International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(9s) 2026 pp. 1279-1296
https://doi.org/10.70917/ijcisim-2026-3553
Article

MMFBA_ESTNet: An Enhanced Swin Transformer – Based Multi-Model Finger Biometric Authentication Framework

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

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

Abstract: Biometric identification is currently the subject of active research to prevent unauthorized use. Finger biometrics creates spoofing attacks by creating fake biometric finger images. Distinguishing a true biometric image from a fake one with the naked eye is a difficult process. Therefore, the proposed work has introduced a highly effective multi-model-finger biometric authentication (MMFBA) based on finger vein, fingerprint, and knuckle. The proposed model has relied on a Densely connected semantic dilated multi-head attention conventional-based Enhanced Swin Transformer (EST) network i.e. named MMFBA_ESTNet. This provides the best architecture for low-level (edges, corners, lines, and texture), mid-level (combined features), and high-level (complex patterns) feature learning. In addition, multi-head attention helps to establish priorities to improve learning efficiency. The implementation of the proposed profile is performed on the Python platform and evaluated using parameters such as accuracy, precision, F1, area under the curve, false positive rate, false rejection rate, and false acceptance rate. The results of the experiment show that the proposed strategy achieved 0.986 accuracy at a 70% learning rate and 0.99 accuracy at an 80% learning rate.
Keywords: Biometric Identification, Finger Vein Recognition, Finger Print Recognition, Knuckle Biometrics, Swin Transformer, and Generative AI.

1. Introduction

Multimodal biometric systems offer improved security compared to PIN-based authentication. They employ multiple biometric modalities, increasing accuracy, resilience, and protection [1]. These systems operate at numerous fusion levels, ensuring reliable and valid authentication while guarding against credential forging and spoofing attacks. Physiological characteristics enhance biometric security by making them unique to each individual and difficult to falsify [2]. By collecting certain angles, sizes, and finger patterns, fingerprint (FP-FV) identification enhances biometric security when paired with deep learning algorithms and computer vision sensor technologies [3]. Biometrics has broader applications in personalized healthcare patient-specific data for tailored medical interventions and predictive analytics. Digital transactions and cybercrime risks increase, and biometric systems are crucial for financial transactions, institutional security, and safe access control in digital environments [4].

Biometric recognition systems face challenges in security, privacy, and flexibility. Fingerprint-based biometric systems are vulnerable to spoofing attacks, necessitating advanced liveness detection and anti-spoofing measures [5]. Privacy and data security concerns arise due to the irreparable nature of biometric information, making encryption, secure storage, and privacy-preserving mechanisms crucial [6]. Achieving uniform precision is challenging due to variations in ambient conditions, sensor quality, and physiological characteristics. Balancing security with user convenience is challenging, with false rejection rates sensor breakage, and delays during authentications detracting from user experience [7,8]. The widespread usage of biometric recognition has ethical and legal ramifications, including consent, surveillance, and misuse of biometric data. To guarantee appropriate data distribution and security, stricter regulations and compliance frameworks are required [9].

The Adaptive Histogram-Based Gamma and Sharpening Filter Combination (AHistGSFC) [10] technique enhances image quality by integrating histogram equalization, gamma correction, image sharpening, multi-filtering, and contrast adjustment. It uses finger vein biometrics and finger texture to create a multimodal biometric system based on near-infrared finger images [11]. A fuzzy method is used to combine single biometric results with improved recognition performance. The Finger Asymmetric Backbone Network (FAB-Net) [12] To minimise network width while retaining high identification accuracy by learning discriminative intra-modal characteristics. Additionally, the method integrates 3D hand geometry with 3D palmprint features [13].

The paper introduces Multimodal Fingerprint and Finger Vein Biometric Authentication with Enhanced Swin Transformer Network (MMFBA_ESTNet), a multimodal biometric authentication scheme that combines fingerprint and finger vein biometrics through Near-Infrared (NIR) imaging. This approach provides spoofing and environmental robustness, while also enhancing recognition performance and computational speed [14]. The Dense Semantic Dilated Multi-Head Attention module supports feature extraction, identifying local and global relationships among different biometric modalities. The extracted features are processed by a Swin Transformer-based network through hierarchical representations enhancing recognition performance and computational speed. The biometric features are combined with an adaptive decision strategy for enhanced authentication robustness [15].

The major contributions of the paper are as follows:

- The proposed method combines fingerprint, knuckle, and finger vein biometrics to improve security and resilience against spoofing attempts, offering a more thorough and trustworthy authentication solution.
- The MMFBA_ESTNet ensures superior feature extraction and classification performance by efficiently learning low-level (edges, textures), mid-level (combined features), and high-level (complex patterns) biometric features through the use of a densely connected semantic dilated architecture, multi-head attention, and Swin Transformer.
- The segmentation (EST) and feature extraction (SIFT, I-AE) processes are carried out concurrently in the parallel flow that the proposed MMFBA_ESTNet applies. HoG features are produced using segmentation, and VGG16 is used to classify the features that are extracted. An MLP processes the concatenated HoG and classification outputs, improving biometric authentication speed and accuracy.

The following sections are organized as follows: Section 2 explores relevant research and literature reviews, Section 3 introduces the proposed framework, Section 4 provides a detailed analysis of the observed results and discussions, and Section 5 offers the final assessment of this study.

2. Literature Survey

Some of the recent research works related to Finger Biometric Authentication were reviewed in this section.

In 2024, Alshardan et al. [16] explored three fusion strategies used in security applications: early fusion, late fusion, and score-level fusion. Early fusion integrates raw images at the input layer of a single CNN, combining information at the initial stages. Late fusion, in contrast, merges feature after individual learning from each CNN model. Score-level fusion employs weighted aggregation to combine scores from each modality, leveraging the complementary information they provide. they also use contrast limited adaptive histogram equalization (CLAHE) to enhance fingerprint contrast and vein pattern features, improving feature visibility and extraction.

In 2024, Nair and Dharan [17] proposed a Secure and Efficient Face-Fingerprint Convolutional Neural Network with a Bat Optimization (SEFF-CNN-BO) model to counter high error rates and spoofing attacks by combining Bat Optimization Secure Hash Algorithm 256-bit (SHA-256) hashing, and Attribute-Based Encryption (ABE) for secure key generation. However, managing encryption key complexity remains a challenge for large-scale applications necessitating a trade-off between security and computational efficiency.

In 2023, Abdullahi et al. [18] proposed filtered spatial and temporal multimodal fingerprint and finger vein network (FS-STMFPFV-Net) is achieved from two-channel independent learning to improve image variabilities. In the first channel, the image sequence is generated by aligning the fingerprint and finger vein images together inside the image generator. The sequences are built into the five layers of a deep convolution neural network fusion model to extract spatial sequence wise features. The second channel is where the extracted sequence-wise features are remembered in the long short-term memory for interactions between temporal and sequence dimensions which finally generate their long-term temporal dependencies as complementary information.

In 2024, Salturk and Kahraman [19] introduced a low-cost biometric verification system using signature and face recognition, utilizing models like CNN, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Temporal Convolutional Network (TCN). However, this method is limited to online environments and struggles in offline security situations. Research explores multimodal methods applicable in both online and offline environments, enhancing flexibility for various authentication scenarios.

In 2025, Gizachew Yirga et al. [20] proposed a novel approach to cryptographic key generation using biometric data from face and finger vein modalities enhanced by deep learning techniques. Using pretrained models FaceNet and VGG19 for feature extraction and employing a Siamese Neural Network (SNN), the study demonstrates the integration of multimodal biometrics with fuzzy extractors to create secure and reproducible cryptographic keys. Feature fusion techniques, combined with preprocessing and thresholding, ensure robust feature extraction and conversion to binary formats for key generation.

In 2024, Singh and Barde [21] developed a Gabor filter to reduce computational expenses in multimodal biometric systems. This method enhances feature extraction and recognition accuracy but faces challenges with varied biometric inputs. Enhancements focus on increasing versatility for wider applications and creating adaptive feature extraction methods to further optimize performance.

In 2024, Singh and Tiwari [22] proposed a dual multimodal biometric authentication system that uses Electrocardiogram (ECG), sclera, and fingerprint verification. However, system complexity limits large-scale deployment. Novel techniques are needed for seamless implementation without performance loss. Research should focus on optimizing fusion approaches for real-time verification and reducing hardware dependency for practical viability.

In 2025, Sumalatha et al. [23] proposed a multimodal biometric system that uses ECG, fingerprint, and FKP for security in risky areas. The system resists spoofing and enhances data quality through preprocessing and augmentation. The score-level fusion favors fingerprint and Finger-Knuckle Print (FKP) for greater accuracy achieving low error rates.

In 2023, Hsia et al. [24] proposed model incorporates a diverse branch block (DBB), adaptive polyphase sampling (APS), and coordinate attention mechanism (CoAM) with the aim of improving the model's performance in accurately identifying finger vein features. To evaluate the effectiveness of the model in finger vein recognition, we employed the finger-vein by university sains malaysia (FV-USM) and PLUSVein dorsal-palmar finger-vein (PLUSVein-FV3) public database for analysis and comparative evaluation with recent research methodologies.

In 2024, Nour et al. [25] proposed system begins with the preprocessing mechanism that is applied to enhance the quality of the acquired images, including contrast enhancement and noise reduction, by using some filters such as Median and Contrast Limited Adaptive Histogram Equalization (CLAHE). Next, a deep learning model, such as a convolutional neural network (CNN), is employed to automatically abstract discriminative features from the preprocessed vein images. The empirical outcomes prove the influence and reliability of the proposed technique for vein recognition, making it a promising solution for biometric authentication systems.

2.1 Research Gap

Multimodal biometric authentication schemes improve security and accuracy but face limitations like computational complexity, memory requirement, vulnerability to spoofing attacks, and hardware requirements. They are limited in real-time processing and non-adaptive to evolving user behavior, highlighting the need for more resilient, efficient, and adaptable biometric security solutions. To overcome these issues, researchers focused on developing ultra-light, secure biometric authentication platforms using deep learning architectures, behavioral biometrics, liveness checks, and crypto-boosting technologies. Hybrid authentication schemes integrating physiological, behavioral, and contextual information can enhance security and privacy retention. This leads to more reliable real-time and large-scale biometric authentication systems.

3. Proposed Methodology

The proposed MMFBA_ESTNet integrates fingerprint, knuckle, and finger vein biometrics to present a reliable multi-model finger biometric authentication system. To efficiently extract hierarchical features, it uses a Densely Connected Semantic Dilated Multi-Head Attention-based EST Network. To improve accuracy, low-level features (textures, edges), mid-level features (combined structures), and high-level features (complex patterns) are all effectively captured. By giving priority to crucial traits, the multi-head attention mechanism increases learning effectiveness and resilience to spoofing assaults based on generative artificial intelligence. The overall architecture of the proposed methodology is shown in Figure 1.

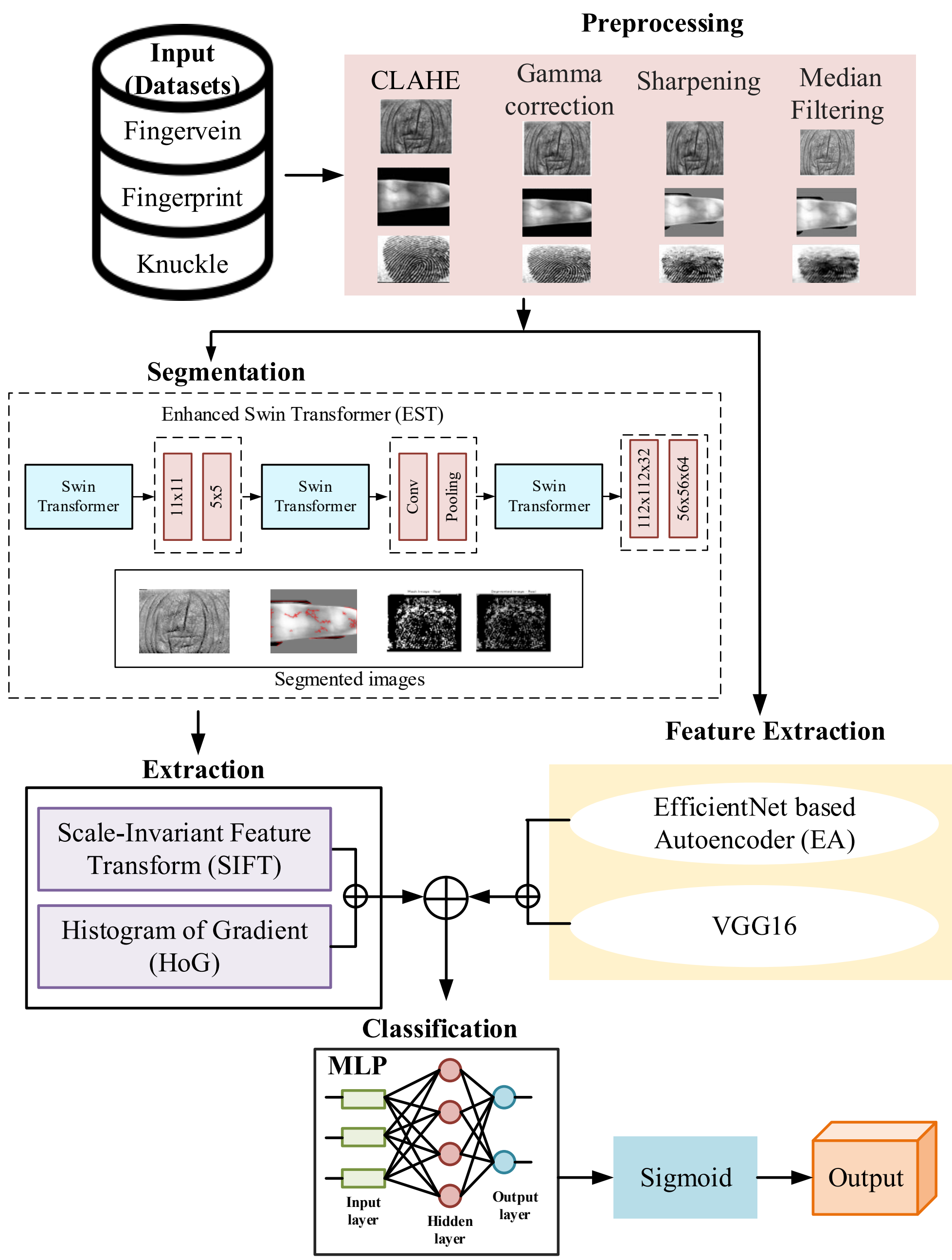
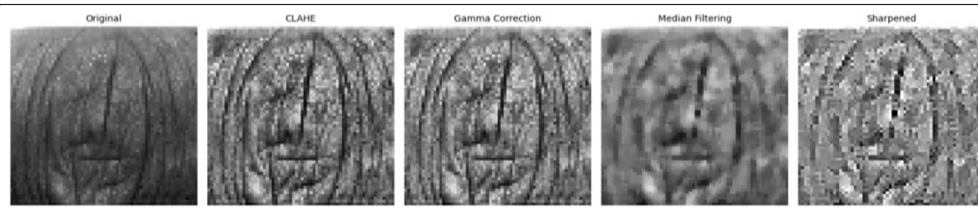


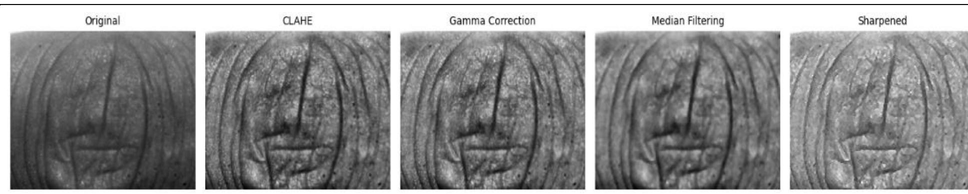
Figure 1: Overall architecture of the proposed methodology

3.1 Preprocessing

Medical images are improved by preprocessing to aid in diagnosis and analysis. Sharpening enhances edges, median filtering lowers noise, gamma correction modifies brightness, and contrast-limited Adaptive Histogram Equalization (CLAHE) improves contrast. By improving image quality, these methods enable deep learning models to diagnose diseases and extract features accurately.

- **CLAHE:** It is an improved contrast enhancement method that reduces noise amplification while enhancing image features. In contrast to global histogram equalization, it distributes pixel intensities to improve local contrast by processing tiny sections (tiles) independently. It applies a contrast limit, clipping histogram values and redistributing extra pixels to prevent excessive noise amplification. CLAHE enhances visibility in low-light conditions while maintaining crucial structural features, making it a popular choice for low-light photography, satellite photos, and medical imaging.
- **Gamma Correction:** A nonlinear image processing method for adjusting contrast and brightness is called gamma correction. Pixel intensity is changed using where darkens the image and brightens it. This method accounts for human visual perception, which is more sensitive to dark areas and displays non-linearity. It is essential for photography, video processing, and medical imaging because it makes images look natural and well-lit on a variety of display screens.
- **Median Filtering:** A nonlinear noise reduction method called median filtering substitutes the median value of each pixel's immediate vicinity for each pixel. It works very well to eliminate salt-and-pepper noise while keeping fine details and significant edges. Median filtering preserves sharp transitions, which makes it advantageous for object detection, remote sensing, and medical imaging in contrast to mean filtering, which blurs images. To improve clarity and lessen noise artifacts, the filter is applied using a kernel (such as 3x3 or 5x5) that scans the entire image.
- **Sharpening:** makes features more prominent and improves image details by boosting edge contrast. It is accomplished by employing methods that highlight high-frequency components, such as the Laplacian filter or unsharp masking. Laplacian filter detects variations in intensity, whereas unsharp masking enhances edges by subtracting a fuzzy version from the original image. This method is frequently used to enhance the clarity of blurry images in computer vision, images, and medical imaging. However, over-sharpening necessitates careful parameter calibration because it might generate noise and artifacts.

<table>
<tr><th>Categories</th><th>Images</th></tr>
<tr><td rowspan="2">Knuckle images</td><td>

Fake</td></tr>
<tr><td>

Real</td></tr>
</table>

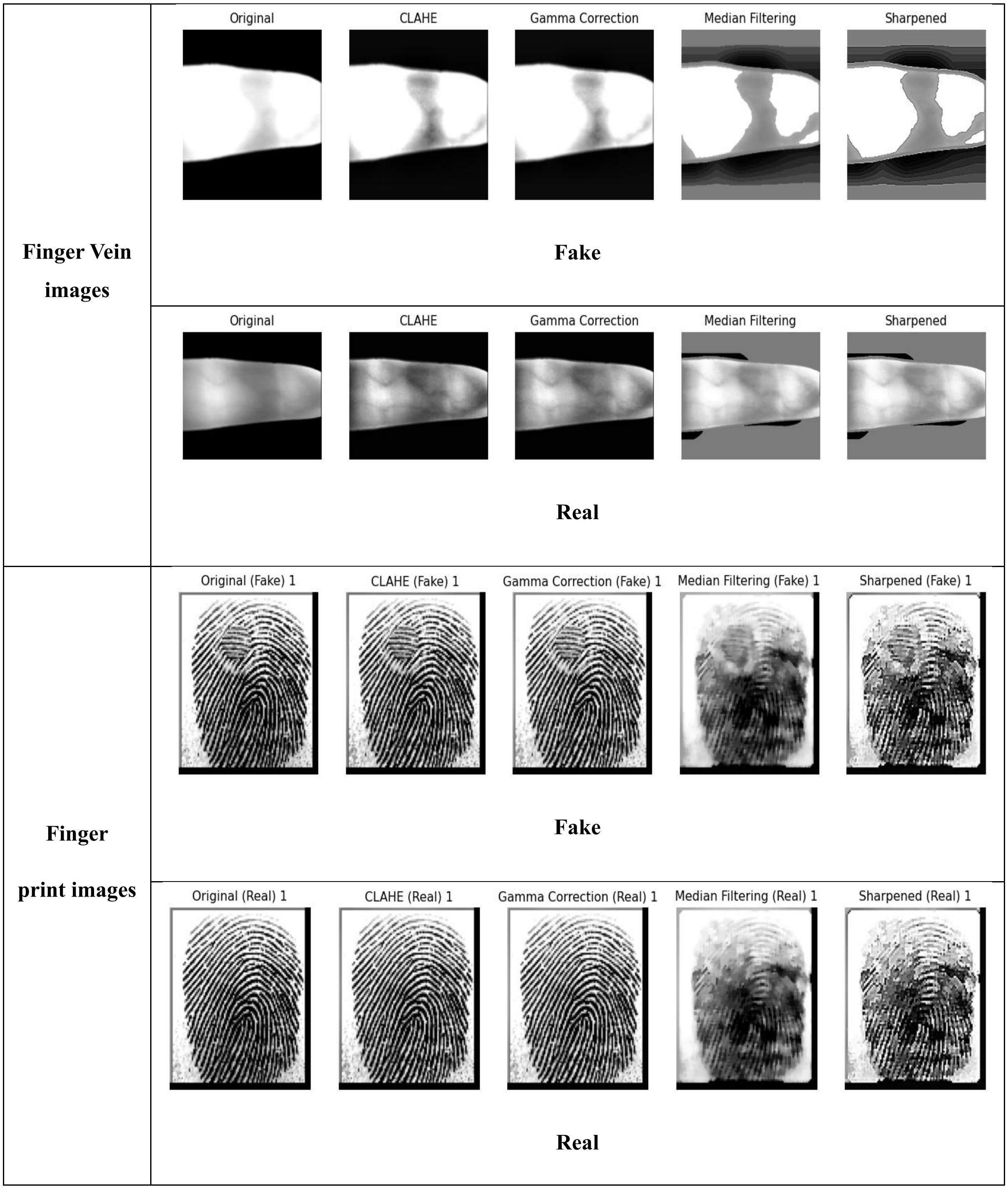


Figure 2: Pre-processed images for Knuckle, Finger Vein, and Fingerprint

3.2 Feature Extraction

Feature extraction, which extracts significant patterns from images for precise classification, is essential to biometric authentication. Two sophisticated feature extraction methods are used in the suggested method such as EfficientNet based Autoencoder (EA) and VGG16.

3.2.1 EfficientNet based Autoencoder (EA)

The EfficientNet based Autoencoder (EA) ensures effective feature extraction and reconstruction by utilizing EfficientNet for both encoding and decoding. The decoder reconstructs the input images with the least amount of loss after the encoder uses EfficientNet to compress them into a latent space. By maintaining crucial characteristics while lowering computing complexity, this work improves performance in medical imaging tasks like denoising and anomaly identification.

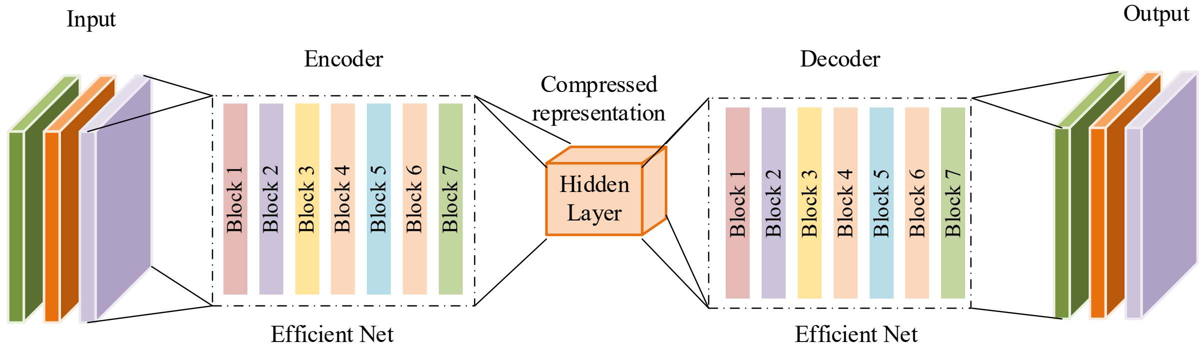


Figure 3: Architecture of the EA

Figure 3 shows the autoencoder architecture makes use of EfficientNet for both the encoding and decoding procedures. Neural networks called autoencoders are made to learn effective representations of input data, usually for tasks like anomaly detection, compression, and denoising. In this configuration, EfficientNet serves as the backbone for both encoding and decoding, and the autoencoder is composed of an encoder, a bottleneck (hidden layer), and a decoder. This makes it a very effective feature extraction method. While preserving the data's key patterns and structures, this stage shrinks the spatial dimensions. After that, the extracted features are transmitted to a hidden layer, which acts as the network's bottleneck and compresses the input into a lower-dimensional latent representation while keeping just the most important elements and eliminating unnecessary ones. The decoder then processes the compressed latent representation, upsampling and fine-tuning the recovered features to reconstitute the original input. The model ensures that the compressed representation contains enough information for precise reconstruction by training the autoencoder to minimize reconstruction loss. In finger biometric authentication, where accurate feature extraction and reconstruction are essential for identity verification. The model's capacity to learn reliable fingerprint patterns while preserving computational efficiency is improved by the usage of EfficientNet. The technique improves authentication accuracy by capturing distinct biometric information by condensing fingerprint photos into a small latent space. The autoencoder is a useful tool for safe biometric systems since it also helps with fingerprint image improvement, denoising, and spoof detection.

Efficient Net Block

The CNN family EfficientNet is created to achieve excellent accuracy with fewer parameters and less computational expense. It effectively balances network depth, width, and resolution using a compound scaling method. The MBConv blocks (Mobile Inverted Bottleneck Convolutions), which employ depth-wise separable convolutions to improve efficiency, come after the initial convolutional layer () at the start of the design. While hierarchical feature extraction takes place across scaling blocks, swish activation enhances gradient flow. Middle blocks identify intricate patterns, later blocks extract high-level features, and early blocks record low-level textures. EfficientNet architecture is shown in Figure 4.

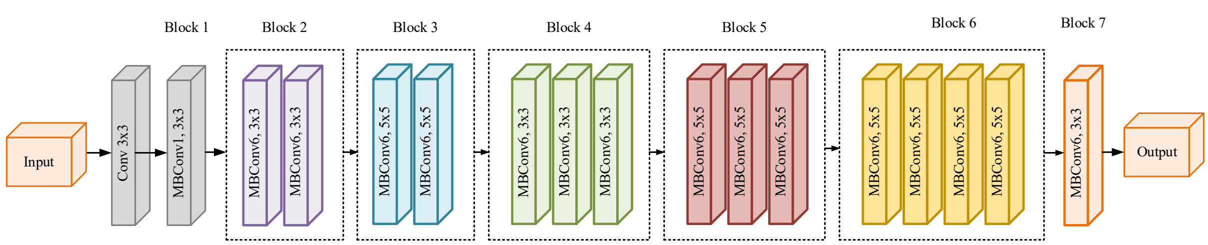


Figure 4: EfficientNet architecture

3.2.2 VGG16

The deep convolutional neural network (CNN) architecture VGG16 is frequently utilized for image classification applications because of its straightforward yet efficient design. Convolutional layers make up the majority of its 16 layers, which are followed by fully connected layers for categorization. With the help of a sequence of tiny 3x3 convolutional filters, VGG16 processes the features that were recovered from SIFT and EA. These filters record spatial hierarchies ranging from low-level edges to high-level patterns. Max pooling layers and ReLU activation for non-linearity come after each convolutional block to downsample the feature maps while keeping important details. The SoftMax classifier, which gives probability scores to various biometric classes (fingerprint, vein, and knuckle) for authentication, and fully linked layers make up the last layers. To ensure secure authentication, VGG16's deep feature extraction capacity makes it extremely effective at differentiating genuine biometric features from spoofing assaults based on generative artificial intelligence. Figure 5 demonstrates the VGG16 architecture.

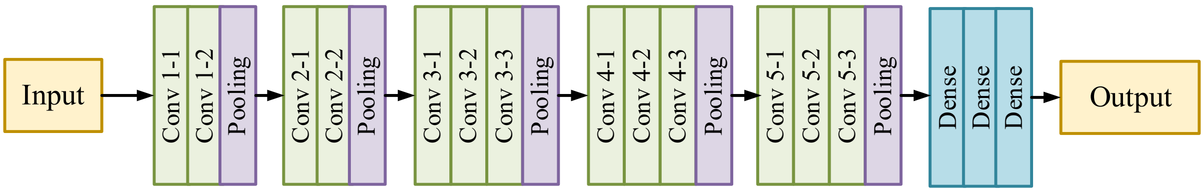


Figure 5: VGG16 architecture

3.3 Segmentation

The process begins with the pre-processed images. The MMF-BA_ESTNet model that has been suggested combines an Enhanced Swin (Shifted Window) Transformer (EST) network with a Densely Connected Semantic Dilated Multi-Head Attention mechanism for medical image segmentation. To improve hierarchical feature learning, it combines Swin Transformers with CNN-based feature extraction (AlexNet, VGG16, and MobileNet). Through the use of patch merging and shifting window self-attention mechanisms, the Swin Transformer is applied in four stages, gradually improving the extracted features. To ensure accurate segmentation, each step records both local and global dependencies. MMF-BA_ESTNet is shown in Figure 6.

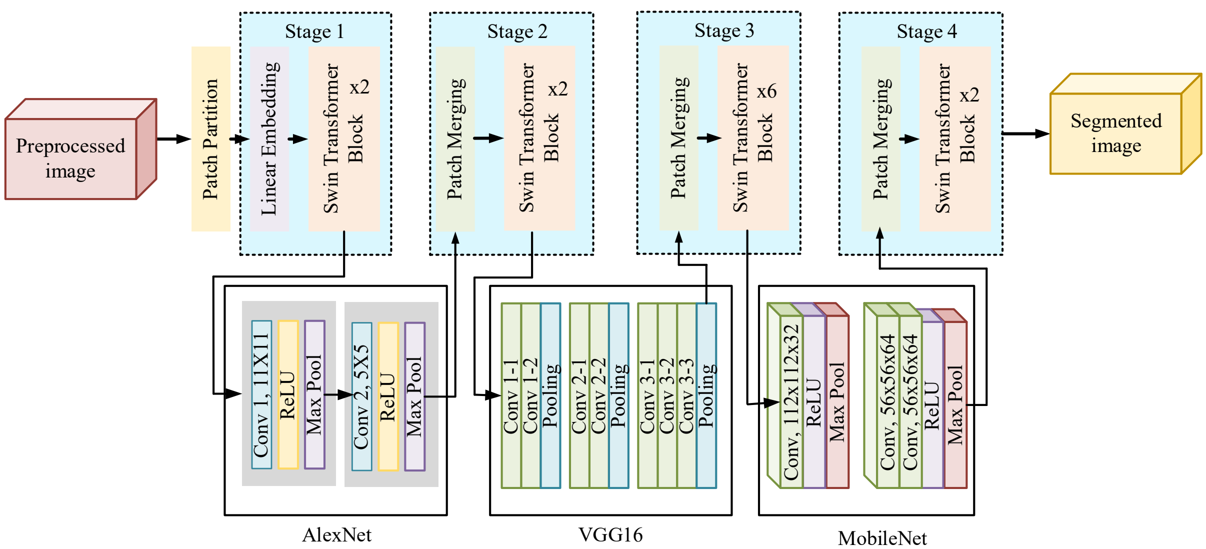


Figure 6: MMF-BA_ESTNet architecture

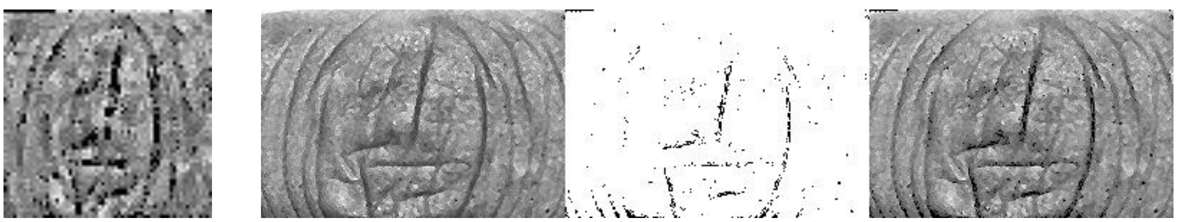
Figure 6 demonstrates First, the preprocessed image is split into smaller patches in Stage 1, which is followed by linear padding and projection into feature embeddings. After that, the patches are processed using CNN backbones for initial feature extraction, including AlexNet, VGG16, or MobileNet. For efficient processing, MobileNet uses depthwise-separable convolutions, VGG16 uses multiple 3×3 convolutions, and AlexNet uses 11×11 and 5×5 convolution filters. Then, to efficiently capture local characteristics, a Swin Transformer Block (K5) is used with a window size of K=5. In Stage 2, CNN-extracted features can be gradually abstracted by a patch-merging technique that increases the number of channels while decreasing the spatial resolution. The shifted window self-attention mechanism, which facilitates information sharing across non-overlapping windows and improves contextual learning, is included in a second Swin Transformer Block (K5).

Using more patch merging and downsampling feature maps while maintaining important information, stage three further refines the features. This approach improves hierarchical feature representation while lowering computational expenses. The model can then represent global dependencies at a lower resolution by applying a third Swin Transformer Block. The last patch merging procedure in Stage 4 makes sure that only the most significant high-level features are kept by further reducing the spatial size and deepening the features. These fine-tuned characteristics are processed by the last Swin Transformer Block, which is essential for precisely segmenting items in the medical image. The segmentation procedure is accurate and effective when the final segmented image output is produced, emphasizing the target medical regions like tumors, tissues, or organs.

Multi-head Attention

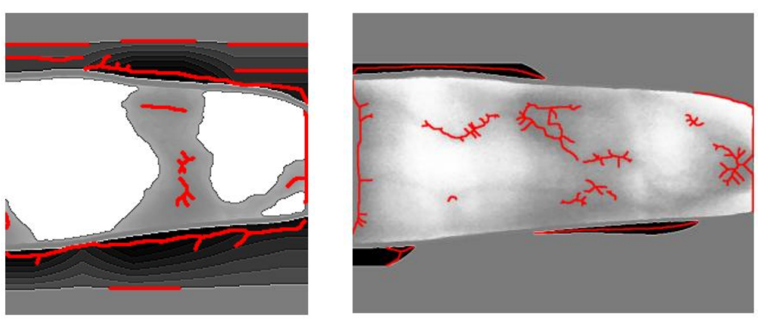
Multi-head attention enhances the fundamental Scaled Dot-Product Attention mechanism by conducting multiple attention processes concurrently. It comprises several fundamental processes that begin with input representation. Query (), Key (), and Value () are the three matrices that make together the input. While the Key matrix includes every word in the input and assesses its relevance to the query, the Query matrix shows the word (or token) that is now attending to other words. The real information that must be communicated is included in the Value matrix. The input embedding matrix is the source of each of these matrices, which are then converted using learned weight matrices to produce numerous attention heads. Multi-head attention improves contextual awareness in deep learning models like transformers by allowing the model can capture a variety of word associations by independently computing attention scores across many heads and then concatenating the findings. The mathematical formula for the attention mechanism is expressed in Eq. (1)

$$Attention\ (Q, K, V) = SoftMax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (1)$$



Fake Real

Figure 7: Segmented Knuckle image



Fake Real

Figure 8: Segmented Finger Vein image

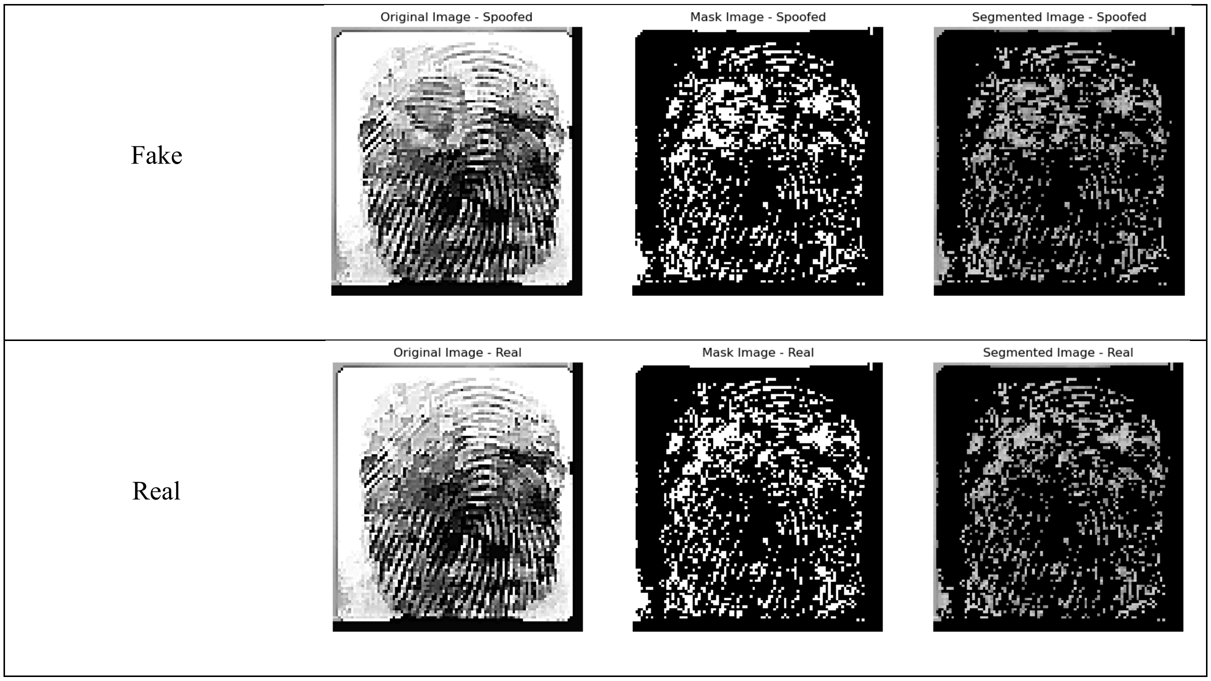


Figure 9: Segmented Finger Print image

3.4 Extraction

The segmented images are subsequently submitted to further extraction procedures, such as Histogram of Gradient (HoG) and Scale-Invariant Feature Transform (SIFT), for further image enhancement.

3.4.1 **Scale-Invariant Feature Transform (SIFT)**

It is a powerful key point-based feature extraction method that is frequently employed in image processing because it can detect unique characteristics that don't change when scale, rotation, or lighting changes. Scale-space creation is the first step in the procedure, which involves convolving the image with Gaussian filters at various scales to identify key points at various resolutions. The Difference of Gaussians (DoG) approach is then used for keypoint detection to locate areas with notable intensity variations. Key point localization uses contrast and edge response analysis to weed out unstable key points to improve dependability. Following the identification of stable key points, orientation assignment is used to ensure rotation invariance by assigning a dominant orientation to each key point based on local gradients. Lastly, a descriptor computation phase examines gradient patterns in the key point’s locality to produce a 128-dimensional feature vector for each key point. SIFT is a strong technique for accurate recognition under a variety of settings because it is very good at extracting important features from fingerprint points, finger vein patterns, and knuckle textures in biometric authentication.

3.4.2 **Histogram of Gradient (HoG)**

The HOG is used to process the segmented biometric images (fingerprint, finger vein, and knuckle) to extract features. HOG is a potent method that uses gradient orientation analysis in certain image regions to capture the shape and texture of objects. To be able to identify variations in intensity, Sobel filters are first used to calculate the gradient's magnitude and direction. Gradient directions are then grouped into predetermined boxes within each of the image's tiny cells to form a histogram. Block normalization reduces the impact of lighting and contrast variations, improving robustness. The edge and texture patterns, which are essential for biometric identification, are represented by these histograms. Block normalization is used to lessen the impact of contrast and lighting changes to ensure robustness. The proposed method improves recognition accuracy and strengthens its defenses against spoofing assaults produced by Generative AI-based methods by applying HOG to the segmented images and extracting unique spatial features from the biometric attributes.

3.5 Classification

A multi-layer perceptron is an MLP, a multi-layered neural network that models complex patterns for classification or regression purposes. MLP stands for a multi-layer neural network comprising input, hidden layers, and the output layer. It is often used in classification and regression tasks. They model complex relationships in data and are trained with backpropagation. In a feed-forward neural network, an MLP converts inputs to an appropriate set of outputs. An input layer, one or more hidden layers, and an output layer are often included in a network, as shown in Figure 10. The hidden node hid_j output, given an input node inp_j, is expressed in Eq. (2)

$$hid_j = \emptyset_1 + \left(\sum_{j=1}^{n} W_{ij} + \theta_j\right) \quad (2)$$

where θ_j is the bias value and W_{ij} is the weight between ith input and jth hidden node. Conversely, results will be provided in Eq. (3)

$$out_j = \emptyset_2 + \left(\sum_{j=1}^{n} W_{jk} + \theta_k\right) \quad (3)$$

Weights W_{jk} are updated at each iteration of the mapping inputs to outputs iterative process. The common algorithm used in the technique is the Back Propagation which updates the weights using the formula in Eq. (4).

$$W_{ji}(t+1) = W_{ji}(t) - \in \frac{\partial Ef}{W_{ji}} \quad (4)$$

Updating weights is done via the difference in the output received and the actual output desired.

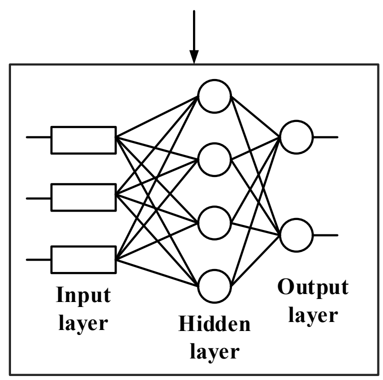


Figure 10: MLP layer

Sigmoid

Binary classification tasks, such as identifying real and fake biometric photos, use the sigmoid function. It is perfect for decision-making since it transforms input values into probabilities between 0 and 1. The result is categorized as real if it is*≥ 0.5; if not, it is fake. By ensuring smooth gradient updates, sigmoid aids in the efficient learning of neural networks. For binary classification tasks, it is frequently utilized in the last layer of deep learning models and logistic regression.

4. Result and Discussion

This section contains results and a discussion of the proposed model. This paper presents A Multi-Model Finger Biometric Authentication Scheme Based on a Densely Connected Sematic Dilated Multi-Head Attention Conventional Swin Transformer Network. The performance effectiveness of the recommended technique is assessed using a variety of metrics, like Specificity, Accuracy, Sensitivity, Precision, Computational time, Dice score, F-measure, False Negative Rate (FNR), False Positive Rate (FPR), Mathews Correlation Coefficient (MCC), Negative Predicted Value (NPV), and Intersection Over Union (IOU). To assess how much the newly built framework's performance has improved, it is compared to existing models, such as TCN, LSTM, AlexNet, SEFF-CNN-BO, and Proposed (MMFBA_ESTNet).

4.1 Evaluation Setup

The Python platform has been used to implement the proposed framework. It contains an Intel (R) Core (TM) i3-8100 CPU configured to run at 3.60 GHz. Of the 16.0 GB of installed RAM, 15.8 GB are available for use. The system is powered by an x64 architecture CPU and runs on a 64-bit operating system.

4.2 Performance Evaluation

Numerous matrices are used in performance evaluation, including Specificity, Accuracy, Sensitivity, Precision, Computational time, Dice score, F-Measure, FNR, FPR, MCC, NPV, and IOU.

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

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{5}$$

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.

$$Precision = \frac{TP}{FP+TP} \tag{6}$$

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

$$F - Measure = 2 \times \frac{Presision.Recall}{Presision + Recall} \tag{7}$$

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

$$Specificity = \frac{TN}{FP+TN} \tag{8}$$

- **Sensitivity:** It is calculated by dividing the total number of positive predictions by the percentage of correct positive predictions.

$$Sensitivity = \frac{TP}{TP+FN} \tag{9}$$

- **MCC:** It 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 predictor.

$$MCC = \frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)(TP+FN)(FP+TN)(TN+FN)}} \tag{10}$$

- **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 negative predicted value.

$$NPV = \frac{TN}{TN+FN} \tag{11}$$

- **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.

$$FNR = \frac{FN}{FN+TP} \tag{12}$$

- **FPR:** It can be characterized by dividing the total quantity of negative data by the portion of positive data that was incorrectly labeled.

$$FPR = \frac{FP}{TN+FP} \tag{13}$$

- **Dice Score:** Compares the segmentation masks that were expected and those that had been created.

$$Dice\ Score = \frac{2 \times TP}{2 \times TP+FP+FN} \tag{14}$$

- **Computational time:** The entire amount of time needed for the model to process and categorize biometric data, which affects the viability of real-time authentication.

IoU: It calculates the overlap between the expected and real region and is frequently used in segmentation work.

4.3 Performance Results

Table 1: Comparative analysis of the performance metrics for learning rate 70/30

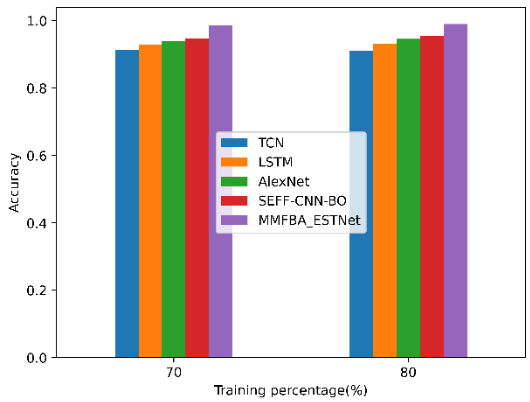
	TCN [19]	LSTM [19]	AlexNet [25]	SEFF-CNN-BO [17]	Proposed (MMFBA_ESTNet)
Accuracy	0.912	0.929	0.940	0.947	0.986
Precision	0.928	0.942	0.946	0.958	0.995
Sensitivity	0.935	0.964	0.950	0.945	0.972
Specificity	0.939	0.938	0.936	0.963	0.995
F_measure	0.941	0.953	0.937	0.931	0.986
MCC	0.930	0.947	0.958	0.959	0.981
NPV	0.927	0.943	0.962	0.949	0.981
FPR	0.091	0.082	0.064	0.052	0.009
FNR	0.093	0.071	0.062	0.045	0.009
Computational Time	0.016	0.017	0.011	0.019	0.007
Dice Score	0.859	0.869	0.875	0.864	0.915
IOU	0.794	0.799	0.808	0.786	0.855

The performance of several deep learning models TCN, LSTM, AlexNet, SEFF-CNN-BO, and MMFBA_ESTNet on biometric authentication tasks is compared in Table 1. MMFBA_ESTNet is the most resilient and effective of these models, outperforming all others on a variety of standards. With the highest accuracy (0.986), it performs better overall in classification. Its capacity to accurately detect authentic biometric inputs while reducing false positives is demonstrated by its precision (0.995) and specificity (0.995). Furthermore, a high F-measure (0.986) and MCC (0.981) validate balanced performance across all categories, while sensitivity (0.972) guarantees efficient detection of true positives. Additionally, MMFBA_ESTNet greatly lowers misclassification errors by achieving the lowest FPR (0.009) and FNR (0.009). It is the most effective model for real-time authentication since it records the minimum computational time (0.007s). MMFBA_ESTNet demonstrates the highest Dice Score (0.915) and Intersection over Union (IoU) (0.855) in segmentation-based evaluation, confirming its accuracy in biometric feature extraction.

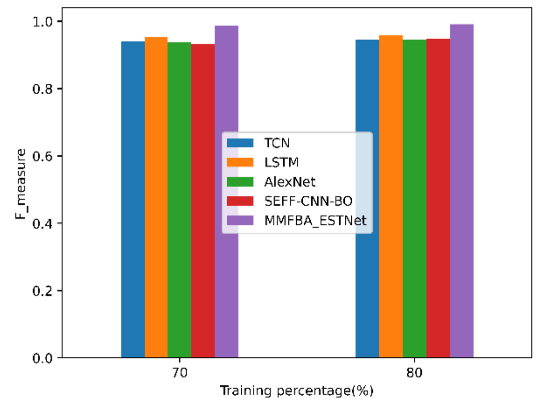
Table 2: Comparative analysis of the performance metrics for learning rate 80/20

	TCN [19]	LSTM [19]	AlexNet [25]	SEFF-CNN-BO [17]	Proposed (MMFBA_ESTNet)
Accuracy	0.910	0.932	0.946	0.955	0.990
Precision	0.957	0.951	0.952	0.947	0.991
Sensitivity	0.937	0.940	0.940	0.956	0.991
Specificity	0.957	0.953	0.952	0.936	0.99
F_measure	0.946	0.958	0.944	0.948	0.991
MCC	0.949	0.951	0.961	0.942	0.979
NPV	0.936	0.944	0.936	0.961	0.99
FPR	0.063	0.081	0.052	0.073	0.011
FNR	0.081	0.060	0.033	0.041	0.011
Computational Time	0.018	0.018	0.013	0.02	0.008
Dice Score	0.850	0.858	0.866	0.847	0.922
IOU	0.780	0.785	0.795	0.770	0.865

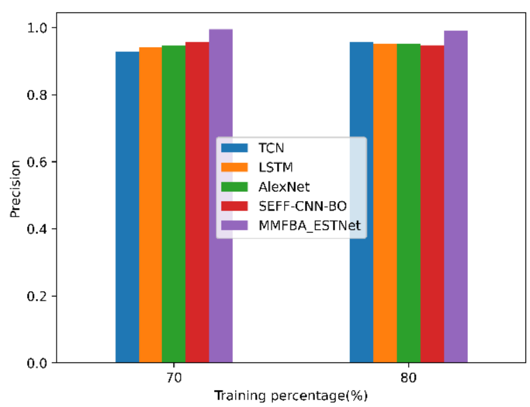
The performance of several deep learning models TCN, LSTM, AlexNet, SEFF-CNN-BO, and MMFBA_ESTNet for biometric identification with an 80/20 learning rate split is compared in Table 2. MMFBA_ESTNet continuously performs better than the others in all important measures. Its remarkable ability to accurately classify biometric inputs demonstrated by its greatest accuracy of 0.990. Furthermore, accuracy (0.991) and specificity (0.990) show that MMFBA_ESTNet successfully reduces false positives, guaranteeing trustworthy biometric identification. It is also very successful in identifying real biometric cases, and lowering false negatives, with a sensitivity of 0.991. The model's robust and balanced classification abilities are confirmed by its MCC (0.979) and F-measure (0.991). The lowest FPR (0.011) and FNR (0.011) are also recorded by MMFBA_ESTNet, confirming its great dependability in separating authentic biometrics from possibly spoof inputs. The model is the most effective option for real-time applications since it also has the quickest computing time (0.008s). MMFBA_ESTNet achieves the highest Dice Score (0.922) and IoU (0.865) in segmentation-based evaluation, indicating excellent biometric feature extraction accuracy.



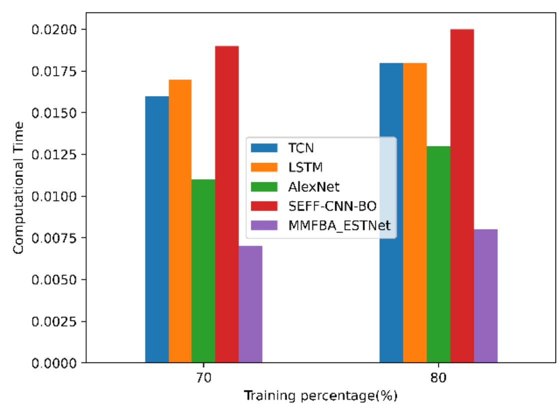
a. Accuracy



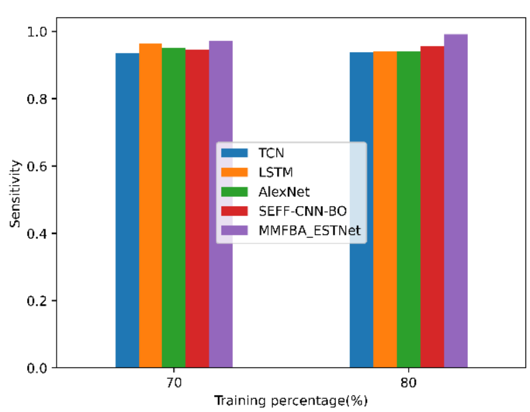
b. F_Measure



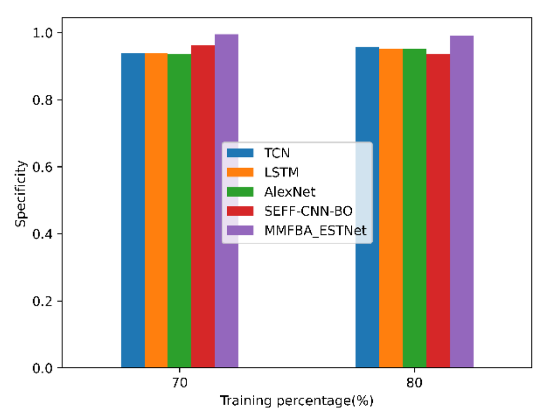
c. Precision



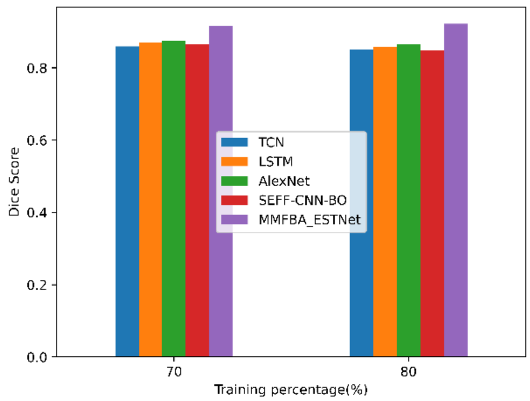
d. Computational time



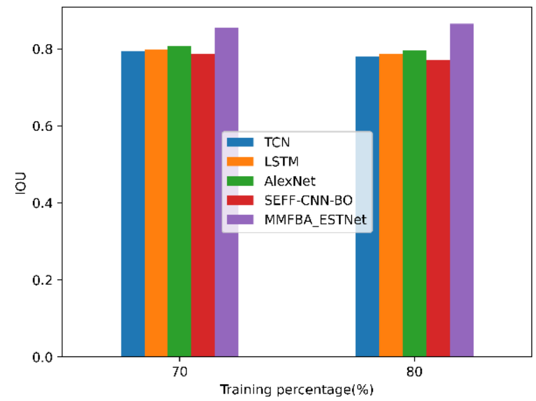
e. Sensitivity



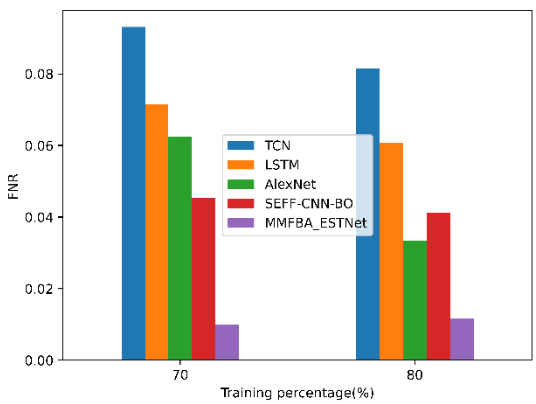
f. Specificity



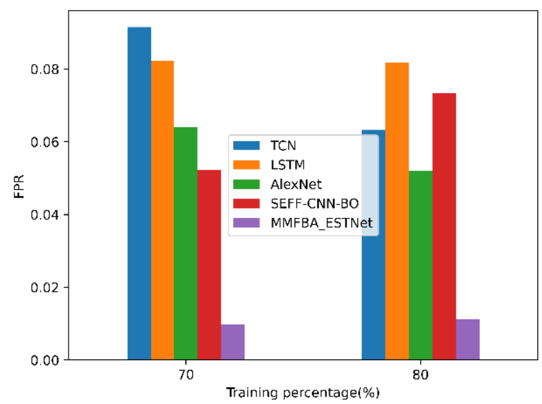
g. Dice score



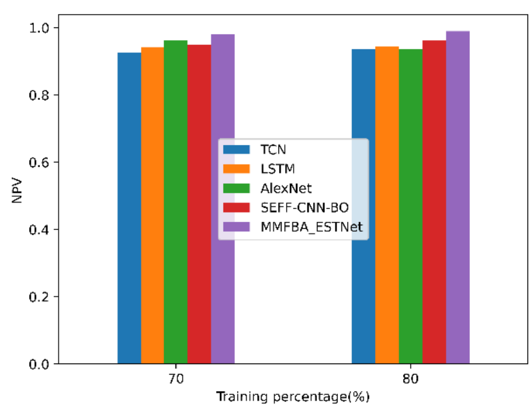
h. IOU



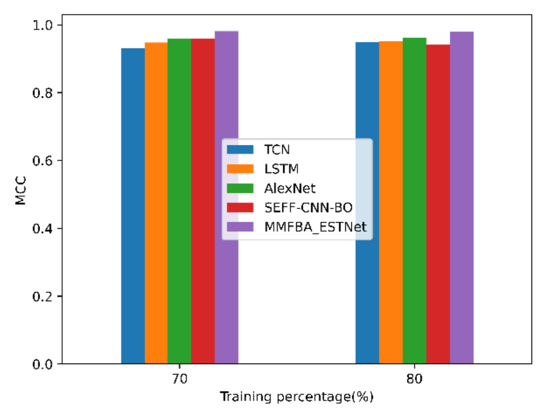
i. FNR



j. FPR



k. NPV



l. MCC

Figure 7 (a)-(l): Visual representation of several performance metrics with proposed and existing works for the learning rates 70/30 and 80/20 respectively.

The performance metrics, which include Specificity, Accuracy, Sensitivity, Precision, Computational time, Dice score, F-Measure, FNR, FPR, MCC, NPV, and IOU are represented graphically in Figures 7 (a) to (l). These metrics are compared with several approaches, including TCN, LSTM, AlexNet, SEFF-CNN-BO, and Proposed (MMFBA_ESTNet).

4.4 Ablation Study

Table 3: Performance analysis for the Dataset comparison for learning rate 70%

Metric	Proposed	Finger Print Dataset	Finger Vein Dataset	3D Finger Knuckle Dataset
Accuracy	0.9863	0.9721	0.9634	0.9756
Precision	0.9951	0.9803	0.9698	0.9785
Sensitivity	0.9724	0.9605	0.9512	0.9583
Specificity	0.9952	0.9819	0.9705	0.9828
F-measure	0.9869	0.9704	0.9592	0.9687
MCC	0.9812	0.9683	0.9556	0.9625

A performance analysis of various biometric datasets at a 70% learning rate is shown in Table 3. The Finger Vein, Finger Print, and 3D Finger Knuckle datasets are all surpassed by the suggested method in every metric. Accuracy (0.9863), precision (0.9951), sensitivity (0.9724), specificity (0.9952), F-measure (0.9869), and MCC (0.9812) are all at their greatest levels.

Table 4: Performance analysis for the Dataset comparison for learning rate 80%

Metric	Proposed	Finger Print Dataset	Finger Vein Dataset	3D Finger Knuckle Dataset
Accuracy	0.9902	0.9825	0.9758	0.9803
Precision	0.9913	0.9801	0.9726	0.9784
Sensitivity	0.9914	0.9812	0.9745	0.9796
Specificity	0.99	0.9798	0.9719	0.9775
F-measure	0.9911	0.9805	0.9734	0.9789
MCC	0.9798	0.9703	0.9607	0.9682

The performance of several biometric datasets at an 80% learning rate is compared in Table 4. The Finger Print, Finger Vein, and 3D Finger Knuckle datasets are outperformed by the suggested technique, which offers the highest accuracy (0.9902), precision (0.9913), sensitivity (0.9914), specificity (0.99), F-measure (0.9911), and MCC (0.9798).

Table 5: Performance analysis for the proposed and existing works for learning rate of 70%

Metric	Proposed	Without Extract SIFT Features	Without Extract VGG16	Without Autoencoder with EfficientNet
Accuracy	0.9863	0.9725	0.9651	0.9583
Precision	0.9951	0.981	0.9732	0.9658
Sensitivity	0.9724	0.9587	0.9504	0.9423
Specificity	0.9952	0.9808	0.9725	0.9617
F-measure	0.9869	0.9695	0.9602	0.951
MCC	0.9812	0.9654	0.9558	0.9476

At a 70% learning rate, Table 5 compares the proposed strategy to current techniques. The suggested approach routinely performs better than VGG16, the autoencoder with EfficientNet, and models that do not use SIFT features. Accuracy (0.9863), precision (0.9951), sensitivity (0.9724), specificity (0.9952), F-measure (0.9869), and MCC (0.9812) are all at their greatest levels.

Table 6: Performance analysis for the proposed and existing works for learning rate 80%

Metric	Proposed	Without Extract SIFT Features	Without Extract VGG16	Without Autoencoder with EfficientNet
Accuracy	0.9863	0.9725	0.9651	0.9583
Precision	0.9951	0.981	0.9732	0.9658
Sensitivity	0.9724	0.9587	0.9504	0.9423
Specificity	0.9952	0.9808	0.9725	0.9617
F-measure	0.9869	0.9695	0.9602	0.951
MCC	0.9812	0.9654	0.9558	0.9476

At an 80% learning rate, Table 6 compares the effectiveness of the suggested strategy to current techniques. The highest accuracy (0.9863), precision (0.9951), sensitivity (0.9724), specificity (0.9952), F-measure (0.9869), and MCC (0.9812) are all attained by the suggested model. Models lacking VGG16, SIFT features, or an autoencoder with EfficientNet perform worse, demonstrating the importance of these elements in improving the robustness and accuracy of biometric detection.

5. Conclusion

The proposed MMFBA_ESTNet integrated fingerprint, knuckle, and finger vein characteristics for multi-model authentication, effectively addressing the problem of biometric spoofing. The model greatly enhanced feature extraction across low, mid, and high levels by utilizing a Densely Connected Semantic Dilated Multi-Head Attention-based Enhanced Swin Transformer (EST) network. Accuracy and computational efficiency were improved by the parallel processing pipeline that combined segmentation, feature extraction, and classification. The model's resilience was shown by experimental evaluations, which showed 0.986 accuracy at a 70% learning rate and 0.99 accuracy at an 80% learning rate. Learning effectiveness and biometric authentication performance were further enhanced by the incorporation of concatenated HoG classification characteristics and multi-head attention. The outcomes confirmed that the suggested method is a dependable alternative for secure authentication since it can effectively differentiate real biometric images from artificial intelligence-generated ones.

Reference

1. A. El_Rahman, S., & Alluhaidan, A. S. (2024). Enhanced multimodal biometric recognition systems based on deep learning and traditional methods in smart environments. *Plos one*, *19*(2), e0291084.
2. Abdullahi, S. B., Bature, Z. A., Chophuk, P., & Muhammad, A. (2023). Sequence-wise multimodal biometric fingerprint and finger-vein recognition network (STMFPFV-Net). *Intelligent Systems with Applications*, *19*, 200256.
3. Hammad, M., Wani, M. A., Shakil, K. A., Shaiba, H., & Abd El-Latif, A. A. (2024). Deep cancelable multibiometric finger vein and fingerprint authentication with non-negative matrix factorization. IEEE Access, 12, 120638-120660.
4. Alshardan, A., Kumar, A., Alghamdi, M., Maashi, M., Alahmari, S., Alharbi, A. A., ... & Alzahrani, Y. (2024). Multimodal biometric identification: leveraging convolutional neural network (CNN) architectures and fusion techniques with fingerprint and finger vein data. *PeerJ Computer Science*, *10*, e2440.
5. Ammour, N., Bazi, Y., & Alajlan, N. (2023). Multimodal approach for enhancing biometric authentication. *Journal of Imaging*, *9*(9), 168.
6. Aouid, F. R., Lati, A., & Chaa, M. (2025). Enhancing Security through Autonomous Learning Multi-Model Classification of 3D Palmprints: The ALMMo-0 Framework. *Informatica*, *49*(5).
7. Naitik, S. T., & Gorabal, J. V. (2023). Design and development of model using fingervein authentication by convolution neural network. International Neurourology Journal, 27(4), 709-719.
8. Gao, C., Yang, Z., Teoh, A. B. J., & Zhu, M. (2024). Beyond first-order: A multi-scale approach to finger knuckle print biometrics. *arXiv preprint arXiv:2406.19672*.
9. Mathew, A., & Amudha, P. (2024). Improved finger vein recognition using generative adversarial network and transfer learning. Journal of Nanoelectronics and Optoelectronics, 19(10), 1042-1052.
10. Guesmi, H. A. (2024). A lightweight Approach to finger-vein authentication based on Con-trast-Limited Adaptive Histogram Equalization. *Journal of Computing & Biomedical Informatics*, *8*(01).
11. Haider, S. A., Ashraf, S., Larik, R. M., Husain, N., Muqeet, H. A., Humayun, U., ... & Khan, M. F. (2023). An Improved Multimodal Biometric Identification System Employing Score-Level Fuzzification of Finger Texture and Finger Vein Biometrics. *Sensors*, *23*(24), 9706.
12. Huang, Y., Ma, H., & Wang, M. (2023). Multimodal finger recognition based on asymmetric networks with fused similarity. *IEEE Access*, *11*, 17497-17509.
13. Sumalatha, U., Prabhu, S., & Nayak, V. C. (2025). Multimodal biometric authentication: a novel deep learning framework integrating ECG, fingerprint, and finger knuckle print for high-security applications. Engineering Research Express, 7(1), 015207.

14. Khatri, M., & Sharma, A. (2023). A Computational Approach for Score-Level Fusion Decision-Making of Multi-biometric Recognition System Using Ant Colony Optimisation. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *36*(1), 147-158.
15. Kyeremeh, G. K., Abdul-Al, M., Qahwaji, R., Ali, N. T., & Abd-Alhameed, R. A. (2025). Fusion of Hand Biometrics for Border Control Involving Fingerprint and Finger Vein. *IEEE Access*.
16. Alshardan, A., Kumar, A., Alghamdi, M., Maashi, M., Alahmari, S., Alharbi, A. A., ... & Alzahrani, Y. (2024). Multimodal biometric identification: leveraging convolutional neural network (CNN) architectures and fusion techniques with fingerprint and finger vein data. PeerJ Computer Science, 10, e2440.
17. Nair, M., & Dharan, S. (2024). Design a Multimodal Biometric Based Protection System by Generation of a Revocable Cryptographic Key Using Separately Extracted Feature Fusion-based Convolutional Neural Network With Bat Optimization.
18. Abdullahi, S. B., Bature, Z. A., Chophuk, P., & Muhammad, A. (2023). Sequence-wise multimodal biometric fingerprint and finger-vein recognition network (STMFPFV-Net). Intelligent Systems with Applications, 19, 200256.
19. Salturk, S., & Kahraman, N. (2024). Deep learning-powered multimodal biometric authentication: integrating dynamic signatures and facial data for enhanced online security. *Neural Computing and Applications*, 1-12.
20. Gizachew Yirga, T., Gizachew Yirga, H., & Addisu, E. G. (2025). Cryptographic key generation using deep learning with biometric face and finger vein data. Frontiers in Artificial Intelligence, 8, 1545946.
21. Singh, K. K., & Barde, S. (2024). A feasible adaptive fuzzy genetic technique for face, fingerprint, and palmprint-based multimodal biometrics systems. *Journal of Current Science and Technology*, *14*(1), 1-1.
22. Singh, S. P., & Tiwari, S. (2023). A dual multimodal biometric authentication system based on woa-ann and ssa-dbn techniques. *Sci*, *5*(1), 10.
23. Sumalatha, U., Prabhu, S., & Nayak, V. C. (2025). Multimodal biometric authentication: a novel deep learning framework integrating ECG, fingerprint, and finger knuckle print for high-security applications. *Engineering Research Express*, *7*(1), 015207.
24. Hsia, C. H., Ke, L. Y., & Chen, S. T. (2023). Improved lightweight convolutional neural network for finger vein recognition system. *Bioengineering*, *10*(8), 919.
25. Nour, R., Moustafa, H. E. D., Abdelhay, E. H., & Ata, M. M. (2024). Improved unsupervised deep Boltzmann learning approach for accurate hand vein recognition. IEEE Access, 12, 18488-18507.
26. Fingervein Dataset:https://huggingface.co/datasets/luyu0311/MMCBNU_6000
27. Fingerprint Dataset:https://gts.ai/dataset-download/sokoto-coventry-fingerprint-dataset-socofing/
28. Knuckle Dataset: https://www4.comp.polyu.edu.hk/~csajaykr/fn2.htm