# Dorsal Hand Vein Recognition by Convolutional Neural Networks: Feature Learning and Transfer Learning Approaches

**Norah A. Al-johania[1]\***      **Lamiaa A. Elrefaei[1,2]**

[1]*Computer Science Department, Faculty of Computing and Information Technology,*
*King Abdulaziz University, Jeddah, Saudi Arabia*
[2]*Electrical Engineering Department, Faculty of Engineering at Shoubra,*
*Benha University, Cairo, Egypt*
\* Corresponding author's Email: naljuhany0001@stu.kau.edu.sa

**Abstract:** In this paper, we propose a dorsal hand vein recognition system using Convolutional Neural Network (CNN). This system is automatically learned how to extract features from original image without preprocessing. The proposed system has two approaches: the first one is using the pre-trained CNN models (AlexNet, VGG16 and VGG19) for extracting features from 'fc6','fc7' and 'fc8' layers then using Error-Correcting Output Codes (ECOC) with Support Vector machine (SVM) and K-Nearest Neighbor (K-NN) algorithms for the classification. The second approach is using transfer learning with CNN (AlexNet, VGG16 and VGG19) models for both features' extraction and classification. The experiments applied on two different datasets: Dr. Badawi hand veins dataset that contains 500 image and BOSPPHORUS dorsal vein dataset that contains 1575 images. In the first approach experiment, the recognition accuracy of all models gives best result when features are extracted from 'fc6'. Also, the accuracy rate of the models that use ECOC with SVM for classification is higher than the models that use ECOC with KNN and the VGG19 model achieves better results in models that use ECOC with SVM. In the second approach experiment, the recognition accuracy for all models give best result when epoch number is 50 where Dr. Badawi dataset in VGG16 and AlexNet reaches to 100% recognition rate and BOSPPHORUS dataset reaches to 99.25 % recognition rate in VGG19. Finally, the discussion concluded that using transfer learning is giving more accuracy rate than using the pre-trained CNN models for extracting features.

**Keywords:** Identification, Biometric, Hand vein, Patterns recognition, CNN, SVM, KNN.

## 1. Introduction

Biometric systems are developing technologies that can be used in automatic systems for identifying individuals uniquely and effectively and it becomes a good alternative to the traditional methods such as passwords. The biometric characteristics could be divided into two types. The first type contains physical features (fingerprint, face, iris, retina, palmprint, hand, face and vein, ETC.) and the second type contains behavioral features (e.g. voice, signature, movement characteristics (gait, hands movement or lips, key press, etc.)) [1]. According to the classification in [1], palm prints, fingerprints, hand vein and ear canal are the four biometric modalities possessing all of the following properties:

- Universality: means all persons must have the characteristic;
- Uniqueness: means the characteristic has to be single to all persons;
- Permanence: the characteristic does not change with aging;
- Measurability: using easy method to obtain image from the persons;
- Performance: exactly how good it is at identifying persons;
- Acceptability: the people must be agreeable to deliver the characteristic;
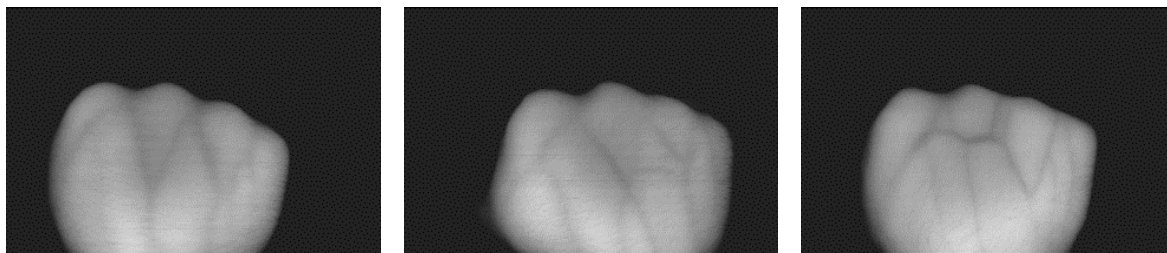- Circumvention: how simply can it be copied.

Figure. 1 Hand vein images [3]

Dorsal hand vein pattern, shown in Fig. 1, is the network of blood vessels under body's skin. The vein pattern was used as a biometric technology in 1992. According to some earlier studies [2], vein patterns are sufficiently dissimilar through persons, also they are stable unaffected via ageing and no important changed in adults through observing. The patterns of blood vein are unique for all parsons, even between twins. Contrasting other biometric characters, like face or fingerprint, vein patterns provide a detailed pattern that are unseen inside of a person body making them unaffected with the conditions of the outside skin (such as dirty hand) [2].

Due to the deep learning characteristics which include extracting the aim feature automatically, it does fit to recognition systems such as visual identification, speech identification and natural language treating. Convolution Neural Network (CNN) is one of the public models of deep learning that inspired through the mechanism of normal visual perception that also becomes the greatest thorough research. Its new structure was recognized in the research available in the 1990s by LeCun et al [4]. CNN is able to get the actual representation of the main image and is able to identify its visual straight from the picture pixels using small preprocess [4].

In this paper, a dorsal hand vein recognition system using Convolutional Neural Network (CNN) is proposed. The main contributions of this paper are:

1. Investigating the dorsal hand vein recognition using the pre-trained CNN models (AlexNet, VGG16 and VGG19) for extracting feature then using Error-Correcting Output Codes (ECOC) with Support Vector machine (SVM) and K-Nearest Neighbor (K-NN) algorithms for the classification.

2. Investigating the dorsal hand vein recognition using transfer learning with CNN (AlexNet, VGG16 and VGG19) models for both features' extraction and classification.

3. Evaluating the proposed system using two different datasets: Dr. Badawi and

BOSPPHORUS dorsal hand veins datasets and achieved high recognition rates.

This reminder of this paper is ordered as follows: Section 2 presents various dorsal hand vein recognition methods in the literature. In section 3, the proposed dorsal hand vein recognition methods based on CNN are explained. Evaluating the performance of the proposed recognition system and experimentations is introduced in Section 4 where details about the used datasets and comparisons with other existing methods are provided. Finally, section 5 provides the paper conclusion.

## 2. Related work

The important characteristic of hand vein patterns is stability, which means that the hand structure and hand veins configuration continue comparatively stable through the individual's life. For this reason, vein identification systems are considered as a promising and reliable biometric. In this section, some of the vein identification systems are presented.

Huang et al. [5] proposed a method for dorsal hand vein identification. A new process integrating together holistic and local analysis then hierarchically joint with that from the surface modality, born by a reputable texture operator, that Local Binary Patterns (LBP), Binary Coding (BC) and graph for decision production by Factorized Graph Matching (FGM). Consequences attained are greater than the state-of-the-art ones so far described in works, which proves its efficiency.

Lee et al. [6] suggested a directional filter include different alignments that cutting hand vein patterns and encode hand vein features into binary code by the minimum directional filtering response (MDFR) and classification by Hamming Distance (HD). Also, there are many areas that not contain the vein in the image, which are not important for hand vein identification. To increase accurateness, the regions that not contain the vein are identified through calculating the modification of the minimum filtering. Their suggested approach

achieves high accuracy that displays the method is effective for dorsal hand vein identification.

Trabelsi et al. [7] suggested a new hand vein pattern identification process for person recognition. Fixed static texture descriptor known as Circular Difference and Statistical Directional Patterns (CDSDP) is suggested to extract hand vein patterns and Artificial Neural Network (ANN), Feedforward Multilayer Neural Network (FMNN) for classification. The CDSDP is a neighboring circular change with weights combining the statistical directional data of vessels. Experimental display that descriptor depend on CDSDP has improved effective than the earlier descriptors that used in LBP.

Yun et al. [8] presented a novel hand vein identification system that depend on the linking lines of feature points. This way the feature points is the connection points with the endpoints of the dorsal hand vein image. They extracted the reference point from feature points that is connection points and the endpoints. Features that extracted are identified by calculating the comparative distances among the two feature points and the angles among the neighboring connections of this two feature points. Lastly, these two features are joint for hand vein identification. The process efficiently overcome the effect on the identification outcomes produced with image translation and rotation.

Chuang et al. [9] suggested local feature-based hand vein image process depend on minutiae features extraction from venous networks to study the greatest discriminative areas and features of dorsal hand veins for recognition. These minutiae feature contain end points and the curve lines among the two end points as measured beside the edge of the area of attention. In addition, suggest a dynamic pattern tree (DPT) to speed up matching presentation and estimate the feature points discriminatory power for verifying an individual's identity.

Zhu et al. [10] proposes an approach for dorsal hand vein recognition that use the texture with geometry features. They first segment the vein area then compute its skeleton in addition to the Energy Cost take out in the Thinning process (TEC) that used to decrease a number of incorrect candidates. Then texture and geometry clues are shown through Local Binary Patterns (LBP) and the diagram collected with the overpass points and endpoints of vein correspondingly and the two modalities are lastly joint for decision production.

A system that uses deep learning for dorsal hand vein recognition was proposed by Wan et al. [11]

that depends on CNN, and extracts image by the region of interest (ROI) then preprocess this image with histogram equalization and Gaussian smoothing filter. The system extracts feature by using Convolution Architecture for Feature Extraction (Reference-CaffeNet), AlexNet and VGG and using logistic regression for classification.

As a summary, the systems presented in [5-10] extracting features by using machine learning techniques that need to be identified by an expert and then hand-coded as per the domain and data type. As well as in [11] images have been processed before the features extraction by extracting the region of interest (ROI) then applied histogram equalization and Gaussian smoothing filter.

However, the proposed system in this paper is automatically learned how to extract features from original image without preprocessing based on deep learning. The CNN depth model is used to eliminate the work of selecting feature artificially. CNN can select and express the depth feature of the image automatically, thus ensuring the accuracy of the selection of image features and the validity of the representation.

## 3.  Framework of the proposed system

The proposed dorsal hand vein recognition system that use Convolutional Neural Networks (AlexNet, VGG16 and VGG19) models is shown in Fig. 2. This system has two approaches:

- Using the pre-trained CNN models for extracting features from 'fc6', 'fc7' and 'fc8' layers then using Error-Correcting Output Codes (ECOC) with SVM and K-NN algorithms for classification.
- Transfer learning with CNN: pre-trained convolutional neural networks (AlexNet, VGG16 and VGG19) are used for both features' extraction and classification.

### 3.1 Features extracted by convolutional neural network

In this paper, we used the CNN of deep learning and the essential modules and analyzed the general structure of the network. CNN structure was earliest suggested in 1988 by Fukushima. The CNN is a category of image recognition technique in the deep neural network that is has singular structure in the process of several procedures of distortion of the image by a high stability. The general structure of a CNN, shown in Fig. 3, contains two key parts:
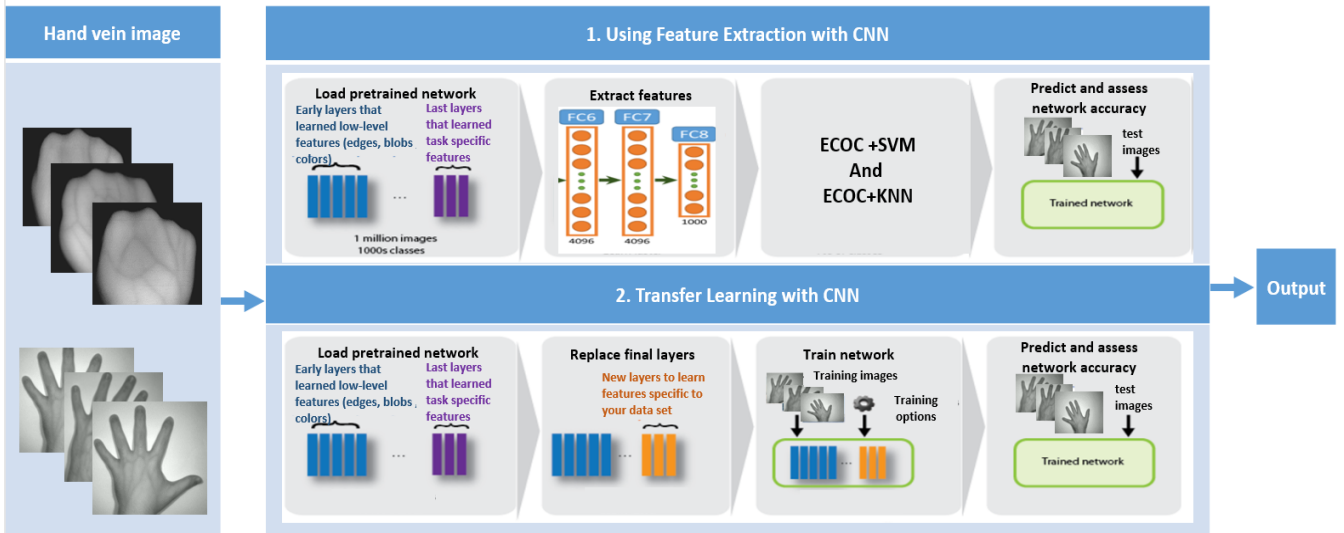
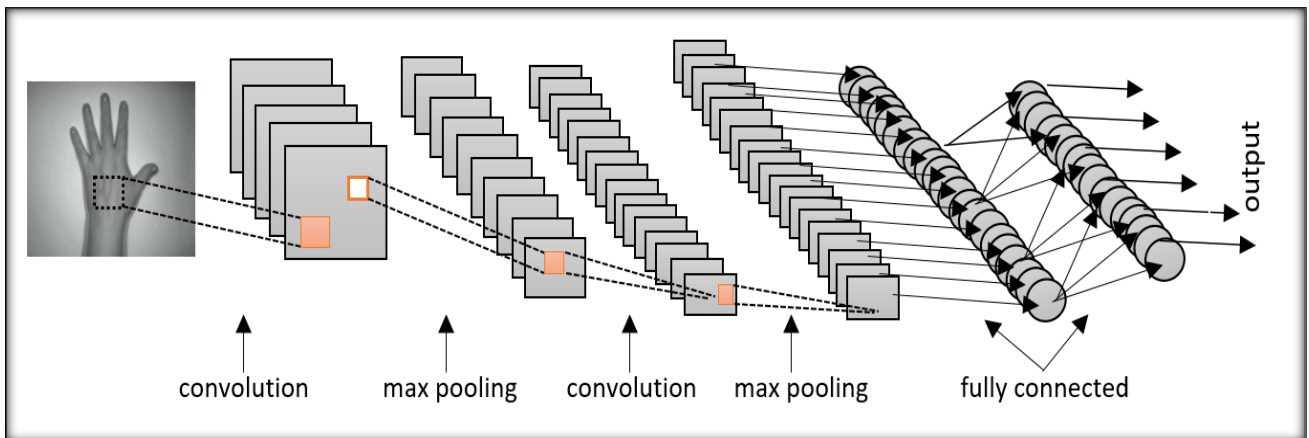Figure. 2 Architecture of dorsal hand vein recognition system



Figure. 3 General structure of a CNN

feature extractors then a classifier. In the feature extraction part, all layer in the network gets the production from its immediate prior layer as its input and get its production as the input to the following layer. In the classification part, the chosen number of features that extracted as inputs with consideration to the measurement of last neural network weight matrix. Then the weight of the separate class in the highest classification layer is calculated, then gives output for the corresponding classes depending on the best weight [12, 13].

The CNN architecture based on some main concepts: Convolutional Layer, Pooling Layer, activation functions (such as Sigmoid, Tanh, ReLU, and so on), and Fully-Connected Layer [14].

- Convolutional Layer: this layer processes the input image and extracts many kinds of features by using filters. The filter applied to image and get a specific map, the result is multiple feature maps. The features change depends on the filter

that applied. The primary convolution layer gets low-level features such as boundaries, lines, and angles. last layers get higher-level features.

- Pooling Layer: this layer is for the purpose of decreasing the image size which is completely independent on respectively depth dimension where depth of the image remains the same. The pooling layer decreases the features resolution. It creates features that are robust to distortion and noise. There are two methods to make pooling are: max pooling and average pooling as show in Fig. 4. The max pooling is most public type of pooling layer.

- Activation functions: this layer set the input in the range of values where the outcome of the earlier layer must be in range of values. That make the input and production sizes of this layer are similar. The usually applied activation functions are Sigmoid, Tanh and ReLU. In this
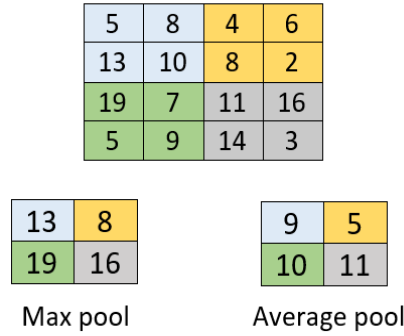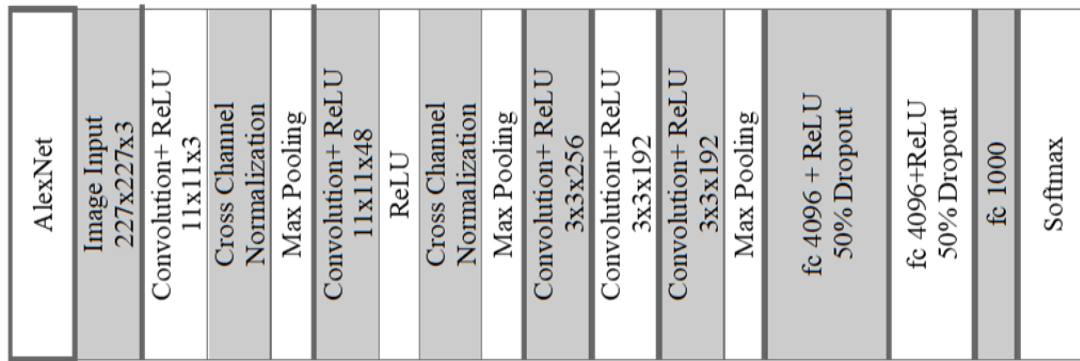
Figure. 4 Pooling methods



Figure. 5 The structure of AlexNet

paper we use ReLU because benefit is the network trains faster Functions of Sigmoid, Tanh and ReLU is defined in Eqs. (1), (2) and (3) respectively [11]:

$$\text{Sigmoid: } \varphi(\mathrm{x}) = \frac{1}{1+e^{-\mathrm{x}}} \qquad (1)$$

$$\text{Tanh: } \varphi(\mathrm{x}) = \frac{1-e^{-2\mathrm{x}}}{1+e^{-2\mathrm{x}}} \qquad (2)$$

$$\text{ReLU: } \varphi(\mathrm{x}) = \max(0.x) \qquad (3)$$

• Fully Connected Layer: is the final layers of a CNN. The major function of this layer is the classification by scientifically adding the weighting of the prior layer of features to decide an exact goal output result.

There are many common CNN architectures that contains of many convolutional, maxpooling then a fully connected and Softmax layers at the final with different parameter. In this research, the pre-trained CNNs are used rather than building new CNN. In this paper a pre-trained CNN used for feature extractions from images by using pre-trained CNNs called AlexNet, VGG-16 and VGG-19 that are used respectively in [15,16 and 17] and features are extracted from fully connected (FC) layers then used for classification task.

### 3.1.1. AlexNet

AlexNet is proposed in 2012, when compared with baggier CNN model, the AlexNet won in the most challenging ImageNet experiment for graphic object identification [18]. AlexNet reached a higher identification accuracy compared to all the machine learning algorithms and computer vision methods. It was important advance in machine learning area also in computer vision aimed at graphic identification.

The structure of AlexNet and the number of filters and its size shown in Fig. 5. The first and second layers implement convolution, ReLU, Cross Channel Normalization and max pooling where the first layer convolution uses 96 different filters of size 11×11×3 with stride size of 4 and at max pooling implement 3×3 filters also using a stride size of 2. The second layer convolution uses 256 different filters of size 5×5×48 with stride size of 1 and at max pooling implement 3×3 filters that using a stride size of 2. The third and fourth layers implement convolution and ReLU where convolution is done with 384 different filters of size 3x3x256 in the third layer and 3x3x192 in the fourth layer. The fifth layer implement convolution, ReLU

Figure. 6 The structure of VGG-16 andVGG-19

and maxpooling where convolution uses 256 different filters of size 3x3x192 with stride size of 1 [15]. Then used three fully connected (FC) layers and dropout, at the end applied a Softmax layer.

### 3.1.2. VGG-16 and VGG-19

In the Visual Geometry Group (VGG) model, the first three layers use dual convolutional layers which following by ReLU activation function with 64,128, and 256 different filters of size 3×3 with stride size of 1 then a single max pooling layer are used. In VGG-16 the fourth and fifth layers use three convolutional layers followed by ReLU activation function with 512 different filters of size 3×3 with stride size of 1 followed by a single max pooling layer. While in VGG-19 the fourth and fifth layers use four convolutional layers followed by ReLU activation function with 512 different filters of size 3×3 with stride size of 1 followed by a single max pooling layer. After that three fully connected layers are used with a ReLU at the end applied a Softmax layer [16, 17]. The structures of VGG-16 andVGG-19 models and the number of filters and its size are shown in Fig. 6.

### 3.2 Classification

In this model we use two methods for classification. The first method is extract learned features from the pretrained network then use those features for classification by ECOC with support vector machine (SVM) and K-Nearest Neighbor (KNN). The second method is fine tuning, that means features extraction and classification by the use of a pre-trained CNN.

### 3.2.1. Error-Correcting Output Codes (ECOC) multiclass model

After extracting features from CNN models, we applied different classifiers to these features. Error-correcting output codes (ECOC) are general classification framework which includes two stages: encoding and decoding. Encoding phase consists of designing Coding Matrix (CM) where columns represent binary classifiers and rows indicate code words for classes. Designing of a coding matrix can be made using binary coding and ternary coding [1]. In binary coding the elements of a coding matrix are $CM_{ij} \in \{1, -1\}$ or $CM_{ij} \in \{0, 1\}$, respectively, i=1, 2, ..., M and j=1,2, …, Nc, where M is the number of classes and Nc is the number of binary classifiers. In the columns the division of values of -1 and 1 or 0 and 1, respectively, shows which classes are grouped together in each classifier. In ternary coding, the elements of the coding matrix belong to the set {-1,0,1}. In this approach values -1 and 1 have the same purpose as in binary coding but now 0 means that this specific class is excluded from the training of an individual binary classifier [18].

There are numerous strategies of how to construct the coding matrix. We chose one-vs-all (OVA) strategies [18] which uses binary coding and each one of the binary classifiers separates one class from the rest. Hence, the number of binary classifiers is only $\varphi(M)$ where M is the number of classes. OVA is a very simple approach but from the computational perspective it is inefficient since the whole training set is needed in the case of each binary classifier.

The ECOC is not attached to any specific classification method, it can be used with various classification methods. For this paper we restricted our focus on the following classification methods to be used with ECOC:

1. K-Nearest Neighbor (K-NN) method: is one of the simplest classification algorithms and it is one of the most used learning algorithms. A test instance is classified by a majority vote of its neighbors. With k-NN the main parameters are to be considered: the number of nearest neighbors, distance measure and distance weighting function. The distance between two J-dimensional vectors x and y is calculated by Euclidean distance that appear in the Eq. (4) where J is the number of variables then select the 5 nearest neighbors and for distance weighting functions we chose equal weighting [18].

$$d_{x.y} = \sqrt{\sum_{j=1}^{J} (x_j + y_j)} \qquad (4)$$

2. Support Vector Machine (SVM) method: in SVM we selected kernel function is linear assuming we have the set of training data $(x_1,y_1),(x_2,y_2),...(x_n,y_n)$. Moreover, we want to classify the set into two classes where $x_i \in R^d$ is the feature vector of length d, and $y_i \in (-1, +1)$ is the label class. The two classes are linearly separable with a hyperplane $wx + b = 0$ With no other previous knowledge about the data, SVM can find the best hyperplane which has the maximum margin [13]. The multi-class Support Vector Machine for N classes is applied as a one-vs-all (OVA) strategies [18] that is equal to merging N Binary Support Vector Machine classifiers, with each classifier distinguished a single class against all other classes.

### 3.2.2. Fine tuning

This approach is based on transfer learning a pre-trained CNN. Transfer learning means to use a model that has already been trained on other images, and now we have a different set of images that we want this model to give a prediction for. The input image is fed to the pre-trained CNN (Alex-Net, VGG16 and VGG19) as an automatic feature extractor. After that obtained the matching score using the generated feature vectors from the last

fully connected layer as the input to the Softmax layer.

A Softmax layer applies a Softmax function, Eq. (5), to the input:

$$P(y = j \mid z^{(i)}) = \frac{e^{z^{(i)}}}{\sum_{j=0}^{k} e^{z_k^{(i)}}} \qquad (5)$$

Where we define the $z$ as

$$z = \sum_{j=1}^{m} w_j \ x_j + b \qquad (6)$$

Where $w$ is the weight vector, $x$ is the feature vector of 1 training sample, and b is the bias unit. Now, this Softmax function calculates the probability that training sample $x^{(i)}$ be appropriate to class $j$ given the weight and net input $z^{(i)}$. So, we compute the probability $P(y = j \mid z^{(i)})$ for each class label j = 1. .... k. The Softmax function is also known as the normalized exponential and can be considered the multi-class generalization of the logistic sigmoid function [19].

## 4. Experiments and results

This part evaluating the performance of proposed dorsal hand vein recognition system. The proposed system has two approaches. first is using the pre-trained CNN models (AlexNet, VGG16 and VGG19) for extracting features from 'fc6', 'fc7' and 'fc8' layers then use ECOC with SVM and K-NN algorithms for classification. Second approach is to use transfer learning with CNN (AlexNet, VGG16 and VGG19) models for features extracting and classification. The experiment applied on two different datasets which are Dr. Badawi hand veins dataset [3] and BOSPPHORUS dorsal vein dataset [20]. The performance of the proposed dorsal hand vein recognition system is represented using the recognition accuracy.

### 4.1 Data set

This paper used two datasets, first dataset is a big dataset related to the field of dorsal hand vein recognition that known in the literature as "Dr. Badawi hand veins dataset" [3]. The dataset contains images for 50 different persons with 10 images per each individual, 5 images for left hand and 5 images for right hands, mirrored to have the same
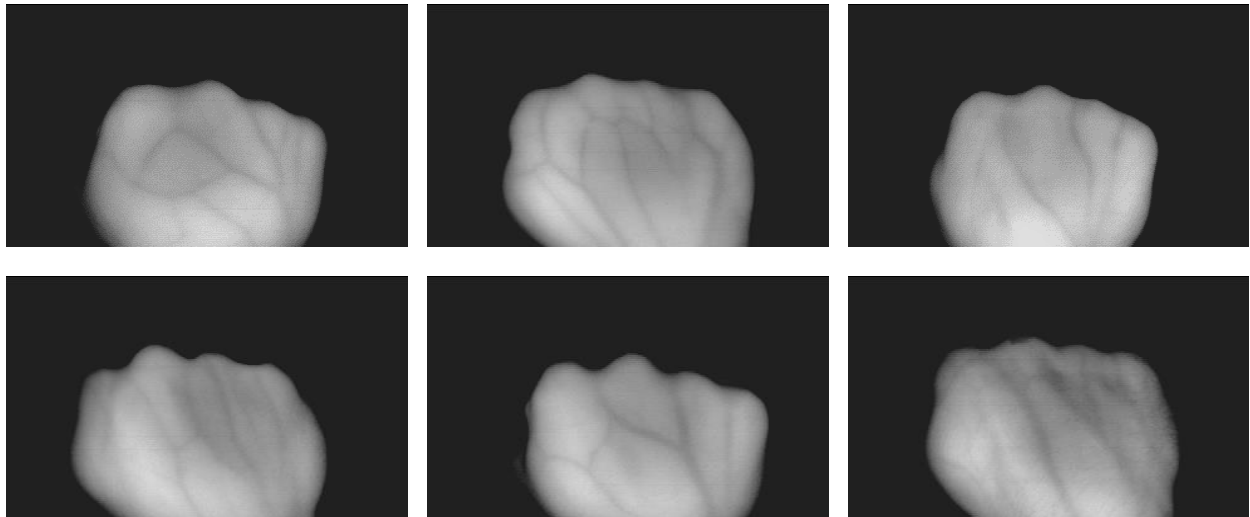
Figure. 7 Some samples from Dr. Badawi hand veins dataset used in this research [3]
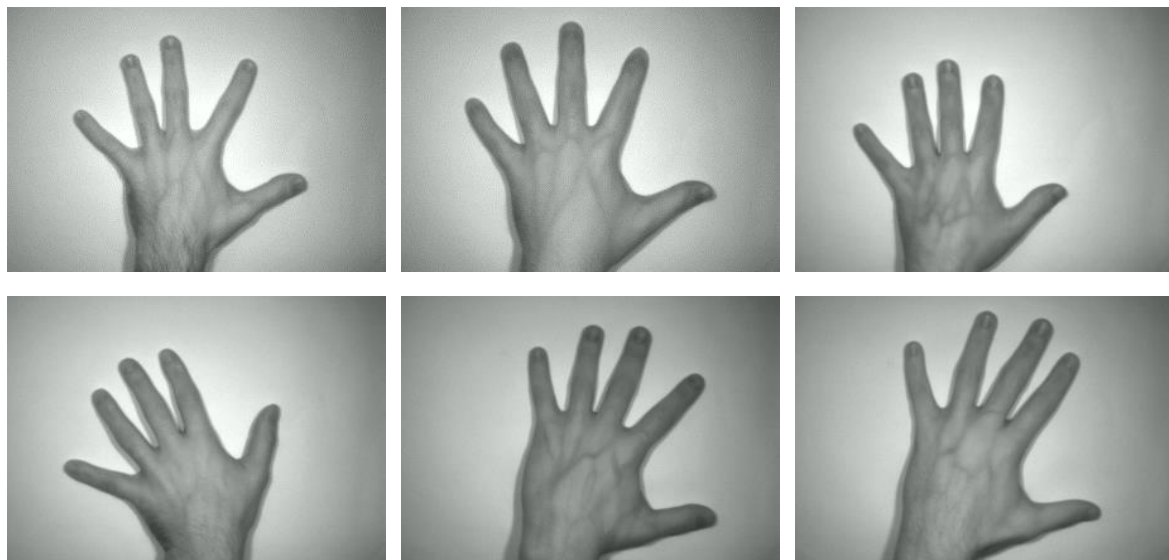


Figure. 8 Some samples from BOSPPHORUS left hand veins dataset used in this research [20]

coordinate basis of patterns localization. This dataset is composed of females and males in the range between 16- and 65-years' age. Persons are in healthy conditions with different affiliations (students, professors, engineers and workers, house wives, etc.), Fig. 7 displays some samples from this database [3, 21, 22].

The second dataset is BOSPPHORUS dorsal vein of left hand that contains 1575 images that taken form 100 persons. Also, this dataset contains 219 images for 25 users obtained after two or five months. In addition, three positions for each customer are use in regular and specific situations (after having carried a bag weighing 3 kg. for one minute and after the hand cooled) [20], Fig. 8 displays some samples from this database.

## 4.2 Results and discussion

The training and testing methods are implemented using a desktop computer environment (Intel(R) Core (MT) i7-7700 HQ CPU @ 2.80 GHz, win10 64-bit operating system, RAM 16 GB, 256GB SSD, NVIDIA GTX 1060 graphics card, and implemented in MATLAB R2018a contains a deep learning toolbox.

### 4.2.1. First approach results: Feature extraction with convolutional neural networks

The CNN network structures a hierarchical representation of input images where higher-level features are extracted from deeper layers, while lower-level features are extracted from the first layers. In the proposed system the feature extracted

Table 1. The recognition accuracy of AlexNet, VGG16 and VGG 19 models for Dr. Badawi and BOSPPHORUS datasets from each layer for feature extraction and ECOC with K-NN classification

| Dataset | model | Feature Extraction Layer | | | | | | | | |
| | | Fc6 | | | Fc7 | | | Fc8 | | |
| | | Split Ratio | | | | | | | | |
| | | 80-20 | 70-30 | 60-40 | 80-20 | 70-30 | 60-40 | 80-20 | 70-30 | 60-40 |
| Dr. Badawi | VGG16 | 97 | 97 | 96.50 | 93 | 93 | 93.500 | 87 | 87 | 86 |
| | VGG19 | 96 | 96 | 91.50 | 92 | 92 | 88.500 | 87 | 87 | 76 |
| | AlexNet | 99 | 99 | 97.50 | 97 | 97 | 94.50 | 91 | 91 | 88.50 |
| BOSPPHORUS | VGG16 | 90 | 85 | 83.048 | 89.500 | 84.500 | 81.80 | 83.500 | 78.250 | 73.681 |
| | VGG19 | 89.50 | 87.50 | 84.314 | 87 | 85 | 81.848 | 82 | 77.75 | 74.781 |
| | AlexNet | 77 | 70 | 65.281 | 76.667 | 67.25 | 62.60 | 72.667 | 60.50 | 57.267 |

Table 2. The recognition accuracy of AlexNet, VGG16 and VGG 19 models for Dr. Badawi and BOSPPHORUS datasets from each layer for feature extraction and ECOC with SVM classification

| Dataset | model | Feature Extraction Layer | | | | | | | | |
| | | Fc6 | | | Fc7 | | | Fc8 | | |
| | | Split Ratio | | | | | | | | |
| | | 80-20 | 70-30 | 60-40 | 80-20 | 70-30 | 60-40 | 80-20 | 70-30 | 60-40 |
| Dr. Badawi | VGG16 | 100 | 100 | 99.500 | 97 | 97 | 97 | 97 | 97 | 96 |
| | VGG19 | 100 | 100 | 98 | 98 | 98 | 96.50 | 96 | 96 | 95 |
| | AlexNet | 99 | 99 | 98.50 | 98 | 98 | 98.50 | 95 | 95 | 94 |
| BOSPPHORUS | VGG16 | 97.50 | 97.25 | 96.60 | 96 | 96.75 | 95.33 | 91 | 92.75 | 90.581 |
| | VGG19 | 98.50 | 96.75 | 94.514 | 97.50 | 94.25 | 92.381 | 93 | 91.75 | 89.248 |
| | AlexNet | 93.50 | 91.250 | 90.867 | 93.500 | 89.250 | 89.733 | 87 | 83.75 | 80.90 |

for training and testing images from deeper layers of a pre-trained model namely 'FC6', 'FC7'and 'FC8' using the "activations" method.

The CNN networks used in this research need the size of the input images to be 227-by-227, but the images in the datasets have different sizes. We resized the images before using the CNNs for training and testing, by generating augmented image datastores, identifying the wanted image size, and using these datastores as input arguments to activations. The used hand vein databases were split by many ways: first 60% training and the last 40% testing, first 70% training and the last 30% testing, and first 80% training and the last 20% testing.

The ECOC with SVM and k-NN classifiers are trained using features extracted from 'FC6', 'FC7'and 'FC8' layers of AlexNet, VGG16 and VGG 19 models. With k-NN the we selected the number of nearest neighbors is set to 5, Euclidean distance measures and for distance weighting functions we chose equal weighting while in SVM we selected kernel function is linear.

Table 1 shows the recognition accuracy of each dataset when classification is done by ECOC with K-NN. Table 2 shows the recognition accuracy when classification is done by ECOC with SVM.

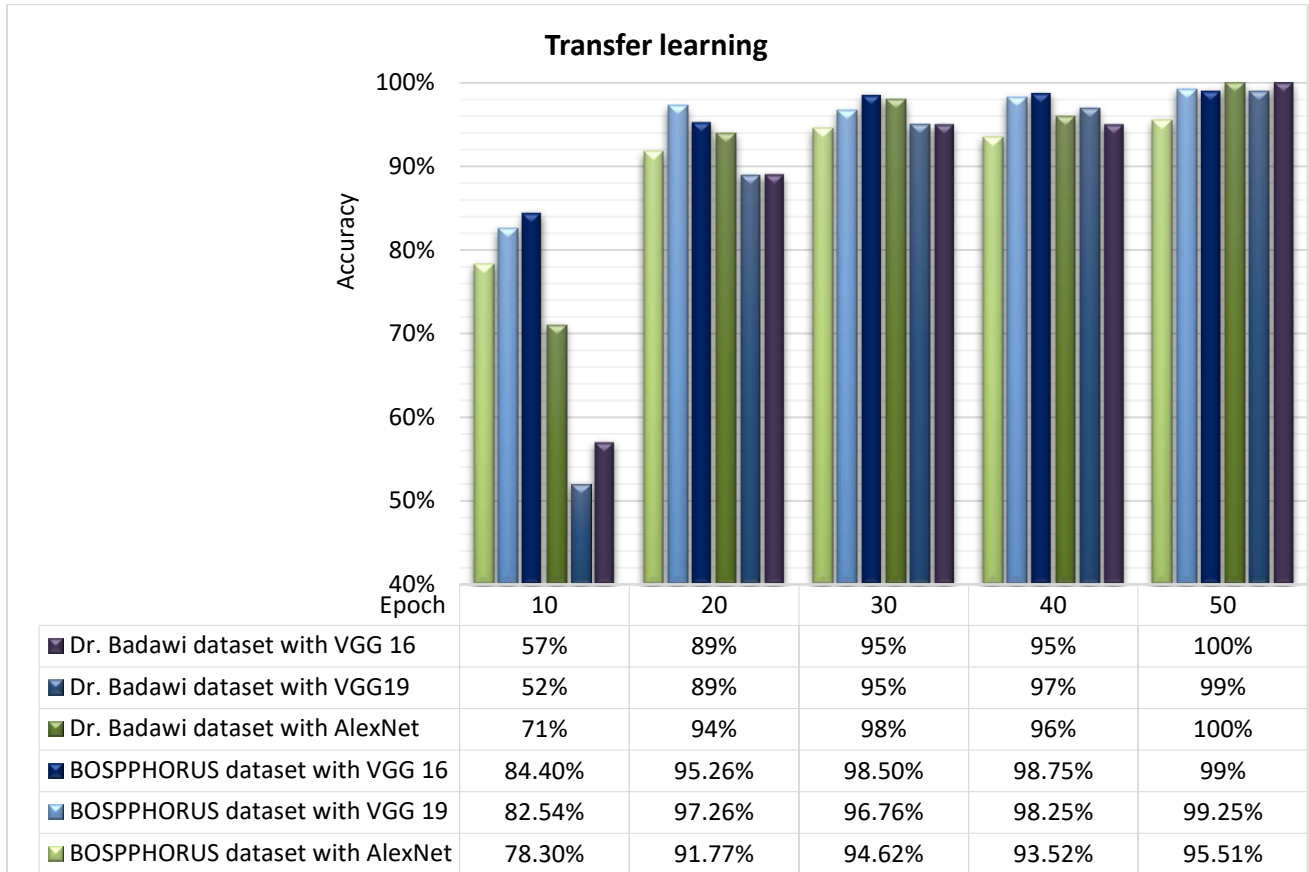It is shown from Table 1 that the k-NN with Dr.

Figure. 9 The recognition accuracy of transfer learning with Alex-Net, VGG16 and VGG19 models

| Epoch | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| ■ Dr. Badawi dataset with VGG 16 | 57% | 89% | 95% | 95% | 100% |
| ■ Dr. Badawi dataset with VGG19 | 52% | 89% | 95% | 97% | 99% |
| ■ Dr. Badawi dataset with AlexNet | 71% | 94% | 98% | 96% | 100% |
| ■ BOSPPHORUS dataset with VGG 16 | 84.40% | 95.26% | 98.50% | 98.75% | 99% |
| ■ BOSPPHORUS dataset with VGG 19 | 82.54% | 97.26% | 96.76% | 98.25% | 99.25% |
| ■ BOSPPHORUS dataset with AlexNet | 78.30% | 91.77% | 94.62% | 93.52% | 95.51% |

Badawi dataset with AlexNet model gives the best accuracy, ranged from 88.5 % to 99 %. Also, the k-NN classifier gives good accuracy when features extracted from 'fc6', ranged from 97.50 % to 99 %, and the split data first 80% training and the last 20% testing gives highest accuracy of about 99%.

While Table 1 shows that the k-NN with BOSPPHORUS dataset produce the best accuracy with VGG 16 model, ranged from 73.681 % to 90 %, and it is best when feature extracted from 'fc6', ranged from 90% to 83.048%, and the split data first 80% training and the last 20% testing gives highest accuracy of about 90%.

It is shown from Table 2 the SVM with Dr. Badawi dataset produce the good accuracy with VGG16 and VGG19 models, whereVGG16 accuracy ranged from 96 % to 100 % and VGG19 accuracy ranged from 95 % to 100%. Also, the SVM classifier gives good accuracy when feature extracted from 'fc6', with VGG16 ranged from 99.50 % to 100 % and with VGG19 ranged from 98 % to 100 %, and the split data first 80% training and the last 20% testing in two models gives highest accuracy of about 100%.

Also, Table 2 shows that SVM with BOSPPHORUS dataset produces the good accuracy in VGG 16 and VGG19 models, where VGG16

ranged from 90.581% to 97.50% and VGG19 ranged from 89.248% to 98.50%. And the SVM classifier gives good accuracy when feature extracted from 'fc6', with VGG16 ranged from 97.50 % to 96.60% and with VGG19 ranged from 98.50% to 94.514%, while VGG19 in the split data first 80% training and the last 20% testing gives higher accuracy of about 98.50%.

In conclusion, from results in Table 1 and Table 2, the SVM and K-NN classifier with AlexNet, VGG19 and VGG16 models gives best results when feature extracted from 'fc6', after that the recognition accuracy drops. Also, the ECOC with SVM classifier has the better classification results than ECOC with KNN and recognition accuracy in the model that use ECOC with SVM classifier give better result with VGG19 model when split data first 80% training and the last 20% testing.

**4.2.2. Second approach results: Transfer learning with convolutional neural networks**

In this approach, the input image is fed to the pre-trained CNN (AlexNet, VGG16 and VGG19) models as automatic feature extractor and classification. these models are trained to recognize 1000 classes but, in our research, we changed it to recognize 100 classes for each dataset. The pre-

trained AlexNet, VGG16 and VGG19 is fine-tuned via changing the number of outputs in the last fully connected layer 'fc8' depend on the number of our dataset classes so that the network changes its output to the new classification task. This approach need to determines an epoch number, the number of complete pass in the training algorithm over the whole training set [23], also it is needed to determines the size of a mini-batch, which means a subset of the training set that is used to evaluate the gradient of the loss function and update the weights [23].

In this paper we tried different numbers of epochs (10 ,20, 30 ,40 and 50) and mini-batch size of 10 with initial learning rate of 0.0001. Transfer learning is applied with Alex-Net, VGG16 and VGG19 models for both Dr. Badawi and BOSPPHORUS datasets. Fig. 9 shows the recognition accuracy of transfer learning in Alex-Net, VGG16 and VGG19 models.

It is shown from Fig. 9 that the recognition accuracy of BOSPHORUS dataset using (10, 20, 30 and 40) epochs is higher than Dr. Badawi dataset where BOSPHORUS dataset ranged from 78.30% to 98.75% and Dr. Badawi dataset ranged from 52% to 98% while in 50 epochs the Dr. Badawi dataset recognition accuracy is higher than that of BOSPHORUS dataset where Dr. Badawi dataset ranged from 99% to 100% and BOSPHORUS dataset ranged from 95.51% to 99.25%.

Also, Fig. 9 shows that Dr. Badawi dataset with VGG16 and AlexNet gives good accuracy reaches to 100% recognition rate in 50 epochs but, on the other hand, using smallest epoch number the AlexNet given best accuracy. While for BOSPHORUS dataset, VGG19 gives the good accuracy reaches to 99.25 % recognition rate in 50 epochs.

In general, as shown in Fig. 9, the 50 epochs are the best over other epoch number, as it ranged from 95.51% to 100 %. The 40 epochs are ranged from 93.52 % to 98.75%, the 30 epochs are ranged from 94.62 % to 98.50 %, the 20 epochs are ranged from 89 % to 97.26 % and the 10 epochs are ranged from 52 % to 84.40 %.

From results presented in Table 1, Table 2, and Fig. 9, it is clear that for Dr. Badawi dataset, using ECOC with SVM, ECOC with K-NN and transfer learning give the good accuracy which is 100% in ECOC with SVM and transfer learning where with SVM produce the 100% accuracy in VGG16 and VGG19 models and in transfer learning produce the 100% accuracy with VGG16 and AlexNet. In the other hand, BOSPHORUS dataset using ECOC with SVM, ECOC with K-NN and transfer learning given

the good accuracy which is 99.25 % in VGG16 transfer learning.

### 4.2.3. Comparison with the state of the art

We compared the proposed system with the dorsal hand vein recognition systems that displayed in literature, Table 3 shows this comparison. These systems use different feature extraction algorithms and classification algorithms. In [5], authors used LBP+BC, and Graph feature extraction algorithms and FGM classifier they get accuracy rate of 99.27%. On the other hand, in [6] authors used MDFR feature extraction algorithms and HD classifier that get accuracy rate of 98.83%. In [7], the authors used CDSDP feature extraction algorithms with ANN and FMNN classifier and got an accuracy rate of 98.8%. In [8] authors used endpoints and the intersections for feature extraction and Relative distances for classification and got accuracy rate of 98%. In [9] authors used minutiae features and the arc line and DPT for classification and got accuracy rate of 99.76%. In [10] the authors used LBP method and graph matching for classification and got accuracy rate of 96.67%. In [11] the authors use deep learning for feature extraction and logistic regression classifier and got accuracy rate of 99.33%. As well as in [11] images have been processed before the features extraction by extracting the region of interest (ROI) then applied histogram equalization and Gaussian smoothing filter. However, the proposed system uses original images without preprocessing then recognition is based on CNN depth models that select and express the depth feature of the image automatically.

From Table 3, it is clear that our proposed system, when using AlexNet with ECOC and K-NN classifier we got accuracy of 99%, while when using VGG16 and VGG19 with ECOC and SVM classifier we got accuracy of 100%. However, with VGG16 transfer learning give accuracy rate of about 100%.

Table 3 shows that the proposed system achieves the highest results in transfer learning that reaches to 100% recognition rate with Dr. Badawi dataset and BOSPPHORUS dataset reaches to 99.25 %. The proposed approaches give better results compared to the literature.

## 5. Conclusions

Advances in biometric systems have attained sufficiently high identification rates, but the need for reliability, robustness and convenience is still a main requirement that remains unsatisfied [1].

Table 3. Comparison of propose system with dorsal hand vein recognition system that displayed in section 2

| Author/ Ref # / Year | Preprocess approach | Feature extraction approach | Matching/ Classification approach | No. of subjects | System Results (Accuracy) |
|---|---|---|---|---|---|
| Huang et al. [5] 2016 | Gaussian filter, global thresholding, morphological opening and closing and thinning technique | LBP+BC, and Graph. | FGM | 102 subjects (NCUT) database | 99.27% |
| Lee et al. [6] 2014 | Gaussian filter | MDFR | HD | 214 persons | 98.83% |
| Trabelsi et al. [7] 2014 | Gaussian filter, | CDSDP | ANN+FMNN | 100 users BOSPHORUS database | 98.8% |
| Yun et al. [8] 2014 | dynamic threshold method | The endpoints and the intersections | Relative distances between the reference point, feature points and the angles among this points. | 20 persons | 98% |
| Chuang et al. [9] 2017 | Gaussian low-pass filter, local threshold through global reduction | Minutiae features and the arc line | DPT | 308 volunteers | 99.76% |
| Zhu et al. [10] 2012 | thinning method | LBP method | graph matching method | 102 subjects NCUT dataset | 96.67% |
| Wan et al. [11] 2017 | contrast limited adaptive histogram equalization and Gaussian smoothing filter | AlexNet | logistic regression | Part A from NCUT dataset with 1020 hand dorsal vein images for 102 subjects | 99.1% |
| | | CaffeNet | | | 99.33% |
| | | VGG-16 | | | 98.43% |
| | | VGG-19 | | | 98.7% |
| Proposed method | Without preprocess | VGG-16 | ECOC with K-NN | Dr. Badawi dataset 100 users with 500 image | 97 |
| | | VGG-19 | | | 96 |
| | | AlexNet | | | **99** |
| | | VGG-16 | | BOSPPHORUS dataset 100user with 1500 images | 90 |
| | | VGG-19 | | | 89.50 |
| | | AlexNet | | | 77 |
| | | VGG-16 | ECOC with SVM | Dr. Badawi dataset 100 users with 500 image | **100%** |
| | | VGG-19 | | | **100%** |
| | | AlexNet | | | 99 |
| | | VGG-16 | | BOSPPHORUS dataset 100user with 1500 images | 97.50 |
| | | VGG-19 | | | 98.50 |
| | | AlexNet | | | 93.50 |
| | | Transfer Learning with VGG-16 | | Dr. Badawi dataset 100 users with 500 image | **100%** |
| | | Transfer Learning with VGG-19 | | | 99 % |
| | | Transfer Learning with AlexNet | | | **100%** |
| | | Transfer Learning with VGG-16 | | BOSPPHORUS dataset 100 user with 1500 images | 99 % |
| | | Transfer Learning with VGG-19 | | | 99.25 % |
| | | Transfer Learning with AlexNet | | | 95.51% |

Identification depending on vein patterns appears as a good applicant where the character of vascular patterns in the dorsal area of the hand is claimed to be unique to a personal even for same twins and it performs to continue stable over lengthy periods of time [2].

This paper proposed a dorsal hand vein identification system that use two approaches. First approach depends on pertaining CNN (AlexNet, VGG16 and VGG19) models for extracting features, then using ECOC with SVM and K-NN for classification. This approach experiments showed that the recognition accuracy of all models gives best results when feature extracted from 'fc6'. Also, the accuracy rate of the models that use ECOC with SVM for classification is higher than the models using ECOC with KNN and the VGG19 model achieve better result in models that use ECOC with SVM.

The second approach is using transfer learning with CNN (AlexNet, VGG16 and VGG19) models for features extraction and classification. This approach experiments showed that the recognition accuracy for all models give best results when epoch number is 50 where Dr. Badawi dataset in VGG16 and AlexNet reaches to 100% recognition rate and BOSPPHORUS dataset reaches to 99.25% recognition rate in VGG19.

In conclusion, using transfer learning is giving more accuracy rate than using the pre-trained CNN models for extracting features only. Finally, we expect this work to be a beneficial initial point for new methods, and a common ground for an extensive range of benefits in the space of other biometrics and dorsal hand vein identification.

## References

[1] G. Neha and S. Tuly, "Palmprint Recognition: A Selected Review", *International Journal of Engineering Science*, Vol.6, No.6, pp 6776-6780, 2016.

[2] B. Sontakke, V. Humbe, and P. Yannawar, "Dorsal Hand Vein Authentication System: A Review", *International Journal of Scientific Research Engineering & Technology*, Vol. 6, No. 5, pp. 511-514, 2017.

[3] M. Shahin, A. Badawi, and M. Rasmy, "Multimodal biometric system based on near-infra-red dorsal hand geometry and fingerprints for single and whole hands", *World Academy of Science, Engineering and Technology*, Vol. 4, No. 4, pp. 268-283, 2010.

[4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition", In: *Proc. IEEE*, Vol. 86, No. 11, pp. 2278-2324, 1998.

[5] D.Huang, X. Zhu, Y. Wang, and D. Zhang, "Dorsal hand vein recognition via hierarchical combination of texture and shape clues", *Neurocomputing*, Vol. 214, pp. 815-828, 2016

[6] J. Lee, T. Lo, and C. Chang, "Dorsal hand vein recognition based on directional filter bank", *Signal Image Video Process*, Vol. 10, No. 1, pp. 145-152,2016.

[7] R. Trabelsi, A. Masmoudi, and D. Masmoudi, "Hand vein recognition system with circular difference and statistical directional patterns based on an artificial neural network", *Springer Multimedia Tools and Applications*, Vol. 75, No. 2, pp. 687-707, 2014.

[8] Y. Hu, Z. Wang, X. Yang, and Y. Xue, "Hand vein recognition based on the connection lines of reference point and feature point", *Infrared Physics & Technology*, Vol. 62, No. 1, pp. 110-114, 2013.

[9] S. Chuang, "Vein recognition based on minutiae features in the dorsal venous network of the hand", *Signal, Image and Video Processing*, Vol. 12, No. 3, pp1-9, 2018.

[10] X. Zhu and D. Huang, "Hand dorsal vein recognition based on hierarchically structured texture and geometry features", In: *Proc. Chin. Conf. Biometric Recognit.*, pp. 157-164, 2012.

[11] H. Wan, L. Chen, H. Song, and J. Yang, "Dorsal hand vein recognition based on convolutional neural networks", In: *Proc. IEEE International Conf. Bioinformatics Biomedicine*, pp. 1215-1221,2017.

[12] M. Alom, T. Taha, C. Yakopcic, S. Westberg, M. Hasan, B. Van Esesn, A. Awwal, and V. Asari, "The history began from alexnet: A comprehensive survey on deep learning approaches", *arXiv preprint arXiv:1803.01164*, 2018.

[13] M. Alaslani and L. Elrefaei, "Convolutional neural network based feature extraction for iris recognition", *International Journal of Computer Science & Information Technology,* Vol. 10, No. 2, 2018.

[14] S. Hijazi, R. Kumar, and C. Rowen, "Using convolutional neural networks for image recognition", *Cadence Design Systems Inc.: San Jose, CA, USA*, 2015.

[15] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks", In: *Proc. Adv. Neural Inf. Process. Syst.*, pp. 1097-1105, 2012.

[16] K. Simonyan and A. Zisserman, "*Very deep convolutional networks for large-scale image recognition"*, *Computer Science*, 2015.

[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, S. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions", In: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1-9, 2015.

[18] H. Joutsijoki, M. Henry, J.Rasku, K. Aalto-Setälä, and M. Juhola, "Error-correcting output codes in classification of human induced pluripotent stem cell colony images", *BioMed Research International*, pp. 1-13, 2016.

[19] M. Nasrabadi, "Pattern Recognition and Machine Learning", *Journal of Electronic Imaging*, Vol. 16, No. 4, pp. 140-155, 2006.

[20] The Bosphorus Database, "http://bosphorus.ee.boun.edu.tr/", Accessed on 1/8/2018.

[21] A. Badawi, "Hand vein biometric verification prototype: A testing performance and patterns similarity", In: *Proc. Int. Conf. Image Process. Comput. Vis. Pattern Recognit.*, pp. 3-9, 2006.

[22] M. Shahin, A. Badawi, and M. Kamel, "Biometric authentication using fast correlation of near infrared hand vein patterns", *International Journal of Biomedical Sciences*, Vol. 2, 2007.

[23] W. Kim, S. Jong, and P. Kang, "Multimodal Biometric Recognition Based on Convolutional Neural Network by the Fusion of Finger-Vein and Finger Shape Using Near-Infrared (NIR) Camera Sensor", *Sensors*, Vol. 18, No. 7, pp.1-34, 2018.