

## Article

# FVR-Net: Finger Vein Recognition with Convolutional Neural Network Using Hybrid Pooling

Lakpa Dorje Tamang  and Byung Wook Kim \* 

Department of Information and Communication Engineering, Changwon National University,  
Changwon 51140, Korea; ld.tamang25@gmail.com

\* Correspondence: bwkim@changwon.ac.kr

**Abstract:** In this paper, we present FVR-Net, which is a novel finger vein recognition network using a convolutional neural network (CNN) with a hybrid pooling mechanism. The scheme is based on the use of a block-wise feature extraction network to extract discrete features from interclass vein image samples, regardless of their visual quality. Input images to FVR-Net are subjected to preprocessing prior to being fed into the network in order to segment the vein patterns from the background. We designed a feature extraction network in which each block consists of a convolutional layer followed by hybrid pooling, whose output activation maps are concatenated before passing features to another block within the network. In the hybrid pooling layer, two subsampling layers of maxpooling and average pooling are placed in parallel where the former activates the most discrete features of the input, and the latter considers the complete extent of the input volume so better localization of features can be accessed. After the features are extracted, they are passed to three fully connected layers (FCLs) for classification. We conduct several experiments on two publicly available finger vein datasets based on visual quality of the images. When compared to conventional studies, the proposed model achieves outstanding recognition performance with accuracies of up to 97.84% and 97.22% for good and poor-quality images, respectively. By varying multiple network hyperparameters, we obtain optimal settings such that the model can guarantee the best recognition accuracy for a finger vein biometric system.

**Keywords:** finger vein; biometrics; Gabor filters; deep neural networks



**Citation:** Tamang, L.D.; Kim, B.W. FVR-Net: Finger Vein Recognition with Convolutional Neural Network Using Hybrid Pooling. *Appl. Sci.* **2022**, *12*, 7538. <https://doi.org/10.3390/app12157538>

Academic Editor: Robert J. Meier

Received: 13 June 2022

Accepted: 25 July 2022

Published: 27 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The proliferation in risks of attack on information have given rise to many security applications for personal identification technologies. In this regard, the ability to intelligently identify individual attributes through user-friendly and secure recognition systems has been a global subject of interest lately. Unlike conventional identification methods, such as secret keys and passwords, biometric techniques [1–3] utilize physiological features such as fingerprints, faces, iris, voices, digital signatures, etc., and they are less susceptible to theft or duplication. Because of this, biometrics has recently been indulged in as a substantially reliable authentication method for recognizing individuals [3].

Techniques based on biometric data can involve automated recognition of a person based on their extrinsic or intrinsic biological data. Recognition with extrinsic characteristics usually utilizes exterior features of the human body, which include fingerprints, the face, the iris, and palm prints. These characteristics bring on security vulnerabilities because they can easily be imitated and can be a threat to the identification system. Different from these modalities, intrinsic characteristics consider concealed features of the skin, such as vein patterns, for recognition purposes. These modalities take several factors into account, such as universality, distinctiveness, and permanence. Therefore, they are not prone to forgery, are minimally influenced by skin conditions, and provide colossal privacy for high-security application scenarios, such as banks, forensics, and legal support [4].

Generally, vein images are acquired by utilizing near infrared radiation (NIR) or thermal infrared-based optical imaging systems [5]. Such imaging techniques do not require close proximity to sensors, which makes them suitable for finger vein imaging systems. In addition, they can facilitate imaging of deeper tissues at the microvascular level. However, during real-time acquisition, the captured image might be inherently susceptible to quality degradation, including image shading due to uneven thickness of muscles inside the finger, irregular shadowing, positional misalignment, scattering of light when imaging finger tissue, and physiological changes such as inaccurate placement of the finger. During real-time recognition, these factors pose nontrivial threats due to irrelevant and ambiguous separability between the vein patterns and the background. Therefore, many consider finger vein recognition with high accuracy a challenging task. Despite these challenges, in recent years, obtaining accurate vein recognition techniques is in demand.

## 2. Related Works

In general, a complete finger vein-based biometric recognition system can be divided into four steps: image acquisition, preprocessing, feature extraction, and feature matching. First, vein images are captured with NIR optical imaging techniques, after which quality enhancement is carried out via several preprocessing steps, such as image filtering [6], region of interest (ROI) extraction [7], and image normalization [8]. The enhanced images are then subjected to feature extraction, in which discriminative features are extracted from the individual vein images to enable good recognition performance during feature matching. To this end, several studies have been conducted to develop finger vein recognition systems based on conventional mathematical models [9–18], machine learning (ML) [19–25], and deep learning (DL) [26–36].

### 2.1. Conventional Finger Vein Recognition

Different types of conventional mathematical models have been utilized for finger vein recognition tasks. The Gabor filter [9] is a popular feature extraction method employed for finger vein recognition. Due to its ability to detect oriented features by tuning in to a specific frequency, the Gabor filter has been used widely for various pattern recognition tasks [10]. Similarly, line tracking [11] methods extract features based on repeated tracking of black pixels in the vein image. The maximum curvature method [12] extracts features based on the fact that the vein patterns appear like a valley, with a high curvature within the cross-sectional area of the vein image. Furthermore, local feature descriptor-based methods, such as local binary pattern (LBP) and local derivative pattern (LDP), are also widely used methods of feature extraction [13]. They extract features based on binary codes obtained by comparing the gray value of the center pixel with the neighboring pixels. However, due to low robustness, high computational complexity, and poor feature extraction capabilities, these techniques are at a great disadvantage when implementing practical finger vein recognition systems. Furthermore, these conventional approaches utilize distance-based, trivial feature-matching algorithms, such as Hamming distance [14], Euclidean distance [15], cross correlation matching [16], template matching [17], and histogram intersection matching [18] to obtain a matching score for verification purposes. These types of feature matching techniques are not suitable for providing reliable recognition performance when the capture devices and image acquisition environment vary according to various applications or the authentication service.

### 2.2. Deep Learning Based Finger Vein Recognition

With the breakthrough of big data and DL in several image processing tasks, such as classification [26], object detection [27], and digital image processing [28], the application of DL has also been considered for fast and automatic feature extraction from finger vein images [29–36]. Due to the robustness of feature representations, a DL technique can be a novel candidate for finger vein recognition, irrespective of the shape and orientation of the vein patterns. Using large amounts of training samples, a DL algorithm such as the

convolutional neural network (CNN) can provide powerful feature extraction capabilities with vein images and can quickly adapt to learning those feature representations. In [29], a densely connected convolutional neural network (DenseNet) was used for finger vein recognition where the matching score was generated by the score level fusion of shape and texture images. A similar study was investigated in [30], where a three-channel composite image was used as input to the DenseNet-based recognition network and provided good recognition performance. To achieve robust recognition performance from finger vein biometrics, the common CNN architecture has been extensively used in recent research [31–33]. Furthermore, to learn more robust features directly from the raw pixels of the finger vein image, the convolutional autoencoder (CAE) [34], recurrent neural networks (RNNs) [35], and generative adversarial networks (GANs) [36] have been researched as well.

Although previous DL-based finger vein recognition techniques have offered promising recognition performance, they still suffer from relevant shortcomings, especially associated with mediocre feature extraction procedures. The finger vein images may include not only vein patterns but also irregular noise and intensity degradation due to the various finger shapes and muscles. Due to infirmly designed feature extraction networks in existing methods, the models cannot properly learn distinctive features of finger vein patterns, and thus, a unique representation of the vein patterns of each subject cannot be accessed, and this ultimately results in poor accuracy. Besides, currently available finger vein image databases acquired via infrared imaging offer significantly poor visual quality in the vein patterns. Thus, for higher recognition performance, poor image quality must be complemented by designing an effective preprocessing pipeline and a robust feature extractor.

To overcome such limitations, in this paper, we propose FVR-Net, which is a novel CNN-based hybrid pooling network for finger vein recognition system that supports exquisite feature extraction to enable accurate recognition of finger vein images, regardless of the acquisition procedures of existing datasets. FVR-Net was designed to capture intricate vein features from the input images through several convolutional and subsampling layers. Unlike previous studies, FVR-Net introduces block-wise feature extraction with hybrid pooling where two subsampling layers (maxpooling and average pooling) in each block are placed in parallel after the convolutional layer, whose output activation maps are concatenated before passing features to another block. In hybrid pooling, the subsampling layers placed in parallel concurrently activate the most discriminative features of each input and provide excellent localization of those features. Upon forward pass, these blocks tend to extract multilevel and discrete interclass features of the input, which means they are able to uniquely represent the vein samples from different categories when subjected to classification with three fully connected layers (FCLs). In the whole process, the finger vein images are first subjected to preprocessing where ROI cropping and quality enhancement is achieved, along with proper segmentation of vein patterns from the background. Subsequently, clear vein images are fed into the feature extraction network, which consists of five blocks for low-level feature extraction followed by three FCLs for the classification task. Each block comprises a 2D convolutional layer and two subsampling operations: 2D average pooling and 2D maxpooling. The features learned from the blocks facilitate robust vein classification through the FCLs, where the probability score of each input is generated through a SoftMax classifier. To verify the effectiveness of the proposed model, it was trained and tested on publicly available finger vein databases. In particular, the image samples from Hong Kong Polytechnic University (HKPU) [3] and Finger Vein Universiti Sains Malaysia (FVUSM) [37] datasets are characterized according to their visual quality (good and poor), with poor-quality images being likely to yield lower recognition performance. During inference, the test images of the corresponding datasets provide discriminating recognition accuracy, compared to the conventional schemes. Overall, FVR-Net enhanced the recognition accuracy with HKPU and FVUSM datasets by up to

13% and 30%, respectively, for good quality images, and 9% and 31%, respectively, for poor quality images.

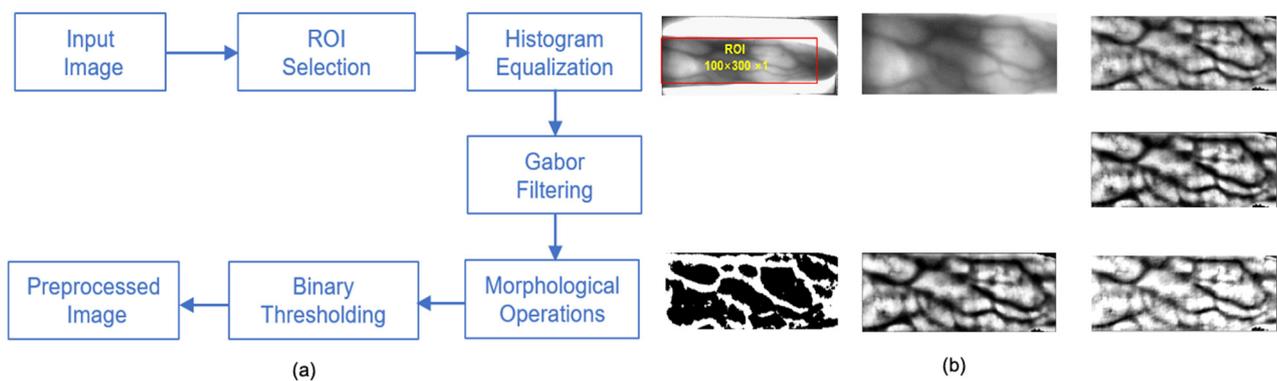
The rest of this paper is organized as follows. Section 2 presents the proposed model, describing data preprocessing and the network architecture. In Section 3, the experiments and results are presented. Finally, concluding remarks are provided in Section 4.

### 3. The Proposed Model

#### 3.1. Data Preprocessing

Good performance from finger vein recognition systems is usually attributed to the quality of the vein images [25]. Therefore, raw vein images acquired from infrared image sensors should be subjected to preprocessing prior to feature extraction. The preprocessing stage is intended to discard any nonideal entities appearing in the vein images, such as noise, shadows, and low contrast due to a rotational or translational variant property of the finger or a faulty acquisition device.

Figure 1a,b, respectively, depict the architectural pipeline of the data preprocessing strategy in our method and the preprocessed images obtained at each corresponding step. Because the input images from the publicly available datasets are presented at various sizes and orientations, the preliminary step is to accomplish ROI selection of the intended part of the raw image. This guarantees dimensional uniformity in all images entering the CNN for feature extraction. At this particular stage, all images are subjected to normalization, where they are subsampled to the desired dimension by cropping them to  $100 \times 300$  pixels. Next, the cropped images undergo image enhancement with contrast limited adaptive histogram equalization (CLAHE) [38], where the intensities of the images are adjusted in such a way that they are better distributed on the histogram. By doing so, the global contrast of the image is improved as the most frequent intensity values are effectively spread, while the lower contrast areas gain higher values.



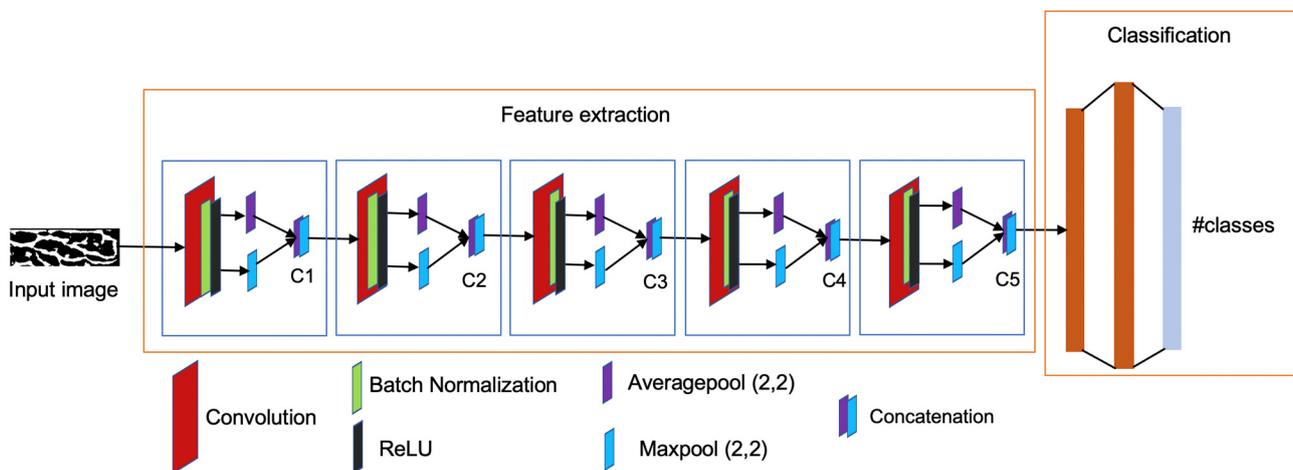
**Figure 1.** Data preprocessing steps. (a) Preprocessing block diagram, and (b) preprocessed images at each step in the order of preprocessing steps.

After that, a Gabor filtering technique [5] is implemented to enhance the concrete vein features. A Gabor filter is a special form of bandpass filter suited for segmentation of texture that possesses optimal localization properties in both the spatial and frequency domain of an image. The filter reduces the impact of noise on the images to enable good feature extraction and ultimately enhances recognition accuracy. Furthermore, to enhance the clarity of patterns, vein images are subjected to morphological operations. These operations tend to discard the irrelevancies while preserving the vital shape characteristics of the image [39]. Within the morphological operations, a grayscale erosion technique is employed, followed by grayscale dilation. The erosion and dilation transformation bear a marked similarity when acquiring perspective imagery of the vein image. When applied iteratively as a pair, they enable smoothing of contour, and they eliminate small gaps in the vein patterns without any geometric distortion of the global features. In the last stage of data preprocessing, to coarsely localize the vein patterns in the images, a

binarization technique with a threshold of 110 is applied. The pixel values greater than 110 represent the vein pattern in white, while the rest represent background pixels in black. After completion of the preprocessing stage, images are immediately passed to the feature extraction procedure, followed by classification of the extracted features.

### 3.2. Network Architecture

As shown in Figure 2, feature extraction comprises a total of five blocks, where each block entails one convolutional layer, followed by batch normalization and a rectified linear unit (ReLU) activation function, which increases the rate of convergence during training. To reduce the computational cost, subsampling layers that decrease the spatial dimension of the feature maps are introduced after each convolutional layer. Specifically, we devise a hybrid pooling layer where two types of pooling layers are utilized: 2D maxpooling and 2D average pooling. Maxpooling outputs the maximum value from the subregion of its input volume by convolving a kernel across its spatial dimension. This allows extracting a high activation value within the feature map and generates sharp features. On the other hand, the average pooling layer considers the average value of the subregion during each pooling operation. It is more generalized computation, which encourages the network to identify the complete extent of its input volume [40]. These layers, however, are not placed sequentially in the network; instead, they are placed in parallel such that the same feature map is distributed to both pooling layers. The intuition behind employing both pooling layers is that finger vein features are translational variant, and therefore, preservation of feature localization is as important as preserving the most discriminative features. The results of both pooling layers are concatenated to ensure that the global textural information of the feature maps generated by each convolutional layer is effectively feed-forwarded to the imminent layers in the network.



**Figure 2.** The proposed FVR-Net architecture.

A detailed depiction of the network structure for the feature extraction process is presented in Table 1. The preprocessed image with a dimension of  $100 \times 300 \times 1$  is fed into the network as input. The first three convolutional layers use a  $(5, 5)$  kernel size, while the remaining two layers use  $(3, 3)$ . The larger kernel sizes are used to significantly increase the receptive field sizes, which improves recognition of large instances in the images. This is an important part of finger vein recognition because the width of the vein pattern is not necessarily the same throughout the image. By doing so, the network not only captures the local patterns but also the global ones that span a large area in the image. The number of channels was chosen with alternate increases and decreases between two consecutive convolutional layers, maintaining the computational complexity that might occur when large filters are used across the network. After hybrid pooling, the concatenation layers (C1–C5) are responsible for concatenating the output features of both subsampling layers

before passing them to the next block. Furthermore, to prevent overfitting while training with big data, an L2 regularizer with a regularization penalty of 0.0001 is used at each convolutional layer. Because each layer processes computationally inexpensive parameters and the L2 regularizer can provide sufficient regularization [41], dropouts are not utilized.

**Table 1.** Detailed Network Hyperparameters of the Feature Extraction Procedure. The abbreviational terminologies in the table are: Conv: Convolution, MaxP: Maxpool, AvgP: Averagepool, C: Concatenation.

Layer	Input	Activation	Regularization	Filter	Kernel Size	Output
Conv-1	Preprocessed image (100 × 300 × 1)	ReLU	L2 (0.0001)	128	(5, 5)	(100, 300, 128)
MaxP-1	Conv-1	-	-	-	(2, 2)	(50, 150, 128)
AvgP-1	Conv-1	-	-	-	(2, 2)	(50, 150, 128)
C1	Concatenate (MaxP-1, AvgP-1)	-	-	-	-	(50, 150, 256)
Conv-2	C-1	ReLU	L2 (0.0001)	64	(5, 5)	(50, 150, 64)
MaxP-2	Conv-2	-	-	-	(2, 2)	(25, 75, 64)
AvgP-2	Conv-2	-	-	-	(2, 2)	(25, 75, 64)
C2	Concatenate (MaxP-2, AvgP-2)	-	-	-	-	(25, 75, 128)
Conv-3	C-2	ReLU	L2 (0.0001)	128	(5, 5)	(25, 75, 128)
MaxP-3	Conv-3	-	-	-	(2, 2)	(12, 37, 128)
AvgP-3	Conv-3	-	-	-	(2, 2)	(12, 37, 128)
C3	Concatenate (MaxP-3, AvgP-3)	-	-	-	-	(12, 37, 256)
Conv-4	C-3	ReLU	L2 (0.0001)	64	(3, 3)	(12, 37, 64)
MaxP-4	Conv-4	-	-	-	(2, 2)	(6, 18, 64)
AvgP-4	Conv-4	-	-	-	(2, 2)	(6, 18, 64)
C4	Concatenate (MaxP-4, AvgP-4)	-	-	-	-	(6, 18, 128)
Conv-5	C-4	ReLU	L2 (0.0001)	128	(3, 3)	(6, 18, 128)
MaxP-5	Conv-5	-	-	-	(2, 2)	(3, 9, 128)
AvgP-5	Conv-5	-	-	-	(2, 2)	(3, 9, 128)
C5	Concatenate (MaxP-5, AvgP-5)	-	-	-	-	(3, 9, 256)

### 3.3. Classification

To learn a nonlinear combination of the features obtained by the feature extraction process, a set of three FCLs are used in the classification block. The output of final concatenation layer C5 of the feature extraction network is flattened to a one-dimension (1D) vector before passing it as input to this block. A detailed architecture of this network is presented in Table 2. Because all input units of an FCL have separate weights associated with each output unit, the total number of learnable parameters is relatively high compared to CNNs. Therefore, to prevent overfitting due to the large number of parameters and to improve generalization of the network, dropouts are exploited after each layer except the final one. The first two layers of the classification network use ReLU, while the last layer uses SoftMax as an activation function. The final layer produces a 1D vector output, which consists of the probabilities that the input data belong to one of the  $S$  classes ( $S$  being the total number of subjects providing sample vein images in the dataset). Among the

obtained probabilities, the highest value of the  $S$  classes are selected for the final finger vein identification.

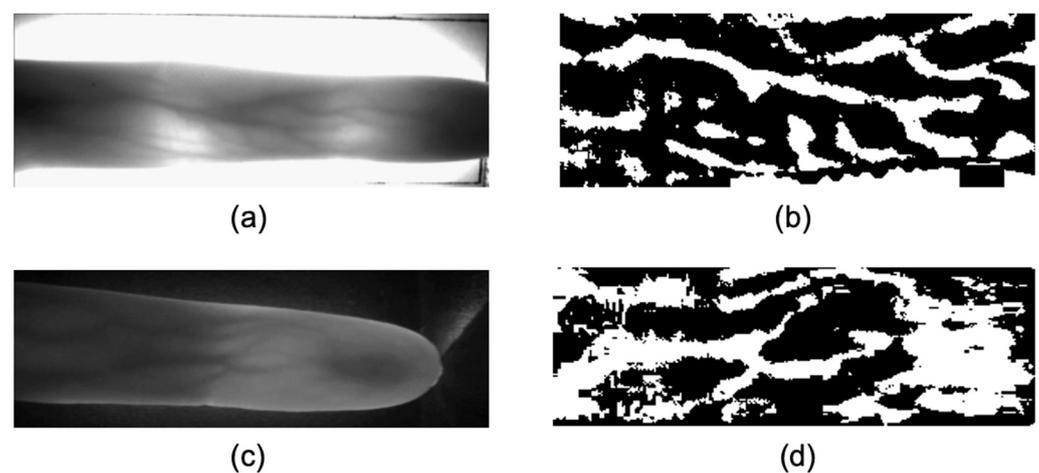
**Table 2.** Detailed Network Hyperparameters of the Classification Procedure. The abbreviation terminologies in the table are: FCL: Fully connected layers, Drop: Dropout.

Layer	Input	Activation	Regularization	Filter	Filter Size	Output
Flatten	C-5	-	-	-	-	(1, 6912)
FCL-1	Flatten	ReLU	-	256	-	(1, 256)
Drop-1	FCL-1	-	Dropout (0.4)	-	-	-
FCL-2	Drop-1	ReLU	-	512	-	(1, 512)
Drop-2	FCL-2	-	Dropout (0.4)	-	-	-
FCL-3	Drop-2	SoftMax	-	$S$	-	(1, $S$ )

## 4. Experiments and Results

### 4.1. Datasets

Our model was evaluated on two publicly available finger vein databases (HKPU [3], and FVUSM [37]); examples are shown in Figure 3. Within all available subjects in the database, every sample is labeled based on visual perception. Specifically, the vein images are consigned into one of two groups: good quality and poor quality. Although manual inspection is time-consuming and tiring for large-scale datasets, the assessed image quality can be genuinely represented. Figure 4 presents example finger vein data used in this research corresponding to visual quality (good and poor). Table 3 presents the database details adopted in our research. All images were resized to a network-specific input size of  $100 \times 300 \times 1$  during the preprocessing stage.



**Figure 3.** Sample images of the finger vein databases used in our research. (a) Original image of HKPU, (b) preprocessed image of HKPU, (c) original image of FVUSM, and (d) preprocessed image of FVUSM.



**Figure 4.** (a) Good quality HKPU. (b) Poor quality HKPU. (c) Good quality FVUSM and (d) Poor quality FVUSM.

**Table 3.** Details about the Publicly Available Finger Vein Image Databases used in this Research.

Database	Sessions/ Session(s) Used	Subjects Per Session	Total Fingers	Number of Images Per Subject	Training/ Validation Split	Subjects According to Image Quality		Original Image Size	Normalized Image Size
						Good	Poor		
HKPU [3]	2/2	156	2	6	7:3	66	245	$513 \times 256$	$100 \times 300$
FVUSM [37]	2/1	494	1	6	7:3	146	348	$640 \times 480$	$100 \times 300$

#### 4.1.1. HKPU Dataset

The HKPU vein database [3] was acquired by using a contactless imaging device on Hong Kong Polytechnic University campus from April 2009 to March 2010. The database consists of 3132 images at  $513 \times 256$  pixels from 156 subjects, where the first 105 subjects were captured in two different sessions. The minimum, maximum, and average intervals between sessions were one month, more than six months, and 66.8 days, respectively. Each of the 105 subjects provided six samples from each of their index and middle fingers. Images from the remaining 51 subjects were captured in a single session.

In our research, all 156 subjects were used: a single session from the first 105 subjects and all sessions from the remaining 51 subjects, which gave us a total of 1872 images ( $156 \text{ subjects} \times 2 \text{ fingers} \times 6 \text{ images per finger}$ ). Each finger was treated as a different subject, and therefore, we obtained a total of 312 subjects, out of which 66 subjects were designated as good quality.

#### 4.1.2. FVUSM Dataset

The Finger Vein Universiti Sains Malaysia [41] database was collected from 123 subjects (83 males and 40 females) with ages ranging from 20 to 52 years old. The images were acquired in two sessions with an interval of more than two weeks. In a single session, each subject provided four samples (left index, left middle, right index, and right middle fingers) which were captured six times each. Therefore, a total of 5904 images ( $2 \text{ sessions} \times 123 \text{ subjects} \times 4 \text{ fingers} \times 6 \text{ images per finger}$ ) at  $640 \times 480$  pixels were collected. In our research, we utilized only one session containing 2952 images in total. In the FVUSM database samples, 146 of 492 subjects were categorized as good quality; the rest, poor.

#### 4.2. Data Augmentation

The total number of samples per subject available in each database is not sufficient to train a model, and therefore, an intraclass data augmentation technique was applied to the training data such that the model became robust to image diversity. The purpose of using

this technique is to extend and enrich the dataset such that deformations in different vein patterns in a real-world scenario can be reflected. In our research, width and height shifts were performed with 0.01 pixels shifting in a left-to-right and up-down manner. To make the model perceive the training image from a different angle, a shear value of 0.01 was also utilized. Since the finger vein pattern and unique characteristics of vein images were already determined by preprocessing, techniques such as flipping (horizontal and vertical), cropping, and zooming were not utilized in the experiment. Table 4 presents details of the data augmentation performed on the different databases. Only training images from each database were subjected to augmentation, where 24 images were generated per class along with the original image. Note that the augmentation was applied to all subjects, regardless of image quality.

**Table 4.** Augmentation Details about the Publicly Available Finger Vein Image Databases used in this Research.

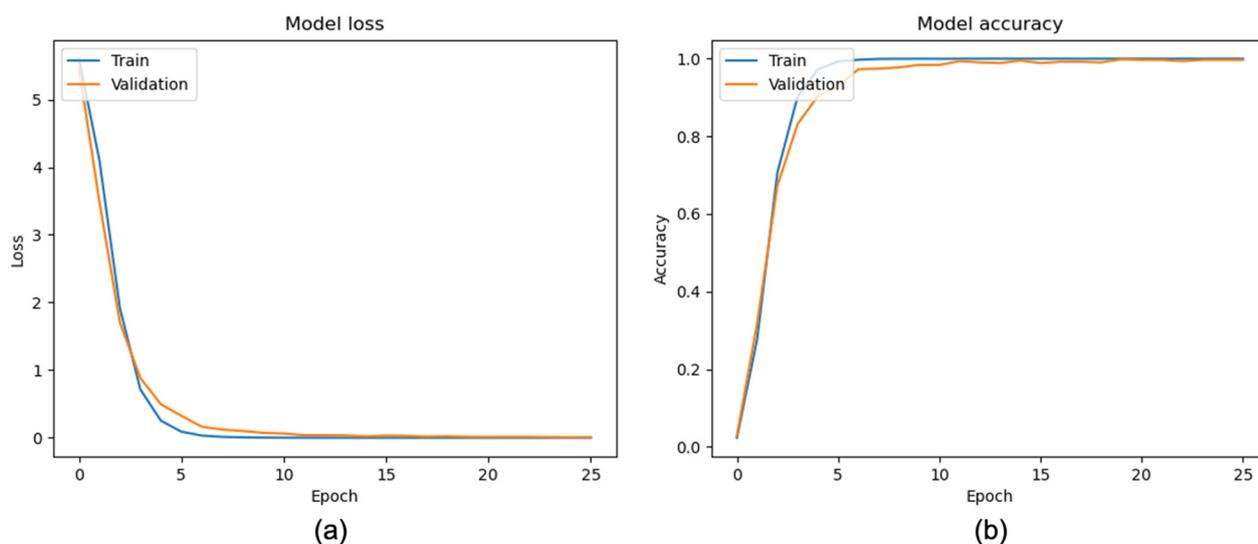
Databases	Augmentation Technique	Total Images Generated Per Class	Total Images after Augmentation
HKPU [3]	Width shift = 0.01 pixels, Height shift = 0.01 pixels, Shear range = 0.01	24	7488
FVUSM [37]		24	11,856

#### 4.3. Training, Validation, and Test Set Generation

In both datasets, each class has a total of six samples per class. To make the model more generalized by training with large variations of data samples, four samples out of six were allocated for training while the remaining two samples were separated for testing purpose. In other words, the total number of image samples per class was subdivided into training and testing sets at a ratio of 4:2 images per class. After that, data augmentation was carried out to construct training samples as mentioned in Section III-B. Furthermore, to provide an unbiased evaluation of model fit on the training data while tuning hyperparameters, all the augmented samples were subjected to training and validation at a 7:3 ratio.

#### 4.4. Training Stage

The finger vein images generated after preprocessing were passed through the network to extract their corresponding features from each block of the network. The whole model was implemented with the Python programming language in a TensorFlow framework on a PC equipped with the Nvidia GeForce RTX 2080 Ti GPU. A cross entropy loss function was used for convergence of the model during training with the objective being to minimize global loss. To achieve faster convergence, the loss function was minimized using the RMS prop optimizer with a learning rate of 0.0001. A batch size of 32 was used, and a total of 100 epochs was used during training. It is necessary to stop training the model at the point where network hyperparameters are optimized and where saturation in training and validation losses is achieved. For this, we employed an early stopping technique where the training was halted when no further improvements to the loss of validation data were observed with respect to the training epochs. Figure 5 demonstrates the generalized learning curves of the proposed model on both training and validation datasets. We can clearly see that although the training epoch was set to 100, saturation in the validation loss was seen at epoch 25. At this point, training stopped, and the model became immune to overfitting from extensive training. Furthermore, we can also observe that the model provided very good convergence in the loss function for classification of the vein images.



**Figure 5.** Learning curves of the proposed model on training and validation datasets. (a) Loss vs. Epochs, and (b) Accuracy vs. Epochs.

#### 4.5. Identification Performance

After successfully training the network, a performance evaluation of the proposed finger vein recognition network was carried out on the test set. As mentioned in Section III-C, two images from each class were allocated for testing purposes to demonstrate the recognition performance of the proposed model. During the evaluation, preprocessed images of test samples from each dataset were used as input to FVR-Net to obtain the corresponding probability scores. To evaluate the performance of the proposed model, we compared the recognition accuracy of FVR-Net with simple P-SVM [42], a neural network (NN) [43], and a state-of-the-art CNN technique [32]. P-SVM is a simple ML-based classifier that is stacked on top of hand-crafted features to generate the probability scores of the corresponding class. Similarly, the NN scheme [42] consists of a single hidden layer and was trained with stochastic gradient descent (SGD). In contrast, the deep neural network (DNN)-based scheme [43] is a multilayer CNN with nine layers (six convolutional layers and three FCLs), which also utilizes the P-SVM classifier at the end to generate the probability of each input belonging to the respective class. The implementation of these studies was not released online, and therefore, the results obtained by these conventional schemes were kept the same as when presented in [32].

Table 5 presents the recognition accuracy obtained by FVR-Net compared with various conventional schemes. We can see that, for good-quality images from the HKPU dataset, the proposed method obtained 12.61%, 11.74%, and 8.59% higher accuracy than the P-SVM, NN, and DNN schemes, respectively. With the FVUSM dataset, recognition accuracy with the proposed method was about 31.36%, 35.49%, and 28.51% higher than the aforementioned schemes. The exquisite recognition performance exhibited by the proposed model can be explained by its robust feature extraction network. The vein patterns of the good-quality images were concisely distributed over the entire spatial location of the image, and thus, subjecting them to the robustly designed feature extraction network activates the most abstract representation of the input at each layer. With distinctive features extracted for each subject, the classification layer is able to accurately classify the image to recognize an individual.

**Table 5.** Comparison of Recognition Accuracy from FVR-Net with Several Baseline Finger Vein Recognition Algorithms.

Methods	Database	Recognition Accuracy (%)	
		Good Quality	Poor Quality
P-SVM [42]	HKPU	80.57	79.04
	FVUSM	66.48	65.71
NN [43]	HKPU	81.44	81.82
	FVUSM	62.35	55.00
DNN [32]	HKPU	84.59	83.64
	FVUSM	69.33	68.57
FVR-Net	HKPU	<b>93.18</b>	<b>88.97</b>
	FVUSM	<b>97.84</b>	<b>97.22</b>

Our proposed model exhibited significantly higher accuracy for low-quality images as well. With the HKPU dataset, we witnessed 9.93%, 7.15%, and 5.33% higher recognition accuracy, whereas with FVUSM, we saw 31.15%, 42.22%, and 28.95% higher accuracy than the baseline P-SVM, NN, and DNN models, respectively. Note that the image acquisition period for HKPU is longer than that of FVUSM, and more intraclass variation exists, leading to more mismatching errors [32]. In addition, because of the thick and accumulated vein patterns in FVUSM, the model can produce more robust activation maps than HKPU, where the vein structures are thinner and more scattered. Since the proposed method makes the best effort to extract the finger vein feature information as much as possible, it shows higher accuracy for the FVUSM dataset, where the within-class variation is slightly lower than that of HKPU. In addition to that, since the vein patterns associated with poor-quality images are extremely vague in nature, the infirmly designed feature extraction network of the existing models cannot extract distinctive features from them, making it exceedingly difficult to obtain higher recognition accuracy. In contrast, the feature extraction network of the proposed model can capture large and ambiguous instances of the input by using substantially larger kernels present in the primary layers. Due to this, a noteworthy accumulation of features at the final layer can be realized, which is sufficient to accurately classify every sample, regardless of its visual quality. Besides, as mentioned, the inclusion of hybrid pooling allows the network to capture the essential textural vein patterns, regardless of size, which significantly contributes to improving the recognition accuracy.

#### 4.6. Impact of Regularization Techniques

As aforementioned in Section 3.2, the original architecture of the proposed model employs mainly two types of regularization techniques, L2 regularization and dropout in feature extraction and classification. The former removes small percentage of the weights at each iteration of training while the latter randomly drops neural units and their connection during training. Both techniques have equal importance while training deep neural network and serve a mutual purpose of reducing the risk of overfitting. In this section, we conduct experiments to examine how the proposed model performs in the presence of one or more regularization techniques.

##### Impact of L2 Regularization and Dropouts

First, we considered removing the L2 regularization from the feature extraction network during training and observed the effects on recognition performance during inference. For this, the number of convolutional layers, subsampling layers, filters, and their sizes in fully convolutional feature extraction were not changed. The dropout penalty used in classification was also kept intact throughout the training; however, the L2 regularization was set with a 0 penalty. In other words, no regularization was utilized in feature extrac-

tion. Table 6 shows the comparison of recognition accuracy exhibited by the model when trained with and without L2 regularization and dropout. In most cases, we can see the obvious degradation in the recognition accuracy when L2 regularization is not adopted. With no regularization, the model becomes extremely overfit to the training data, so its recognition performance with the test data became horrendous. Moreover, the absence of a regularization penalty constrains the model's ability to optimize the weight value. Due to this, the network becomes more sensitive to noise in the training data and incurs degraded recognition performance. A good finger vein recognition model must perform well, regardless of the physiological variations in the acquisition environment, i.e., the trained model must generalize well on varying orientations in the test data. Therefore, the introduction of regularization during the finger vein feature extraction is compulsory.

**Table 6.** Impact of L2 Regularization and dropout on the Recognition Performance of the FVR-Net.

Datasets	Recognition Accuracy (%)					
	with L2 and Dropout		with Dropout and without L2		with L2 and without Dropout	
	Good Quality	Poor Quality	Good Quality	Poor Quality	Good Quality	Poor Quality
HKPU	93.18	88.97	87.87	76.32	87.12	46.73
FVUSM	97.84	97.22	96.61	97.27	95.13	93.67

Second, we considered removing dropout from the classification network of the proposed model. During the experiment, we analyzed the impact of dropout on recognition performance from FVR-Net. As mentioned, dropout is utilized only in the network's classification process (after FCL-1, and FCL-2). Therefore, by removing dropout, the units present in the corresponding layers remain intact, with active incoming and outgoing connections from and to the neighboring layers. The L2 regularization penalty was kept unchanged in feature extraction. The same table shows the recognition performance from FVR-Net with and without dropout layers during classification. We can see significantly higher recognition accuracy when dropout layers are utilized. While training with dropout, the corresponding neuron of the imminent layer remains deactivated during feedforwarding. Therefore, an arbitrary neuron from the same layer is responsible for making predictions for the missing neuron. With input such as finger vein images with translational variant features, this technique allows the network to learn various intrinsic representations.

#### 4.7. Impact of Changes in Network Architecture

Designing a good CNN architecture is itself a tedious task, because several factors (one of them being choosing optimal hyperparameters) need to be considered to traverse every single possibility in finding the best model. The features generated by each layer of the network completely rely upon its architecture and the hyperparameters used. For this reason, we performed multiple experiments by tweaking the hyperparameters to some extent, such that an optimal model is found that generalizes well on the test data in achieving a robust finger vein recognition model.

##### 4.7.1. Varying the Kernel Size

One of the various factors that impact the performance of the network is the kernel size of the network. Kernel, in other words, filters, are the matrix of real valued numbers that are slid across the spatial location of the input image, which represents a convolution operation. Kernel sizes in CNNs have a direct relationship with the number of features detected from the input. In general, smaller kernel sizes tend to detect smaller local features while the large kernels detect the generic and larger ones spread across the input. In this experiment, we analyze the impact of kernel size on recognition accuracy from FVR-Net by varying the kernel sizes of the first three convolutional layers. Table 7 shows the

recognition accuracy obtained by FVR-Net with both databases when kernel sizes varied from as small as  $3 \times 3$  pixels to as large as  $7 \times 7$ . We can see that out of all kernel sizes, the optimum recognition accuracy was obtained with  $5 \times 5$ . Due to diversity in the size of extracted features in the finger vein images (ranging from very small to large), the amount of information extracted by extreme kernel sizes was not enough for representation learning by FVR-Net.

**Table 7.** Impact of Varying Kernel Size on the Recognition Performance of the FVR-Net.

Kernel Size	HKPU		FVUSM	
	Good Quality	Poor Quality	Good Quality	Poor Quality
$3 \times 3$	78.03	76.53	90.97	95.97
$5 \times 5$	93.18	88.97	97.84	97.22
$7 \times 7$	45.45	72.65	90.27	83.90

#### 4.7.2. Increasing the Network Depth

In our study, we defined the network depth as the total number of layers present in the network. In this experiment, we have used variable number of layers ranging from three to five in order to observe the recognition performance. The network with three layers represents the compact network with comparatively less trainable parameters than that with five layers. Table 8 shows the recognition accuracy obtained by FVR-Net with variable network depths. We can see that, for similar image quality, recognition accuracy provided by a network with a larger  $M$  is significantly higher. We also notice gradual enhancement in the recognition accuracy upon increasing the depth of the network. Deeper networks impose threats during training, such as higher computational requirements due to the large number of trainable parameters; however, the model's performance improved significantly. With more layers, the deep networks can create hierarchical representations at each layer, from which the model can learn new and more abstract representations of its input.

**Table 8.** Impact of Increasing Depth of the Network on the Recognition Performance of the FVR-Net.

$M$	HKPU		FVUSM	
	Good Quality	Poor Quality	Good Quality	Poor Quality
3	82.57	68.775	90.97	90.94
4	87.12	76.73	89.23	95.4
5	93.18	88.97	97.84	97.22

We witnessed the best model hyperparameters for FVR-Net when the depth of the network was highest (i.e., at  $M = 5$ ) and at a kernel size of  $5 \times 5$  in the first three convolutional layers, followed by  $3 \times 3$  in the later ones. Taking into account the translational variant nature of vein image samples, the selected hyperparameters fit into FVR-Net in such a way that the multiscale deep features of the input can be exploited without losing any distinctive properties of individual subjects. Therefore, they can be regarded as an optimal hyperparameter choice to determine a robust finger vein recognition system based on our proposed network structure.

## 5. Conclusions

This paper investigated a DL framework called FVR-Net for enhancing the recognition accuracy of the finger vein biometric system. The proposed network consists of a block-wise feature extraction network and a classification network. The images input to FVR-Net were first subjected to preprocessing where ROI cropping and image enhancement were performed to segment the vein patterns from the background. Then, the processed image

was fed into the feature extraction network where each block entailed a convolutional layer followed by hybrid pooling of parallelly placed two subsampling layers (maxpooling and average pooling). Their activation maps were concatenated and fed into the classification network with three FCLs for providing the vein recognition output. The whole model was designed to discreetly identify an individual by extracting distinct interclass features of the vein image sample, regardless of ambiguity. This was possible through the introduction of hybrid pooling, which not only activated the most discriminative features of the input but also took into account feature localization, which is a very important thing to consider when dealing with vein images that have translational variant features. Extensive experiments on two publicly available datasets demonstrated the effectiveness of our model, because maximum recognition accuracies of up to 97.84% and 97.22% were achieved for good- and poor-quality images, respectively, and the performance improved significantly, compared to previous methods. Further experiments were conducted by varying network hyperparameters such as regularization technique, the size of convolutional kernels, and the network depth, from which we obtained the optimal hyperparameter settings to achieve maximum recognition accuracy. The proposed method provides great potential and insights into the development of finger vein recognition models for future biometric systems.

**Author Contributions:** Conceptualization, L.D.T. and B.W.K.; methodology, L.D.T. and B.W.K.; formal analysis, L.D.T. and B.W.K.; investigation, L.D.T.; writing—original draft preparation, L.D.T. and B.W.K.; writing—review and editing, L.D.T. and B.W.K.; supervision, B.W.K.; project administration, B.W.K.; funding acquisition, B.W.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by a project for Industry-Academic Cooperation Based Platform R&D funded Korea Ministry of SMEs and Startups in 2020 (S3014213) and a National Research Foundation of Korea (NRF) grant funded by the Korean government (2022R1A2B5B01001543).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Jain, A.; Hong, L.; Bolle, R. On-line fingerprint verification. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 302–314. [[CrossRef](#)]
2. Zhang, D.; Kong, W.K.; You, J.; Wong, M. Online palmprint identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1041–1050. [[CrossRef](#)]
3. Kumar, A.; Zhou, Y. Human Identification Using Finger Images. *IEEE Trans. Image Processing* **2012**, *21*, 2228–2244. [[CrossRef](#)] [[PubMed](#)]
4. Yang, L.; Gang, G.; Yin, Y.; Xi, X. Finger Vein Recognition with Anatomy Structure Analysis. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 1892–1905. [[CrossRef](#)]
5. Yang, J.; Shi, Y.; Yang, J. Finger-Vein Recognition Based on a Bank of Gabor Filters. In Proceedings of the 9th Asian Conference on Computer Vision, 1, Xi'an, China, 23–27 September 2009; pp. 374–383. [[CrossRef](#)]
6. Yang, J.; Yang, J. Multi-Channel Gabor Filter Design for Finger-Vein Image Enhancement. In Proceedings of the Fifth International Conference on Image and Graphics, Xi'an, China, 20–23 September 2009; pp. 87–91.
7. Yang, L.; Yang, G.; Zhou, L.; Yin, Y. Superpixel based finger vein ROI extraction with sensor interoperability. In Proceedings of the 2015 International Conference on Biometrics (ICB), Phuket, Thailand, 19–22 May 2015; pp. 444–451.
8. Huang, B.; Dai, Y.; Li, R.; Tang, D.; Li, W. Finger-Vein Authentication Based on Wide Line Detector and Pattern Normalization. In Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 1269–1272.
9. Zhang, Y.; Li, W.; Zhang, L.; Ning, X.; Sun, L.; Lu, Y. Adaptive Learning Gabor Filter for Finger-Vein Recognition. *IEEE Access* **2019**, *7*, 159821–159830. [[CrossRef](#)]
10. Turlapaty, C.; Goru, H.K.; Gokaraju, B. Gabor filter based entropy and energy features for basic scene recognition. In Proceedings of the IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 18–20 October 2016; pp. 1–4.
11. Miura, N.; Nagasaka, A.; Miyatake, T. Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. *Mach. Vis. Appl.* **2004**, *15*, 194–203. [[CrossRef](#)]
12. Li, J.; Ma, H.; Lv, Y.; Zhao, D.; Liu, Y. Finger Vein Feature Extraction Based on Improved Maximum Curvature Description. In Proceedings of the Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2019; pp. 7566–7571.

13. Lu, Y.; Xie, S.J.; Yoon, S.; Park, D.S. Finger vein identification using polydirectional local line binary pattern. In Proceedings of the International Conference on ICT Convergence (ICTC), Jeju, Korea, 14–16 October 2013; pp. 61–65.
14. Yang, L.; Yang, G.; Xi, X.; Su, K.; Chen, Q.; Yin, Y. Finger Vein Code: From Indexing to Matching. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 1210–1223. [[CrossRef](#)]
15. Parthiban, K.; Wahi, A.; Sundaramurthy, S.; Palanisamy, C. Finger vein extraction and authentication based on gradient feature selection algorithm. In Proceedings of the Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014), Bangalore, India, 17–19 February 2014; pp. 143–147.
16. Vega, P.; Travieso, C.M.; Alonso, J.B. Biometric personal identification system based on patterns created by finger veins. In Proceedings of the 2014 International Work Conference on Bio-inspired Intelligence (IWOBI), Liberia, Costa Rica, 16–18 July 2014; pp. 65–70.
17. Liu, Y.; Di, S.; Jin, J.; Huang, D. An improved finger-vein recognition algorithm based on template matching. In Proceedings of the International Symposium on Optoelectronic Technology and Application, Beijing, China, 25 October 2016.
18. Xiong, Z.; Liu, M.; Guo, Q. Finger Vein Recognition Method Based on Center-Symmetric Local Binary Pattern. In Proceedings of the IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 29–31 March 2019; pp. 262–266.
19. Wu, J.D.; Liu, C.T. Finger-vein pattern identification using principal component analysis and the neural network technique. *Expert Syst. Appl.* **2011**, *38*, 5423–5427. [[CrossRef](#)]
20. Khan, A.; Farhan, M.; Khursid, A.; Akram, A. A Multimodal Biometric System Using Linear Discriminant Analysis for Improved Performance. Available online: <https://arxiv.org/abs/1201.3720> (accessed on 21 April 2021).
21. Zhou, L.; Yang, G.; Yang, L.; Yin, Y.; Li, Y. Finger vein image quality evaluation based on support vector regression. *Int. J. Signal Process. Image Process. Pattern Recog.* **2015**, *8*, 211–222. [[CrossRef](#)]
22. Wu, J.D.; Liu, C.T. Finger-vein pattern identification using SVM and neural network technique. *Expert Syst. Appl.* **2011**, *38*, 14284–14289. [[CrossRef](#)]
23. Hoshyar, N.; Sulaiman, R. Smart access control with finger vein authentication and neural network. *J. Am. Sci.* **2011**, *7*, 192–200.
24. Khellat-kihel, S.; Abrishambaf, R.; Cardoso, N.; Monteiro, J.; Benyettou, M. Finger vein recognition using Gabor filter and support vector machine. In Proceedings of the International Image Processing, Applications and Systems Conference, Sfax, Tunisia, 5–7 November 2014; pp. 1–6.
25. Wang, K.Q.; Khisa, A.S.; Wu, X.Q.; Zhao, Q.S. Finger vein recognition using LBP variance with global matching. In Proceedings of the Wavelet Analysis and Pattern Recognition (ICWAOR), Guangdong, China, 15–17 July 2012; pp. 196–201.
26. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep Learning-Based Classification of Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2094–2107. [[CrossRef](#)]
27. Hou, Q.; Cheng, M.; Hu, X.; Borji, A.; Tu, Z.; Torr, P. Deeply Supervised Salient Object Detection with Short Connections. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 815–828. [[CrossRef](#)]
28. Tamang, L.D.; Kim, B.W. Deep D2C-Net: Deep learning-based display-to-camera communications. *Opt. Express* **2021**, *29*, 11494–11511. [[CrossRef](#)] [[PubMed](#)]
29. Noh, K.J.; Choi, J.; Hong, J.S.; Park, K.R. Finger-Vein Recognition Based on Densely Connected Convolutional Network Using Score-Level Fusion with Shape and Texture Images. *IEEE Access* **2020**, *8*, 96748–96766. [[CrossRef](#)]
30. Song, J.M.; Kim, W.; Park, K.R. Finger-Vein Recognition Based on Deep DenseNet Using Composite Image. *IEEE Access* **2019**, *7*, 66845–66863. [[CrossRef](#)]
31. Huang, H.; Liu, S.; Zheng, H.; Ni, L.; Zhang, Y.; Li, W. DeepVein: Novel finger vein verification methods based on Deep Convolutional Neural Networks. In Proceedings of the IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), New Delhi, India, 22–24 February 2017; pp. 1–8.
32. Qin, H.; El-Yacoubi, M.A. Deep Representation for Finger-Vein Image-Quality Assessment. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 1677–1693. [[CrossRef](#)]
33. Kuzu, R.S.; Piciuccio, E.; Maiorana, E.; Campisi, P. On-the-Fly Finger-Vein-Based Biometric Recognition Using Deep Neural Networks. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 2641–2654. [[CrossRef](#)]
34. Hou, B.; Yan, R. Convolutional Autoencoder Model for Finger-Vein Verification. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 2067–2074. [[CrossRef](#)]
35. Qin, H.; Wang, P. Finger-Vein Verification Based on LSTM Recurrent Neural Networks. *Appl. Sci.* **2019**, *9*, 1687. [[CrossRef](#)]
36. Zhang, J.; Lu, Z.; Li, M.; Wu, H. GAN-Based Image Augmentation for Finger-Vein Biometric Recognition. *IEEE Access* **2019**, *7*, 183118–183132. [[CrossRef](#)]
37. Asaari, M.S.M.I.; Suandi, S.A.; Rosdi, B.A. Fusion of band limited phase only correlation and width centroid contour distance for finger-based biometrics. *Expert Syst. Appl.* **2014**, *41*, 3367–3382. [[CrossRef](#)]
38. Yadav, G.; Maheshwari, S.; Agarwal, A. Contrast limited adaptive histogram equalization based enhancement for real time video system. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI), Delhi, India, 24–27 September 2014; pp. 2392–2397.
39. Haralick, R.M.; Sternberg, S.R.; Zhuang, X. Image Analysis Using Mathematical Morphology. *IEEE Trans. Pattern Anal. Mach. Intell.* **1987**, *PAMI-9*, 532–550. [[CrossRef](#)] [[PubMed](#)]

40. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
41. Park, S.; Kwak, N. Analysis on the Dropout Effect in Convolutional Neural Networks. Available online: [http://mipal.snu.ac.kr/images/1/16/Dropout\\_ACCV2016.pdf](http://mipal.snu.ac.kr/images/1/16/Dropout_ACCV2016.pdf) (accessed on 14 January 2021).
42. Platt, J.C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv. Large Margin Classif.* **1999**, *10*, 61–74. Available online: <http://www.cs.cornell.edu/courses/cs678/2007sp/platt.pdf> (accessed on 14 January 2021).
43. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554–2558. [[CrossRef](#)] [[PubMed](#)]