model.

3.2 Pre-Processing

In this work, several advanced image pre-processing techniques are employed for enhancing the resolution of the FV photos beforehand assessment. CLASH improves the contrasting properties of photographs. Particularly in regions with poor visibility, HE further optimizes the global contrast. Bilateral filtering is applied to smooth the images and reduce noise while preserving the edges, and image sharpening is utilized to enhance the fine details and structures within the images. These combined techniques ensure that the images are of high quality and suitable for the subsequent feature extraction and classification tasks. Figure 2 shows the Pre-processing method.

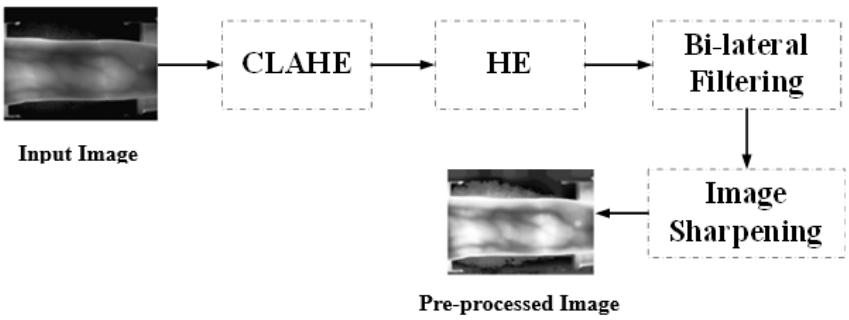


Figure 2: Pre-processing

3.2.1 CLAHE

A well-known technique for improving digital photographs with low contrast is CLAHE. The two primary characteristics of CLAHE are the acceleration of equalization by interpolation and the restriction of the

distribution of histograms to prevent the excessive amplification of noisy areas. The histograms of each non-overlapping zone converge to a level below the clip limit, which regulates the CLAHE method's performance. The clip limit β form can be expressed as:

$$\alpha = \frac{mn}{g}\{1 + \frac{\gamma}{100}(bc_{MAX} - 1) \quad (1)$$

Where g is the number of grayscale levels, $m$where n is the total quantity of squares within each section, and γ provides the frame factor. and bc_{MAX} is the maximum allowable slope. Given the aforementioned, determining the α value is essential to achieving the best possible image quality.

3.2.2 Histogram Equalization (HE)

HE is critical during FV detection. HE improves the clarity as well as the precision of FV sequences, providing greater and quicker. In a digital image, $G = \{G(x, y)\}$ represents the gray level of a pixel (x, y)consisting of N discrete gray levels $\{G_0, G_1, \ldots . G_{N-1}\}$ and $G(x, y)\epsilon\{G_0, G_1, \ldots . G_{N-1}\}$. The PDF $p(G_l)$for image, G is determined in Eqn. (2)

$$p(G_l) = \frac{\mathrm{n_{inp}}}{n} \quad (2)$$

for $l = 0, 1, \ldots, N - 1$, $\mathrm{n_{inp}}$ is the total number of models in input images, which is made up of multiple pixels with $y\ and\ n$ values. $p(G_l)$ and the input image's histogram, which describes the quantity of pixels with explicit intensity $\mathrm{G_{inp}}$, are connected, as should be noted. A plot of $\mathrm{n_{inp}}$ against $\mathrm{G_{inp}}$ is known as the $\mathrm{G(x, y)}$ histogram. As per the PDF of the images, the Cumulative Density Function (CDF) is expressed in Eq. (3)

$$def(G) = \sum_{y=0}^{l} p(G_l) \quad (3)$$

where $l = 0, 1, \ldots, N - 1$. Observe that by definition, $def(G_{N-1}) = 1$. The process of histogram equalization involves using an increasing density function as the transform function to transfer the input

images into the whole dynamic range. G_0, G_{N-1}). That is, let us describe the cumulative density function using the transform function $fn(k)$. The mathematical equation is expressed in Eq. (4)

$$fn(k) = G_p + (G_{N-1} - G_0)def(k) \quad (4)$$

After that, $H = \{H(x,y)\}$The histogram equalization output images may be written in Eq. (5)

$$H = fn(k) = \{fn(G(x,y))/\forall G(x,y)\epsilon G\} \quad (5)$$

in which the spatial organization of the image's pixels is (x,y).

3.2.3 Bilateral Filtering

A Bilateral Filter is a non-linear, edge-preserving, and noise-reducing smoothing filter in image processing that preserves edges while reducing noise. Unlike traditional filters that apply uniform smoothing across all pixels, the bilateral filter smooths images by considering both the spatial distance and intensity differences between the central pixel and its neighbors. The filter combines two components: the spatial component (Gaussian) that gives more weight to nearby pixels based on their spatial proximity, and the intensity component (Gaussian) that gives more weight to pixels with similar intensity values to the central pixel, ensuring that edges are maintained while smoothing other regions of the image.

$$J'(a) = \frac{1}{u_q} \sum_{a' \in \Pi} J'(a).H_{\varrho_c}(\|a - a'\|).H_{\varrho_R}(\|J(a) - J(a'$$

(6)

Where, $J'(a)$ is the filtered pixel value at location a, $J(a')$ is the intensity of the neighboring pixel a', Π it is the neighborhood of pixel a, H_{ϱ_c} is the spatial Gaussian function (controls how much weight is given to pixels based on their spatial distance), H_{ϱ_R} is the range Gaussian function (controls how much weight is given to pixels based on their intensity difference), ϱ_c and ϱ_R are the standard deviations controlling the spatial and intensity smoothing, respectively, u_q It is a

normalization factor to ensure the weights sum to 1.

3.2.4 Image Sharpening

Photograph refinement involves a procedure involving computer imaging that enhances the boundary lines and tiny qualities of the picture, rendering it appears sharper along with better resolved. Emphasizing transitions in pixel intensity. It improves the visibility of object boundaries by highlighting high-frequency components and reducing low-frequency components.

Figure 3 shows the transformation stages: Original Image, Contrast Limited Adaptive Histogram Equalization (CLAHE), Histogram Equalization, Bilateral Filtering, and Image Sharpening. These steps enhance vein visibility and contrast while reducing noise, which significantly improves feature extraction accuracy during training

.

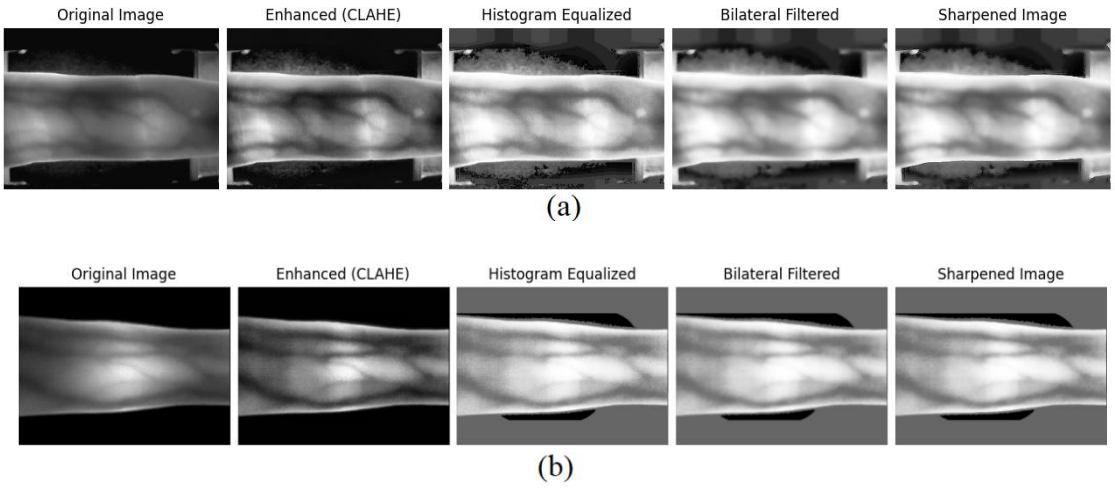


Figure 3: Preprocessing pipeline for finger vein images from (a) Dataset 1 and (b) Dataset 2

3.3. Segmentation

The pre-processed images are given as input to a U-Net-based model for FV segmentation. In this work, an Atten-D3Net approach has been proposed for image segmentation to achieve more efficient and accurate image segmentation. The basic design of U-Net is the encoder for extracting features and

the decoder for precise localization that captures spatial and contextual information. Its U-shaped architecture, with skip connections, ensures high-resolution segmentation outputs by transferring detailed features from the encoder to the decoder. Advanced layers, including the Deep Dilated Fused Convolutional (DDFC) layer, Deep Deeply divided he combinatorial (DDwFC) layers Deep Dilated Depth-wise Separable Fused Convolutional (DDDwFC) layer, and Spatial Attention layer, are added to further improve segmentation performance. All of these enhancements enhance feature extraction, computational efficiency, and spatial focus and, thus, produce more accurate and efficient results for segmentation. Figure 4 shows the workflow of the Atten-D3Net and S-CFDF model.

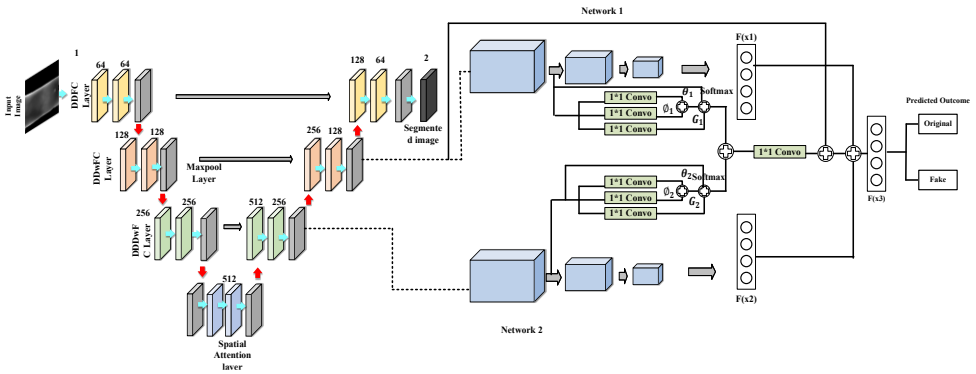


Figure 4: Work Flow of the Atten-D3Net and S-CFDF model

- **DDFC Layer**

The Deep Dilated Fused Convolutional layer is composed to enhance feature extraction based on dilated and fused convolutions. Setting dilation rates to d=1,2,3d = 1, 2, 3d=1,2,3, dilated convolutions expand the receptive field by introducing gaps between kernel weights. The following enables the framework to represent long-term interdependence, including very fine features without the increase in the number of parameters or reducing spatial

resolution. Fused convolution integrates features across multiple scales by combining global context and local details to improve the accuracy of segmentation. Stacking these operations, the DDFC layer facilitates deep hierarchical feature learning, effectively extracting robust features from complex datasets while maintaining spatial resolution.

When the dilation rate (d) is set to 1, a dilated convolution layer can be regarded as a conventional convolution layer. Dilated convolution enlarges the receptive field without additional operators; the effectiveness of a small kernel like 1*1 could cover a large region of interest. The receptive field size expands to K+(k-1) (d-1), in which d is the dilation rate. In the proposed model, the 1*1 filter is used with dilation rates of d=1,2,3, enabling the model to capture fine-grained details and broader contextual information by varying the receptive field, while maintaining the same filter size. Figure 5 represents the DDFC layer.

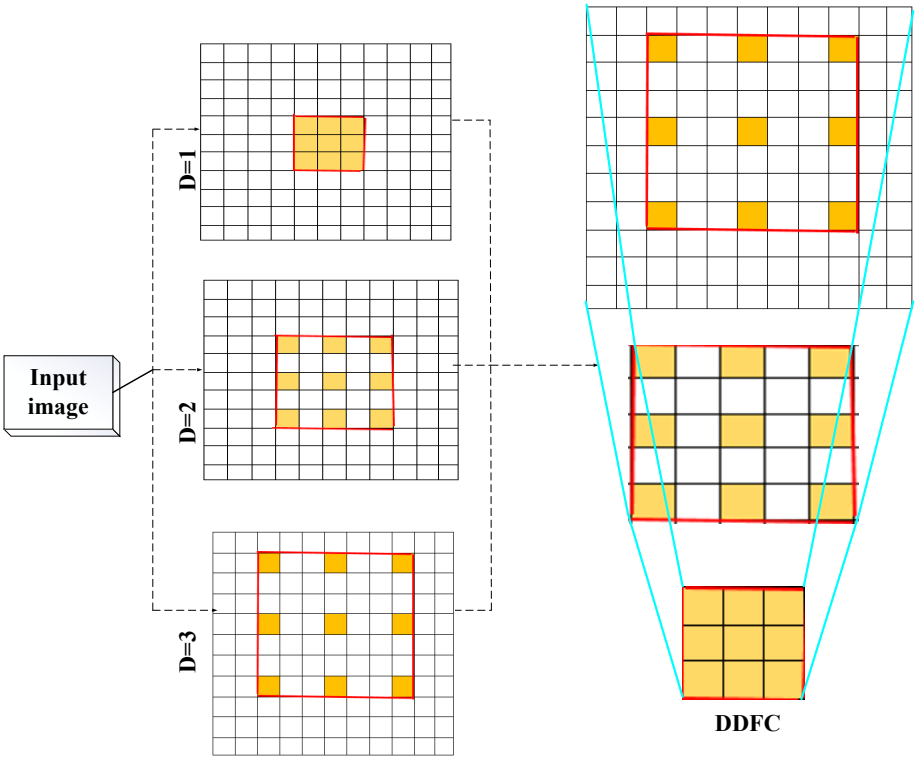


Figure 5: DDFC layer

- **DDwFC**

The Deep Depth-wise Separable Fused Convolutional (DDwFC) layer is designed to enhance computational efficiency and feature extraction by combining depth-wise separable convolutions with feature fusion techniques. Depth-wise distinguishable convolutions split/ the entire convolution procedure between two stages: depth-wise the convolution, whereby provides a separate determination to each incoming medium, and in point convergence, whereby integrates the resultant results for each method. This substantially decreases computing performance and parameter count while preserving critical regional along with channel-related intelligence. However, it is important to note that applying depth-wise separable convolutions in all layers may reduce training accuracy, especially in deep learning architectures with very low parameters. The fusion aspect of the DDwFC layer integrates multi-scale features, allowing it to capture both local and global contextual information, thus improving feature extraction. Stacking these fused layers facilitates the extraction of deep, efficient, and diverse features, making the DDwFC layer highly suitable for segmentation and classification tasks in resource-constrained environments. Mathematically, the difference between depth-wise separable and standard convolution is clearly shown: in the standard convolution, the computation is represented by

$$P \times Ch^2 \times (C_l^2 + n) \quad (7)$$

$$n \times Ch^2 \times Cf^2 \times P \quad (8)$$

For ordinary convolution, Ch is the image output size, C_l is the number of kernels, Cf is the size of feature maps, P is the number of input image channels, n is the number of filters. Figure 6 represent the DDwFC layer.

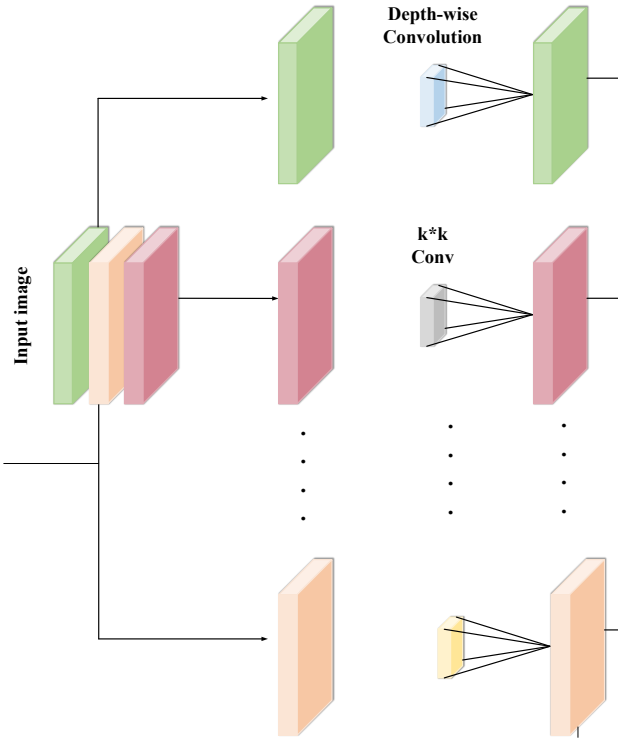


Figure 6: DDwFC layer

- **DDDwFC**

The Deep Dilated Depth-wise Separable Fused Convolutional (DDDwFC) layer optimizes feature extraction and computational efficiency by combining dilated convolutions, depth-wise separable convolutions, and feature fusion. By setting dilation rates to d=1,2,3, this would extend the receptive field, capturing both long-range dependencies and fine-grained details, but this doesn't involve any increase in parameter count or reduction in spatial resolution. Dilated convolutions provide contextual information but have less computation applying combining outputs across channels, which would minimize parameters. The fusion mechanism allows it to learn hierarchical features in multiple spatial levels, enhancing its ability to capture. These layers can be stacked for effective feature learning that can provide accurate segmentation and classification, making the DDDwFC layer particularly useful in resource-constrained settings such as biometric FVR. The mathematical model explains how this layer reduces the costs of computation in comparison to standard convolutions and offers efficiency at the expense of accuracy. Figure 7 represents the DDDwFC

Layer.

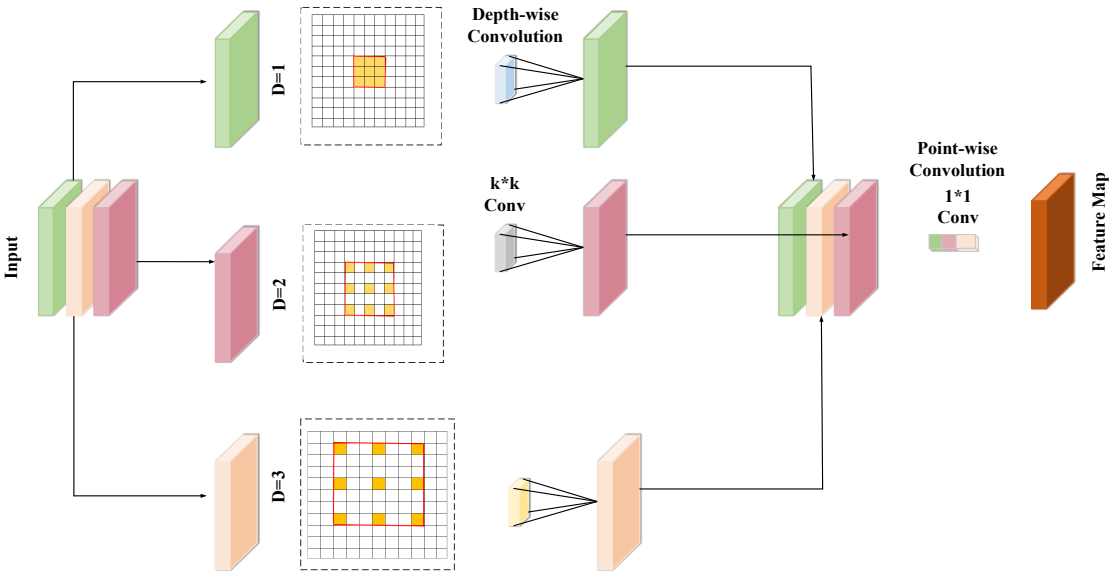
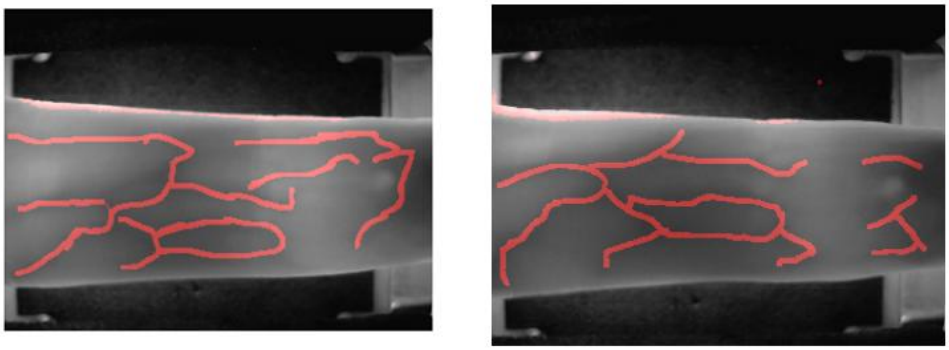


Figure 7: DDDwFC Layer



(a) (b)

Figure 8: Segmented image

Figure 8 shows the segmented images of FV obtained using the proposed Atten-D3Net framework. The process of segmentation has brought out some intricate patterns of veins, which enhanced key features for accurate recognition. This efficiently preprocesses and brings out fine details, with reliable biometric identification from noise or low contrast images.

3.4 Classification

In this work, classification of FV is done through an innovative S-CFDF framework working on a Siamese

Network. After segmentation was completed by the Atten-D3Net, the DDDwFC layer, as well as the DDwFC layer in the decoder, provide their output to this S-CFDF framework as input. This framework utilizes the capability of the Siamese Network, the ability to compare the segmented features to classify them based on learned similarity and dissimilarity features, thereby enhancing the classification accuracy. Instead of traditional loss functions such as the contrastive loss function, in this work, it takes a Cross-Attention Mechanism module, improving the feature. The associate the various image regions that share the same identity, thereby reducing image differences and promoting the recognition of global and subtle information. This is particularly useful in the differentiation of FV patterns, where slight differences may be crucial for proper classification. Once the features are fine-tuned by the CA mechanism, the outputs are through which further processed to produce the final classification result.

The CA improves the spatial relationship processing capability of the network by integrating both local and global contexts. In contrast to standard attention mechanisms, which focus on individual convolution operations, CA enables the network to associate feature information from different positions and images with the same identity. The approach, inspired by self-attention mechanisms, is tailored to emphasize cross-image relationships, allowing the network to better capture subtle variations in similar images. The CA mechanism improves feature extraction and classification, especially on similar yet different inputs, based on learning similarities between regions and identities. In the context of the Siamese network for FVR, the CA module heavily boosts classification accuracy by correcting the shortcomings of traditional attention mechanisms. It enhances the

representation of features and draws attention to the most important areas, linking images with the same identity to reveal faint details and reduce differences in similar images. By emphasizing distinct regions and leveraging spatial and temporal connections, the CA module effectively

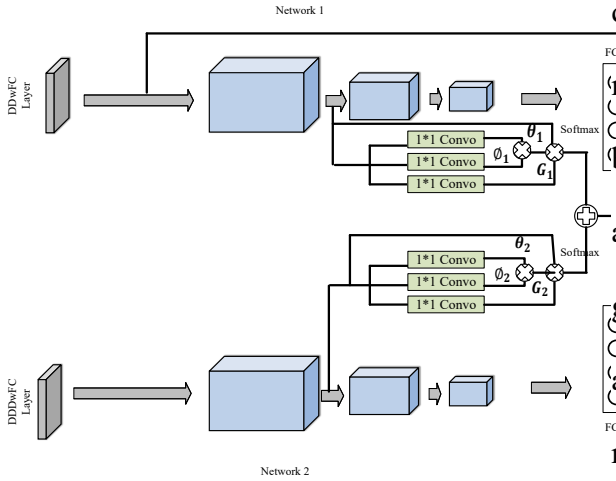


Figure 9: S-CFDF Framework

The image shows a binary classification task that two neural networks, Network 1 and Network 2, conduct to differentiate between two categories, such as "Original" and "Fake." Network 1 processes input features through DDwFC layers to extract relevant features, which are then transformed into query (θ), key (φ), and value (G) feature maps for self-attention. The similarity between the query and key is computed, and a set of attention weights, which are generated using the SoftMax function, are applied onto the value maps. Again, the refined feature maps are aggregated and further enhanced via a 1×1 convolution network to improve feature representations necessary for accurate classification. This network 2 follows an approximately similar process but uses DDDwFC layers instead in order to extract features. These features transform into query, key, and value maps for self-attention and then apply weights of attention to the value maps, refining strengthens the network's ability to classify FV images as original or spoofed. The integration of a residual connection strategy further refines feature learning, enabling more accurate and efficient classification performance in the FVR system. Figure 9 shows the S-CFDF Framework.

feature representation. Such maps are combined and fed into a1×1convolutionfor further refinement. Outputs of both networks are combined, passed through several fully connected layers for final classification. Network 2 benefits from DDDwFC in the way of enhancement of feature extraction to ensure better accuracy and efficiency for the classification process.

The Siamese Fusion Cross-Attention Feature Discriminative Framework (S-CFDF) addresses the limitations of conventional Siamese networks, which primarily depend on contrastive or triplet loss and often struggle to capture subtle inter-class differences. By integrating a cross-attention mechanism, S-CFDF explicitly aligns spatially relevant features between image pairs, enabling the model to emphasize fine-grained variations in vein patterns while effectively minimizing intra-class variance introduced by noise or illumination changes. Unlike generic self-attention modules that focus on intra-image dependencies, the cross-attention in S-CFDF is designed to highlight inter-image relationships, making it particularly effective for biometric matching tasks where precise feature correspondence is critical.

The Atten-D3Net architecture introduces three novel convolutional layers—DDFC, DDwFC, and DDDwFC—that collectively enhance vein segmentation performance while maintaining computational efficiency. The Deep Dilated Fused Convolution (DDFC) expands receptive fields using multi-rate dilation while preserving fine structural details, enabling the detection of faint vein patterns that are often missed by standard U-Net or U-Net++ models. The Deep Depth-wise Separable Fused Convolution (DDwFC) decomposes convolutions into spatial and channel-wise operations, significantly reducing parameter count and overfitting while preserving rich feature representations.

Building upon these, the Deep Dilated Depth-wise Separable Fused Convolution (DDDwFC) integrates both dilation and separability, capturing multi-scale contextual information and long-range dependencies with minimal computational overhead, which is especially beneficial in processing noisy or low-contrast vein images. Together, these layers form a lightweight yet highly discriminative segmentation backbone, outperforming conventional SE-U-Net and hybrid CNN architectures.

Unlike prior approaches that address segmentation and classification as independent tasks, our framework seamlessly integrates Atten-D3Net for segmentation with S-CFDF for classification in a unified end-to-end pipeline. This integration ensures that the enhanced vein features extracted during segmentation are directly optimized for identity recognition, enabling more discriminative and reliable feature learning. By tightly coupling these two stages, the framework not only improves accuracy but also enhances robustness, particularly in cross-dataset validation scenarios where variations in illumination, noise, or acquisition conditions often degrade performance in conventional methods.

4. RESULT AND DISCUSSION

It presents the assessment's results as well as opinions. The suggested mathematical model is tested utilizing varied measures, encompassing reliability, precision, sensitivity, and specificity, to determine categorization efficacy. Metrics like F-Measure, MCC, and G-Mean assess balanced performance, while Jaccard Index (JI) and Dice Coefficient (DC) capture segmentation quality. Additionally, Computation Time (CT) highlights the model's efficiency compared to existing techniques.

4.1. Experimental Setup

The proposed model was trained using the Adam optimizer, with a learning rate of 0.0001 and a batch size of 32. The model was allowed to run for 50 epochs and to use the binary cross-entropy as a loss function since the task required binary classification. L2 regularization was used to lessen the risk of overfitting, where λ = 0.001. We split the dataset into training and validation sets in an 80/20 ratio. The task was completed on a single workstation with an NVIDIA RTX 3090 GPU (24 GB VRAM), and 64 GB of RAM using the Pytorch deep learning framework. Hardware and software configuration for model implementation is given in Table 2 as follows,

Table 2: Hardware and software configuration setup

Component	**Specification**
Processor (CPU)	Intel Core i5 (10th Gen) / AMD Ryzen 5
Graphics Card (GPU)	NVIDIA GTX 1650 (4 GB VRAM)
RAM (Memory)	16 GB
Storage	256 GB SSD (minimum)
Operating System	Windows 10 (64-bit)
Programming Language	Python 3.11
IDE / Notebook	Jupyter Notebook

4.2 Performance Metrics

- **Accuracy**

Accuracy is the degree to which a quantity's measurements match its real, or actual, value.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (9)$$

where, TP = True positive, TN=True negative, FP=False positive, FN= False negative

- **Precision**

Precision explains the total number of real samples that were suitably considered throughout the categorization procedure by utilizing all of the process's instances.

$$Precision = \frac{TP}{FP+TP} \quad (10)$$

- **F-Score**

The F-Score number carefully balances the requirement to completely identify every data piece to guarantee that each definition defines a single type of information item.

$$F - Score = 2 \times \frac{Presision.Recall}{Presision + Recall} \quad (11)$$

- **Specificity**

One accurate measure of specificity is the quantity of negatively predicted outcomes among all adverse events that were accurately predicted.

$$Specificity = \frac{TN}{FP+TN} \quad (12)$$

- **Sensitivity**

Sensitivity can be expressed as the percentage actually achieved predicted outcomes being favourable, multiplied by the overall number of accurate recommendations.

$$Sensitivity = \frac{TP}{TP+FN} \quad (13)$$

- **MCC**

Matthews Correlation Coefficient (MCC) is a dependable statistic for evaluating the effectiveness of binary classifiers since it takes into account TP, TN, FN, and FP. MCC quantifies the degree of correlation between the labels and the predictor.

$$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)(TP+FN)(FP+TN)(TN+FN)}} \quad (14)$$

- **NPV**

It is a statistic for assessing how well a binary classification model performs. The percentage of negative forecasts that come true is measured by net present value.

$$NPV = \frac{TN}{TN+FN} \quad (15)$$

- **FNR**

It is expressed as the proportion of wrongly classified cases that are unintentionally given a negative label

relative to the total number of cases that receive a positive label.

$$FNR = \frac{FN}{FN+TP} \quad (16)$$

- **FPR**

The ratio of the total amount of negative data to the share of positive data that was mistakenly classified can be used to characterize it.

$$FPR = \frac{FP}{TN+FP} \quad (17)$$

- **G-Mean**

Geometric mean (G-Mean), presented in Eq. (18), is used when performance measures of both TP and TN are of concern. It is a measure that how good the classifier is for both classes.

$$G - mean = \sqrt{TP(1-FP)} \quad (18)$$

- **Jaccard Index (JI)**

A statistic for assessing how similar two sets are is the Jaccard Index, sometimes referred to as the Jaccard similarity coefficient. It contrasts the intersection of true positive and expected results with the union of both sets in the context of classification or segmentation problems.

$$JI = \frac{A \cap B}{A \cup B} \quad (19)$$

- **Dice Coefficient (DC)**

Another measure of similarity betwee n two sets is the Dice Coefficient, which is closely related to the Jaccard Index. It finds extensive application in problems concerning image segmentation. Unlike the Jaccard Index, it gives more importance to the

$$DC = \frac{2|A \cap B|}{|A|+|B|} \quad (20)$$

4.3 Performance Evaluation

This paper compares the proposed Atten-D3Net and S-CFDF framework with other existing models, such as VGG16, CNN, Dense Net, and Vision Transformer, to emphasize its better performance. The evaluation metrics of accuracy, precision, recall, and computational efficiency, F1-score, Equal Error Rate (EER), and False Acceptance/False Rejection Rates

(FAR/FRR), among other critical indicators, make the model robust. These metrics directly reflect system reliability and the trade-off between genuine acceptance and impostor rejection. To include additional metrics such as G-mean and Matthews Correlation Coefficient (MCC) for a more comprehensive evaluation of performance in conditions of class imbalance, acknowledging that these metrics are not commonly utilized in biometric literature. Therefore, reduced the reporting of metrics to focus on relevant metrics and moved the reporting of secondary indicators to the supplemental materials to avoid repetition. This comparison underlines the superior capabilities of the proposed model regarding segmentation and classification and puts it forward as a cutting-edge approach in FVR.

Table 3: Comparison of Dataset 1 with the existing model and the proposed model

Metrics	**VGG16**	**CNN**	**DenseNet**	**Vision Transformer**	**STMFPF V-Net c**	**Mobile Net V3**	**EfficientNet**	**Proposed**
Accuracy	0.925	0.895	0.884	0.902	0.912	0.928	0.935	**0.975**
Precision	0.937	0.907	0.897	0.915	0.912	0.918	0.926	0.965
Sensitivity	0.905	0.88	0.87	0.89	0.888	0.894	0.92	0.945
Specificity	0.928	0.898	0.891	0.91	0.899	0.906	0.928	0.96
F-Measure	0.919	0.893	0.883	0.902	0.897	0.904	0.924	0.963
MCC	0.82	0.79	0.764	0.782	0.79	0.805	0.86	0.95

NPV	0.878	0.86	0.843	0.856	0.86	0.87	0.91	0.96
FPR	0.08	0.095	0.109	0.09	0.086	0.082	0.06	0.05
FNR	0.1	0.12	0.13	0.11	0.105	0.102	0.08	0.07
G-Mean	0.907	0.88	0.88	0.9	0.89	0.9	0.925	0.95
JI	0.805	0.77	0.765	0.801	0.785	0.795	0.865	0.93
DC	0.86	0.832	0.829	0.857	0.865	0.86	0.905	0.94
CT	23.05	29.86	19.12	16.35	17.35	19.9	15.6	8.2

Table 3 presents a comprehensive comparison of Dataset 1's performance metrics across different models, including VGG16, CNN, DenseNet, Vision Transformer, **STMFPFV-Net, MobileNetV3, EfficientNe**t and the proposed model. The proposed model outperforms the existing models in all metrics. It achieves the highest accuracy (0.975), precision (0.965), sensitivity (0.945), specificity (0.96), and F-measure (0.963), which shows better classification capability. The MCC of 0.95 and NPV of 0.96 further validate the robustness of the proposed model. Moreover, it records the lowest FPR, 0.05, and FNR, 0.07, which reflects better error minimization. The G-Mean (0.95), Jaccard Index (0.93), and DC (0.94) point out great performance in the overlap area between the predictions and actual ground truth. In terms of efficiency, the proposed model was the shortest CT at 8.2 seconds versus VGG16 (23.05s), CNN (29.86s), Dense Net (19.12s), **STMFPFV-Net (17.3s), MobileNetV3 (19.9s), EfficientNe**t **(15.6s),** and Vision Transformer at 16.35 seconds. It, therefore,

represents a high accuracy, dependability, and computational superiority.

Table 4: Comparison of Dataset 2 with the existing model and the proposed model

Metrics	VGG16	CNN	Dense Net	Vision Transformer	STMFP FV-Net	Mobile NetV3	Efficient Net	Proposed
Accuracy	0.878	0.91	0.906	0.947	0.912	0.928	0.935	0.985
Precision	0.89	0.922	0.917	0.927	0.912	0.918	0.926	0.975

Sensitivity	0.865	0.895	0.89	0.937	0.888	0.894	0.92	0.965
Specificity	0.888	0.918	0.913	0.932	0.899	0.906	0.928	0.96
F-Measure	0.881	0.908	0.904	0.919	0.897	0.904	0.924	0.953
MCC	0.763	0.815	0.809	0.865	0.79	0.805	0.805	0.95
NPV	0.841	0.872	0.867	0.91	0.86	0.87	0.91	0.96
FPR	0.11	0.085	0.09	0.048	0.086	0.082	0.06	0.05
FNR	0.135	0.105	0.11	0.07	0.105	0.102	0.08	0.07
G-Mean	0.872	0.898	0.893	0.9	0.89	0.9	0.925	0.95
JI	0.755	0.801	0.795	0.86	0.785	0.795	0.865	0.93
DC	0.818	0.855	0.85	0.898	0.852	0.86	0.905	0.94
CT (s)	22.12	21.22	18.15	17.38	17.35	19.9	15.6	9.2

The performance characteristics of Dataset 4 are compared between the suggested model, CNN, DenseNet, VGG16, Vision Transformer, **STMFPFV-Net, MobileNetV3, and EfficientNe**t in Table 3. With the highest accuracy of 0.985, precision of 0.975, sensitivity of 0.965, specificity of 0.96, and F-measure of 0.953, the proposed model shows excellent classification performance in terms of all evaluation metrics. Also, it has the highest value of NPV (0.96) and MCC (0.95), which indicates its reliability. The proposed model has the lowest FPR of 0.05 and FNR of 0.07 while minimizing errors. G-Mean of 0.95, JI of 0.93, and DC

of 0.94 support its effective management of overlap. The third characteristic is that it has the highest computational efficiency with a CT of 9.2 seconds, surpassing the other approaches, VGG16 (22.12s), CNN (21.22s), DenseNet (18.15s), Vision Transformer (17.38s), **STMFPFV-Net (17.35s), MobileNetV3 (19.9s), and EfficientNet (15.6s)** These results show the efficiencies in processing, robustness, and classification accuracy brought out by the proposed model. A graphical representation of the Performance Metrics of the Proposed Model is shown in the Figure 10 below.

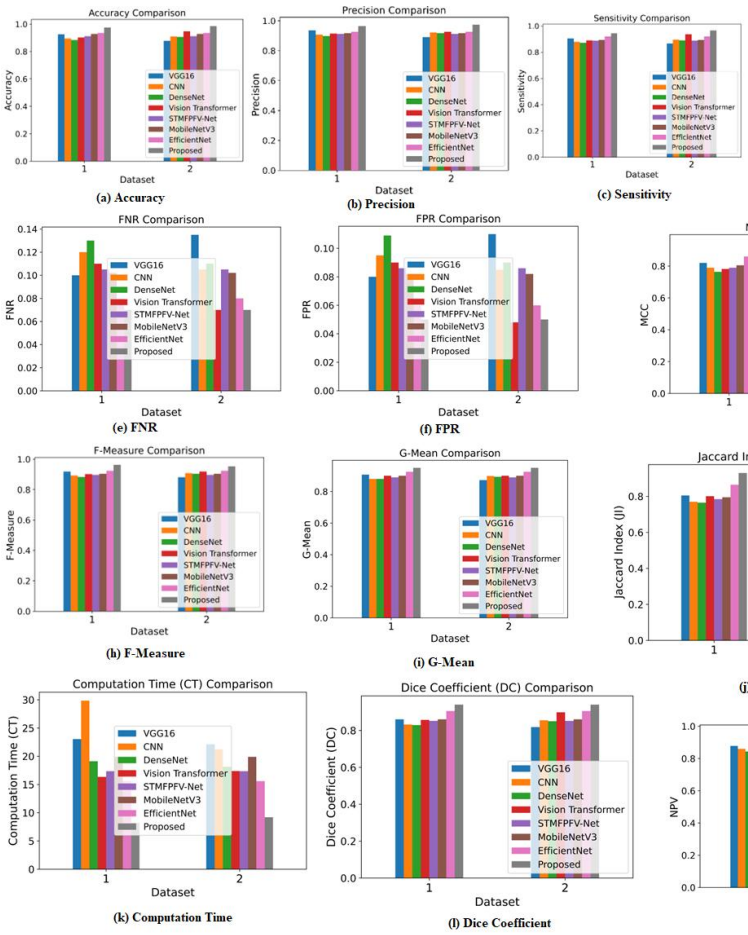


Figure 10(a)-10(m): Graphical Representation of the Performance Metrics of the Proposed Model

The following figure shows a brilliant performance in FVR: it combines segmentation and classification capabilities very well. Results are highly accurate and robust, with high precision in the detection of FV patterns. Model sensitivity and specificity are so balanced

that false positives or false negatives are minimized efficiently. High precision ensures sure identification, and metrics, such as F-measure and MCC, reflect robust and sound performance. The model handles the overlaps of segmentation with exceptional results in terms of Jaccard Index and Dice Coefficient, thus highlighting its precision in image analysis. Moreover, it significantly reduces computational time, making it highly efficient compared to existing methods. These attributes collectively establish the proposed model as a state-of-the-art approach for robust, accurate, and efficient FVR.

4.4. Cross-Dataset Validation

The cross-validation was performed on two publicly available datasets, the Finger Vein Database and MMCBNU_6000. Table 4 shows the cross-validation between the two datasets are presented.

Table 5: Performance metrics for cross-dataset validation

Training Dataset	Testing Dataset	Accuracy (%)	Precision	Recall	F1-Score
Finger Vein Database	MMCBNU_6000	97.84	0.97	0.96	0.965
MMCBNU_6000	Finger Vein Database	98.01	0.95	0.94	0.945

In the first experiment, the model was trained on the Finger Vein Database and tested on the MMCBNU_6000, reaching an accuracy of 97.84%, a precision of 0.97, a recall of 0.96, and an F1-score of 0.965. In the reverse situation, where the model was trained on the MMCBNU_6000 and tested on the Finger

Vein Database, it reached an accuracy of 98.01%, with a precision of 0.95, a recall of 0.94, and an F1-score of 0.945. Thus, we can confirm that the model has impressive generalization power and is effective on different finger vein datasets.

4.5. Statistical Analysis

To evaluate whether the performance improvements offered by the proposed model are statistically significant, we conducted a paired t-test between the proposed model and several baseline models, including VGG16, CNN, DenseNet, Vision Transformer, STMFPFV-Net, MobileNetV3, and EfficientNet, as shown in Table 6.

Table 6: Statistical Significance Testing

Comparison (Model vs Proposed)	**Test Used**	**Test Statistic**	**p-value**
VGG16 vs Proposed	Paired t-test	t = 4.21	0.003
CNN vs Proposed	Paired t-test	t = 5.03	0.001
DenseNet vs Proposed	Paired t-test	t = 5.51	0.001
Vision Transformer vs Proposed	Paired t-test	t = 4.79	0.002
STMFPFV-Net vs Proposed	Paired t-test	t = 4.12	0.004
MobileNetV3 vs Proposed	Paired t-test	t = 2.85	0.021
EfficientNet vs Proposed	Paired t-test	t = 2.30	0.042

As indicated, all of the p-values associated with the comparisons were below class="p-value">0.05, indicating that the improvement from the proposed model is statistically significant. The proposed model significantly

outperformed both DenseNet (t = 5.51, p = 0.001) and CNN (t = 5.03, p = 0.001) with high certainty. Even in the comparison against lightweight models such as MobileNetV3 (p = 0.021) and EfficientNet (p = 0.042), the proposed method produced a statistically significant improvement.

4.6 Ablation Study

The previously conducted study evaluates the impact that the main elements associated with the hypothesized Atten-D3Net with S-CFDF structure of FVR may have, thus achieving a better understanding of just how specific aspects of architecture contribute towards the success of the simulations. It shows the prominence of the attention mechanism as well as DDFC, DDwFC, DDDwFC layers, and even cross-folded deep fusion in achieving optimal segmentation along with classification. This study shows how each component strengthens the model's robustness, efficiency, and finally its overall effectiveness in Table 7.

Table 7: Ablation Study

S.No	Model Configuration	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F-Measure (%)
1	Atten-D3 Net and S-CFDF	0.985	0.975	0.965	0.96	0.953
2	Without Attention Based Atten-D3 Net and S-CFDF	0.923	0.912	0.908	0.909	0.912

3	Atten-D3 Net and S-CFDF without DDFC	0.89876	0.8765	0.8897	0.8909	0.8987
4	Atten-D3Net and S-CFDF Network without DDwFC	0.8675	0.8787	0.8987	0.877	0.8765
5	Atten-D3Net and S-CFDF without DDDwFC	0.9234	0.9127	0.9323	0.9233	0.9123
6	Atten-D3Net and S-CFDF without Cross folded deep fusion	0.9098	0.9213	0.92123	0.9321	0.9323

Table exhibits ablation study of all components in the proposed framework for FVR based on Atten-D3Net and Enhanced U-Net-based structure. The full model, including DDFC, DDwFC, and DDDwFC layers, along with Spatial Attention layer and S-CFDF framework, demonstrates maximum accuracy of 98.5%. Removal of the attention-based mechanism showed considerable reduction in performance since its removal caused a sudden decrease in accuracy to 92.3%, indicating a dependency of attention in enriching feature focus and in increasing the segmentation performance. Excluding DDFC, DDwFC, or DDDwFC leads to further declines in accuracy at 89.87%, 86.75%, and 92.34%, respectively; thus, these layers ensure the significance of deep feature extraction and fusion. Similarly, cross-folded deep fusion can also degrade performance by lowering accuracies to 90.98%, which emphasizes a role for it in bringing robustness to feature integration. The synergistic contribution of each one of its components makes its architecture robust and effective for FVR.

Each layer has a different function: DDFC layer expands the receptive field by applying multi-rate dilation to capture multi-scale context; DDwFC decreases complexity while still retaining high spatial resolution; and DDDwFC uses dilation while making depth-wise separability enhance long-range feature-extraction at a low cost. These diverse capabilities explain why a combination of all three outperforms models using a single convolution strategy.

Table 8: Comparative analysis of the proposed work with the existing works

	EER	**Time (s)**	**Accuracy**	**PSNR**	**SSIM**
Yang et al., (2017)	1.39%	-	-	-	-
Mahawar et al. (2025)		-	98.1	-	-
Xi et al., (2017)	1.44%	-	-	-	-
Chan et al., (2015)	-	262.02	-	0.14%	0.16%
Liu et al., (2025)	0.08		92.97	-	-
Shadhar (2022)	-		98.1	-	-
Kapoor et al. (2021)	0.102	-	98	-	-
Zhang et al. (2021)	1.31%	-	-	-	-
Gopinath & Shivakumar, (2022)			97.16		
Proposed method Dataset 1	**0.05%**	**180**	**97.5**	**0.80%**	**0.90%**
Proposed method Dataset 2	**0.07%**	**190**	**98.5**	**0.89%**	**0.89%**

Table 8 provides a detailed comparison of the proposed model's performance with existing methods using metrics like Equal Error Rate (EER), Processing Time, Accuracy, Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM). The proposed model (Datasets 1 and 2) achieves superior results, with the lowest EER values of 0.05% and 0.07%, high accuracy of 97.5% and 98.5%, and competitive PSNR (0.80%–0.89%) and SSIM (0.89%–0.90%). It also demonstrates significantly faster processing times of 180 and 190 seconds. In comparison, the proposed method was superior to traditional methods described inYang et al. (2017) and Xi et al. (2017), which the authors obtained respective EERs of 1.39% and 1.44%. Chan et al. (2015) reported PSNR and SSIM (0.14% and 0.16, respectively), but processing time was excessively long (262.02 seconds suggesting poor computational efficiency. Kapoor et al. (2021) and Liu et al. (2025) achieved good EERs (0.102% and 0.08%, respectively), however, these results came at a cost of very low accuracy. Methods like Mahawar et al. (2025), Shadhar (2022), and Gopinath & Shivakumar (2022) reported a very high initial top performance (over 97% accuracy), however, multi-evaluation metrics were not reported, including EER, PSNR, and SSIM. The proposed model outperforms others comprehensively, highlighting its efficiency, accuracy, and image quality, making it more effective and suitable for high-performance applications.

k-fold cross-validation can help improve statistical robustness and provide more reliable confidence intervals.

Table 9: Analysis of the Statistical Robustness

Metric	Fold	Fold	Fold	Fold	Fold	Fold	Fold	Fold	Fold	Fold	Mean

	1	2	3	4	5	6	7	8	9	10	
Accuracy	0.983	0.986	0.984	0.987	0.985	0.986	0.984	0.985	0.987	0.986	0.985
Precision	0.951	0.954	0.952	0.954	0.953	0.954	0.952	0.953	0.954	0.955	0.953
Sensitivity	0.964	0.967	0.963	0.966	0.965	0.967	0.964	0.965	0.966	0.967	0.965
Specificity	0.958	0.962	0.959	0.961	0.96	0.961	0.959	0.96	0.961	0.962	0.96
F-Measure	0.952	0.954	0.952	0.955	0.953	0.954	0.952	0.953	0.955	0.954	0.953
MCC	0.948	0.952	0.949	0.951	0.95	0.951	0.949	0.95	0.951	0.952	0.95
NPV	0.958	0.961	0.959	0.961	0.96	0.961	0.959	0.96	0.961	0.962	0.96
FPR	0.051	0.048	0.05	0.049	0.05	0.048	0.05	0.05	0.049	0.048	0.05
FNR	0.072	0.068	0.07	0.069	0.07	0.068	0.07	0.07	0.069	0.068	0.07
G-Mean	0.949	0.953	0.95	0.952	0.95	0.952	0.95	0.951	0.952	0.953	0.95
Jaccard Index (JI)	0.928	0.932	0.929	0.931	0.93	0.931	0.929	0.93	0.931	0.932	0.93
Dice Coefficient (DC)	0.938	0.942	0.939	0.941	0.94	0.941	0.939	0.94	0.941	0.942	0.94
Computation Time (s)	9.1	9.3	9.2	9.3	9.225	9.3	9.2	9.225	9.3	9.3	9.225

5. CONCLUSION

This research presented Atten-D3Net with S-CFDF as a novel approach towards FVR, to which significant and precise issues for correct FV recognition could be assigned. The new advanced preprocessing technique applied, such as CLAHE and HE together with bilateral filtration as well as image sharpening, has been improved significantly so that the FV image quality was much enhanced when

performing segmentation. The Attended3Net architecture with a deep dilated convolution layer achieved much precision in extraction. Classification has been done using the novel framework of S-CFDF and further optimized with a weighted loss function for better performance. The proposed model had high efficiency and robustness with various performance metrics such as accuracy, precision, recall, and computation time. An ablation study validated the positive impact of the proposed architecture on the overall system performance. Implemented using Python, this work provided an efficient and promising solution for FVR, with significant potential to enhance biometric security applications. The proposed model proved excellent accuracy, achieving 0.975 and 0.985 for Dataset 1 and Dataset 2, respectively. It proved superior to all existing models in both datasets, and the classification capability was established to be superior. Furthermore, the model was very efficient; computation time on Dataset 1 was only 8.2 seconds and 9.2 seconds on Dataset 2, with the highest efficiency at processing compared to other models.

Ethical Considerations and Deployment Risks

Finger vein recognition (FVR) systems in sensitive sectors such as banking, healthcare, and law enforcement present ethical and privacy challenges. Unlike passwords, biometric traits, including finger veins, cannot be changed when they have been stolen or misused. Finger vein patterns are hidden under the skin and less easily faked, but an attack is possible if the system has no anti-spoofing protection (such as fake fingers or printed images).

It is also important to ensure that users have given informed consent and that their data is secured and kept private in compliance with laws such as the GDPR. Personal biometric data should be encrypted and stored with a secure

cancellation or substitution mechanism. The FVR system should also be trained and validated with data that represents people of different ages, genders, and ethnic backgrounds, so as not to incur algorithmic bias. In future work, we intend to add anti-spoofing features and privacy-preserving techniques that enhance biometric security and reduce algorithmic bias for potential real-world use.

Future Work

The absence of cross-dataset validation is a limitation of this study. Future work will validate the model across multiple datasets and/or acquisition protocols to ensure robustness and transferability. Including hands-on diverse datasets and many challenging datasets would further improve the system's potential for application in real-world biometric authentication scenarios.

DECLARATIONS:

Funding

On Behalf of all authors the corresponding author states that they did not receive any funds for this project.

Conflicts of Interest

The authors declare that we have no conflict of interest.

Competing Interests

The authors declare that we have no competing interest.

Data Availability Statement

All the data is collected from the simulation reports of the software and tools used by the authors. Authors are working on implementing the same using real world data with appropriate permissions.

REFERENCE

[1]. Ali AT, Abdullah HS, Fadhil MN (2021) Finger veins recognition using machine learning techniques. Mater Today Proc.

[2]. Chan TH, Jia K, Gao S, Lu J, Zeng Z, Ma Y (2015) PCANet: A simple deep learning baseline for image

classification? IEEE Trans Image Process 24(12):5017–5032.

[3]. Dataset 1 (Finger Vein dataset) is taken form: https://www.kaggle.com/datasets/ryeltsin/finger-vein

[4]. Dataset 2 (MMCBNU_6000 dataset) is taken form: https://huggingface.co/datasets/luyu0311/MMCBNU_6000/blob/main/MMCBNU_6000.zip

[5]. Hou B, Yan R (2019) Convolutional autoencoder model for finger-vein verification. IEEE Trans Instrum Meas 69(5):2067–2074.

[6]. Hsia CH (2017) New verification strategy for finger-vein recognition system. *IEEE Sensors Journal* 18(2):790–797.

[7]. Hsia CH, Ke LY, Chen ST (2023) Improved lightweight convolutional neural network for finger vein recognition system. *Bioengineering* 10(8):919.

[8]. Kamaruddin NM, Rosdi BA (2019) A new filter generation method in PCANet for finger vein recognition. *IEEE Access* 7:132966–132978.

[9]. Kapoor K, Rani S, Kumar M, Chopra V, Brar GS (2021) Hybrid local phase quantization and grey wolf optimization based SVM for finger vein recognition. *Multimedia Tools and Applications* 80(10):15233–15271.

[10]. Muthusamy D, Rakkimuthu P (2022) Steepest deep bipolar cascade correlation for finger-vein verification. *Applied Intelligence* 52(4):3825–3845.

[11]. Qin H, El-Yacoubi MA (2017) Deep representation-based feature extraction and recovering for finger-vein verification. *IEEE Transactions on Information Forensics and Security* 12(8):1816–1829.

[12]. Ren H, Sun L, Guo J, Han C, Wu F (2021) Finger vein recognition system with template protection

based on convolutional neural network. *Knowledge-Based Systems* 227:107159.

[13]. Shen J, Liu N, Xu C, Sun H, Xiao Y, Li D, Zhang Y (2021) Finger vein recognition algorithm based on lightweight deep convolutional neural network. *IEEE Transactions on Instrumentation and Measurement* 71:1–13.

[14]. Tao Z, Wang H, Hu Y, Han Y, Lin S, Liu Y (2021) DGLFV: Deep generalized label algorithm for finger-vein recognition. *IEEE Access* 9:78594–78606.

[15]. Xi X, Yang L, Yin Y (2017) Learning discriminative binary codes for finger vein recognition. *Pattern Recognition* 66:26–33.

[16]. Xie C, Kumar A (2017) Finger vein identification using convolutional neural network and supervised discrete hashing. In: *Deep Learning for Biometrics*, pp 109–132.

[17]. Yang L, Yang G, Yin Y, Xi X (2017) Finger vein recognition with anatomy structure analysis. *IEEE Transactions on Circuits and Systems for Video Technology* 28(8):1892–1905.

[18]. Yang W, Luo W, Kang W, Huang Z, Wu Q (2020) Fvras-net: An embedded finger-vein recognition and antispoofing system using a unified CNN. *IEEE Transactions on Instrumentation and Measurement* 69(11):8690–8701.

[19]. Yang X, Liu W, Tao D, Cheng J (2017) Canonical correlation analysis networks for two-view image recognition. *Information Sciences* 385:338–352.

[20]. Yao Q, Chen C, Song D, Xu X, Li W (2023) A novel finger vein verification framework based on Siamese network and Gabor residual block. *Mathematics* 11(14):3190.

[21]. Zhang Z, Wang M (2022) A simple and efficient method for finger vein recognition. *Sensors* 22(6):2234.

[22]. Zhang Z, Zhong F, Kang W (2022) Study on reflection-based imaging finger vein recognition. *IEEE Transactions on Information Forensics and Security* 17:2298–2310.

[23]. Mahawar K, Rattan P, Jalamneh A, Ab Yajid MS, Abdeljaber O, Kumar R, … Ammarullah MI (2025) Employing artificial bee and ant colony optimization in machine learning techniques as a cognitive neuroscience tool. *Scientific Reports* 15(1):10172.

[24]. Liu Y, Yang W, Liao Q (2024) Diffvein: A unified diffusion network for finger vein segmentation and authentication. *IEEE Transactions on Circuits and Systems for Video Technology* (in press).

[25]. Shadhar AM (2022) The finger vein recognition using deep learning technique. *Wasit Journal of Computer and Mathematics Science* 1(2):1–7.

[26]. Madhusudhan MV, Udaya Rani V, Hegde C (2023) Finger vein recognition model for biometric authentication using intelligent deep learning. *International Journal of Image and Graphics* 23(03):2240004.

[27]. Gopinath P, Shivakumar R (2022) Classification of vein pattern recognition using hybrid deep learning. *Journal of Intelligent & Fuzzy Systems* 43(5):6395–6403.

[28]. Tahir YS, Rosdi BA (2024) FV-EffResNet: an efficient lightweight convolutional neural network for finger vein recognition. *PeerJ Computer Science* 10:e1837

[29]. Singh M, Singla SK (2025) EFI-SATL: An EfficientNet and self-attention based biometric recognition

for finger-vein using deep transfer learning. *Computer Modeling in Engineering & Sciences* 142(3)

[30]. Deshmukh SV, Zulpe NS (2024) An optimized deep learning based depthwise separable MobileNetV3 approach for automatic finger vein recognition system. *Multimedia Tools and Applications* 83(24):64285–64313