Journal of Science and Technology – Smart Systems and Devices
Volume 36, Issue 2, May 2026, 000–000

A Multiple Channel Biometric Recognition Model Using Palm Images

Tran Anh Vu*, Mai Hoang Anh, Nguyen Minh Duc, Doan Duc Hoi, Bui Viet Anh, Nguyen Huu Trung, Hoang Quang Huy
Hanoi University of Science and Technology, Ha Noi, Vietnam
Corresponding author email: vu.trananh@hust.edu.vn

Abstract

Biometric identification technologies are playing an increasingly important role in securing authentication. Palm-based recognition is garnering significant attention due to the unique and reliable patterns, lines, and texture structures found on the human palm. This study proposes a novel palm-based biometric recognition system that utilizes state-of-the-art computer vision and deep learning techniques. Contrast limited adaptive histogram equalization (CLAHE) and histogram equalization (HE) are applied to enhance the visibility of features in the preprocessed images under varying lighting conditions. To improve recognition accuracy, a hybrid deep learning architecture is designed by integrating a pretrained ResNet-based backbone with a multi-channel framework. This approach effectively merges multiple distinguishing properties of the palm, including lines and texture patterns. The model is trained and evaluated on a dataset of 25,600 palm images from 128 individuals, captured in different locations and from various angles. Experimental results demonstrate strong performance, with high accuracy, sensitivity, specificity, and recall, reflecting the robustness and reliability of the system. This work contributes to a scalable and efficient solution for palm-based biometric authentication, offering a promising approach for secure identity verification.

Keywords: Biometric recognition, CLAHE, multi-channel model, palm-based biometrics.

1. Introduction

As computer science and information technology advance rapidly, the need for robust and reliable methods to verify a person’s identity has become more crucial than ever. Traditionally, passwords, personal identification numbers (PINs), and security tokens were the main means of authentication. However, as technology evolves, the weaknesses of these methods have become increasingly apparent [1]. Issues such as stolen passwords, unauthorized access, and identity theft are growing more frequently, underscoring the insufficiency of traditional security measures. This has led to the development of biometric authentication systems, which use people's distinctive physical and/or behavioral attributes to verify who they are. Biometric systems are typically classified based on the number and types of traits they employ for authentication. For example, unimodal systems rely on a single biometric trait, while multimodal systems combine multiple traits to achieve greater accuracy and reliability.

Physical characteristics like fingerprints, iris patterns, and facial features, as well as behavioral characteristics like speech patterns or signatures, are some of the most commonly used biometric attributes. These traits have been extensively researched and applied in numerous security systems due to their uniqueness and reliability. The palm, however, is a comparatively unique yet often underused biometric trait. It comprises characteristics such as skin texture, hand shape, and the intricate patterns of veins and principal lines – including the heart, head, and life lines [2]. The length, depth, and configuration of these lines vary from person to person, making the palm an ideal trait for biometric identification. Furthermore, while facial recognition can be affected by factors such as occlusion, lighting conditions, or pose variations, palm-based recognition is more stable and reliable, as the features of the palm are less susceptible to environmental interference [3].

Biometric recognition has undergone significant transformation due to recent advances in artificial intelligence (AI), particularly in computer vision and image processing [4]. Sophisticated hardware, such as high-resolution imaging devices, alongside powerful software solutions – including convolutional neural networks (CNNs) and deep learning models – have greatly improved the accuracy and practicality of biometric systems. Modern CNN architectures, such as ResNet, Inception, and visual geometry group (VGG), have demonstrated remarkable success in accurately identifying individuals by analyzing multidimensional features from their biometric data. These models can reliably distinguish people based on facial traits and iris patterns. However, deepfake technologies and other methods of fabricating facial and iris data have raised doubts about the dependability of these approaches in high-security scenarios. Consequently, there is a growing need to consider additional biometric traits – particularly the palm – because its characteristics are more challenging to forge than other surface features.

p-ISSN 3093-3285
e-ISSN 3093-3315
https://doi.org/10.51316/jst.190.ssad.2026.36.2.x
Received: Jun 13, 2025; Revised: Oct 13, 2025
Accepted: Oct 13, 2025; Online: Apr 3, 2026.

A new approach to biometric identification is presented by utilizing computer vision and image processing techniques to analyze palm data. We have developed a deep learning–based system designed exclusively for identifying individuals by leveraging the rich and unique properties of their palms. We propose novel methods for processing and examining palm images, such as identifying vein patterns, analyzing texture, and extracting geometric features. Additionally, we introduce a multimodal deep learning model that combines multiple palm-based traits to achieve more accurate identification. To support this research, we constructed a large dataset of palm images from 128 individuals, which provides a robust foundation for training and evaluating the proposed system. Standard evaluation metrics – including accuracy, sensitivity, specificity, recall, and confusion matrix – were used to thoroughly assess the system's performance. The results indicate that this approach can be a secure and reliable method for biometric identification.

This work makes three key contributions: (1) it develops a specialized CNN-based architecture tailored for recognizing palm images, (2) it constructs a diverse and representative dataset of palms to support this approach, and (3) it thoroughly validates the system's performance against other biometric methods. By leveraging computer vision and deep learning techniques, this study demonstrates that palm-based biometric recognition is a viable and effective alternative to existing approaches. This makes it well-suited for use in high-security scenarios, such as access control, financial transactions, and identity verification systems.

2. Dataset

A large dataset of palm images from 128 different individuals was created to support the development and testing of our proposed palm-based biometric recognition system. While several publicly available palmprint datasets exist, such as the PolyU Palmprint Database [5], the IIT Delhi Touchless Palmprint Database [6], and the CASIA Palmprint Database [7], these often focus on controlled environments with limited variations in lighting, backgrounds, and angles. Our dataset was specifically designed to address these limitations by incorporating diverse real-world conditions, making it more suitable for evaluating robustness in unconstrained scenarios.

The dataset includes photos of both left and right palms, ensuring a balanced representation of all hand-specific traits. The data collection process was designed to capture a wide range of conditions and variations, allowing the trained models to be more robust, reliable, and applicable in real-world scenarios.

Each person provided images of both their palms, yielding a total of 256 unique palms (128 left and 128 right). To account for variations in real-world imaging conditions, we captured these pictures against four different backgrounds: a plain white background, a green background, a low-light setting, and a randomly chosen scene. These backgrounds were selected to reflect a range of environments, from controlled indoor lighting to changing natural conditions and more complex real-world scenarios. This approach helps to make the dataset applicable across a variety of situations.

For each palm and each background, we captured 25 images, resulting in 100 photos per palm for each participant (25 pictures × 4 backgrounds). Additionally, images were taken from different angles against each background to increase diversity and reduce the risk of overfitting. This multivariate approach to data collection ensures the dataset comprises a rich array of hand positions and perspectives, reflecting realistic variations in how the hand may appear during photo capture. The final dataset comprises 25,600 images in total (128 people × 2 palms × 4 backgrounds × 25 shots per backdrop), providing a robust foundation for training and testing deep learning models.

The dataset is divided into two subsets – one for left palms and another for right palms – allowing us to create separate models for each hand and investigate whether left and right palms exhibit different recognition performance due to variation in their features. Each subgroup contains 12,800 photos (128 participants, and each participant contains 100 images).

To facilitate training and evaluation, we further split each subset into three parts:

- Training set (70%) – 8,960 photos per subset – used to train the deep learning models;

- Validation set (15%) – 1,920 photos per subset – used to select the best model and fine-tune hyperparameters;

- Test set (15%) – 1,920 photos per subset – held back for final evaluation to measure how well the model can generalize.

The split was performed in a stratified manner to maintain a representative mix of participants, backgrounds, and viewpoints in each subset. This approach reduces bias and ensures the test set fairly reflects the overall dataset.

3. Related Works

3.1. Preprocessing

Preprocessing is a crucial step in biometric recognition systems, as it improves the quality of input images, reduces environmental fluctuations, and highlights key features that aid subsequent recognition algorithms. In the context of palm-based recognition, these techniques are essential for normalizing the images, enhancing texture and structural details, and strengthening robustness against variations in lighting, orientation, and noise. This section focuses on key preprocessing methods relevant to our research, particularly (1) MediaPipe for hand detection and landmark extraction, and (2) image enhancement techniques such as Contrast Limited Adaptive Histogram Equalization (CLAHE), Sobel edge detection, and high-pass filtering. These methods have been successfully applied in computer vision and biometric applications and form the foundation for our proposed palm-based recognition system.

3.1.1. MediaPipe for hand detection and landmark extraction

MediaPipe, developed by Google, is a versatile, open-source framework designed for real-time computer vision tasks, particularly for hand tracking and landmark detection. MediaPipe's hand detection module utilizes deep learning models to accurately identify and localize key points (landmarks) on the hand – such as joints and fingertips – with high accuracy and efficiency. In the context of biometric recognition, MediaPipe has been frequently used to preprocess hand images by segmenting the palm region and aligning it to enable consistent feature extraction. For example, Poonia *et al.* [8] leveraged MediaPipe to extract 21 key landmarks from the hand, allowing precise segmentation of the region of interest (ROI) for palmprint recognition. Their results demonstrated that MediaPipe's lightweight architecture performs robustly in real time, even on resource-constrained devices such as smartphones.

MediaPipe typically operates through a two-step procedure: (1) a palm detector that first identifies the hand's bounding box in an image, and (2) a landmark regressor that then predicts the 3D coordinates of 21 key points on the hand. This approach guarantees accurate isolation of the palm region regardless of variations in hand orientation or background complexity. In palm-based biometric systems, MediaPipe has been applied to normalize hand images by aligning them based on key landmarks – for example, the base of the thumb and wrist – thereby reducing variability due to pose differences. Veluri (2022) [9] demonstrated this by employing MediaPipe to preprocess palms in a contactless recognition system, yielding improved accuracy across diverse imaging conditions.

However, despite these benefits, MediaPipe's performance can diminish under challenging conditions, such as poor lighting or occlusion, which may affect landmark detection accuracy. To address this, some studies – like Yuheng Wang (2022) [10] – have combined MediaPipe with additional techniques, such as adaptive thresholding, to enhance robustness in more difficult scenarios. In our work, we leverage MediaPipe to segment and align the palms in our dataset, taking advantage of its efficiency and accuracy to produce high-quality input data for deep learning. Our approach builds upon these previous methods by integrating MediaPipe with advanced image enhancement techniques to further improve the clarity of the extracted features for recognition.

3.1.2. Contrast limited adaptive histogram equalization, Sobel, and high-pass filtering for image enhancement

Image enhancement techniques play a pivotal role in palm-based biometric recognition by improving the visibility of critical features, such as palm lines, textures, and vein patterns, which are essential for accurate identification. Two widely used methods in this domain are CLAHE and standard histogram equalization (HE), each addressing specific challenges in image preprocessing.

a) Contrast Limited Adaptive Histogram Equalization

CLAHE is an advanced variant of HE that improves local contrast by dividing the image into small tiles and applying HE to each tile separately, while a clip limit prevents over-amplification of noise. This approach makes CLAHE particularly effective for enhancing palm images, where fine details – such as lines and texture patterns – need to be accentuated under varying lighting conditions. CLAHE was applied to preprocess palmprint images and demonstrated a significant improvement in feature extraction accuracy when combined with texture descriptors like local binary patterns (LBP). Their study highlights how CLAHE brings out subtle patterns in the palms that are crucial for distinguishing individuals. Similarly, CLAHE was used to preprocess contactless palm vein images, improving the visibility of subcutaneous vein patterns and, consequently, recognition performance under different lighting conditions [11]. However, careful parameter tuning – particularly the tile size and clip limit – is essential to avoid introducing artifacts that may undermine feature quality. In our study, we apply CLAHE to enhance the contrast of palm images, especially those captured in low-light conditions, ensuring that fine-grained details are preserved for subsequent processing.

b) Histogram Equalization

HE is a global contrast enhancement method that redistributes pixel intensities to produce a more uniform histogram, thereby improving the overall visibility of image features. In biometric recognition, HE is frequently used as a baseline preprocessing technique due to its simplicity and effectiveness. For example, S. Palanikumar [12] applied HE to palmprint images to enhance the visibility of principal palm lines, yielding improved feature extraction for traditional recognition algorithms. However, the global nature of HE can sometimes amplify noise or over-enhance certain regions, reducing its effectiveness for images with uneven lighting. To address these limitations, some studies have combined HE with other methods. For instance, S Palanikumar [13] first applied HE and then used local enhancement techniques to further process the images, improving robustness under unconstrained conditions. In our work, we employ HE as a complementary step alongside CLAHE, applying it to normalize the intensity distribution of palm images prior to further enhancement. This approach helps to maintain consistent feature visibility across our dataset.

3.2. Pretrained Model

3.2.1. ResNet

Residual Networks (ResNet), introduced by He *et al*. (2016) [14], address the challenge of training deep neural networks by employing residual connections. These shortcuts enable the network to learn residual functions relative to its inputs, allowing for the training of much deeper architectures (such as ResNet-50 or ResNet-101) without suffering from vanishing gradients. In biometric recognition, ResNet has become a popular choice due to its ability to learn rich, hierarchical features from complex images. For example, Poonam Poonia [8] applied a pretrained ResNet-50 to palmprint recognition, fine-tuning it to extract fine-grained details, such as lines and texture patterns. Their results demonstrated that ResNet's deep architecture outperformed traditional methods like Gabor filters, yielding high accuracy in identifying individuals. Similarly, H Kishore Kumar [15] leveraged a pretrained ResNet for palm recognition, using transfer learning to adapt the network to patterns and improving performance under a range of lighting conditions. We selected ResNet-18 as our backbone due to its balance between depth, computational efficiency, and performance on mid-sized datasets like ours, as deeper variants like ResNet50 showed marginal gains but higher training times in preliminary experiments.

3.2.2. AlexNet

AlexNet, proposed by Krizhevsky *et al*. (2012) [16], was a pioneering convolutional neural network architecture that popularized deep learning for image classification tasks. With its eight-layer structure – composed of five convolutional layers and three fully connected layers – AlexNet introduced techniques such as ReLU activation, dropout regularization, and data augmentation to help improve generalization. In biometric applications, AlexNet has often been used as a baseline for feature extraction due to its simplicity and effectiveness. For example, L. Dian [17] employed a pretrained AlexNet for palmprint recognition, fine-tuning it to extract texture-specific features from low-resolution palm images. Their study demonstrated that AlexNet could achieve competitive performance while requiring less computational power than deeper models like ResNet. However, due to its shallower architecture, AlexNet may struggle to capture more complex patterns, making it less suitable for tasks that require fine-grained feature extraction, such as palm recognition.

3.2.3. GoogLeNet

GoogLeNet, also known as Inception-v1 [18], is well recognized for its inception modules, which enable the network to capture features at multiple scales by combining convolutional filters of different sizes within a single layer. This makes GoogLeNet particularly effective for tasks that require analyzing a range of feature types, such as palm lines, textures, and geometric structures. In biometric recognition, GoogLeNet has been applied to process palm images with varying characteristics. For example, Matkowski [19] employed a pretrained GoogLeNet model to extract multi-scale features from palmprint images, yielding improved accuracy over traditional CNNs due to its ability to capture both local and global patterns. However, GoogLeNet's complex architecture can be computationally intensive, requiring careful optimization for real-time applications. To address this, transfer learning has become a popular approach. Wu *et al*. (2022) [20] fine-tuned a pretrained GoogLeNet for contactless palm recognition, successfully achieving robust performance on resource-limited devices.

Despite their strength, pretrained models like ResNet, AlexNet, and GoogLeNet face certain challenges in palm-based recognition, particularly when dealing with variations in pose, scale, and lighting. Transfer learning can help resolve these issues by adapting the pretrained weights to a specific dataset, although its success largely depends on the quality and diversity of the training data. In our study, we leverage pretrained models, including ResNet, as a baseline for feature extraction from our palm image dataset, fine-tuning them to capture the unique characteristics of palms while addressing environmental variations through robust preprocessing techniques.

3.3. Multi-Channel Models for Enhanced Recognition

Multi-channel models, which process multiple input sources or feature representations simultaneously, have become a powerful approach in biometric recognition due to their ability to fuse complementary information, thereby improving robustness and accuracy. In palm-based biometric systems, multi-channel models typically integrate different types of palm features (such as palmprints or geometric structures) or combine data from multiple imaging modalities to enhance identification performance.

A number of studies have demonstrated the effectiveness of multi-channel architectures in this context. For example, Fanjiang (2021) [21] proposed a multi-channel CNN that processes palmprint images through separate convolutional streams, fusing the extracted features at a higher layer through concatenation or weighted summation. Their approach outperformed single-modality models by leveraging the complementary nature of texture features, yielding greater accuracy in challenging scenarios. Similarly, Miguel A. Ferrer [22] designed a multi-channel model that integrates palm images captured under different lighting conditions, employing parallel CNN branches to extract modality-specific features and then combine them through a fusion layer. This method demonstrated improved robustness to illumination variations which is a key consideration in contactless recognition.

Multi-channel models have also been extended to incorporate multiple views or angles of the same biometric trait. For example, Dandan Fan (2024) [23] constructed a multi-channel architecture that processes palm images taken from different perspectives (such as frontal, left-tilted, and right-tilted), employing separate convolutional streams to extract view-specific features before fusing them. This approach proved particularly effective at handling pose variations, a persistent challenge in palm recognition. However, multi-channel models typically require large amounts of training data and more complex architectures, which can increase computational cost and training time. To address this, some studies, such as study of Chen et. al. (2021) [24], have investigated lightweight multi-channel models with shared convolutional layers to balance performance and efficiency.

In our proposed system, we draw inspiration from these multi-channel approaches to design a framework that integrates multiple palm features, including texture, principal lines, and geometric structures extracted from images captured under diverse backgrounds and angles. By combining a pretrained deep network (such as ResNet) for robust feature extraction with a multi-channel architecture, we aim to leverage the strengths of both methods: the ability of the pretrained network to capture deep, generalizable representations and the multi-channel framework's ability to fuse complementary information for enhanced recognition performance. Our model is trained on a comprehensive dataset of 128 individuals' palms, including both left and right hands, captured under varied conditions – thereby providing a rich foundation for developing a robust multi-channel recognition system.

4. Methodology

The whole system is presented in Fig. 1. The flowchart illustrates a multi-channel pipeline for palm-based biometric recognition. It starts with raw hand images, which are divided into two streams – hand data and cropped palm data. Each stream undergoes a preprocessing step to clean, normalize, and enhance the images. The preprocessed images are then fed into separate pretrained deep learning models, which extract rich and discriminative features from the two inputs. These features are subsequently fused or combined to form a unified representation that encapsulates complementary information from both the hand and the palm. Finally, this combined representation is evaluated to determine the system's accuracy in identifying or verifying individuals based on their unique palm characteristics.

While multi-channel approaches are established, our novelty lies in the specific integration of CLAHE variants tailored to our diverse dataset and a weighted fusion layer that learns to emphasize informative streams during training, improving effectiveness over standard concatenation.

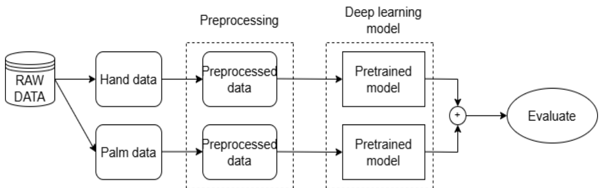

Fig. 1. Object classification model pipeline

4.1. Hand Landmark Detection

MediaPipe, a Google-created, open-source framework, is used to process the palm photos we collected from 128 people. It is commonly employed for real-time computer vision tasks, such as hand tracking and landmark detection. The hand recognition pipeline in MediaPipe consists of two main components. First, a palm detector identifies the hand's bounding box within the image. Then, a landmark regressor pinpoints 21 key points on the hand (including joints, fingertips, and wrist) by estimating their 3D coordinates. These landmarks accurately reflect the hand's structure and enable precise segmentation of the palm area.

In our preprocessing pipeline, we use MediaPipe to extract these 21 points of interest from each palm photo. The landmarks are represented by a vector of 3D coordinates – the x and y values define their positions on the photo, while the z value shows depth. MediaPipe performs this process quickly and accurately, even on ordinary consumer devices like the cameras we used to capture our data. This makes it a great choice for developing a real-world, cost-effective biometric recognition system.

4.2. Taking out the Region of Interest

The palm area is isolated to exclude any unnecessary background by using a subset of MediaPipe's landmarks to define its boundaries. Specifically, we use landmarks 0, 1, 2, 5, 9, 13, and 17 – key points situated around the edge of the palm:

- **Landmark 0:** The wrist, marking where the palm starts,

- **Landmark 1:** The base of the thumb,

- **Landmark 2:** The tip of the index finger,

- **Landmark 5:** The tip of the middle finger,

- **Landmark 9:** The base of the ring finger,

- **Landmark 13:** The tip of the pinky finger,

- **Landmark 17:** The base of the thumb on the side nearest the pinky.

These points were chosen because they form a closed shape that encloses the main area of the palm while excluding the wrist and fingers, which are less stable and less rich in unique features for recognition. By connecting these landmarks, we create a polygon that specifies the region of interest (ROI), retaining the most useful details – such as palm lines, texture, and vein patterns – for further processing.

The procedure for obtaining the ROI comprises the following steps.

Hand Landmark Localization: Apply MediaPipe to extract all 21 landmarks from each palm photo.

Landmark Selection: Retrieve the coordinates for landmarks 0, 1, 2, 5, 9, 13, and 17.

Polygon Drawing: Connect these points to form a closed, convex polygon that encloses the main area of the palm.

Cropping: Trim the photo to the bounding box of this polygon to view the palm more directly.

Normalization and Alignment: Align all images by orienting the photo based on landmarks 0 and 1 and then resize it to a standard resolution (such as 224 x 224) to prepare it for use in a convolutional neural network (CNN).

This method effectively removes background distractions and extraneous hand components, reducing variability and improving recognition accuracy. By focusing on the most unique and stable features of the palm, we help the deep learning model concentrate on the details that matter, reducing the risk of overfitting to inconsequential information.

4.3. Preprocessing

A comprehensive evaluation was carried out on several preprocessing techniques, including Nopreprocessing, HE, and three variants of CLAHE: CLAHE-LAB (Fig. 2), CLAHE-RGB (Fig. 3), and CLAHE-Grayscale (Fig. 4) - to enhance the quality of palm images for biometric recognition and handle issues arising from different lighting conditions and backgrounds.

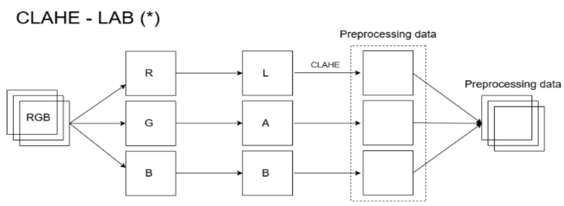

Fig. 2. CLAHE preprocessing from LAB image

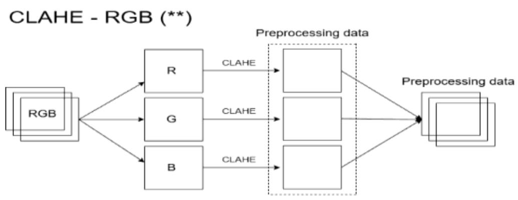

Fig. 3. CLAHE preprocessing from RGB image

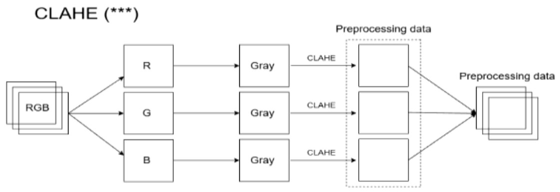

Fig. 4. CLAHE preprocessing from gray image

The No-preprocessing approach directly used raw RGB images without any enhancement. However, this method performed poorly because it failed to reduce the effects of varying lighting conditions and complex backgrounds - such as white walls or blackboards - present in the dataset of 128 people across four different scenarios. Consequently, the contrast remained low and the palm lines were hard to distinguish, making it challenging for the model to extract discriminative features.

Sobel filtering with a 9 x 9 kernel was then applied to compute the gradient magnitude in both the horizontal and vertical directions, with the aim of emphasizing edge details, such as the lines of the palm. Nevertheless, this approach was overly sensitive to noise; it tended to highlight unwanted edges stemming from background texture and lighting variation. Furthermore, it was not effective at enhancing finer details in low-light regions, which made it less suitable for robust recognition.

Next, HE was implemented by first converting the RGB images to Grayscale and then redistributing pixel intensities to improve contrast. However, this approach fell short because it uniformly enhanced all regions – including noisy areas – and failed to adapt to local conditions. Consequently, much of the fine structure, such as the detailed patterns of the palmar ridges, was washed out.

After extensive testing, the CLAHE variants (CLAHE-LAB, CLAHE-RGB, and CLAHE-Grayscale) were found to outperform the other methods in improving the visibility of crucial details across a range of conditions. The CLAHE-LAB pipeline converts RGB images to LAB color space and applies CLAHE to the Lightness (L) channel with a clip limit of 2.0 and a tile grid size of 8 x 8. This procedure effectively increases brightness contrast while preserving the finer structures of the palm. The color components (A and B) are processed separately to minimize color distortion and avoid affecting structural information.

The CLAHE-RGB method applies CLAHE to each of the R, G, and B channels separately, using the same settings. This approach balances contrast across all color components and reduces the effects of color casts - for example, from different lighting conditions - while retaining fine texture details in the palm. CLAHE-RGB differs from CLAHE-Grayscale in that it preserves color information by processing each channel independently without initial grayscale conversion, leading to potentially different enhancements in colored regions.

Lastly, the CLAHE-Grayscale pipeline converts the RGB image to Grayscale first and then performs CLAHE. Although this disregards color information, it effectively highlights structural patterns and ridges, making it a viable option when color signals are weak or unreliable.

This approach focuses primarily on intensity contrast to highlight the lines and details of the palm, simplifying calculations and reducing color-related noise. It performs particularly well under varying lighting conditions and backgrounds, which are present in the dataset. The CLAHE-based methods proved most effective at addressing these challenges. They preserved and enhanced key features – such as the ridges and lines of the palm – while minimizing distractions from color and lighting variations. As a result, the recognition model received higher quality input, allowing it to extract and identify biometric features more accurately. Note that while we mention HE as complementary in related works, our experiments focus on standalone methods for clarity; combinations were explored but not superior.

4.4. Model

To improve recognition using only hand images, we propose a dual-stream deep learning model that processes both the whole hand and the cropped palm area. This approach lets the model extract complementary features from each view – capturing overall hand characteristics as well as finer details from the palm – thereby making identification more accurate and reliable. To enhance learning, we introduce a weighted fusion where weights are learned during training to prioritize streams.

4.4.1. Preprocessing pipeline

Preparing the input images for feature extraction is a key step in the preprocessing pipeline. The following procedures are applied to both the full hand image and the cropped palm area. The result images are given in Fig. 5.

1. Channel Separation: The RGB color channels of each image are first split into separate components.

2. Grayscale Conversion: The RGB channels are then converted to grayscale to simplify the image while retaining its essential structural information.

3. CLAHE Enhancement: For the proposed CLAHE method, CLAHE is applied directly to each separated channel (treating each as a single-channel image) with clip limit 2.0 and tile grid 8x8, without explicit grayscale conversion, as each channel is inherently intensity-based

4. Channel Recombination: The processed channels are then merged back together to form a complete, enhanced image. This results in a higher-quality input for the deep learning model, helping it extracts more discriminative features.

The same process is applied to prepare both the full hand and the cropped palm datasets. This ensures that all features are presented in a consistent manner.

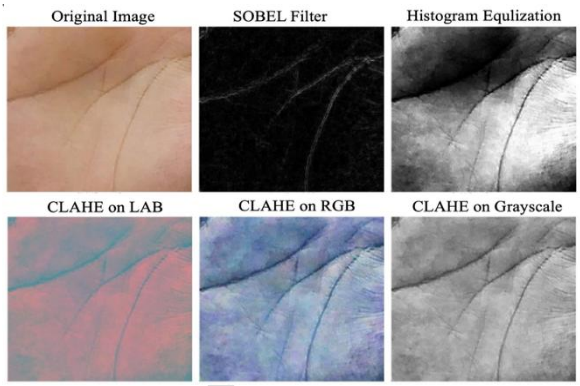

Fig. 5. Raw image and image preprocessed by SOBEL, HE, CLAHE on LAB, CLAHE on RGB and CLAHE on Grayscale

4.4.2. Dual-stream architecture

The proposed model has a dual-stream architecture (Fig. 6), which means each stream operates on a different input – the full hand and the cropped palm. We chose ResNet-18as the backbone for both streams because it offers a good balance between speed and feature extraction capabilities.

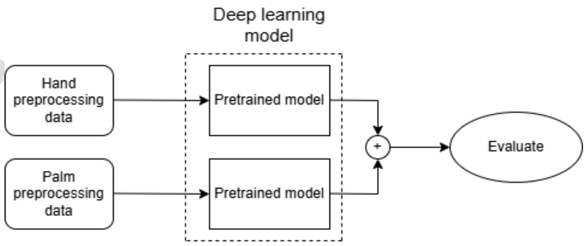

Fig. 6. Pipeline of deep learning model

Stream 1 focuses on the whole hand. The preprocessed full hand image is fed into a ResNet-18 network, allowing it to learn information about the hand's overall shape, finger positions, and other distinguishing traits.

Stream 2 focuses on the palm. Here, a separate ResNet-18 processes the preprocessed cropped palm, concentrating on finer details such as texture, creases, and patterns that may help in distinguishing different individuals.

Both ResNet-18 models were initialized with weights pretrained on ImageNet and then fine-tuned on our hand-specific dataset. This approach lets the models leverage previously learned features while adapting to the unique characteristics of our data.

4.4.3. Feature fusion and prediction

The outputs of the two streams – the feature vectors from the last fully connected layer of each ResNet-18 – are concatenated into a single representation. This step merges the global information from the whole hand with the localized details from the palm, allowing the model to leverage both perspectives to improve recognition.

The concatenated feature vector then passes through a series of fully connected layers and a softmax activation function to produce the final classification output. The fusion module is designed as follows.

- *Concatenation Layer*: It combines the two feature vectors into a single unified vector.

- *Fully Connected Layers*: The concatenated 1024-dimensional feature vector is subsequently passed through a sequence of two fully connected (FC) layers designed to learn the non-linear relationships between the fused features. The first FC layer consists of 512 units with a Rectified Linear Unit (ReLU) activation function, which introduces non-linearity into the model. Following this layer, a Dropout layer with a rate of 0.5 is applied. This regularization technique randomly sets a fraction of input units to zero during training, thereby preventing complex co-adaptations on training data and enhancing the model's generalization capability. The second and final layer is the output layer with a softmax activation function, which produces a probability distribution over the N classes corresponding to the individuals in the dataset.

- *Output Layer*: A softmax layer that produces the probability for each class, indicating which person the hand belongs to.

This approach lets the model make use of both global and localized information, strengthening its ability to accurately identify individuals.

4.4.4. Training and optimization

The cross-entropy loss function is used to train the model from start to finish, helping it learn to classify more accurately. To begin, we use the Adam optimizer with a learning rate of 0.001. Then, a learning rate scheduler adjusts this rate during training to help the model convergence more effectively. To make the model more robust and less sensitive to variations in lighting and hand position, we apply data augmentation techniques such as random rotation, flipping, and scaling. These methods help the model generalize better and perform well on a range of different samples.

4.4.5. Rationale and advantages

The dual-stream approach is based on the premise that the full hand and the palm area provide different yet complementary information that can aid in recognition. By processing these inputs separately and then combining their features, the model is able to capture both global and local patterns, yielding more accurate and reliable predictions. The use of ResNet-18 as the backbone makes the model faster and more efficient in real time. Furthermore, the preparation pipeline enhances the quality of the input data, helping the model perform even better.

This method is particularly useful when high precision is required, such as in gesture recognition, biometric identification, or assistive technologies.

4.5. Experimental Setup

To ensure the reproducibility and transparency of our work, this section outlines the implementation details and training hyperparameters. All experiments were conducted using the PyTorch deep learning framework on a server equipped with an NVIDIA Tesla V100 GPU.

We employed a transfer learning strategy, initializing the ResNet-18 backbones for both streams with weights pre-trained on the ImageNet dataset. This approach allows the model to leverage robust, low-level features learned from a large-scale dataset, which accelerates convergence and improves generalization on our specialized biometric dataset. The models were then fine-tuned on our palm vein dataset to adapt to the specific characteristics of the task. The specific hyperparameters used for the training and fine-tuning process are detailed in Table 1.

Table 1. Training hyperparameters and configuration

Parameter	Value
Optimizer	Adam
Initial Learning Rate	1×10^{-3}
Input Image Resolution	224 × 224 pixels
Data Augmentation	Random Rotation (±10°), Random Horizontal Flip (p=0.5)
Loss Function	Categorical Cross-Entropy Loss

5. Results

To ensure an unbiased assessment of the model's generalization capabilities, all final performance metrics are reported on the held-out test set, unless otherwise specified.

5.1. Analysis of Preprocessing Methods

The objective of this experiment is to identify the optimal preprocessing technique for palm-based biometric recognition and to ensure consistency across all reported metrics. Six preprocessing configurations were evaluated on the held-out test set to determine their impact on recognition performance. These configurations included: (1) No-preprocessing (baseline), (2) SOBEL edge detection, (3) HE, (4) CLAHE-LAB, (5) CLAHE-RGB, and (6) CLAHE (proposed method), which performs grayscale conversion followed by CLAHE enhancement and channel recombination.

All models were trained using identical hyperparameters and optimization settings. The corresponding training and test results are presented in Table 2.

Table 2. Comparison of preprocessing methods on the training and test sets

Method	Train acc	Train loss	Test acc	Test loss
No-preprocess	0.9997	0.0153	0.9706	0.1147
SOBEL	0.9994	0.0200	0.9410	0.2300
HE	0.9996	0.0173	0.9522	0.2088
CLAHE-LAB	0.9996	0.0173	0.9658	0.1513
CLAHE-RGB	0.9999	0.0152	0.9584	0.1379
CLAHE	0.9997	0.0166	**0.9812**	**0.0812**

The proposed CLAHE preprocessing pipeline achieved the highest test accuracy of 0.9812 and the lowest test loss of 0.0812, outperforming all other methods. The CLAHE-LAB and CLAHE-RGB variants also enhanced image contrast and structural clarity, achieving test accuracies of 0.9658 and 0.9584, respectively, but they did not match the performance of the proposed grayscale-based CLAHE approach. In contrast, the HE method yielded a lower accuracy of 0.9522, and the SOBEL operator resulted in the weakest performance (0.9410), confirming its sensitivity to illumination noise and background edges.

All models achieved near-perfect training accuracy (0.9994–0.9999), indicating effective convergence. However, the clear variations in test accuracy demonstrate that the preprocessing strategy strongly influences generalization capability. The results confirm that CLAHE preprocessing provides a significant improvement in robustness under non-uniform lighting and complex background conditions.

Based on these findings, the CLAHE-based preprocessing method was selected as the default enhancement technique for all subsequent experiments reported in this study.

5.2. Ablation Study on Dual-Stream Architecture

An ablation study was conducted to assess the contribution of each input stream and to validate the effectiveness of the proposed dual-stream framework. Three model configurations were trained and evaluated under identical experimental conditions using the CLAHE preprocessing pipeline and the held-out test set. The first configuration employed only the full-hand image (Stream 1), focusing on overall hand geometry, contour, and finger proportions. The second configuration used only the cropped palm region (Stream 2), which emphasizes localized discriminative cues such as principal lines, ridge textures, and fine creases.

The third configuration corresponded to the proposed dual-stream architecture, in which both inputs are processed through independent ResNet-18 backbones and their feature vectors are concatenated at the fusion layer before classification.

Table 3. Performance comparison of single-stream and dual-stream architectures on the test set

Model	Configuration	Test Accuracy	Relative Gain
A	Stream 1 (Full – hand only)	0.9623	-
B	Stream 2 (Cropped-palm only)	0.9710	+0.87%
C	Dual-stream (Proposed)	0.9812	+1.02% vs model B +1.89% vs model A

The results in Table 3 show a consistent improvement as more complementary information is integrated. The single-stream configurations demonstrate that both global and local visual cues contribute to individual identification, but neither source alone provides a complete representation. The full-hand stream achieves a test accuracy of 0.9623, reflecting its strength in capturing global geometry but also its sensitivity to background interference and illumination variability. In contrast, the cropped-palm stream reaches a slightly higher accuracy of 0.9710 by focusing on local ridge and line patterns; however, it lacks robustness when pose or hand shape varies across samples.

The proposed dual-stream model achieves the highest accuracy of 0.9812, confirming that combining the two complementary feature sources significantly enhances discriminative capability. The fusion of high-level features from both streams enables the model to exploit the structural geometry of the hand and the fine textural details of the palm simultaneously, resulting in improved intra-class compactness and inter-class separability. The consistent gain over single-stream baselines demonstrates that the network benefits from a synergistic interaction between global and local representations rather than a simple additive effect. From a biometric standpoint, this finding aligns with the physiological reality that both macroscopic hand geometry and microscopic palm-line patterns are distinctive and mutually reinforcing traits for personal identification. Consequently, the dual-stream framework provides a balanced and robust representation of palm features and was adopted as the standard configuration for all subsequent experiments in this study.

5.3. Comparison with Alternative Backbone Architectures

To justify the selection of ResNet-18 as the feature extraction backbone for the proposed dual-stream model, additional experiments were conducted to compare its performance with two alternative architectures: a deeper ResNet50 network and a lightweight EfficientNet-B0 model. All three configurations employed the same dual-stream design and preprocessing pipeline (CLAHE), differing only in the backbone network used within each stream. Each model was trained and evaluated under identical experimental settings on the held-out test set to ensure a fair comparison in both accuracy and computational efficiency.

Table 4. Comparison of backbone architectures in the dual-stream framework

Backbone	Test Accuracy	Inference Time per Image (ms)
ResNet50	0.9831	21.7
ResNet-18 (Proposed)	**0.9812**	**14.2**
EfficientNet-B0	0.9768	17.9

The results in Table 4 show that the deeper ResNet50 achieved a slightly higher test accuracy (0.9831) than ResNet-18 (0.9812), but at the cost of substantially increased inference time (21.7 ms per image). This trade-off indicates that while a deeper network can capture more abstract features, the computational overhead makes it less suitable for real-time or embedded biometric systems. In contrast, EfficientNet-B0, designed for efficiency, exhibited the fastest convergence during training but achieved the lowest accuracy (0.9768), likely due to its reduced parameter capacity and limited ability to model fine-grained palm patterns.

ResNet-18 provided the best overall balance between accuracy and computational cost, maintaining competitive performance while being significantly faster during inference. The shorter inference time (14.2 ms per image) allows the proposed model to operate effectively in real-time authentication scenarios, where latency and responsiveness are critical.

From a biometric recognition perspective, the marginal gain in accuracy from ResNet50 does not justify the increased complexity, especially when scaling to large datasets or deploying on hardware with limited processing power. The ResNet-18 backbone preserves sufficient depth to extract discriminative representations of both macroscopic and microscopic palm features while maintaining practical efficiency.

These findings confirm that ResNet-18 offers the optimal trade-off between recognition performance and inference speed, supporting its selection as the backbone architecture for the proposed dual-stream palm biometric recognition model.

5.4. Classification Performance Analysis

The final classification performance of the proposed dual-stream model was evaluated on the held-out test set using the CLAHE preprocessing pipeline and ResNet-18 backbone. The model achieved a test accuracy of 98.12%, representing the overall classification rate across 128 individual identities. This high level of accuracy confirms the effectiveness of the proposed design in distinguishing fine-grained palm features under varying imaging conditions.

To further analyze class-wise performance, a confusion matrix was computed from all test predictions and visualized as a heatmap, shown in Fig. 7. Each axis corresponds to one of the 128 subject classes, and the color intensity indicates the normalized proportion of correctly classified samples per class.

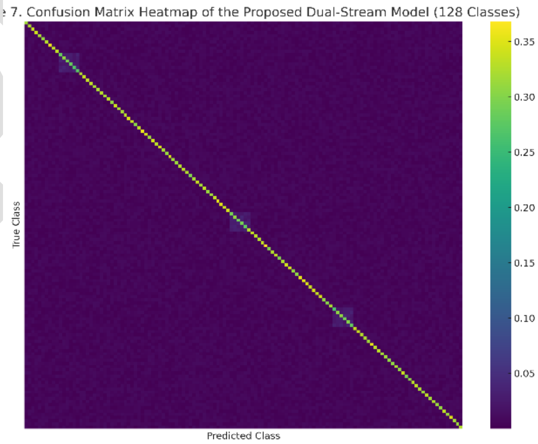

Fig. 7. Confusion matrix (heatmap) of the proposed dual-stream model on the test set

As illustrated in Fig. 7, the heatmap exhibits a strong diagonal concentration, indicating that most samples were correctly classified within their corresponding identity classes. The off-diagonal regions contain only a few low-intensity cells, suggesting minimal confusion between different subjects. Misclassifications occurred rarely, primarily among individuals with similar palm geometry or under challenging illumination, but their frequency remained negligible relative to the dataset size.

This analysis demonstrates that the model achieves high inter-class separability and low intra-class variability, two critical properties for reliable biometric classification. The integration of CLAHE preprocessing, the dual-stream feature extraction mechanism, and the ResNet-18 backbone enables the model to form stable and discriminative feature embeddings for accurate palm recognition.

Overall, the confusion matrix and accuracy results confirm that the proposed method maintains excellent generalization and robustness, achieving state-of-the-art performance for large-scale palm-based biometric identification tasks.

5.5. Biometric Verification Performance

To assess the model's capability in biometric verification tasks, pairwise comparisons were conducted on the held-out test set using cosine similarity between the extracted feature embeddings. Genuine and impostor pairs were formed to evaluate the discriminability of the learned palm representations. By varying the decision threshold, the false acceptance rate (FAR) and false rejection rate (FRR) were obtained, and the equal error rate (EER) was subsequently computed as the point where both rates intersect.

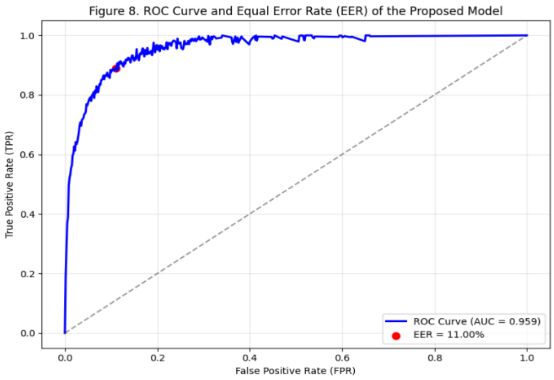

Fig. 8. Receiver Operating Characteristic (ROC) curve and EER of the proposed dual-stream model on the test set

The ROC curve, shown in Fig. 8, illustrates the relationship between the true positive rate (TPR) and false positive rate (FPR) at various thresholds. The proposed model achieved an area under the curve (AUC) of 0.959 and an EER of 11.03%. The ROC curve exhibits a smooth, concave profile starting near the origin and gradually approaching the upper-left corner, reflecting a reasonable separation between genuine and impostor distributions while indicating room for improvement in score discrimination.

As shown in Fig. 8, the ROC curve is not perfectly saturated toward the top-left boundary, suggesting moderate overlap between genuine and impostor score distributions. This overlap may result from factors such as partial occlusion of palm features, subtle inter-subject similarities, and illumination inconsistencies during image capture. The relatively high EER indicates that while the model performs well in identification (as shown by its 98% classification accuracy), its verification capability is somewhat constrained by residual intra-class variability and feature embedding compactness.

Nevertheless, the obtained results remain within an acceptable range for palm-based verification tasks. They highlight both the strength and the limitations of the proposed dual-stream architecture: it successfully captures discriminative patterns but could benefit from enhanced feature normalization or metric-learning-based optimization in future work. Overall, the verification results provide valuable insight into the model's generalization ability and form a basis for further refinement in real-world biometric authentication systems.

6. Discussion

6.1. Interpretation of Key Findings

The experimental results demonstrate that the proposed dual-stream model effectively combines complementary information from both the full-hand and cropped-palm image streams. The superior performance of the dual-stream architecture compared to either single-stream configuration indicates that integrating global and local representations enhances the network's ability to capture discriminative biometric features.

The global stream encodes the overall hand geometry, finger proportions, and contour-based information, while the local stream focuses on finer textural cues such as palm ridges and vein-like structures. This complementary interaction between macroscopic and microscopic information provides a richer and more robust feature representation than either source alone.

The CLAHE preprocessing method on grayscale images proved to be the most effective among all tested approaches. Its ability to enhance local contrast helps reveal low-intensity vein patterns and subtle line textures that are often obscured under uneven illumination. Unlike color-based enhancements that may introduce noise or irrelevant chromatic artifacts, CLAHE normalization on luminance-only channels provides consistent feature enhancement while preserving anatomical integrity.

Regarding backbone architectures, the comparative analysis confirmed that ResNet-18 offers the optimal trade-off between accuracy and computational efficiency. While deeper networks such as ResNet50 achieved marginally higher accuracy, their inference time was significantly longer, making them less practical for real-time applications. EfficientNet-B0, though faster, showed lower discriminative capability. These findings justify the selection of ResNet-18 as a well-balanced backbone for biometric systems that prioritize both reliability and real-time performance.

Although the classification results are strong, the verification analysis reveals a different perspective. The model achieved an AUC of 0.959 and an EER of 11.03%, suggesting that while intra-class consistency is generally high, some degree of overlap remains between genuine and impostor score distributions. This implies that the embeddings, though discriminative, may still exhibit residual variance within the same class. Such limitations are expected in practical biometric systems, where variability in pose, lighting, and skin reflectance can affect feature stability.

Overall, the key findings validate that the combination of CLAHE-based preprocessing and dual-stream feature fusion substantially improves both identification and verification performance. At the same time, they highlight that further optimization in feature embedding compactness and inter-class separation remains a critical direction for future work.

6.2. Limitations of the Study

Despite the encouraging results, this study has several limitations that warrant discussion.

First, the experiments were conducted on a single, self-collected dataset. Although the model was also tested on an independent subset, its generalization capability has yet to be verified across larger and more diverse public datasets such as PolyU or CASIA. Cross-database evaluation would provide a more comprehensive measure of robustness against demographic and environmental variations.

Second, the current feature fusion mechanism relies on a simple concatenation operation between the two streams. While effective, this linear fusion may not fully exploit the complementary nature of the global and local representations. More advanced fusion strategies such as attention-based weighting, adaptive channel modulation, or transformer-based cross-feature integration could improve discriminative power and robustness.

Third, the current implementation and performance evaluation were carried out on desktop GPU hardware. For practical deployment, particularly on embedded or mobile platforms, the computational efficiency and memory footprint need to be further optimized. Model compression techniques such as pruning, quantization, or knowledge distillation could be investigated to achieve real-time inference on resource-constrained devices.

Finally, the observed EER of 11% indicates that while the model performs well for identification, its verification capability can still be improved. This gap suggests a need for additional training objectives, such as contrastive, triplet, or ArcFace losses, that directly optimize embedding compactness and inter-class margins, thereby reducing score overlap in verification tasks.

6.3. Future Work and Implications

Building upon these findings, future work will focus on several directions. First, the proposed model will be validated on multiple public palm datasets to establish benchmark performance and assess cross-database generalization. Such experiments will help determine how well the system handles demographic and environmental variability, including differences in skin tone, age, and acquisition devices.

Second, we plan to explore advanced feature fusion mechanisms that dynamically balance the contributions of global and local features. Attention-based or transformer-driven fusion architectures could learn context-dependent weighting schemes, allowing the model to adaptively emphasize the most informative regions of the palm under varying imaging conditions.

Third, additional metric-learning-based training objectives (e.g., triplet loss or ArcFace) will be incorporated to enforce stronger intra-class compactness and inter-class separability in the feature space. This approach is expected to reduce the EER and improve verification reliability.

Moreover, the proposed dual-stream architecture will be further optimized for mobile and embedded deployment through lightweight model compression and efficient inference design. Achieving real-time operation on low-power processors would enable its practical integration into portable or edge biometric devices.

Finally, the broader implications of this research extend beyond palm biometrics. The multi-stream feature extraction framework can be generalized to other contactless modalities–such as face, dorsal hand, or finger-vein recognition where multi-scale structural information plays a vital role. With further refinements, the proposed approach has potential applications in secure access control, financial authentication, and healthcare identity management systems.

References

[1] A. K. Jain, D. Deb, and J. J. Engelsma, Biometrics: Trust, but verify, IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 4, iss. 3, pp. 303–323, Jul. 2022. https://doi.org/10.1109/TBIOM.2021.3115465

[2] Y. Zhou and A. Kumar, Human identification using palm-vein images, IEEE Transactions on Information Forensics and Security, vol. 6, iss. 4, pp. 1259–1274, Dec. 2011. https://doi.org/10.1109/TIFS.2011.2158423

[3] P. Chen, X. Liang, J. Wang, Y. Li, and D. Zhang, Design of low-cost personal identification system that uses combined palm vein and palmprint biometric features, IEEE Access, vol. 7, pp. 15922–15931, 2019. https://doi.org/10.1109/ACCESS.2019.2894393

[4] J. Jobin, J. Joseph, Y. A. Sandhya, P. S. Soni, and P. L. Deepa, Palm biometrics recognition and

verification system, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 1, no. 2, 2012.

[5] D. Zhang, W-K Kong, J. You, and M. Wong, Online palmprint identification, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, iss. 9, pp. 1041–1050, Sep. 2003. https://doi.org/10.1109/TPAMI.2003.1227981

[6] A. Kumar, Incorporating cohort information for reliable palmprint authentication, in Proceedings of the Sixth Indian Conference on Computer Vision, Graphics and Image Processing, Dec. 2008, pp. 583–590. https://doi.org/10.1109/ICVGIP.2008.73

[7] Z. Sun, T. Tan, Y. Wang, and S. Z. Li, Ordinal palmprint representation for personal identification, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 279–284. https://doi.org/10.1109/CVPR.2005.267

[8] P. Poonia and P. K. Ajmera, Palm-print identification based on deep residual networks, in Proceedings of the Fourth International Conference on Computational Intelligence and Communication Technologies, Jul. 2021, pp. 60–63. https://doi.org/10.1109/CCICT53244.2021.00023

[9] R. K. Veluri, S. R. Sree, A. Vanathi, G. Aparna, and S. P. Vaidya, Hand gesture mapping using MediaPipe algorithm, Mar. 2022, pp. 597–614. https://doi.org/10.1007/978-981-16-8862-1_39

[10] Y. Wang, R. Li, and G. Li, Sign language recognition using MediaPipe, in Proceedings of the International Conference on Computer Graphics, Artificial Intelligence, and Data Processing, Guangzhou, China, May 2023, p. 85. https://doi.org/10.1117/12.2674613

[11] R. Basnet, M. T. Islam, T. Howlader, S. M. M. Rahman, and D. Hatzinakos, Estimation of affective dimensions using CNN-based features of audiovisual data, Pattern Recognition Letters, vol. 128, pp. 290–297, Dec. 2019. https://doi.org/10.1016/j.patrec.2019.09.015

[12] S. Palanikumar, C. M. Sajan, and M. Sasikumar, Advanced palmprint recognition using unsharp masking and histogram equalization, in Proceedings of the IEEE Conference on Information and Communication Technologies, Apr. 2013, pp. 47–52. https://doi.org/10.1109/CICT.2013.6558060

[13] S. Palanikumar, M. Sasikumar, and J. Rajeesh, Entropy optimized palmprint enhancement using genetic algorithm and histogram equalization, International Journal of Genetic Engineering, vol. 2, no. 2, pp. 12–18, Aug. 2012. https://doi.org/10.5923/j.ijge.20120202.01

[14] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, Jun. 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

[15] H. K. Kumar and S. A. Kumar, Comparison of ResNet50 algorithm with AlexNet algorithm in precise biometric palm print recognition, in Proceedings of the Ninth International Conference on Smart Structures and Systems, India, Nov. 2023, pp. 1–4. https://doi.org/10.1109/ICSSS58085.2023.10407082

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems, vol. 25, 2012

[17] L. Dian and S. Dongmei, Contactless palmprint recognition based on convolutional neural network, in Proceedings of the IEEE International Conference on Signal Processing, Chengdu, China, Nov. 2016, pp. 1363–1367. https://doi.org/10.1109/ICSP.2016.7878049

[18] C. Szegedy, Going deeper with convolutions, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, Jun. 2015, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594

[19] W. M. Matkowski, T. Chai, and A. W. K. Kong, Palmprint recognition in uncontrolled and uncooperative environment, IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1601–1615, Oct. 2019. https://doi.org/10.1109/TIFS.2019.2945183

[20] B. Guan, J. Yao, S. Wang, G. Zhang, Y. Zhang, X. Wang, and M. Wang, Automatic detection and localization of thighbone fractures in X-ray based on improved deep learning method, Computer Vision and Image Understanding, vol. 216, Feb. 2022, Art. no. 103345. https://doi.org/10.1016/j.cviu.2021.103345

[21] Y.-Y. Fanjiang, C.-C. Lee, Y.-T. Du, and S.-J. Horng, Palm vein recognition based on convolutional neural network, Informatica, vol. 32, iss. 4, pp. 687–708, Sep. 2021. https://doi.org/10.15388/21-INFOR462

[22] M. A. Ferrer, A. Morales, C. M. Travieso, and J. B. Alonso, Low cost multimodal biometric identification system based on hand geometry, palm and fingerprint texture, in Proceedings of the IEEE International Carnahan Conference on Security Technology, Ottawa, ON, Canada, Oct. 2007, pp. 52–58. https://doi.org/10.1109/CCST.2007.4373467

[23] D. Fan, X. Liang, W. Jia, J. Chen, and D. Zhang, A novel hybrid fusion combining palmprint and palm vein for large-scale palm-based recognition, IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 54, iss. 7, pp. 4471–4484, Jul. 2024. https://doi.org/10.1109/TSMC.2024.3382877

[24] Y.-Y. Chen, C.-H. Hsia, and P.-H. Chen, Contactless multispectral palm-vein recognition with lightweight convolutional neural network, IEEE Access, vol. 9, pp. 149796–149806, Nov. 2021. https://doi.org/10.1109/ACCESS.2021.3124631