Computer Modeling in Engineering & Sciences

Tech Science Press

Doi:10.32604/cmes.2025.060863

ARTICLE

EFI-SATL: An EfficientNet and Self-Attention Based Biometric Recognition for Finger-Vein Using Deep Transfer Learning

Manjit Singh and Sunil Kumar Singla*

Department of Electrical and Instrumentation Engineering, Thapar Institute of Engineering and Technology, Patiala, 147004, India

*Corresponding Author: Sunil Kumar Singla. Email: ssingla@thapar.edu

Received: 12 November 2024; Accepted: 06 February 2025; Published: 03 March 2025

ABSTRACT: Deep Learning-based systems for Finger vein recognition have gained rising attention in recent years due to improved efficiency and enhanced security. The performance of existing CNN-based methods is limited by the puny generalization of learned features and deficiency of the finger vein image training data. Considering the concerns of existing methods, in this work, a simplified deep transfer learning-based framework for finger-vein recognition is developed using an EfficientNet model of deep learning with a self-attention mechanism. Data augmentation using various geometrical methods is employed to address the problem of training data shortage required for a deep learning model. The proposed model is tested using K-fold cross-validation on three publicly available datasets: HKPU, FVUSM, and SDUMLA. Also, the developed network is compared with other modern deep nets to check its effectiveness. In addition, a comparison of the proposed method with other existing Finger vein recognition (FVR) methods is also done. The experimental results exhibited superior recognition accuracy of the proposed method compared to other existing methods. In addition, the developed method proves to be more effective and less sophisticated at extracting robust features. The proposed EffAttenNet achieves an accuracy of 98.14% on HKPU, 99.03% on FVUSM, and 99.50% on SDUMLA databases.

KEYWORDS: Biometrics; finger-vein recognition (FVR); deep net; self-attention; EfficientNets; transfer learning

1 Introduction

Vein patterns in the fingers can be utilized to verify an individual's identity, making finger vein biometrics a promising technology in terms of security. Finger vein biometrics has grown in popularity in recent years due to its significant advantages over conventional biometrics such as fingerprints, palm prints, and others: High security—the vein pattern is beneath the skin, so it is hard to steal or forge. High stability—the vein pattern is nearly unaffected by scars, oil, or sweat [1,2]. Liveness detection—only living people's finger veins can be captured and recognized [3]. As an emerging technology, biometric technology based on finger veins is far from flawless. Finger-vein verification can be hindered by internal and external factors, such as ambient or uneven lighting, light dispersion inside finger tissues, user behavior, and others [4]. Hence, it is crucial to develop a method for extracting attributes of finger-vein network patterns that is precise and reliable for recognition and verification. As compared to other vein biometrics such as palm vein and dorsal vein, the finger vein has some added advantages [5], such as less cross-section area; hence, the size of the instrument will be less for image capture, making it suitable for both user and employer for smaller areas. Also, more finger vein images of the same subject, i.e., 4 to 6 fingers of both hands for one subject, resulted in enhanced security. In addition, finger vein patterns are more confined and complex in a smaller area,

making it challenging to spoof or forge, providing reasonable security and high accuracy compared to palm or dorsal veins.

Initially, input image acquisition, data pre-processing, and feature extraction and matching are typical components of a finger-vein recognition system. The finger vein pattern is imaged using near-infrared (NIR) technology. The non-invasive, contactless method of getting a finger vein is easy and clean for users. The most common way to reprocess images is the contrast-limited adaptive histogram equalization (CLAHE) method [6], which enhances image contrast and noise reduction. One of the most crucial processes, feature extraction, is inextricably linked to recognition accuracy. Presently, methods employing feature extraction are broadly of three types: template-based methods that employ finger-vein pattern extraction in the form of networks [7,8] or template generation form minutia features [9,10] of finger-vein images; representation based methods employing robust codes generated by using descriptors [11,12], transformers [13,14], or textures [15,16] to represent vein patterns; and feature learning-based methods that adaptively optimize the extracted features. However, these methods rely primarily on past knowledge, are profound to noise, and are difficult to calibrate. Deep learning methods, especially CNN-based, have been widely used in the biometric field, such as finger-vein verification, due to their superior ability to simultaneously perform extraction of features, dimension reduction, and classification. Methods such as 05 layer CNN design using fused convolutional and subsampling [16] on four open-source datasets, deep densely connected CNN [17] employing composite images of two fingers and input to finger-vein recognition (FVR), and a competitive order CNN [18] method have demonstrated the CNN model's potent feature extraction and classification ability. However, these methods suffer from problems such as vast experiments for designs and training, the incapability of real-time models running, and less feature extraction capability due to a lack of training data. In addition, recent works employing vision transformers [19], CNN with correlation-based matching [20], and large kernel-attention mechanism [21] have shown advantages such as improved feature understanding capability, generalization, low parameter count, and others, but suffer from challenges such as the need to focus on the open scenario, low quality and low contrast images misclassification, diverse lighting condition and higher resource requirement, and others.

An implementation for finger-vein recognition using deep transfer learning and EfficientNet with self-attention has been done in this study to enhance the dependability and effectiveness of the Finger vein recognition methods and to address problems in the current literature. The use of self-attention will make the system focus on salient features to capture the visual structure of finger vein images and reduce the losses by suppressing the unnecessary features, resulting in increased recognition performance. The self-attention mechanism gains valuable knowledge regarding the dependencies and complex patterns among the input data, resulting in an increased understanding of the model. Extensive experiments have been performed on three publicly available finger vein databases to assess the effectiveness of the proposed method. The results established superior performance achievement by the proposed method regarding Area Under the Curve (AUC) metrics. The motivation of this work involves researching potential performance enhancement deep learning techniques and researching the feasibility of FVR employing a deep learning approach. This proposed method provides a new direction for deep nets as it is a state-of-the-art model and has not been researched much. Also, using data augmentation with K-fold cross-validation helps to tackle the problem of data shortage by maintaining a balance between efficiency and computation cost. This method can assist the scientific community in better understanding the importance of K-fold cross-validation and attention mechanisms.

The significant contributions of the study are as follows:

1. A simplified Transfer Learning-based framework, EFI-SATL, is proposed with the application of EfficientNets for finger-vein recognition and classification. This is the first time EfficientNets and the self-attention mechanism for finger-vein biometric recognition are being used.
2. Development of a self-attention based on the network EffAttenNet for identifying prominent features for improved recognition performance of the network is done, which is a significant accomplishment.
3. K-fold cross-validation is employed for better testing and fair comparison of the results obtained using the proposed method with modern deep nets.
4. Extensive experimentation of the proposed method on three public databases achieved promising results with significant recognition accuracy on small datasets.

The paper is organized as follows: Section 2 describes the related work for finger vein recognition. Section 3 illustrates the background knowledge on transfer learning and EfficientNets architecture. Section 4 describes the methodology employed, followed by results and analysis in Section 5. Section 6 includes the discussion, followed by Section 7, which concludes the study with future considerations.

2 Related Work

Research on deep finger vein recognition systems is classified broadly into various categories based on the significant objective for which they are employed. One such category classification is presented here, i.e., systems focusing on feature extraction, verification, identification, Generative Adversarial Network (GAN), and attention-based. Methods that use feature extractors are designed to create feature vectors from images of veins on the finger so that distances between the two can be compared during authentication. Some of the feature extraction works include the Finger Vein Verification Network (FV-Net) [22], supervised discrete hashing [23], center loss with dynamic regularisation, and the 3D Finger Vein Verification Network [24]. Image search [25], real-time vein detection [26]. The segmentation-based CNN methods employ learning semantic segmentation to extract the vein patterns. Recent research on this method is proposed by [27,28]. The main limitation of this approach is the scarcity of labeled data necessary to train the model. Finger vein image pairs belonging to the same individual are analyzed using verification-based approaches. In order to distinguish between real and fake pairs of finger vein samples, a CNN network is typically fed two pairs of finger vein samples as input [29,30]. However, these methods will need the creation of image pairs in order to perform verification learning.

Methods aiming for identification assign a specific identity to a finger vein. It typically trains the CNN as a classification problem having more than one class, and then end-to-end identification is performed with the help of the trained classification network. With the help of template matching algorithms, authors in [31] achieved a Percentage of Correct Classification (PCC) of 90.72 percent. Kuzu et al. [32] recently researched a new acquisition architecture that allows for on-the-fly capture of finger vein patterns based on a set of low-cost cameras and, hence, has proposed a CNN and Recurrent Neural Network (RNN) based recognition framework. These methods, in general, have been employed for close-set scenarios, i.e., the classifier can only identify the subjects with which it was trained. GAN-based techniques have recently gained the interest of the research community. Despite positive results, GAN-based methods have frequently encountered problems in convergence and training instability [33]. Research is still needed to generate rich quality and miscellaneous finger vein samples employing GAN with restricted or less training data. Yang et al. [34] proposed FV-GAN to extract robust patterns from vein images employing CycleGAN, and they achieved promising results.

Lightweight attention networks have gained much limelight recently. These types of models include spatial, channel, and multi-head attention. They have achieved superior quantitative performance compared

to other deep learning-based techniques. A semantic similarity learning scheme was introduced in the literature [35] to learn preserved discrete binary feature learning. Fang et al. [36] proposed a self-attention-based scheme to enhance authentication. It was a type of Siamese network that gave better results. Their method involved a 3-layer CNN to model feature maps in the global context. A network using multiple CNNs (Merge CNN) with different input images was proposed in the literature [37]. The final network that emerged from the experiment, which gave a superior performance, was based on an enhanced image by CLAHE and the original image. Lastly, a pre-trained model was used by authors in [38] to extract robust features for classification. It was based on a depth-wise separable convolution layer. They employed multi-scale data augmentation to enhance Fingerprint images and curve let transform. Mustafa et al. [39] presented a self-attention-based verification model for finger verification with their custom-made CNN comprising Residual blocks with an attention mechanism. However, their method did not perform well on low-contrast images as they did not use pre-processing or data augmentation techniques. Abdullahi et al. [40] proposed a filtered spatial and temporal sequence-wise multimodal fingerprint and finger vein network (STMFPFV-Net) for finger vein and fingerprint recognition. Their method achieved good results by selecting discriminative contextual sequence features. However, their method did not employ image augmentation before the fusion of different sequences, which affected recognition performance, and their RelieFS method required expensive hyperparameter tuning for network optimization.

This study aims to enhance the FVR performance from a logical view of transfer learning, K-fold cross-validation, and network architecture selection. Although these angles have been conferred in the past, a thorough examination of these methods is required. As an illustration of how transfer learning techniques can overcome data scarcity, many studies start their neural networks with ImageNet pre-trained weights to gain additional prior knowledge. However, the effectiveness of transfer learning mainly relies upon the source knowledge and the target task. There can be some limitations to using wide-ranging object ImageNet samples to help identify finger vein images because their data distribution differs significantly from that of available object ImageNet samples.

3 Theoretical Background

3.1 Transfer Learning

Transfer Learning (TL) has been a crucial part of research since 1995 and is employed in various domains with various titles such as lifelong learning, knowledge transfer, multitask learning, incremental learning, meta-learning, and others [41]. Most research and scientific communities believe that transfer learning, which is a subset of deep learning, can take computational intelligence tasks to further advancement. The critical difference between Machine Learning (ML) and transfer learning is knowledge transfer, i.e., the transfer learning model trained on one dataset can be utilized to accurately classify another dataset with minimum computation. However, in the case of ML, one model can learn for one task only and cannot perform classification for another task.

Transfer learning in CNNs has been in the limelight recently. CNN models require ample training and testing data, which is sometimes a difficult task, especially for real-world applications, which is one of the critical factors owing to the popularity of transfer learning applications in CNNs. Transfer learning is one of the most well-known ideas in machine learning [42]. Various transfer learning strategies depend upon the availability of labeled data. Since a pre-trained model is employed in this study, that comes under the inductive transfer learning method. It takes the background knowledge needed to solve one task and applies it to other tasks that are similar but not the same despite a similar domain. Initially, the base model, which is pre-trained for a specific problem and dataset, is taken. The model is then applied to the target dataset to solve the goal issue [43]. The inductive TL is similar to multitask learning when the source and target domain are the same and many labeled datasets are available. It is similar to self-taught learning when the

source and target domain are somehow related but different, with no labeled data available. In addition, TL can be employed to transfer generic specific knowledge (broader) that is not tied to any specific task or domain, such as spatial relationships, image texture details, language syntax, and others. In addition, most TL-based methods employ feature extraction and fine-tuning techniques [44]. In feature extraction, a pre-trained model with previous weights is utilized to extract fixed features from the input data, which can be fed to any classifier or model as per the target task. Fine-tuning comprises taking a pre-trained model and training it further on a target task. During this process, the model weights are updated using the target task's data while holding some knowledge from the source task. Fine-tuning is mainly employed in tasks where the source and target are narrowly related.

3.2 EfficientNets

Recently, Tan et al. [45] investigated the relationship between the width and depth of CNN models. They proposed a competent way to design CNN models with better classification accuracy and fewer parameters. They termed their new model EfficientNet and proposed seven new models, namely EfficientNetb0 and EfficientNet-b7. The EfficientNets outpaced all the previous models in terms of parameters and accuracy for the ImageNet database [46] by reducing Floating-point Operations per Second (FLOPS) and several parameters. The EfficientNet architecture is much smaller and has few parameters compared to other modern deepnets with similar ImageNet accuracy. For example, EfficientNetb0 has only 5,330,564 parameters compared to 23,534,592 of ResNet50 in the Keras application. Still, EfficientNetb0 underperforms ResNet50.

The key building blocks of EfficientNet include MBConv, Inverted residual connection, and Linear bottleneck. The primary block of the EfficientNet family is mobile inverted bottleneck convolution (MBConv), which is based on the MobileNet model [47]. One key idea of MobileNets is to use depth-wise separable convolutions, which comprise depth-wise and point-wise convolution layers connected in a cascade. The ResidualNets [48] employed skip connections between layers with many channels in the original residual blocks. The theory of squeeze and excitation (SE) is also used in the MBConv layers, further enhancing performance. Each channel in a convolution layer output is given different weightage by the SE block technique instead of treating them all as equal. The EfficientNets use the Swish activation function instead of other commonly used activations. The Swish function shares good performance advantages with Rectified Linear Unit (RELU) and LeakyRelu as it has a similar shape to these activation functions but is flatter comparatively. The Swish activation function equation is defined as follows:

$$f_{\text{Swish}} = \frac{x}{1 + e^{-\beta x}} \tag{1}$$

where $\beta \geq 0$ is a learnable parameter during the CNN training. When $\beta = 0$, f_{Swish} will be a linear activation function, and with $\beta \to \infty$, f_{Swish} behaves similarly to the ReLU function.

As stated previously, the model scaling concept is extremely reliant on the baseline network. The EfficientNetb0 [45] is the baseline network of the EfficientNet family, and the automatic machine learning (AutoML) MNAS framework is employed for its creation, as listed in Table 1. AutoML searches automatically for a CNN model that maximizes precision and efficiency (FLOPS) parameters. The EfficientNetb0 repeatedly uses MBConv1 and MBConv6 layers, which are diverse types of MBConv blocks. The illustration [49] of the MBConv block types (MBConv1 and MBConv6) is shown in Fig. 1. The MBConv blocks include all the concepts discussed above, including inverted residual connection, depth-wise separable convolution, linear bottleneck, Swish activation, and Squeeze and Excitation block. MBConv1 was used at the beginning of the EfficientNet model, and MBConv6 was employed in between several times with different kernel sizes. Also, each block type varies depending on the size of the filter used in the convolution layers inside and whether the block contains an inverted residual connection.

Table 1: EfficientNetb0 architecture

Stage	Operation	#channels	Size
1	Conv 3 × 3	32	224 × 224
2	MBConv1, k 3 × 3	16	112 × 112
3	MBConv6, k 3 × 3	24	112 × 112
4	MBConv6, k 5 × 5	40	56 × 56
5	MBConv6, k 3 × 3	80	28 × 28
6	MBConv6, k 5 × 5	112	14 × 14
7	MBConv6, k 5 × 5	192	14 × 14
8	MBConv6, k 3 × 3	320	7 × 7
9	Conv 1 × 1/Pooling/FC	1280	7 × 7

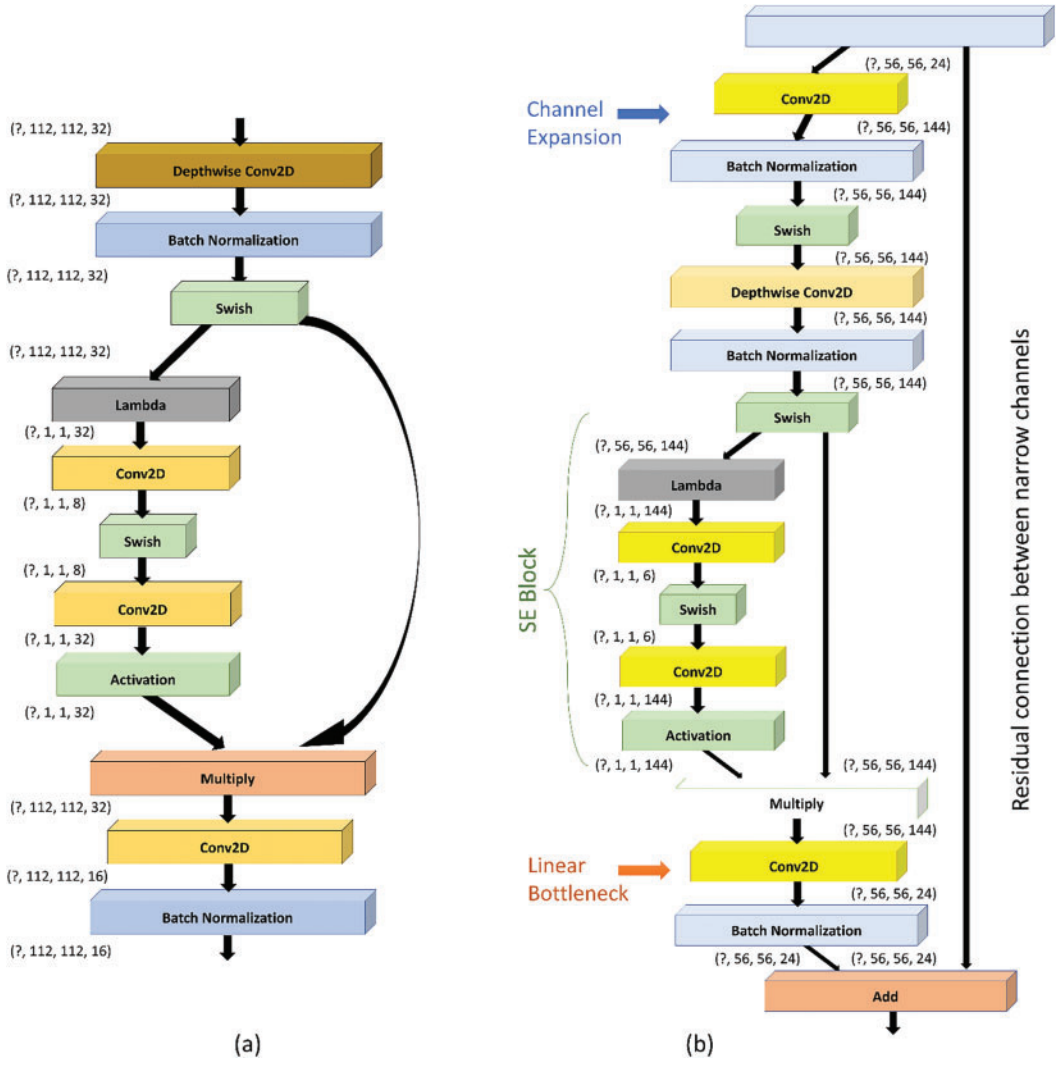

Figure 1: Illustration of MBConv blocks (a) MBConv1 block, (b) Design of MBConv6 block employing SE block with inverted residual connection (24 → 144 → 24)

This study investigates EfficientNets architectures for feature extraction and classification of a finger-vein biometric. EfficientNet incorporates a simple, extraordinarily effective compound coefficient for scaling up CNN models in terms of three dimensions, i.e., depth, width, and resolution. The performance of a model improves by scaling individual dimensions; however, effective performance improvement comes with a balance between all the dimensions. The scaling of each dimension is done by parameter φ, as per Eq. (2). This parameter φ is a user-defined coefficient related to the resources available in model scaling, where the variables α, β, γ are constants experimentally found by a grid search. Other networks can be obtained by varying φ. For example, φ = 1 gives the EfficientNet-b1, for φ = 2 provides the EfficientNet-b2, and others

$$\text{depth: } d = \alpha^{\varphi} \tag{2}$$

$$\text{width: } w = \beta^{\varphi} \tag{3}$$

$$\text{resolution: } r = \gamma^{\varphi} \tag{4}$$

$$\text{s.t. } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \tag{5}$$

$$\alpha \geq 1, \beta \geq 1, \text{and } \gamma \geq 1 \tag{6}$$

Based on Eq. (5) [45], there is a decent trade-off between computational cost and performance. In this work out of the EfficientNets family, EffecientNet-b3 was implemented to strike a balance between computing resources and accuracy.

4 Methodology

4.1 Proposed Framework

The proposed EFI-SATL framework block diagram and the modified EfficientNetb3 model for finger vein recognition are shown in Fig. 2. This is the first time EfficientNet has employed a self-attention mechanism for finger vein recognition. Initially, the input image is fed to the network with pre-processing and data augmentation finger vein images, which increases the database size and facilitates efficient network training. K-fold cross-validation is employed along with data augmentation to balance out the problem of less training data for finger vein biometrics. The database is split into K equal parts before inputting the network for training and testing. Pre-processing, data augmentation, and K-fold cross-validation methods employed in this study are discussed in the following parts.

The designed framework utilizes transfer learning and fine-tuning of the modified EfficientNetb3 model. Like EfficientNetb0, it uses MBConv1 and MBConv6 layers, as discussed in the previous section. The model architecture contains one 3 × 3 Conv layer followed by Batch normalization and Swish activation, then two 3 × 3 MBConv1 layers followed by MBConv6 layers with ten 3 × 3 and fourteen 5 × 5 kernel sizes. Finally, one 1 × 1 Conv layer is used before the Global Average Pooling (GAP) layer. Then, one dropout layer is added, followed by one dense layer. ImageNet weights are used for the initial training of the model with a modified top to assign knowledge obtained from previous training on the ImageNet dataset to the intermediate layers. The attention module is added, and the new model is known as 'EffAttenNet'. The attention module starts by inputting extracted flattened features from modified EfficientNetb3 into one self-attention layer, followed by one dense layer, and finally, a dense output layer. The detailed model information about the architecture of EffAttenNet is shown in Table 2. The total parameters of modified EfficientNetb3 and self-attention are 12,357,423 and 1,367,402, respectively. The use of self-attention layers enhances the feature selection for classification and helps the network focus on more prominent features that can get skipped without any attention mechanism, as it gives a weighted sum of values based on the query and key vector similarities.

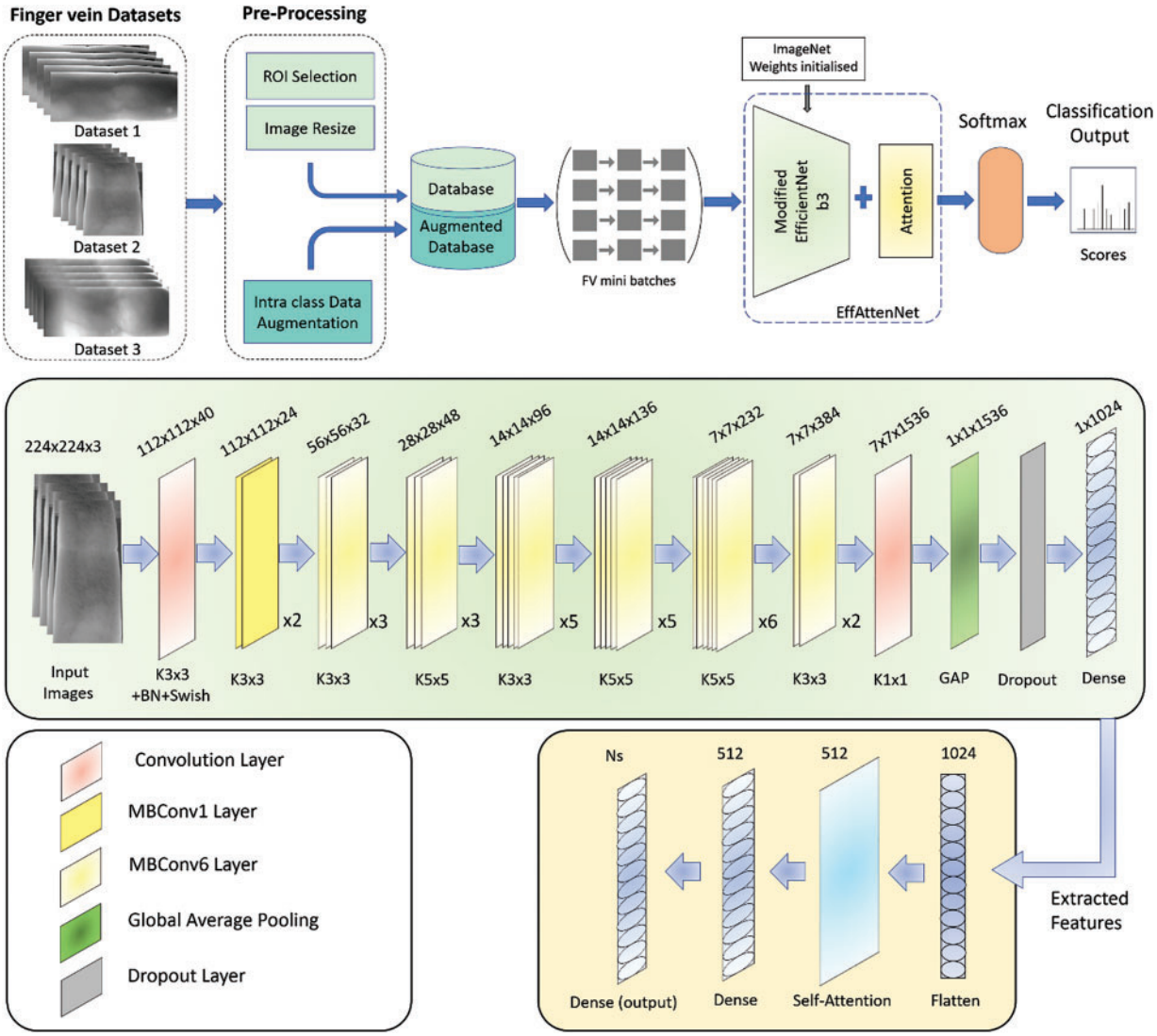

Figure 2: The block diagram of the proposed EFI-SATL framework with the modified EfficientNetb3 architecture

Table 2: The EffAttenNet model summary with self-attention mechanism

Modified EfficientNetb3			**Attention**		
Layer (type)	Output shape	Param #	Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 224, 224, 3)	0	Self-attention_1	(None, 512)	1,050,368
efficientnetb3 (Functional)	(None, 7, 7, 1536)	10,783,535	dense_8 (Dense)	(None, 512)	262,656
global_average_ pooling2d_1(Global AveragePooling2D)	(None, 1536)	0	dense_9 (Dense)	(None, 106)	54,378
dropout_1 (Dropout)	(None, 1536)	0			

(Continued)

Table 2 (continued)

Modified EfficientNetb3	Attention
dense_2 (Dense) (None, 1024) 1,573,888	
Total params: 12,357,423 (47.14 MB)	Total params: 1,367,402 (5.21 MB)
Trainable params: 12,270,120 (46.81 MB)	Trainable params: 1,367,402 (5.21 MB)
Non-trainable params: 87,303 (341.03 KB)	Non-trainable params: 0 (0.00 Byte)

The complete process of the EFI-SATL framework is shown in Algorithm 1. All the database images are initially resized for Region of Interest (ROI) selection using the per-processing method stated above, followed by intra-class data augmentation operations. These operations involve random (*Rand*) rotation, scaling, shear, reflection, and translation of Region of Interest (ROI) images to enhance data generalization. Then, the data-augmented images are prepared to be fed for model training using *ImagedataAugmentor* and K-fold cross-data division. The model is then trained in two configurations, with and without attention mechanism, i.e., modified EfficientNetb3 and EffAttenNet. Initially, modified EfficientNetb3 is trained, and features are extracted. Thereafter, the self-attention layer is added to attain more useful information from the extracted features, followed by fine-truing and optimizing hyperparameters. Hyperparameter tuning is essential to achieve a trade-off between the complexity and simplicity of the model. Lastly, the models are tested on test data in two configurations, as mentioned earlier. Then, the scores regarding AUC metrics and receiver operating curve (ROC) plots for all the databases are computed.

Binary Cross-Entropy, frequently termed the cost function, is utilized as the loss function for the architecture. Typically, binary cross-entropy is stated by the average cross-entropy across all data samples, as defined by the equation below:

$$L_{CE} = -\frac{1}{N}\sum_{k=0}^{N}\left[t_k \log\left(sp_k\right) + \left(1 - t_k\right)\log\left(1 - sp_k\right)\right] \tag{7}$$

where sp_k is the softmax probability for the kth data points, t_k is the truth value taking a value 0 or 1, and the number of scalar values is represented by N in the model output.

Algorithm 1: Proposed framework: EFI-SATL

Require: Input Dataset *(ID)* ▷FVUSM, HKPU, SDUMLA
Ensure: Finger vein recognition *(Y)*, Performance Parameters *(P)* ▷ *Y* (class variable)
1: ROI selection *(RS)* ⇐ *ID* ▷ Input Data is passed for cropping
2: Image resize *(IRS)* ⇐ *RS* ▷ Cropped Data is passed to flipping
3: ImagedataAugmentor: *Rand* rotation, scale *(RRS)* ⇐ *IRS* ▷ Intra-class DataAugmentation started
4: *Rand* shear, reflection *(RSR)* ⇐ *RRS*
5: *Rand* translation *(RT)* ⇐ *RSR*
6: Final data prepared *(D)* ⇐ *RT* ▷ ImagedataAugmentor output
7: *K-fold* division (K_fD) ⇐ *D* ▷ *D* is divided using K-fold cross validation for training and testing
8: ***for*** *i* ◂ *1 to 9* ***do*** ▷ Repeat the experiment 9 times
9: Modified *EfficientNetb3* ⇐ K_fD_i ▷ Pass dataset to deepnet employed
10: Extract finger vein features F_i using modified *EfficientNetb3*
11: Self−attention ⇐ F_i ▷ Add self-attention layer

(Continued)

Algorithm 1 (continued)

12: Dense $\Leftarrow F_i$ ▷ Add final dense layer
13: Perform hyper-parameters optimization ▷ Adam is used
14: Compute Y by passing test data to *EFI-SATL* using optimal parameters
15: Evaluate P
16: ***end***
17: Compute average P
18: Plot *ROC*

4.2 Self-Attention

The features extracted by the modified EfficientNetb3 deepnet are fed to the attention mechanism consisting of one self-attention layer. The self-attention mechanism has a significant advantage in working with computer vision tasks as it gains valuable knowledge regarding the dependencies and complex patterns among the input sequences, helping the network understand better. Self-attention captures long-range dependencies by allowing each element to attend to all other elements in a sequence for image classification tasks between images, producing better results than other attention modules [50]. This helps the model understand a more comprehensive context while performing parallel computation for efficient and fast processing, unlike simple attention, which works sequence-to-sequence. Also, self-attention's ability to handle variable length sequence data provides extended flexibility to the model to learn various factors in the same sequence and enhance the model's overall performance [51]. The major components of the self-attention mechanism [52] are Query (q), Key (k), and Value (v). These vectors help in calculating weights and attention output; these are computed as follows:

$$q = x.w_q;\ k = x.\,w_k \text{ and } v = x.w_v \tag{8}$$

where x is the input sequence and w_q, w_k, w_v are the learnable weight matrices. With the help of q and k, the attention weights A_{tt} are calculated as follows:

$$A_{tt} = Softmax\left(\frac{qk^t}{\sqrt{d_k}}\right) \tag{9}$$

where d_k is the dimensionality of the key vectors. The output O of the attention layer is calculated as:

$$O = A_{tt}.v \tag{10}$$

The performance classification can be enhanced using attention in CNNs focusing on important regions after the deepness cannot extract meaningful information anymore [53]. This study employs one self-attention layer with 512 units for feature selection purposes, followed by a dense layer. The Adam optimizer with an initial learning rate of 0.0001 is utilized to optimize the learning process. Adam's algorithm integrates the momentum principles of feature learning with adaptive learning rates, hence achieving better results.

4.3 Pre-Processing and Data Augmentation

One of the crucial steps in computer vision tasks is pre-processing. This method can help eliminate unwelcome noise and emphasize features of the image that are beneficial for the recognition work or training phase to learn image features. The proposed work employs minimum pre-processing tasks to maintain the original features of the finger vein images. First, ROI extraction from finger vein images is performed using the Lee Mask [54] method, and then the extracted images are centered horizontally using the Huang

Normalization [55] method. Secondly, all the images are resized to 224 × 224 to sustain compatibility per the network architecture. For the HKPU dataset, segmented masks released by the authors [22] are employed for ROI extraction and alignment of finger vein images. Finally, before giving input to the CNN network, the pre-processed images are normalized to [−1, 1].

The data augmentation technique is employed to upsurge the training data in case of original data scarcity, which also helps the network to train better. Data samples are increased by transforming the original images to keep semantic information intact in the original data images. Deep nets based on CNN necessitate many data samples to achieve high-performance accuracy. A smaller dataset can result in overfitting problems where the network performs too well during training and is very poor while testing. Hence, data augmentation is one of the salient steps. Table 3 shows the various intra-class data augmentation methods/properties using the ImagedataAugmentor used in this work. Some sample images of all three databases are shown in the Fig. 3.

Table 3: Data augmentation with parameters

Method/Properties	Parameters/Range
RandRotation	[−20 20]
RandScale	[1 1]
RandXShear	[0 0]
RandYShear	[0 0]
RandXReflection	01
RandYReflection	01
RandXTranslation	[−3 3]
RandYTranslation	[−3 3]

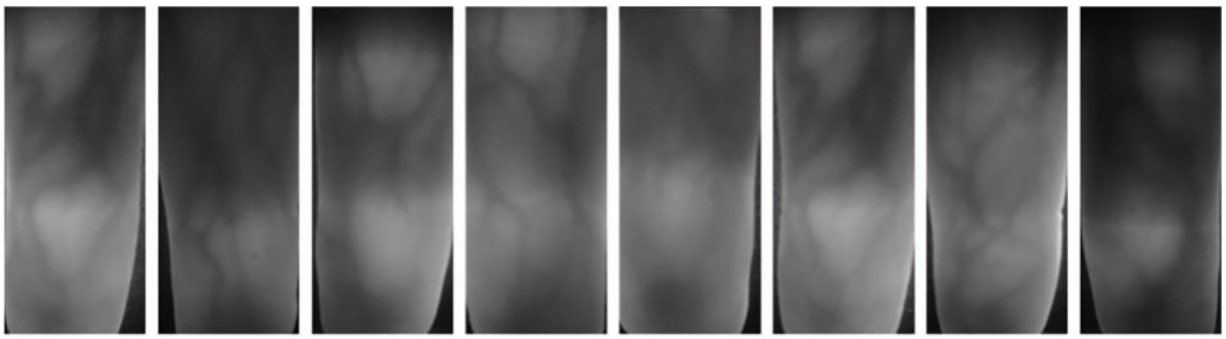

Figure 3: (Continued)

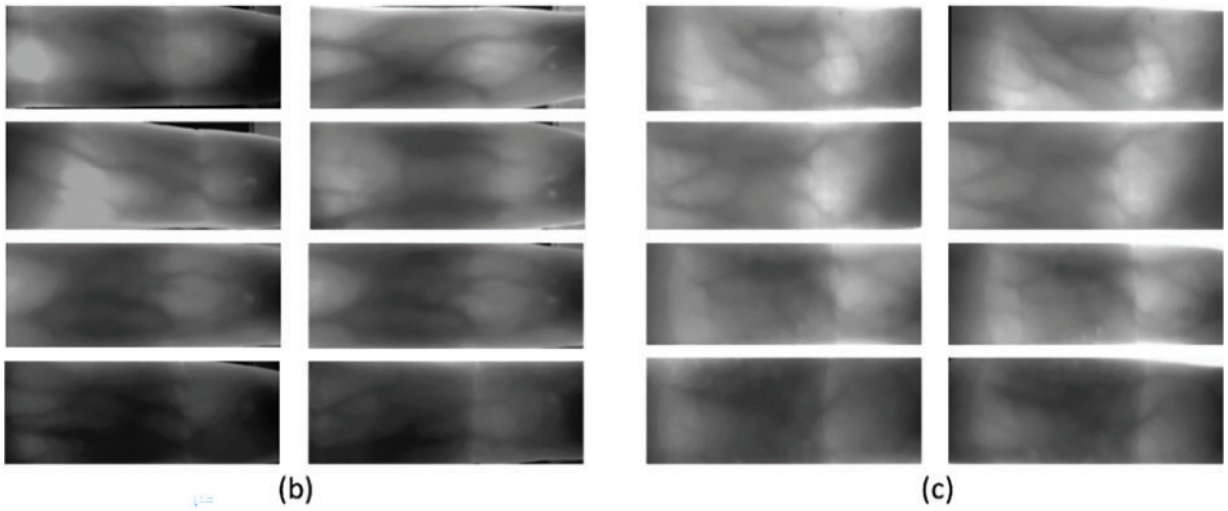

Figure 3: Sample finger-vein images of various databases, (a) FVUSM, (b) SDUMLA, and (c) HKPU

4.4 Evaluation Metrics

Standard assessment metrics were utilized to calculate the performance of the proposed model, which are precision (*Precision*), recall (*Recall*), Sensitivity (*Sen*), Specificity (*Spe*), f-score (*F*1), accuracy (*Acc*), and *ROC*. The ratio of correctly predicted positives *vs*. all true positives is known as Sensitivity, and the proportion of correctly predicted negatives *vs*. all true negatives is termed Specificity. Accuracy gives the rate of correctly classified samples. Precision is the proportion of correctly predicted positives out of all the positives. These standard metrics are based on the fundamental indicators, namely, true positive (*tp*), true negative (*tn*), false positive (*fp*), and false negative (*fn*). *tp* provides images classified correctly in each category, whereas *tn* represents the sum of correctly classified images in all categories other than the applicable category. *fn* is the figure of misclassified images in the related category, while *fn* represents the number of misclassified images in all categories other than the relevant. These metrics are defined as follows:

$$Precision = \frac{tp}{(tp + fp)} \tag{11}$$

$$Recall = \frac{tp}{(tp + fn)} \tag{12}$$

$$Senstivity = \frac{tp}{(tp + fn)} \tag{13}$$

$$Specificity = \frac{tn}{(tn + fp)} \tag{14}$$

$$F1 - Score = 2x\left(\frac{precision\,x\,recall}{precision + recall}\right) \tag{15}$$

$$Accuracy = \frac{(tp + tn)}{(tp + tn + fp + fn)} \tag{16}$$

4.5 K-Fold Cross-Validation

The statistical technique for estimating machine learning models' competency is called Cross-validation. It is frequently employed with machine learning algorithms to compare and select models for a given predictable modeling problem because it is simple to learn and implement and generates estimates with lower bias than other techniques. This is because the mean estimate of any parameter is less biased than a one-shot estimate [56]. This type of validation process involves dividing the data into k portions, with one portion serving as testing data and the remaining k-1 serving as training data. The popular practices by researchers in the literature use K = 2, 3, 5, 10, and other fold values, but due to the added computation cost of cross-validation and to maintain balance in computation and performance, K = 3 is used. For K = 3 initially, fold 1 will be used as test data, and folds 2 and 3 will be utilized to train the model for the first prediction. After that, fold 2 will be used as test data and folds 1, 3 will be used as train data for another prediction. Finally, fold 3 will be used as test data, and folds 1 and 2 will be used as train data for the last prediction. The obtained average performance and standard deviation will be testified. This algorithm is computationally expensive but does not leave unwanted data behind to validate the results. This is a significant advantage for solving some problems, such as inverse inference, where the sample number is small.

4.6 Network Initialization and Optimization

Binary Cross entropy (BCE) loss is employed for the model training and validation to define the preferred illustration of features. Adam optimizer was employed with a 32 as the batch size for training optimization. The initial learning rate was kept at 0.0001 for all datasets and optimized with a reduction of 10 on the plateau or after 20 epochs. The momentum was set to 0.9 value for accelerating the convergence of gradient vectors. The max epochs were set as 50, and the training halted with the minimization of the validation loss. The weights of the pre-trained Imagenet model were employed to initialize the considered deepnet and assess the effectiveness of transfer learning for this application. Glorot uniform initialization was employed for the fully connected layer.

5 Results and Analysis

5.1 Employed Databases and Experimental Setup

Experiments were conducted on three different types of publicly available vein databases, namely the SDUMLA finger vein dataset [57], the HKPU finger vein dataset [58], and the FVUSM database [59], to assess the strength of the considered architectures under various input types. Before being fed into the recognition system, all the considered samples were resized to 224 × 224 pixels and normalized to 0 means with 1 variance. The main reason for choosing these particular databases is the current literature, with most researchers using them, which allows for an accurate comparison of these tried-and-true techniques. Table 4 provides an overview of the two databases under consideration, and the following subsections provide more information about them. Different fingers and hands of the same user were categorized as separate classes.

There are 156 male and female volunteers in the HKPU finger vein image database. Hong Kong Polytechnic University used a contactless imaging device between April 2009 and March 2010. It comprises 3132 Bitmap (BMP) images with a resolution of 513 × 256 pixels, taken from 156 subjects. Ninety-three percent of the subjects in this dataset are under the age of 30, and images of the finger veins of 105 people were collected over 02 sessions separated by an average of 66.8 days, with a month of minimum interval and six months of maximum interval. Subjects were asked to bring six images of the index and middle fingers on their left hands for each session. The remaining 51 subjects have only one data collection session to their name.

Table 4: Various publicly available finger-vein databases

Database	Image size	# Subjects	# Fingers	Images per finger	Finger details	Sessions	Total images
SDUMLA	320 × 240	106	6	6	Index, middle and ring fingers of both hands	1	3816
HKPU	513 × 256	156	2	12	Index and middle finger of left hand only	2	3132
FVUSM	640 × 480	123	4	12	Index, and middle finger of both hands	2	5904

The Shandong University of China compiled the SDUMLA database. It includes finger vein images of 636 fingers from 106 people. Six images were captured in grey-level BMP format with 320 × 240 pixel resolution from the index, middle, and ring fingers of each person's left and right hands. The University of Sains Malaysia maintained the FV-USM database. It consists of vein images from 123 subjects' left and right index and middle fingers. There are 83 males and 40 females, with an average age of 2052. Images were collected in two separate sessions, each with six images per finger. The images are all in grayscale BMP format and have a resolution of 640 480 pixels.

This study performed training and testing with a desktop workstation comprising an Intel Core^{TM} Xeon Gold CPU with 64 GB RAM and an NVIDIA Quadro 5000 graphics processing unit (GPU) card with a graphics memory of 16 GB. Windows 10 OS with MATLAB 2021b was employed for pre-processing. Training and validation were done using the Tensorflow AI Python framework along with the Anaconda package manager.

5.2 Performance of the Proposed Method on Various Databases

The effectiveness of the proposed method on the considered databases is discussed in this section. These datasets were tested as per the K-fold cross-validation process, and the evaluated metrics were reported. State-of-the-art deepnets, i.e., ResNet18, ResNet50, Xception, Inception, and Dense Net 121, were considered as stated above for comparison to the proposed method to assess the effectiveness of transfer learning applied to the model. Two separate networks, EfficientNetb3 and EffAttenNet, are considered to understand the performance difference of the proposed method with and without attention. As mentioned earlier, EfficientNetb3 is without attention and EffAttenNet is with the attention mechanism.

The network's recognition accuracy is evaluated first. The input image is scaled down to 224 × 224 pixels for the network model. The state-of-the-art models: ResNet18, ResNet50, Xception, Inception, Dense Net 121, and EfficientNetb3 outperform the proposed method EffAttenNet with 98.14% for the HKPU dataset, 99.03% for FVUSM dataset and 99.50% accuracy for SDUMLA dataset, as shown in Fig. 4. It can also be stated that clear difference in performance can be achieved with the use of attention mechanism, as the accuracy differ by 1.26% for HKPU, 1.47% for FVUSM and 1.2% for SDUMLA databases, in between EfficientNetb3 and EffAttenNet. These results show that the proposed approach accurately identifies the finger vein images. In addition, the proposed method's training and validation accuracy plots for all the databases are shown

in Fig. 5. The method performed well in the training and validation process except for the HKPU dataset, with some fluctuations due to low-quality images. In addition, the same can be observed for the loss curves shown in Fig. 6, which show that the losses are more for the HKPU database than for the SDUMLA and FVUSM databases.

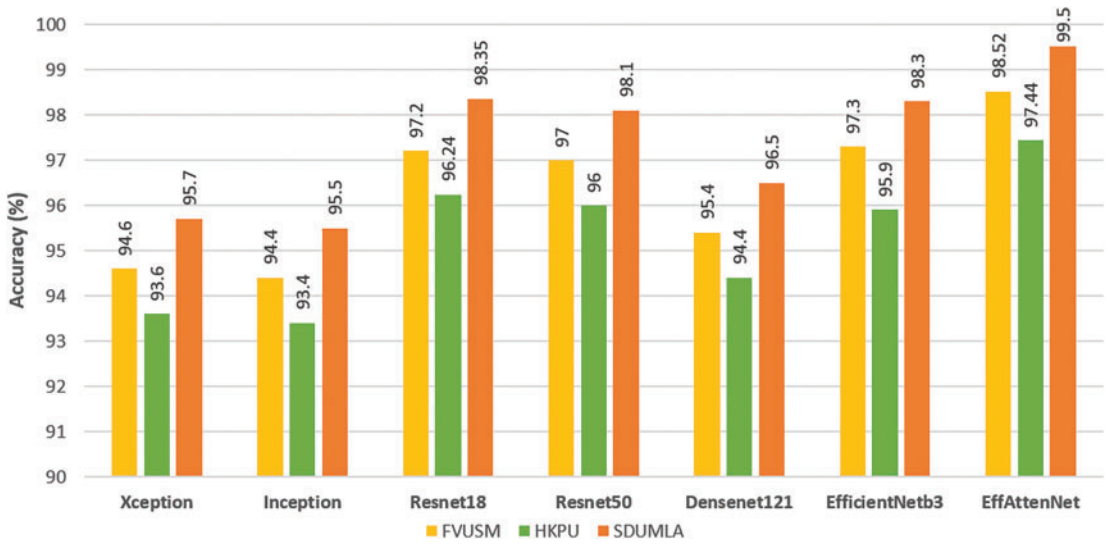

Figure 4: Recognition accuracy of various deep nets and proposed net on the databases

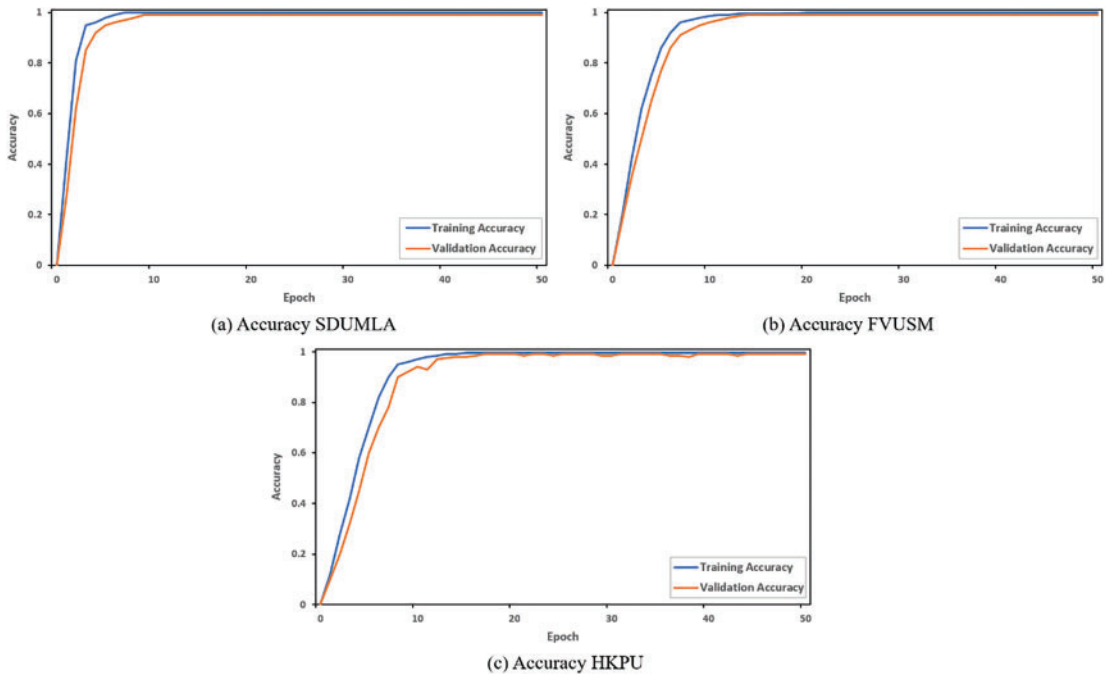

Figure 5: Training accuracy of the proposed method on all the databases

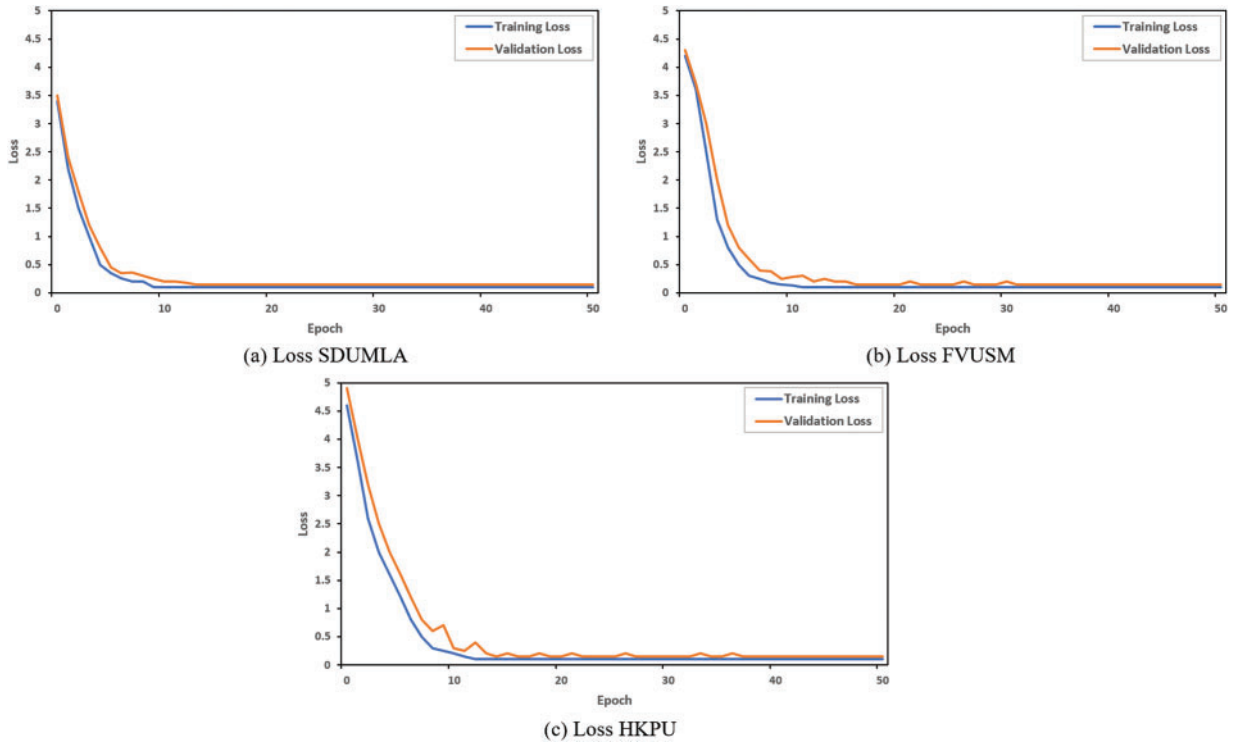

Figure 6: Training loss of the proposed method on all the databases

The proposed network model was tested using the K-fold cross-validation dataset and performed excellently on training data. When performing experiments with the proposed scheme, the dataset is divided into three equal parts, as K = 3. Each fold alternates between randomly selecting two to be used during training and one to be used during validation. Compared to other network architectures, the proposed method metrics, namely, precision, recall, F1-score, Sensitivity, Specificity, and accuracy, are shown in Table 5, which consists of average values of the three folds for all the databases.

Table 5: The comparative analysis of the proposed method on all the databases

Database	Model	Precision	Recall	F1	Accuracy
FVUSM	ResNet18	97.11	97.32	97.35	97.24
	ResNet50	96.95	97.18	97.20	97.00
	Xception	94.57	94.84	94.80	94.63
	InceptionV3	94.33	94.55	94.46	94.40
	Dense Net 121	95.37	95.57	95.60	95.44
	EfficientNetb3	97.27	97.46	97.50	97.56
	EffAttenNet	**99.18**	**99.25**	**99.45**	**99.03**
HKPU	ResNet18	96.18	96.30	96.33	96.28
	ResNet50	95.97	96.17	96.28	96.10
	Xception	93.53	93.76	93.67	93.60
	InceptionV3	93.36	93.47	93.60	93.42
	Dense Net 121	94.41	94.57	94.36	94.40

(Continued)

Table 5 (continued)

Database	Model	Precision	Recall	F1	Accuracy
	EfficientNetb3	96.27	96.53	96.56	96.88
	EffAttenNet	**98.24**	**98.46**	**98.21**	**98.14**
SDUMLA	ResNet18	98.21	98.40	98.43	98.35
	ResNet50	98.16	98.26	98.30	98.11
	Xception	95.67	95.87	95.90	95.70
	InceptionV3	95.46	95.69	95.73	95.55
	Dense Net 121	96.48	96.66	96.70	96.50
	EfficientNetb3	98.39	98.59	98.61	98.40
	EffAttenNet	**99.47**	**99.67**	**99.70**	**99.50**

Also, the various AUC metrics achieved by EffAttenNet on various datasets are shown in Fig. 7, with the help of a box plot. The box plot is made for all the folds of the experiment and hence shows the average value of all the folds. The box plot clearly shows that EffAttenNet performed efficiently and accurately identified the subjects with a Precision, Recall, and F1-score value of 99.18%, 99.25%, 99.45% for FVUSM dataset and 98.24%, 98.46%, 98.21% for HKPU dataset and finally 99.47%, 99.67%, 99.70% for SDUMLA dataset, respectively.

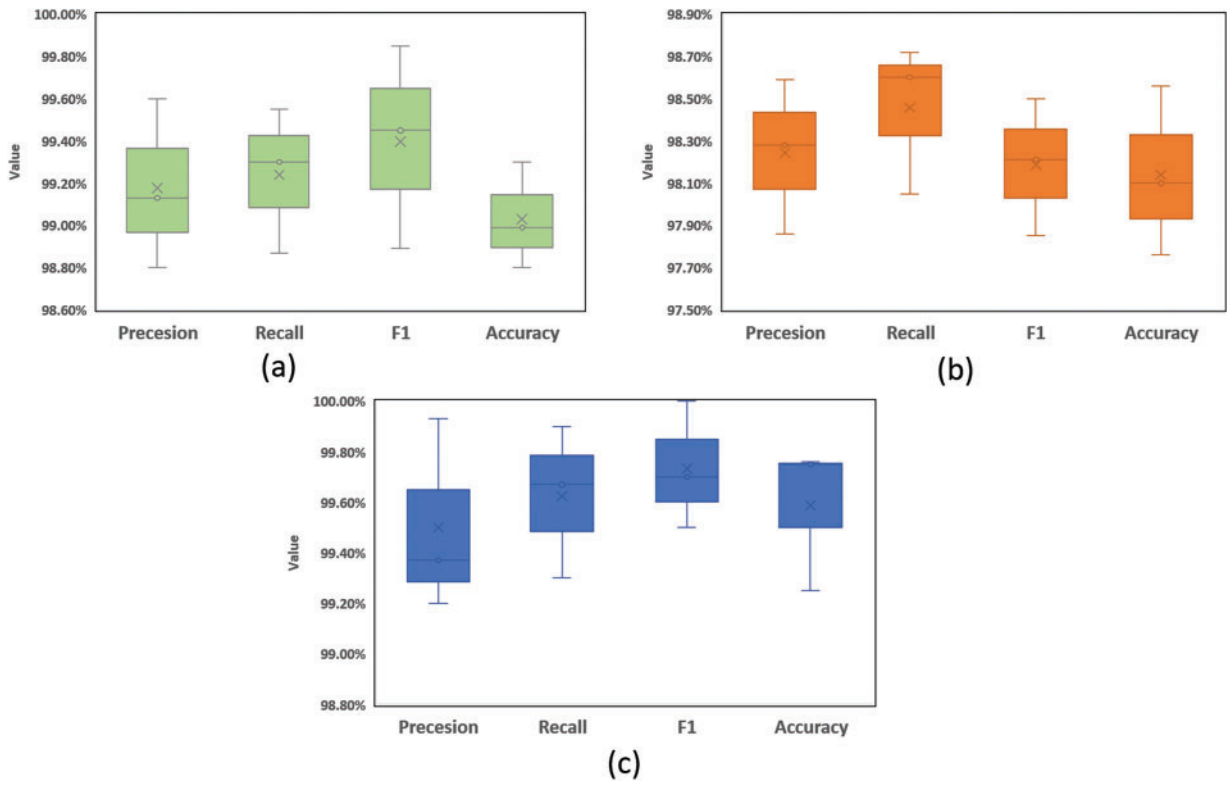

Figure 7: The AUC metrics using box plot of various datasets (a) FVUSM, (b) HKPU, and (c) SDUMLA

In addition, the performance of the proposed EFI-SATL framework is evaluated by plotting the ROC on various datasets along with state-of-the-art deepnets in Fig. 8. The ROC curve includes the performance of the proposed method with and without attention mechanism as EfficienNetb3 and EffAttenNet, respectively. The inclusion of the attention mechanism increases the model's recognition performance. The main motive for including deepnets for ROC is to get a clear picture of the method's performance against the state-of-the-art deepnets.

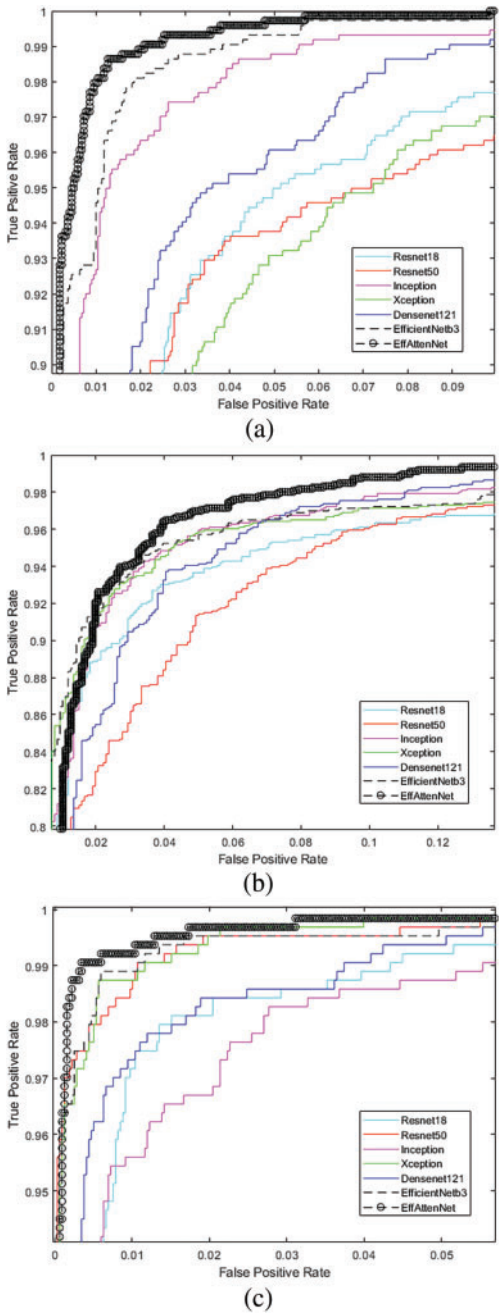

Figure 8: The ROC plot obtained of EFI-SATL along with various deepnets on (a) FVUSM, (b) HKPU, and (c) SDUMLA databases

In addition, the Cumulative Curve (CMC) plot for various databases is shown in Fig. 9. It clearly shows the enhanced performance of the proposed method for all the databases and proves its effectiveness for large variable datasets.

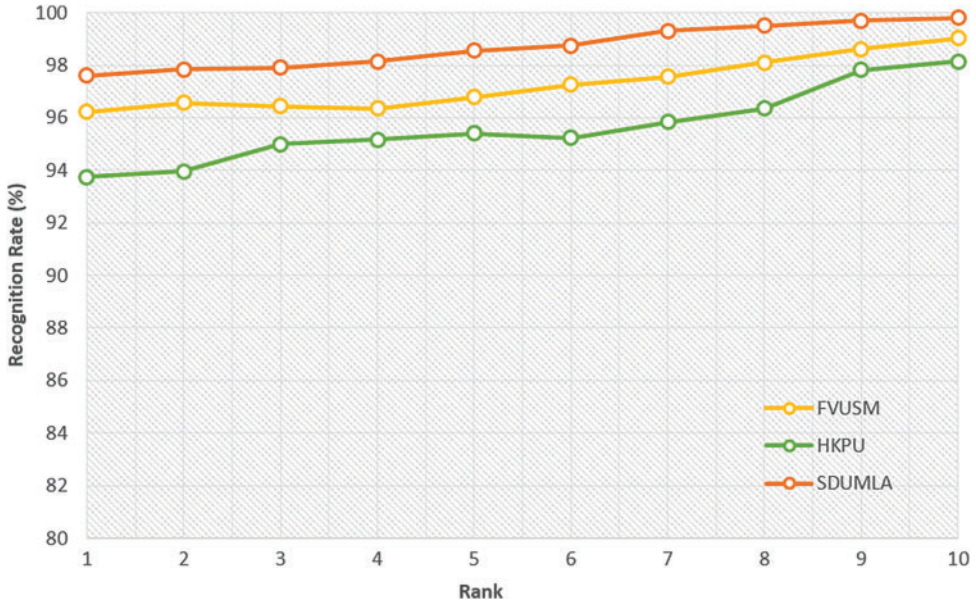

Figure 9: CMC curves showing recognition accuracy *vs.* rank for all the databases

5.3 Comparing the Existing State-of-the-Art Algorithm

The proposed method was compared in terms of accuracy with other Deep Learning (DL) based finger vein recognition systems shown in Table 6 to assess the strength of the proposed method and comparatively analyze the outcome. All three datasets were utilized to compare the performance of the proposed method and existing finger vein recognition methods. The state-of-the-art existing methods, i.e., CNN, ResNet18, Finger Vein Generative Adversarial Network (FV-GAN), Semantic Similarity Preserved Discrete Binary Feature Learning (SSP-DBFL), Self-Attention Convolution (SAC) Siamese Net, Deeply fused CNN, Xception Net, Self-Attention based Convolutional Neural Network (SA-CNN), Convolutional Auto Encoder (CAE), and Fusion of Global and Local Feature Network (FGL Net) recognition comparison is listed in Table 5 which elaborate the enhanced performance of the method over the others.

In addition, the column plots are plotted to compare the existing works with the proposed methods, as shown in Fig. 10. The plots clearly show that EffAttenNet performed efficiently on SDUMLA, FVUSM, and HKPU datasets. However, the SSP-DBFL method performed much closer for the HKPU dataset than the proposed method. Except for this case, for all the other cases, the proposed EFFAttenNet achieved remarkable performance on the rest of the databases. In addition, the training and testing time analysis are listed in Table 7, comparing the existing works with the proposed EFI-SATL framework. The proposed method is computationally efficient compared to the state-of-the-art works and outperforms all work in computation time.

Table 6: Finger vein recognition comparison to existing deep learning-based methods on all the datasets

Ref.	Year	Method	HKPU	FVUSM	SDUMLA
[18]	2019	CNN	96.35%	98.12%	98.00%
[30]	2021	ResNet18	97.05%	98.20%	—
[34]	2019	FV-GAN	—	—	97.30%
[35]	2023	SSP-DBFL	98.08%	98.10%	99.07%
[36]	2023	SAC-Siamese Net	—	98.14%	99.06%
[37]	2022	Deeply fused CNN	—	96.15%	96.30%
[38]	2022	Xception Net	—	—	96.10%
[39]	2023	SA-CNN	97.10%	—	98.56%
[60]	2020	CAE	—	97.69%	98.60%
[61]	2023	FGL Net	—	97.78%	—
		Proposed method	**98.14%**	**99.03%**	**99.50%**

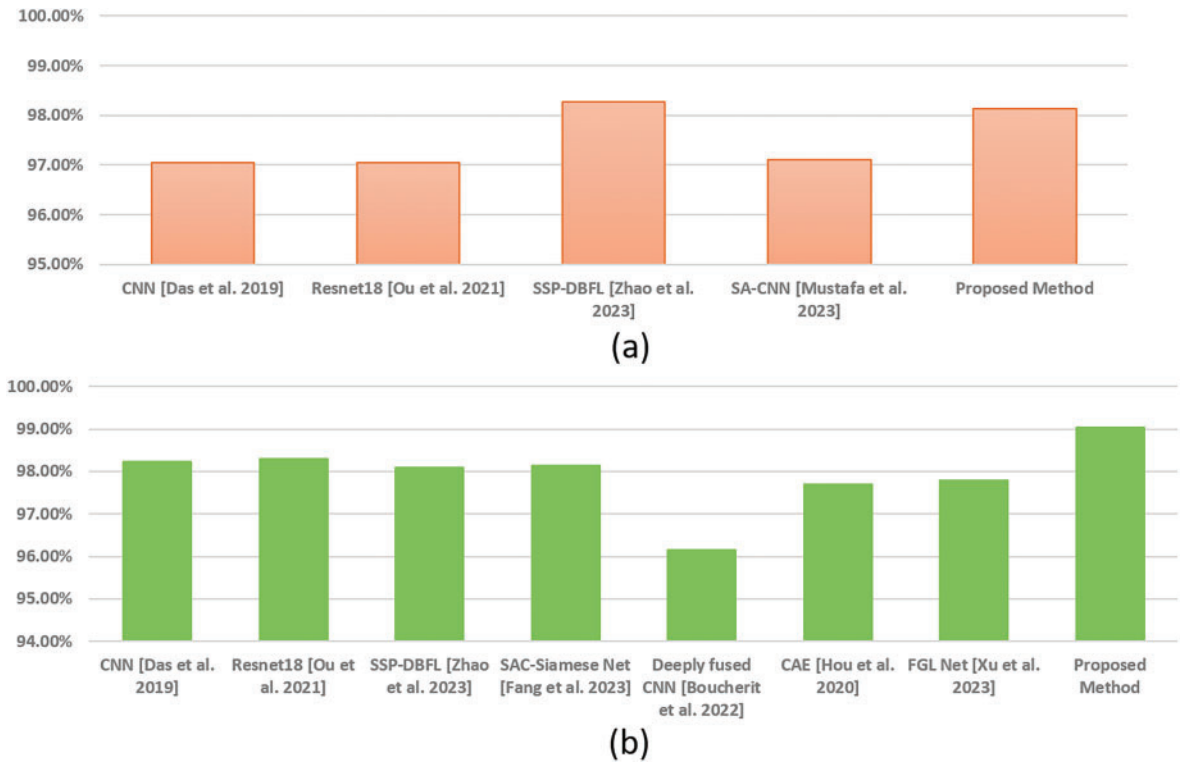

Figure 10: (Continued)

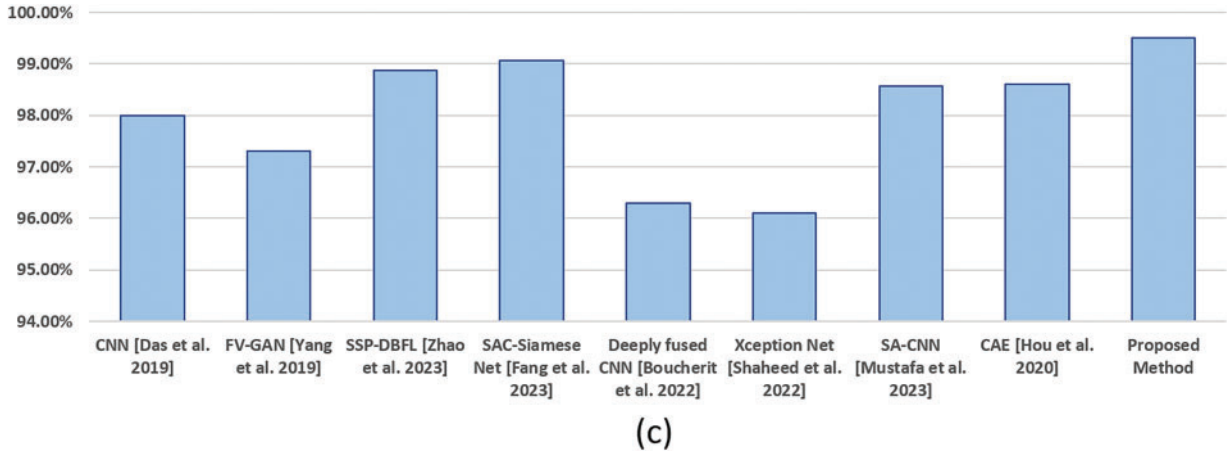

Figure 10: Comparison of the existing works with the proposed method for (a) HKPU, (b) FVUSM and (c) SDUMLA databases [18,30,34–39,60,61]

Table 7: Training and testing time analysis of existing deep learning-based methods on all the datasets

Ref.	Year	Method	HKPU		FVUSM		SDUMLA	
			Training	Prediction	Training	Prediction	Training	Prediction
[18]	2019	CNN	145.8 s	2.01 s	147.3 s	1.9 s	148.1 s	2.2 s
[30]	2021	ResNet18	128.3 s	1.8 s	127.3 s	1.65 s	—	—
[34]	2019	FV-GAN	—	—	—	—	158.7 s	2.6 s
[35]	2023	SSP-DBFL	124.2 s	1.31 s	126.2 s	1.62 s	134.7 s	1.78 s
[36]	2023	SAC-Siamese Net	—	—	125.4 s	1.53 s	128.4 s	1.84 s
[37]	2022	Deeply fused CNN	—	—	134.0 s	1.86 s	137.2 s	1.95 s
[38]	2022	Xception Net	—	—	—	—	136.8 s	1.60 s
[39]	2023	SA-CNN	126.6 s	1.22 s	—	—	129.3 s	1.62 s
[60]	2020	CAE	—	—	133.5 s	2.15 s	137.5	2.38 s
[61]	2023	FGL Net	—	—	127.4 s	1.44 s	—	—
		Proposed method	**120.5 s**	**1.0 s**	**122.0 s**	**1.24 s**	**124.1 s**	**1.36 s**

6 Discussion

In this part, the performance of the proposed EFI-SATL method relative to alternative methods is elaborated using the data from Section 5. The superiority of the approach for identifying and classifying finger vein images is established in this section. Several factors influence the decision to use deep transfer learning with a pre-trained EfficientNetb3 model with attention mechanism, data augmentation, and K-fold cross-validation: (i) inspired by state-of-the-art approaches based on deep transfer learning neural networks, (ii) data on the minimal size of finger veins available, (iii) to attain an authentication accuracy level suitable for small datasets. The proposed work began by enhancing and normalizing the raw data using a data pre-processing method. Then, the data augmentation method expands the sample size of the vein input image for training. Then, K-fold cross-validation is employed to train and test the model. Lastly, the pre-trained and modified EfficientNetb3 model is employed for feature information extraction and the self-attention mechanism to classify or identify finger vein images. The outcomes from Table 5 demonstrate that

the proposed method (EffAttenNet) outperformed all other methods considered in this work, employing three distinct databases. The employed method produced good results for all K-fold (K = 3) parameters examined in this work. For example, the F1 and accuracy are not less than 98.21% and 98.14% on 03 data fold using three standard FVUSM, HKPU, and SDUMLA datasets. In addition, the ROC curve is elaborated and demonstrated for model performance; computing the ROC curve indicates the best model out of the 03 folds. Also, it is confirmed that the proposed method beats all other models, as illustrated in Fig. 8, which shows various ROC plots.

In addition, applying the proposed EfficientNetb3 model in conjunction with the self-attention mechanism and data augmentation with K-fold cross-validation improves the best accuracy achieved for identification by 1.2% for SDUMLA, 1.26% for HKPU and 1.47% for FVUSM database. Thus, the proposed model EffAttenNet can accurately recognize finger vein images provided with limited data. One of the reasons is that the EfficientNets architecture incorporates MBConv blocks, inverted residual connections, and squeeze and excitation blocks, significantly improving recognition accuracy [49]. Using the self-attention mechanism helps to learn more detailed representations from vein images. EfficientNets architecture is relatively simple to implement compared to other deep Net architectures. It should be considered that the pre-trained EfficientNetb3 model is initialized using ImageNet weights, a deep transfer learning approach. This facilitated training the proposed model on finger vein datasets and obtaining good scores. In addition, the attention mechanism can help to focus on a single context rather than multiple ones to achieve better results for the cases where CNNs reach a plateau and cannot extract any useful data. Table 6 demonstrates the superior performance of the proposed method with an increase in recognition accuracy of 0.06% for HKPU, 0.83% for FVUSM, and 0.43% increase for SDUMLA databases compared to the existing works. Also, the proposed framework is proved to be computationally efficient compared to existing works in Table 7 in terms of training and prediction time analysis. The proposed model EffAttenNet correctly identifies the individuals with images showing a clear difference in the vein and non-vein regions for most images. However, some of the low-quality images with unclear vein regions were misclassified by the approach, as shown in Fig. 11. The proposed method achieved less accuracy on the HKPU dataset than FVUSM and SDUMLA databases as compared to existing works, possibly due to the fewer fingers considered while forming the database and its lower data size.

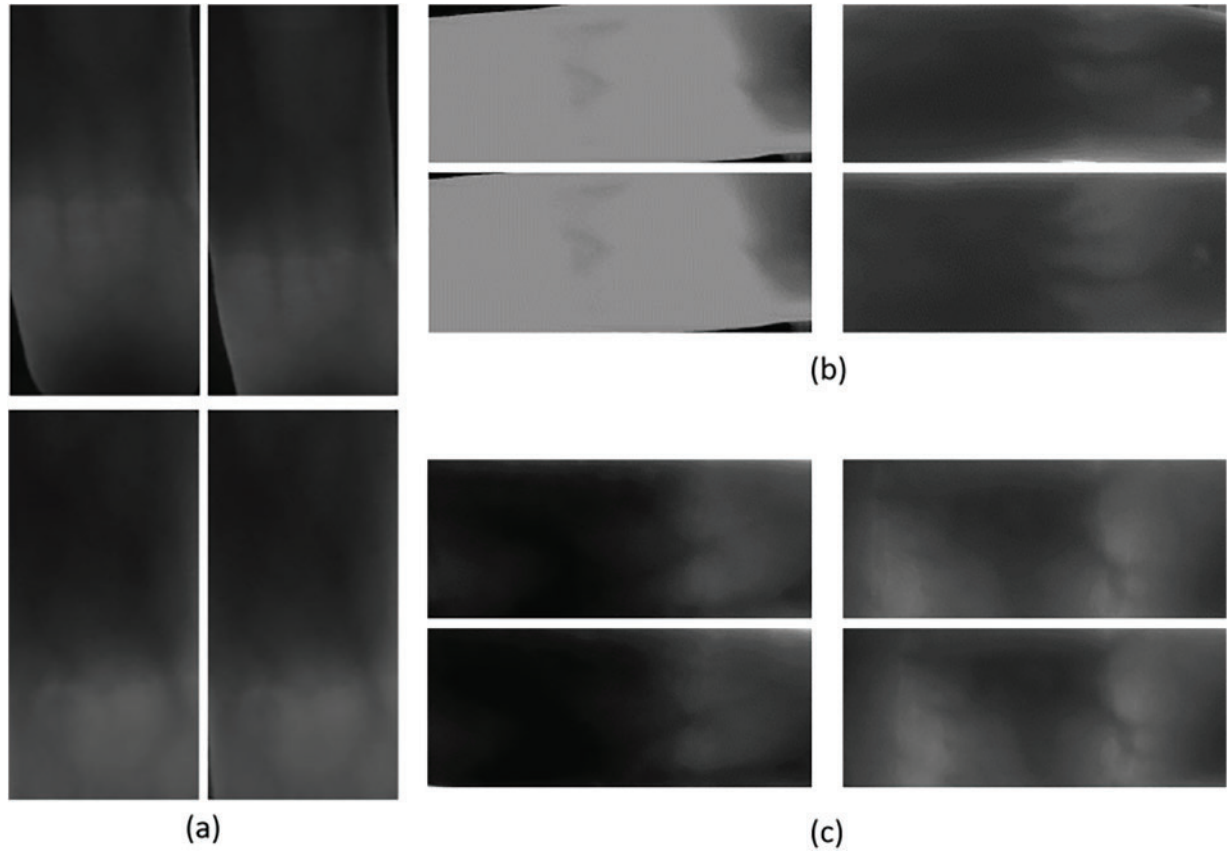

Figure 11: Finger vein image pairs with no apparent difference between vein and no-vein regions, (a) FVUSM, (b) SDUMLA, and (c) HKPU datasets

7 Conclusion and Future Scope

This research provides a novel and simplified method for finger vein biometric detection based on deep transfer learning. The proposed EFI-SATL framework employs feature extraction and classification with the help of the deep CNN EfficientNetb3 model, transfer learning to enhance feature extraction of finger vein biometric images combined with the self-attention mechanism for improved biometric recognition. In addition, the data augmentation concept is employed to broaden the database size and enhance the proposed model's usefulness. The proposed framework is competed with other deep architectures based on performance using three databases with 3-fold cross-validation. Also, the proposed work is compared to the state-of-the-art deep learning-based finger vein recognition system. The projected method achieves good results in terms of accuracy without the need for training from scratch, i.e., achieved an accuracy of 98.14% on HKPU, 99.03% on FVUSM, and 99.50% on SDUMLA databases with a 0.06% increase for HKPU, 0.83% for FVUSM and 0.43% increase for SDUMLA databases compared to the existing works, which establishes its superiority compared to other deep learning systems. However, the proposed model classified some low-quality images incorrectly due to no apparent difference between vein patterns. Hence, for future work, the proposed model architecture can employ more advanced deep learning algorithms for pre-processing to achieve better performance, as the ROI selection with the movement of a finger while image capture is still an issue. In addition, the proposed method can be researched to fuse with various diverse biometrics for authentication applications (face, palmprint, finger anatomy, and others) to achieve robust performance.

Acknowledgement: None.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Manjit Singh implemented the work and wrote the manuscript; Sunil Kumar Singla supervised and reviewed the manuscript. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The three public datasets used in this paper can be accessed by request from the original authors available at the following links: FVUSM: drfendi.com/fv_usm_database (accessed on 05 February 2025); SDUMLA: time.sdu.edu.cn/kycg/gksjk.htm (accessed on 05 February 2025); HKPU: www4.comp.polyu.edu.hk/~csajaykr/fvdatabase.htm (accessed on 05 February 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Hemis M, Kheddar H, Bourouis S, Saleem N. Deep learning techniques for hand vein biometrics: a comprehensive review. Inf Fusion. 2025;114:102716. doi:10.1016/j.inffus.2024.102716.
2. Kashif S, Aihua M, Imran Q, Munish K, Sumaira H, Xingming Z. Recent advancements in finger vein recognition technology: methodology, challenges and opportunities. Inf Fusion. 2022;79:84–109. doi:10.1016/j.inffus.2021.10.004.
3. Hoshang K, Shiva A, Kayode AA, Mohd SR. Finger vein recognition techniques: a comprehensive review. Multimed Tools Appl. 2023;82:33541–75. doi:10.1007/s11042-023-14463-5.
4. Orru G, Rattani A, Rida I, Marcel S. Recent advances in behavioral and hidden biometrics for personal identification. Pattern Recognit Lett. 2024;185:108–9. doi:10.1016/j.patrec.2024.07.016.
5. Borui H, Huijie Z, Ruqiang Y. Finger-Vein biometric recognition: a review. IEEE Trans Instrum Meas. 2022;71:1–26. Art no. 5020426. doi:10.1109/TIM.2022.3200087.
6. Yadav G, Maheshwari S, Agarwal A. Contrast limited adaptive histogram equalization based enhancement for real time video system. In: 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI); 2014; Delhi, India. p. 2392–7. doi:10.1109/ICACCI.2014.6968381.
7. Yang L, Yang G, Yin Y, Xi X. Finger vein recognition with anatomy structure analysis. IEEE Trans Circ Syst Video Technol. 2018;28(8):1892–905. doi:10.1109/TCSVT.2017.2684833.
8. Liu Y, Ling J, Liu Z, Shen J, Gao C. Finger vein secure biometric template generation based on deep learning. Soft Comput. 2018;22:2257–65. doi:10.1007/s00500-017-2487-9.
9. Hengyi Ren H, Sun L, Guo J, Han C, Wu F. Finger vein recognition system with template protection based on convolutional neural network. Knowl-Based Syst. 2021;227:107159. doi:10.1016/j.knosys.2021.107159.
10. Liu F, Yang G, Yang Yin Y, Wang S. Singular value decomposition based minutiae matching method for finger vein recognition. Neurocomputing. 2014;145:75–89. doi:10.1016/j.neucom.2014.05.069.
11. Yang L, Yang G, Xi X, Meng X, Zhang C, Yin Y. Tri-branch vein structure assisted finger vein recognition. IEEE Access. 2017;5:21020–8. doi:10.1109/ACCESS.2017.2728797.
12. Li X, Xi X, Yin Y, Yang G. Finger vein recognition based on personalized discriminative bit map. Appl Math Inform Sci. 2014;8(6):3121–7. doi:10.12785/amis/080653.
13. Lu Y, Yoon S, Wu S, Park DS. Pyramid histogram of double competitive pattern for finger vein recognition. IEEE Access. 2018;6:56445–56. doi:10.1109/ACCESS.2018.2872493.
14. Noh KJ, Choi J, Hong JS, Park KR. Finger-Vein recognition based on densely connected convolutional network using score-level fusion with shape and texture images. IEEE Access. 2020;8:96748–66. doi:10.1109/ACCESS.2020.2996646.
15. Wang D, Li J, Memik G. User identification based on finger-vein patterns for consumer electronics devices. IEEE Trans Consum Electron. 2010;56(2):799–804. doi:10.1109/TCE.2010.5506004.

16. Zhou L, Yang G, Yin Y, Yang L, Wang K. Finger vein recognition based on stable and discriminative superpixels. Int J Pattern Recognit Artif Intell. 2016;30(6):1650015. doi:10.1142/S0218001416500154.
17. Kang W, Lu Y, Li D, Jia W. From noise to feature: exploiting intensity distribution as a novel soft biometric trait for finger vein recognition. IEEE Trans Inf Forensics Secur. 2019 Apr;14(4):858–69. doi:10.1109/TIFS.2018.2866330.
18. Das R, Piciucco E, Maiorana E, Campisi P. Convolutional neural network for finger-vein-based biometric identification. IEEE Trans Inf Forensics Secur. 2019;14(2):360–73. doi:10.1109/TIFS.2018.2850320.
19. Gurunathan V, Sudhakar R, Sathiyapriya T, Soundappan J. Finger vein authentication using vision transformer. In: International Conference on Science Technology Engineering and Management (ICSTEM); 2024; Coimbatore, India. p. 1–5. doi:10.1109/ICSTEM61137.2024.10560933.
20. Zhao P, Song Y, Wang S, Xue JH, Zhao S, Liao Q, et al. VPCFormer: a transformer-based multi-view finger vein recognition model and a new benchmark. Pattern Recognit. 2024 Apr;148:110170. doi:10.1016/j.patcog.2023.110170.
21. Li M, Gong Y, Zheng Z. Finger vein identification based on large kernel convolution and attention mechanism. Sensors. 2024 Feb;24(4):1132. doi:10.3390/s24041132.
22. Hu H, Kang W, Lu Y, Fang Y, Liu H, Zhao J, et al. FV-Net: learning a finger-vein feature representation based on a CNN. In: 24th International Conference on Pattern Recognition (ICPR), Beijing, China; 2018. p. 3489–94. doi:10.1109/ICPR.2018.8546007.
23. Xie C, Kumar A. Finger vein identification using convolutional neural network and supervised discrete hashing. Pattern Recognit Lett. 2019;119:148–56. doi:10.1016/j.patrec.2017.12.001.
24. Kang W, Liu H, Luo W, Deng F. Study of a full-view 3D finger vein verification technique. IEEE Trans Inf Forensics Secur. 2019;15:1175–89. doi:10.1109/TIFS.2019.2928507.
25. Wang X, Han X, Huang W, Dong D, Scott MR. Multi-similarity loss with general pair weighting for deep metric learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2019; Long Beach, CA, USA. p. 5022–30.
26. Ibrahim MM, Sujith C, Florence SM, Rajagopal S. Enhancing ATM security: a finger vein biometrics approach. In: Proceedings of the 2nd International Conference on Networking and Communications (ICNWC); 2024; Chennai, India. p. 1–4. doi:10.1109/ICNWC60771.2024.10537537.
27. Jalilian E, Uhl A. Enhanced segmentation-CNN based finger-vein recognition by joint training with automatically generated and manual labels. In: IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA); 2019; Hyderabad, India. p. 1–8. doi:10.1109/ISBA.2019.8778522.
28. Jalilian E, Uhl A. Finger-vein recognition using deep fully convolutional neural semantic segmentation networks: the impact of training data. In: IEEE International Workshop on Information Forensics and Security (WIFS); 2018; Hong Kong, China. p. 1–8. doi:10.1109/WIFS.2018.8630794.
29. Fang Y, Wu Q, Kang W. A novel finger vein verification system based on two stream convolutional network learning. Neurocomputing. 2018;290:100–7. doi:10.1016/j.neucom.2018.02.042.
30. Ou WF, Po LM, Zhou C, Rehman YAU, Xian PF, Zhang YJ. Fusion loss and inter-class data augmentation for deep finger vein feature learning. Expert Syst Appl. 2021;171:114584. doi:10.1016/j.eswa.2021.114584.
31. Banerjee A, Basu S, Nasipuri M. ARTeM: a new system for human authentication using finger vein images. Multimed Tools Appl. 2018 Mar;77(5):5857–84. doi:10.1007/s11042-017-4501-8.
32. Kuzu RS, Piciucco E, Maiorana E, Campisi P. On-the-fly finger-vein-based biometric recognition using deep neural networks. IEEE Trans Inf Forensics Secur. 2020;15:2641–54. doi:10.1109/TIFS.2020.2971144.
33. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6:60. doi:10.1186/s40537-019-0197-0.
34. Yang W, Hui C, Chen Z, Xue JH, Liao Q. FV-GAN: finger vein representation using generative adversarial networks. IEEE Trans Inf Forensics Secur. 2019;14(9):2512–24. doi:10.1109/TIFS.2019.2902819.
35. Zhao P, Zhao S, Xue JH, Yang W, Liao Q. The neglected background cues can facilitate finger vein recognition. Pattern Recognit. 2023;136:109199. doi:10.1016/j.patcog.2022.109199.
36. Fang C, Ma H, Li J. A finger vein authentication method based on the light weight Siamese network with the self-attention mechanism. Infrared Phys Technol. 2023;128:104483.

37. Boucherit I, Zmirli MO, Hentabli H, Rosdi BA. Finger vein identification using deeply-fused convolutional neural network. J King Saud Univ-Comput Inf Sci. 2022;34:646–56. doi:10.1016/j.jksuci.2020.04.002.
38. Shaheed K, Mao A, Qureshi I, Kumar M, Hussain S, Ullah I, et al. DS-CNN: a pre-trained Xception model based on depth-wise separable convolutional neural network for finger vein recognition. Expert Syst Appl. 2022;191:116288. doi:10.1016/j.eswa.2021.116288.
39. Mustafa K, Adem A, Nurettin A. Automated vein verification using self-attention-based convolutional neural networks. Expert Syst Appl. 2023;230:20550. doi:10.1016/j.eswa.2023.120550.
40. Abdullahi SA, Bature ZA, Chophuk P, Muhammad A. Sequence-wise multimodal biometric fingerprint and finger-vein recognition network (STMFPFV-Net). Intell Syst Appl. 2023;19:200256. doi:10.1016/j.iswa.2023.200256.
41. Aurangzeb K, Javeed K, Alhussein M, Rida I, Haider SI, Parashar A. Deep learning approach for hand gesture recognition: applications in deaf communication and healthcare. Comput Mater Contin. 2024;78:127–44. doi:10.32604/cmc.2023.042886.
42. Kuzu RS, Maiorana E, Campisi P. Vein-based biometric verification using transfer learning. In: 2020 43rd International Conference on Telecommunications and Signal Processing (TSP); Milan, Italy; 2020. p. 403–9. doi:10.1109/TSP49548.2020.9163491.
43. Hosna A, Merry E, Gyalmo J, Alom Z, Aung Z, Azim MA. Transfer learning: a friendly introduction. J Big Data. 2022;9:102. doi:10.1186/s40537-022-00652-w.
44. Iman M, Arabnia HR, Rasheed K. A review of deep transfer learning and recent advancements. Technologies. 2023;11(2):40. doi:10.3390/technologies11020040.
45. Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. In: Proceedings of the International Conference on Machine Learning (ICML); 2019; Long Beach, CA, USA. p. 6105–14.
46. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017 May;60(6):84–90. doi:10.1145/3065386.
47. Sandler M, Howard A, Zhu M, Zhmoginov M, Chen LC. MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun; Salt Lake City, UT, USA. p. 4510–20. doi:10.1109/CVPR.2018.00474.
48. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun; Las Vegas, NV, USA. p. 770–8. doi:10.1109/CVPR.2016.90.
49. Alhichri H, Alswayed A, Bazi Y, Ammour N, Alajlan NA. Classification of remote sensing images using EfficientNet-B3 CNN model with attention. IEEE Access. 2021;9:14078–94. doi:10.1109/ACCESS.2021.3051085.
50. Zhuang B, Liu J, Pan Z, He H, Weng Y, Shen C. A survey on efficient training of transformers. arXiv:2302.01107. 2023.
51. Tay Y, Dehghani M, Bahri D, Metzler D. Efficient transformers: a survey. ACM Comput Surv. 2023 Jun;55(6):28. doi:10.1145/3530811.
52. Zhang H, Goodfellow I, Metaxas D, Odena A. Self-attention generative adversarial networks. In: International Conference on Machine Learning; 2019; Long Beach, CA, USA: PMLR. p. 7354–63.
53. Zhao H, Jia J, Koltun V. Exploring self-attention for image recognition. In: Proceedings of the EEE/CVF Conference on Computer Vision and Pattern Recognition; 2020; Seattle, WA, USA. p. 10076–85.
54. Lee EC, Lee HC, Park KR. Finger vein recognition using minutia-based alignment and local binary pattern-based feature extraction. Int J Imaging Syst Technol. 2009;19(3):179–86. doi:10.1002/ima.20193.
55. Huang B, Dai Y, Li R, Tang D, Li W. Finger-vein authentication based on wide line detector and pattern normalization. In: 20th International Conference on Pattern Recognition; 2010; Istanbul, Turkey. p. 1269–72. doi:10.1109/ICPR.2010.316.
56. Brownlee J. A gentle introduction to k-fold cross-validation. 2023 [cited 2024 Mar 02]. Available from: https://machinelearningmastery.com/k-fold-cross-validation/.
57. Yin Y, Liu L, Sun X. Sdumla-hmt: a multimodal biometric database. Biom Recognit. 2011;7098:260–8. doi:10.1007/978-3-642-25449-9.

58. Kumar A, Zhou Y. Human identification using finger images. IEEE Trans Image Process. 2012 Apr;21(4):2228–44. doi:10.1109/TIP.2011.2171697.
59. Asaari MSM, Suandi SA, Rosdi BA. Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics. Expert Syst Appl. 2014 Jun 1;41(7):3367–82. doi:10.1016/j.eswa.2013.11.033.
60. Hou B, Yan R. Convolutional autoencoder model for finger-vein verification. IEEE Trans Instrum Meas. 2020;69(5):2067–74. doi:10.1109/TIM.2019.2921135.
61. Xu W, Shen L, Wang H, Yao Y. A finger vein feature extraction network fusing global/local features and its lightweight network. Evol Syst. 2023;14:873–89. doi:10.1007/s12530-022-09475-9.