


Palmprint And Dorsal Hand Vein Multi-Modal Biometric Fusion Using Deep Learning

Norah Abdullah Al-johani, Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Lamiaa A. Elrefaei, Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia & Electrical Engineering Department, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt

 <https://orcid.org/0000-0001-5781-2251>

ABSTRACT

Advancements in biometrics have attained relatively high recognition rates. However, the need for a biometric system that is reliable, robust, and convenient remains. Systems that use palmprints (PP) for verification have a number of benefits including stable line features, reduced distortion and simple self-positioning. Dorsal hand veins (DHVs) are distinctive for every person, such that even identical twins have different DHVs. DHVs appear to maintain stability over time. In the past, different features algorithms were used to implement palmprint (PP) and dorsal hand vein (DHV) systems. Previous systems relied on handcrafted algorithms. The advancements of deep learning (DL) in the features learned by the convolutional neural network (CNN) has led to its application in PP and DHV recognition systems. In this article, a multimodal biometric system based on PP and DHV using (VGG16, VGG19 and AlexNet) CNN models is proposed. The proposed system is uses two approaches: feature level fusion (FLF) and Score level fusion (SLF). In the first approach, the features from PP and DHV are extracted with CNN models. These extracted features are then fused using serial or parallel fusion and used to train error-correcting output codes (ECOC) with a support vector machine (SVM) for classification. In the second approach, the fusion at score level is done with sum, max, and product methods by applying two strategies: Transfer learning that uses CNN models for features extraction and classification for PP and DHV, then score level fusion. For the second strategy, features are extracted with CNN models for PP and DHV and used to train ECOC with SVM for classification, then score level fusion. The system was tested using two DHV databases and one PP database. The multimodal system is tested two times by repeating PP database for each DHV database. The system achieved very high accuracy rate.

KEYWORDS

Biometric, Convolutional Neural Network (CNN), Dorsal Hand Vein (DHV), Error-Correcting Output Codes (ECOC), Identification, Palmprints (PP), Patterns Recognition, SVM

1. INTRODUCTION

Biometric systems have gained popularity for authentication purposes due to the dire need for protecting personal identity. These types of systems are deemed the best in terms of security (Orság & Drahanský, 2003). Their advancement in technology allows for unique and effective identification

DOI: 10.4018/IJAIML.2020070102

This article, originally published under IGI Global's copyright on June 12, 2020 will proceed with publication as an Open Access article starting on January 18, 2021 in the gold Open Access journal, International Journal of Artificial Intelligence and Machine Learning (converted to gold Open Access January 1, 2021), and will be distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

in automated systems; which is a great replacement to conventional approaches such as passwords. Users prefer biometrics, since they provide better security, as opposed to using password protection (Haghighat, Zonouz, & Abdel-Mottaleb, 2015).

Biometric systems rely either on physical (e.g., fingerprint, facial features, palm) or behavioral features (e.g., voice, movements, handwriting) (Bolle & Pankanti, 1998).

The palm print, fingerprint, hand vein, and ear canal are four biometric modalities that possess each of the following seven biometric characteristics: universality (i.e., it must be present in everyone), uniqueness (i.e., it must be distinct), permanence (i.e., it must be consistent over long periods of time), measurability (i.e., it must be measurable), performance (i.e., its efficiency in correct identification), acceptability (i.e., people need to be inclined to provide the feature), and circumvention (i.e., the ease to which it can be duplicated or imitated) (Bolle & Pankanti, 1998).

Palmprint recognition, which is the process of using PPs in the verification or identification process, has been the focus of numerous researches. Researchers have found that using the palm for identification purposes is more advantageous than using fingerprints or the iris (Jia, Hu, Lei, Zhao, & Gui, 2013; Ding, Zhuang, & Wang, 2005). Such advantages include stable line features, decreased distortion, that it is easier to manually position, and its ability to reach a higher recognition rate quicker (Jia, Hu, Lei, Zhao, & Gui, 2013). Hand veins are gaining popularity among recognition systems since their patterns possess the following four biometric characteristics: universality, uniqueness, permanence, and circumvention. The vein patterns are hard to copy, or forge given that they are beneath the skin and are generally not visible. The intricate hand-vascular patterns provide an effective feature set for identification (Sanchit, Ramalho, Correia, & Soares, 2011).

Biometrics are applied in several areas, the most prevalent of which are the following areas: logical/physical access control, tracking time/attendance, law enforcement, and surveillance (Orság & Drahanský, 2003). As shown in Figure 1, some applications that use the DHV and PP for recognition are ATM machines, bank transactions, computers, and entry systems (e.g., home, school, hospital, airport).

Several features from previous researches were hand crafted to handle occlusion, scale variation and lighting. Designing hand crafted features usually entails identifying the right compromise between accuracy and computational efficiency. The use of deep learning (DL) allows for complex networks to be created using convolutional neural networks (CNNs). The deeper layers within the network serve as feature extractors (LeCun, Bengio, & Hinton, 2015). This implies that DL is able to obtain a set of features just by observing the input images (Bora, Chowdhury, Mahanta, Kundu, & Das, 2016), which may have been pre-processed with pyramidal method (Han, Lei, & Chen, 2016). This method aims at identifying numerous representation levels so the semantics of the data may be represented by higher-level features. This subsequently may result in increased robustness to intraclass variance (Chan, et al., 2015). While feature extraction in CNNs requires DL, hand crafted features are designed in advance by humans for extracting a specific set of select characteristics; as shown in Figure 2.

Figure 1. Example technologies that use dorsal hand vein and palmprint patterns



When comparing biometric systems, those that rely on a single trait do not achieve adequate results in terms of recognition. On the other hand, those that rely on multiple traits are found to enhance the system's performance through error rate reduction and an improved recognition rate (Wild, Radu, Chen, & Ferryman, 2016).

The robustness and resistance to spoof attack increase with biometric systems that rely on multiple traits (Gopal, Srivastava, Bhardwaj, & Bhargava, 2016). Improving multimodal biometric system through fusion is done by four ways: sensor level fusion, feature level fusion (FLF), score level fusion (SLF) or decision level fusion (Byahatti & Hatture, 2017).

We noticed that many researchers have proposed different feature extraction and classification methods for PP and DHV recognition systems in the past years. Most of these features are hand-crafted features which often depend on expert knowledge, require expensive human labor and often do not generalize well. To overcome these issues, the researchers made much effort recently for feature-learning algorithms, especially the CCN, where the algorithm automatically gets the most excellent features of image fed to the CNN.

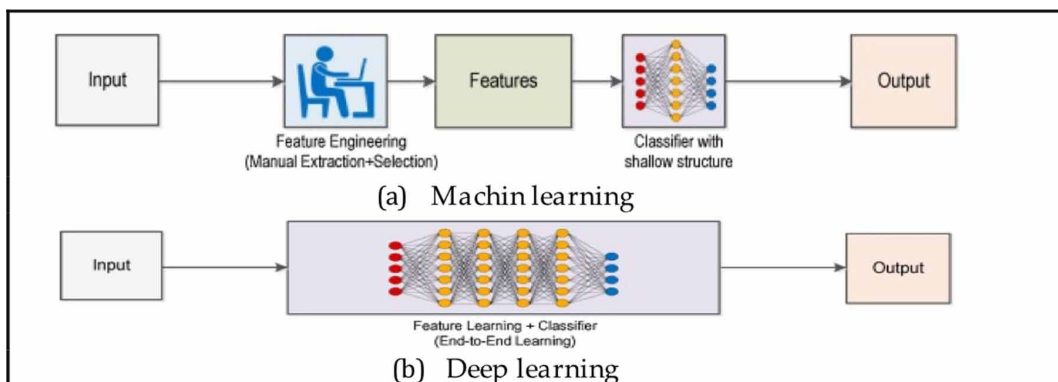
This paper goal to apply pre-trained CNN methodologies for fusing PP and DHV (DHV) using two approaches: feature level fusion with serial and parallel methods and second approach score level fusion by sum, max and product methods. Multimodal recognition accuracy is investigated when extracting features by the pre-trained CNN models (AlexNet, VGG16 and VGG19) then classification by Error-Correcting Output Codes (ECOC) with Support Vector machine (SVM), and when both feature extraction and classification are done by transfer learning (TL) with CNN models (AlexNet, VGG16 and VGG19). the proposed system is evaluated using two DHV databases and one PP database and achieved high recognition rates.

This paper is organized as follows: Section 2 introduces the literature review focusing on fusing PP and DHV with other biometrics. Section 3 explains methods and methodologies we used in the proposed approach. Section 4 explains the proposed multimodal biometric system approached based on PP and DHV. Section 5 provides the experimental results of the system with two stages: FLF experiments and SLF experiments. Also, it compares the outcomes with literature review studies. Section 6 concludes the paper.

2. RELATED WORK

Palm print recognition has drawn wide attention from many researchers (Jia, Hu, Lei, Zhao, & Gui, 2013; Li, Zhang, Zhang, & Yan, 2012; Palma, Montessoro, Giordano, & Blanchini, 2015; Kabir & Ahmad, 2017; Chaki, 2018; Dubey, Kanumuri, & Vyas, 2017) where recognition systems based on PP

Figure 2. Machin learning and deep learning framework (Schmidhuber, 2015)



using different features algorithms are presented. The important characteristic of hand vein patterns is stability, which means that the hand structure and DHVs configuration continue comparatively stable through the individual's life. For this reason, vein identification systems are considered as a promising and reliable biometric. Some of the vein identification systems are in (Huang, Zhu, Wang, & Zhang, 2016; Lee, Lo, & Chang, 2016; Trabelsi, Masmoudi, & Masmoudi, 2016; Hu, Wang, Yang, & Xue, 2014; Chuang, 2018;Zhu & Huang, 2012, Belean, Streza, Crisan, & Emerich, 2017; Wan, Chen, Song, & Yang, 2017)

Multimodal biometric system increases the robustness and improve the resistance to spoof attack (Gopal, Srivastava, Bhardwaj, & Bhargava, 2016). There are various studies in multimodal biometric fusion either PP or DHV biometric system with other biometric.

Shahin, Badawi, & Rasmy (2008) propose to fuse the hand's biometrics: ring fingerprints, DHV patterns and hand geometry shape at decision level. Sanchit, Ramalho, Correia, and Soares (2011) fused hand palm vein and DHV at SLF, they used a two-dimension Gabor filter Feature extraction. For matching, biometric models kept in the database are compared with the new picked up ones, at the registration stage. The metric used depends on the Hamming distance.

Wang, Yau, Suwandy, and Sung (2008) fused palm vein with PP at FLF. Their system fused the PP with palm vein images using edge-preserving and contrast-enhancing wavelet fusion process in which it joins the changed multiscale boundaries of the PP and palm vein images. Wang, Yau, and Suwandy (2008) fused PP with palm vein images at FLF. The system first extracted PP and palm vein then display the fused features by extracting junction points (JPs) that contains their location and direction.

Bharathi, Sudhakar, and Balas (2015) presented a biometric recognition system using finger vein, palm vein and DHV and extract feature by the Shearlet transform and Scale-invariant feature transform. Ben Khalifa, Gazzah, and BenAmara (2013) studied the performance of many normalization techniques of the face, PP and fingerprint using several fusion methods. Nageshkumar, Mahesh, and Swamy (2009) designed a system for fusing face and PP at SLF. Also, Nousath, Imran, Jetly, Rao, and Kumar (2013) fused face and PP biometric at each level: sensor, feature, decision and score level. Feature are extracted by linear discriminant analysis (LDA) and Local phase quantization (LPQ), respectively.

Gopal, Srivastava, Bhardwaj, & Bhargava (2016) presented a multimodal individual verification system fusing PP, palm-phalanges print (PPP) and DHV at SLF of PP-PPP, PP-DHV, PPP-DHV, and PP-PPP-DHV approaches. Matching is done by using K-nearest neighbor, SVM and random forest (RF).

From reviewed literature we noticed that: Multimodal biometric systems increase the robustness of the system. Many researchers have proposed different feature extraction algorithms for PP and DHV recognition systems in the past years. Most of these features are hand-crafted features which often depend on expert knowledge, require expensive human labor and often do not generalize well. Also, there is a lack of researches that fuse PP and DHV by deep learning.

3. METHODS AND METHODOLOGIES

This section describes methods and methodologies we use in the proposed system: CNN basic architecture (section 3.1), the used CNN models (section 3.2), Different ways CNN is used for biometric systems (section 3.3), Multimodal Biometric fusion levels applied in the proposed system (Section 3.4).

3.1. Convolutional Neural Network (CNN) Basic Architecture

Deep Learning (DL) is a new category of machine learning approaches that has recently emerged and attracted the attention of both researchers and the industry (Alom, et al., 2018; Al-johania & Elrefaei, 2019). One the main DL approaches is the CNN. It was initially proposed by Fukushima in

1988. CNNs are feedforward networks that are able to extract topological properties able to withstand distortion, translation, scaling, rotation, and squeezing. CNNs are generally trained similar to standard neural networks, using a backpropagation system. The main architecture of a CNN consists of three major layers:

- **Convolutional Layer:** It deals with the input image and extracts various features by using filters. The filter is applied to the image and then gets a specific map, the result in this case would be multiple feature maps. The features change depending on the filter that is applied. The main convolution layer gets low-level features like boundaries, lines, and angles. The last layers get higher-level features.
- **Pooling layer:** usually following the convolutional layer. This layer focuses on reducing the image size which is completely independent from respectively depth dimension where the depth of the image stays the same. The pooling layer reduces the features resolution. It creates features that are robust to distortion and noise. The pooling layer uses a sliding window that is shifted in stride across the input transforming the values into representative values (Hijazi, Kumar, & Rowen, 2015).
- **Activation functions:** The input is set in the range of values by this layer where the outcome of the precedent layer must be in the range of the values. That makes the input and production sizes of this layer similar. Sigmoid is the usual applied activation functions, Tanh and ReLU. In this research we used ReLU because the network trains faster by Sigmoid, Tanh and ReLU is defined in (Krizhevsky, Sutskever, & Hinton, 2012).

Fully-connected layer: previous layer output is fully connected with this layer. It is the final stages of the CNN connect to the output layer (Bengio, 2009).

3.2. The Used CNN Models

The main stage in models that are CNN-based is feature extraction and classification. One of the benefits of DL models is their ability to automatically learn hierarchical feature representations. In these types of models, features that are learned on the lower levels are general and may be used in more than one domain; features that are learned on the outer layers are more fine-tuned and rely on the chosen dataset and task (Yosinski, Clune, Bengio, & Lipson, 2014). There are numerous architectures in the field of CNN model. The most common models are: AlexNet, VGG16 and VGG19 that are used respectively in (Krizhevsky, Sutskever, & Hinton, 2012; Simonyan & Zisserman, 2014; Szegedy et al., 2015).

The structure of AlexNet and the number of filters and its size are presented in detail in (Al-johania & Elrefaei, 2019). 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 $11 \times 11 \times 3$ 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 $5 \times 5 \times 48$ 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 $3 \times 3 \times 256$ in the third layer and $3 \times 3 \times 192$ in the fourth layer. The fifth layer implement convolution, ReLU and maxpooling where convolution uses 256 different filters of size $3 \times 3 \times 192$ with stride size of 1 (Krizhevsky, Sutskever, & Hinton, 2012). Then used three fully connected (FC) layers and dropout, at the end applied a Softmax layer.

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 VGG16 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 VGG19 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 (Simonyan & Zisserman, 2014; Szegedy, et al., 2015). The structures of VGG16 and VGG19 models and the number of filters and its size are in presented in detail in (Al-johania & Elrefaei, 2019).

The main difference between AlexNet and VGGNet is the architecture depth. Additionally, VGGNet assigns very small 3×3 convolutional kernels against the 7×7 (Conv 1), 5×5 (Conv 2), and 3×3 (others) kernels in AlexNet (Hertel, Barth, Käster, & Martinetz, 2015).

3.3. CNN For Biometrics System

Biometric systems use the CNN pre-trained model using three steps. The first step consists of selecting the most suitable pre-trained model. Then, the problem is classified based on the size similarity matrix (Marcelino, 2019), as shown in Figure 3 (a). Problems are classified using the matrix based on the dataset's size and its resemblance to the dataset from the pre-trained model. The final step consists of fine-tuning the model. This involves using the matrix to repurpose the pre-trained model (Marcelino, 2019). Figure 3 (b) presents a summary of how to repurpose the pre-trained model.

The first quadrant contains a large dataset that differs from that of the pre-trained model. In such cases, the pre-trained model's architecture is used and trained to suit the dataset. Given the dataset's size, the model may be trained from scratch.

The second quadrant contains a large dataset which resembles that of the pre-trained model. In such cases, not every layer is trained. The amount that is trained depends on the dataset's size and total amount of parameters used. It is important to note that frozen layers may not be trained later.

The third quadrant contains a small dataset which differs from that of the pre-trained model. In such cases, not every layer is trained. However, determining the number of layers to be trained is difficult.

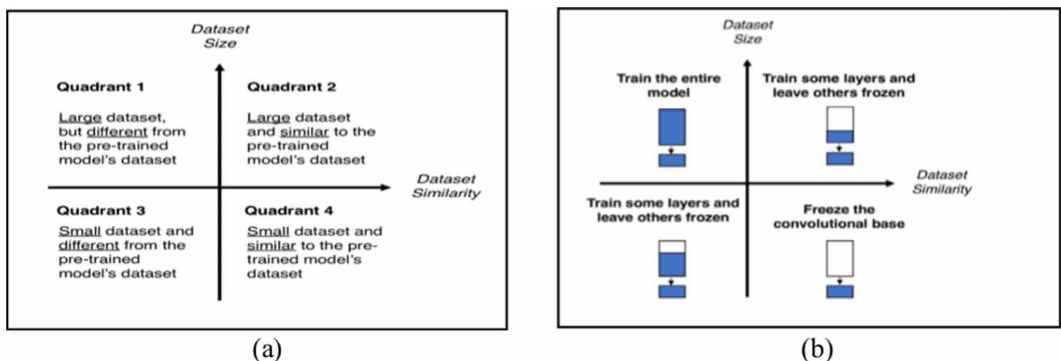
The fourth quadrant contains a small dataset which resembles that of the pre-trained model. In such cases, the original convolutional base is kept, and its output is fed to the classifier. This process entails the following: removing the last output layer, using the pre-trained model as a fixed-feature extractor and finally, using the extracted features in training a new classifier.

In this paper we choose the pre-trained model in quadrant 4 that keeps the convolutional base without change then uses its outputs as input for the classifier. Pre-trained model is used by two approaches: transfer learning with CNN and feature extraction with CNN.

3.3.1. Transfer Learning (TL) With CNN

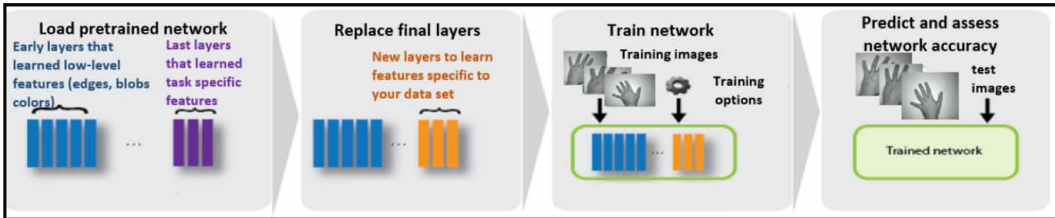
Transfer learning is another machine learning approach. In this method, knowledge that was learned when solving one problem is applied to another. As such, this method utilizes previous models that

Figure 3. (a) Size-Similarity Matrix (b) Summary of how select repurposing a pre-trained model (Marcelino, 2019)



were developed. In these ‘pre-trained’ models, a large dataset is used to train the model in solving a problem similar to the problem at hand. To repurpose pre-trained models, the original classifier is

Figure 4. Fine-tuning with CNN



removed, a new classifier that suits the needs of the system is added and then the model is fine-tuned. A fine-tuned system is shown in Figure 4.

The most recent layers of the pre-trained network are constructed for 1000 classes these layers fine-tuned for the new classification problem. The classification by scientifically adding the weighting of the prior layer of features to decide an exact goal output result is the major function of this layer (Huang, Pan, & Lei, 2017).

When transferring the layers to the new classification problem, only the last three layers are not extracted. Instead, they are replaced with the following three layers: fully connected, a Softmax, and classification. The fully-connected layer’s options are specified based on the new dataset, and its size is set to correspond to that of the new dataset’s class size.

A Softmax layer get the output of Fully Connected (fc) layers then applies a Softmax function, equation (1), to the input (Bishop & Christopher, 2006):

$$p(y = j | z^{(i)}) = \frac{e^{z_j^{(i)}}}{\sum_{k=0}^k e^{z_k^{(i)}}} \quad (1)$$

where define the z as equation (2),

$$z = \sum_{j=1}^m w_j x_j + b \quad (2)$$

Where w is the weight vector, x is the feature vector of 1 training sample, and b is the bias unit.

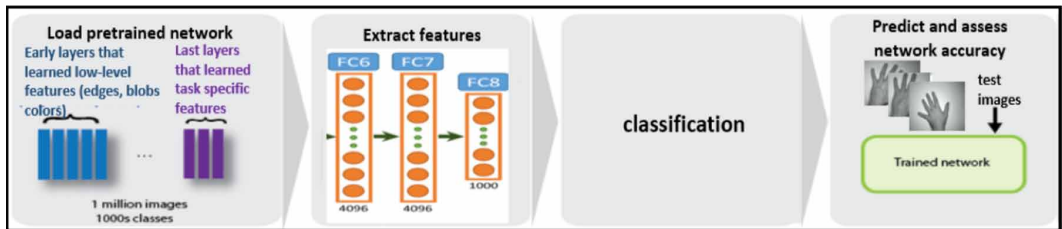
Now, the probability is calculated by this Softmax function that training sample $x^{(i)}$ be appropriate to class j given the weight and net input $z^{(i)}$. So, the probability is computed by $P(y = j | z^{(i)})$ for each class label $j = 1, \dots, k$. The Softmax function also known as the normalized exponential can be considered the multi-class generalization of the logistic sigmoid function (Bishop & Christopher, 2006).

3.3.2. Feature Extraction With CNN

In this method, learned image features are extracted from pre-trained CNNs and used in training the image classifier, as shown in figure 5. A hierarchical representation is constructed using the input images. Lower layers consist of high-level features that are made using low-level features of previous layers. While the ‘fc7’ layer is used to obtain feature representations for the training and test images, earlier layers are used to obtain low-level representations. Different classification approaches are able to use ECOC because it is not confined to one approach.

In this paper, we restricted our focus to ECOC uses in SVM models. In a SVM model with the following given data: a training dataset of $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ which needs to be classified into two, linearly separable, classes where length d 's feature vector is $(x_i \in R^d)$, the label class is $[y_i$

Figure 5. Feature extraction with CNN



$\in (-1, +1]$ and the hyperplane is $(wx + b = 0)$; without any other prior knowledge about the dataset, the model is able to obtain the hyperplane with the most margin (Zhang, Gao, Chen, Shan, & Zhao, 2006).

In the ECOC algorithm, when the coding matrix (CM) is designed, the binary classifiers and classes are represented by the columns and rows, respectively. Either binary or ternary coding may be used to design the CM. With binary coding, the elements of the CM are $[CM_{ij} \in \{-1, 1\}]$ or $[CM_{ij} \in \{0, 1\}]$, where $(i = 1, 2, \dots, M)$ and $(j = 1, 2, \dots, N)$. M and N are the number of classes and binary classifiers, respectively. To represent which classes are grouped together in every classifier, values are divided in the columns into $(-1$ and $1)$ or $(1$ and $0)$, respectively. With ternary coding, the elements of the CM belong to the set $\{-1, 0, 1\}$; -1 and 1 signify the same as they do in binary coding and 0 signifies that the class is not included when training the individual binary classifier (Joutsijoki, Haponen, Rasku, Aalto-Setälä, & Juhola, 2016).

After creating and train multi-binary classifiers then the data divided into two super groups based on zeros and ones in respective columns, there are numerous strategies of how to construct the coding matrix. We chose one-vs-all (OVA) strategies (Liu & Wechsler, 2001) which uses multi-binary coding and all classes with one are assembled in the first group, and the rest of classes are collected in the second group.

Once the CM is built and the individual classifiers have finished training, test examples need to be classified. Every classifier is given every test sample to predict its output. After output collection is completed, a ‘codeword’ is constructed for the text example. The decoding phase then begins. This phase searches the codewords in the classes to determine which is closer to the one from the test example.

3.4. Multimodal Biometric Fusion

Combining data from individual modalities requires an efficient fusion structure. In biometric fusion, where biometric features are combined, the system's accuracy, robustness and efficiency may be increased. This research utilizes two fusion levels: the feature and scoring levels.

3.4.1. Feature Level Fusion (FLF)

In this level of fusion, the features are extracted from each biometric and used to calculate a multimodal feature vector. For feature fusion some methods have been proposed (Yang & Yang, 2002) (Yang, Yang, Zhang, & Lu, 2003) (Mi, Wang, & Qi, 2016). We used two methods for fusing the different types of features.

Serial feature fusion strategy: the fused features dimension is equivalent to the summation of the two sets of features dimensions. Feature vectors of the first biometric, $U = \{u'_1, u'_2, \dots, u'_m\}$ and feature vectors of the second biometric $V = \{v'_1, v'_2, \dots, v'_n\}$ the new feature vector, $UV = \{u'_1, u'_2, \dots, u'_m, v'_1, v'_2, \dots, v'_n\}$, $UV \in R^{m+n}$ is created. The purpose is to merge these two features sets to create a new feature vector.

Parallel feature fusion strategy: is to combine the two sets of features represented by F1 and F2. The final fused feature representation is formulated as in equation (3),

$$F_f(I) = F_1(I) + iF_2(I) \quad (3)$$

where i the weight unit.

3.4.2. Score Level Fusion (SLF)

The involvement of fusion at the score level is the second approach. For each biometric, the user is validated and a matching score indicating the feature vector proximity is calculated with the trained model. These scores are fused to verify the identity.

SLF consists of two main steps: score normalization and fusion. In the first step, a comparator (Si) calculates scores. The scores are then mapped to another score scale, (domain Si'). For example, if comparator (X) was to produce scores on the domain [1;200], and comparator (Y) was to generate scores on the domain [1;2500], then, normalization and mapping of both scores to a common scale is required (Heenaye & Khan, 2012).

Score normalization is conducted so that the comparator's parameters and data types are transformed to a common domain. Min-max normalization (MM) methods were used in the proposed system. MM linearly transforms the original data. It is a simple method that is suitable to use when the generated score limits are given. While it is efficient and performs adequately, its generated results may not be accurate if there are outliers in the data. It maps raw scores to the [0,1] range and uses the maximum (\max_s) and minimum (\min_s) score to assign the score range's endpoints (Heenaye & Khan, 2012). Given a set (S) which contains the scores for the matcher, (s) and (s') are denoted as the raw matching score and the normalized score which corresponds to it, respectively as in equation (4).

$$s' = \frac{s - \min_s}{\max_s - \min_s} \quad (4)$$

Score Fusion Techniques

There are two types of score fusion methods: classification and combination. While the first type does not require normalization since scores may be heterogenous, it is required in the second type before score fusion occurs. When using the classification methods, for the problem to be formulated, the decision space is divided into an authentic class and an imposter class. The extent to which this method is reliable and effective base on the size and quality of the dataset used in training the classifier. Classification approaches that have been studied include neural networks, nearest neighborhood algorithms and tree-based classifiers (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014).

The second score fusion method, combination, is not only more frequently used than the first method, but it also effectively combines biometric scores. Scores from multiple comparators are combined in order to generate one score. The following three SLF methods are analyzed in this research: maximum score, simple sum and weighted product.

- **Maximum Score:** In this technique, the posteriori probabilities' mean is estimated by the largest value, equation (5).

$$\text{maximum score} = \max(i=1\text{to}N)s' \quad (5)$$

- **Simple Sum:** This technique reflects the raw scores' weighted average. It sums the scores from comparisons without normalization. This technique is simple and operates under the assumption that raw scores inputted using biometric approaches have a scale, distribution and strength that are able to be compared. This technique is applicable in classifications with data ambiguity due to high noise levels, equation (6).

$$\text{simple sum} = \sum(i=1\text{to}N)s' \quad (6)$$

- **Weighted Product:** In this technique, scores are combined by utilizing the weighted multiplication of all modality scores, equation (7).

$$\text{weighted product} = \prod(i = 1 \text{ to } N)s' \quad (7)$$

The Proposed Multimodal Biometric Fusion of Dorsal Hand Vein (DHV) and Palmprint (PP)

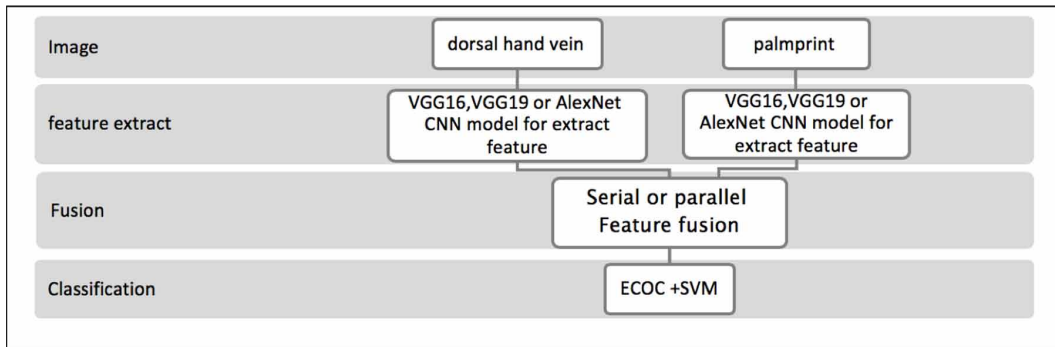
In this paper, the proposed multi-modal biometric system is based on PP and DHV that use two approaches for fusion:

4.1. First Approach: Fusing Dorsal Hand Vein (DHV) and Palmprint (PP) at Feature Level

In this approach, we applied one strategy where PP and DHV are fused at feature level by using CNN for features extraction and train ECOC with SVM classifier for classification, Figure 6 and Figure 7. Features are Extracted using the pre-trained convolutional neural networks (VGG16, VGG19 and AlexNet) and train ECOC with SVM classifier for classification. Fusion at feature level is investigated by serial or parallel techniques. We analyzed four methods in this approach:

1. VGG16 for extracting features from PP and DHV.

Figure 6. Fusion palmprint and dorsal hand vein at feature level



2. VGG19 for extracting features from PP and DHV.
3. AlexNet for extracting features from PP and DHV.
4. VGG16, VGG19 and AlexNet for extracting features from PP and DHV, as shown in Figure 7.

4.2. Second Approach: Fusing Palmprint (PP) And Dorsal Hand Vein(DHV) at Score Level

In this approach, we applied two strategies where PP and DHV are fused at score level by using CNN for feature extraction and train Softmax layer or ECOC with SVM classifier for classification (Figure 8):

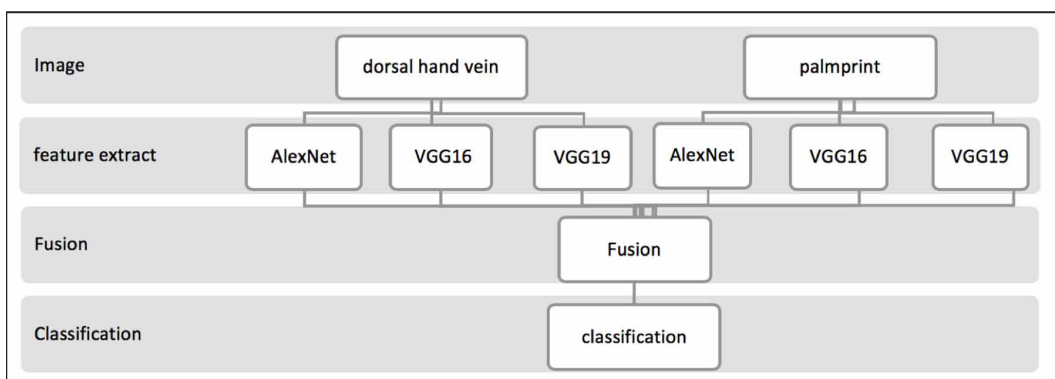
Strategy1: Transfer Learning with CNNs where we fine-tuned a pre-trained convolutional neural network (VGG16, VGG19 and AlexNet) for features extraction and classification then fusing with Sum, max and product methods.

Strategy2: use the pre-trained CNN (VGG16, VGG19 and AlexNet) for features extraction and train ECOC with SVM classifier for classification then fusing at score level by Sum, max and product methods.

Features are extracted in this approach for each strategy using the three methods:

1. VGG16 for extracting features from PP and DHV.

Figure 7. VGG16, VGG19 and AlexNet for extract feature



2. VGG19 for extracting features from PP and DHV.
3. AlexNet for extracting features from PP and DHV.

There are number of techniques used in the proposed system to avoid overfitting in the learning process and increase the generalization ability of the neural network. These techniques are: dropout method (Al-Waisy, Qahwaji, Ipson, & Al-Fahdawi, 2018), data augmentation and The ReLU (Krizhevsky, Sutskever, & Hinton, 2012).

5. THE EXPERIMENTAL RESULTS

This section presents a detailed description of the experimental results. The experimental setup is presented in Section 5.1. In Section 5.2. we present the datasets. In Section 5.3. the results of the experiments in the two approaches: Fusion at feature level and Fusion at score level is illustrated.

Section 5.4 shows a comparison between the two approaches results. Then, we provide in Section 5.5 the comparison between the performance of the proposed multimodal biometric fusion of PP and DHV with the state-of-the-art.

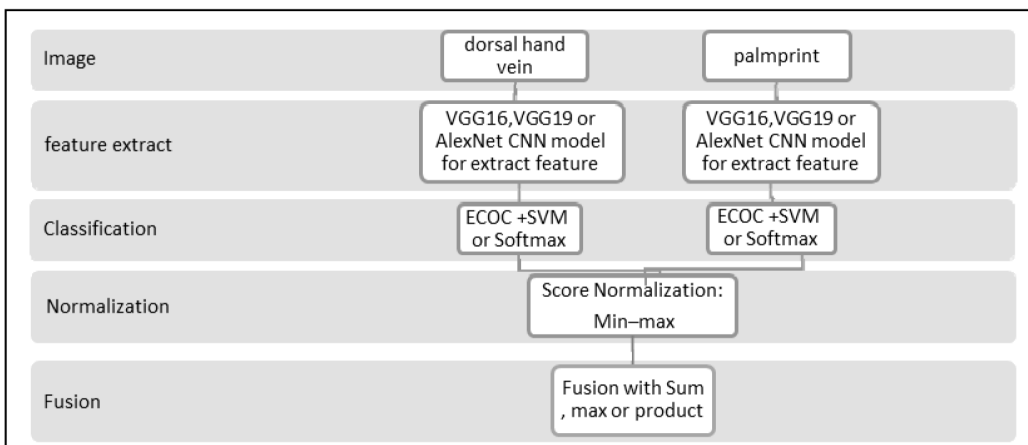
5.1. The Experimental Setup

The system and its stages are implemented using the desktop computer environment (Intel(R) Core (MT) i7-7700 HQ CPU @ 2.80 GHz, windows10 64-bit operating system, RAM 16 GB, 256GB SSD, NVIDIA GTX 1060 graphics card, our proposed system was designed and programmed using MATLAB R2018a with Computer Vision and deep learning toolboxes.

During this experiment, dataset was randomly divided into two parts: 70% for training and the remaining 30% for testing. The AlexNet, VGG16 and VGG19 CNN networks used in this system 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.

The proposed system performance is calculated by the recognition accuracy rate using equation (8),

Figure 8. Fusion palmprint and dorsal hand vein at score level



$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \times 100 \quad (8)$$

5.2. Datasets

The proposed architecture is tested on three datasets, two for DHV and one dataset for PP. Additional augmentation operations is performed on the training images: randomly translate the images up to 30 pixels horizontally and vertically. Random rotation in range from 35 to -35. Data augmentation avoid overfitting and remembering the exact details of the training images. Dataset is described in following sub-section:

- **DHV Datasets:** The 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” (Shahin, Badawi, & Rasmy, 2011). The dataset contains images for 50 different persons with 10 images per everyone, 5 images for left hand and 5 images for right hand, mirrored to have the same 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, housewives, etc.), Figure 9 part (a) displays some samples from this database (Shahin, Badawi, & Rasmy, 2011).

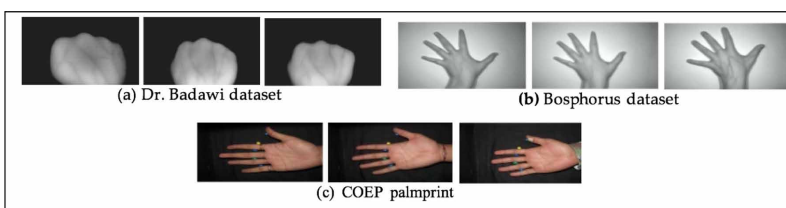
The second DHV dataset is Bosphorus of left hand that contains 1575 images that taken form 100 persons. Also, this dataset contains 219 images for 25 users obtained after two or five months. In addition, three positions for each customer are use in regular and specific situations (after having carried a bag weighing 3 kg. for one minute and after the hand cooled) (The Bosphorus Hand Database, 2010), Figure 9 part (b) displays some samples from this database.

- **PP Dataset:** We use COEP palmprint database which consists of 8 various palm images of each person. The database involves 1344 images pertaining to 168 people. The database was collected in one year. The images were taken by digital camera and the resolution is 1600 × 1200 pixels (Autonomous Institute of Government of Maharashtra, 2018). Figure 9 part (c) displays some samples from this database.

5.3. Data Augmentation

It is well known that to achieve satisfactory prediction DNNs need to be trained on a large number of training samples and prevent overfitting (Krizhevsky, Sutskever, & Hinton, 2012). Data augmentation is a simple and commonly used method to artificially enlarge the dataset by methods such as random crops, intensity variations, horizontal flipping.

Figure 9. Some samples from hand veins datasets (Autonomous Institute of Government of Maharashtra, 2018) (The Bosphorus Hand Database, 2010)



In this work, the CNN networks require input images of size 227-by-227-by-3, but the images in the image datastores have different sizes. Use an augmented image datastores to automatically resize the training images. Specify additional augmentation operations to perform on the training images: randomly translate the images up to 30 pixels horizontally and vertically. Random rotation in range from 35 to -35. Data augmentation helps prevent the network from overfitting and memorizing the exact details of the training images.

5.4. Experiments and Results

This section displays the result of our experiments and the accuracy of the proposed multimodal in FLF and SLF.

5.4.1. First Approach: Feature Level Fusion (FLF) Experiment

The multimodal system was tested two times by repeating the PP database for each DHV system database: Dr. Badawi with COEP palmprint and Bosphorus with COEP palmprint. The fusion between DHV dataset and COEP palmprint applied in FLF by serial and parallel methods.

5.4.1.1. The Fusing Dorsal Hand Vein (DHV) and Palmprint (PP) Datasets at Feature Level by Serial Method

In this part, the feature is extracted by four methods. The first three methods are extracting features from the models (AlexNet, VGG16 or VGG19) for each dataset then the result of the two datasets are fused using serial fusion. The fourth method extracting features from all model (AlexNet, VGG16, and VGG19) for each dataset then the fusion result of the six features. Then classification by ECOC with SVM.

Result of fusing Dr. Badawi dataset with COEP dataset by the first three methods is shown in Figure 10 part (a). Experiment shows the result of fusing Bosphorus dataset with COEP dataset by the first three methods is shown in Figure 10 part (b).

As can be shown from Figure 10, fusing Dr. Badawi dataset with COEP dataset achieved range from 100% to 96.33%. While Bosphorus dataset with COEP dataset achieved accuracy range from 100% to 98%. Overall, we note that the highest accuracy when fusing Dr. Badawi dataset with COEP dataset and Bosphorus dataset with COEP dataset is achieved in VGG16 model of 100%.

Experiment shows the result of fusing Dr. Badawi dataset with COEP dataset and Bosphorus with COEP dataset by feature extracted from all models (AlexNet, VGG16 and VGG19) are shown in Figure 11, the result achieved accuracy results range from 100% to 99.67%. We note this approach get accuracy about 100% in the most result and 'fc6', 'fc7' layers in this approach get high accuracy.

5.4.1.2. The Fusing Dorsal Hand Vein (DHV) and Palmprint (PP) at Feature Level by Parallel Method

The experiment result of fusing Dr. Badawi dataset with COEP dataset by parallel method is shown in Figure 12 (a). Experiment shows the result of fusing Bosphorus with COEP dataset by parallel fusion is shown in Figure 12 (b). As can be seen from Figure 12, fusing Dr. Badawi dataset with COEP dataset achieved accuracy range from 100% to 99%. While fusing Bosphorus dataset with COEP dataset achieved accuracy range from 100% to 96.67%. Overall, we note that the highest accuracy when fusing Dr. Badawi dataset with COEP dataset and Bosphorus dataset with COEP dataset is achieved using VGG16 model.

In FLF approach, we note that when feature extracted from one model (AlexNet, VGG16 or VGG19) for each dataset then the result of two datasets is serial fused, achieved highest accuracy when fusing Dr. Badawi dataset with COEP dataset of 100% in VGG16 while the highest accuracy when fusing Bosphorus dataset with COEP is about 99.67% in VGG16. When feature extracted from all models (AlexNet, VGG16 and VGG19) for each dataset then the result of two datasets is serial fused, we get accuracy of 100% in most of the results. The feature extraction from one model (AlexNet, VGG16 or VGG19) for each dataset then the result is parallel fused achieved highest accuracy when fusing Dr. Badawi dataset with COEP dataset of about 100% in VGG16 while highest accuracy when fusing Bosphorus dataset with COEP is about 100% in VGG16.

Figure 10. Experiment result of serial method

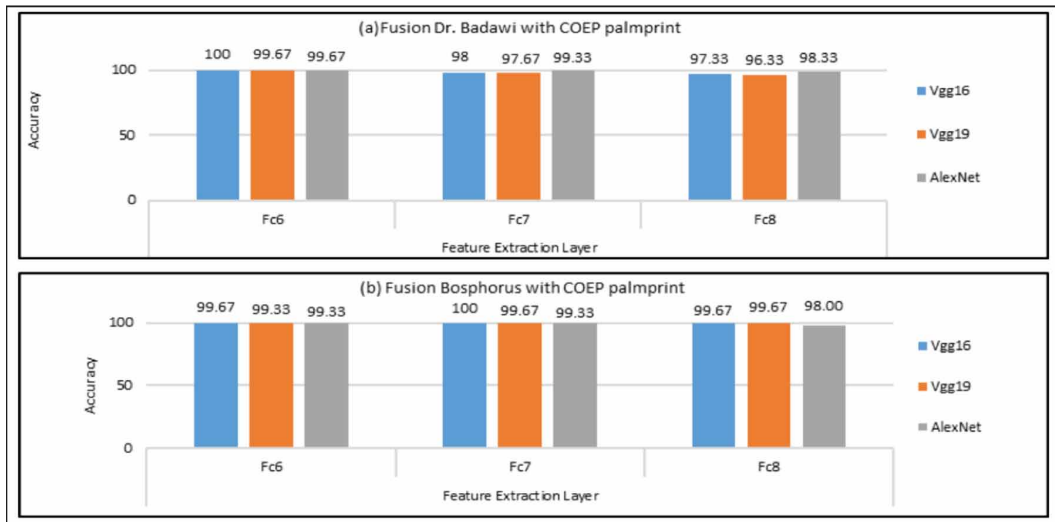
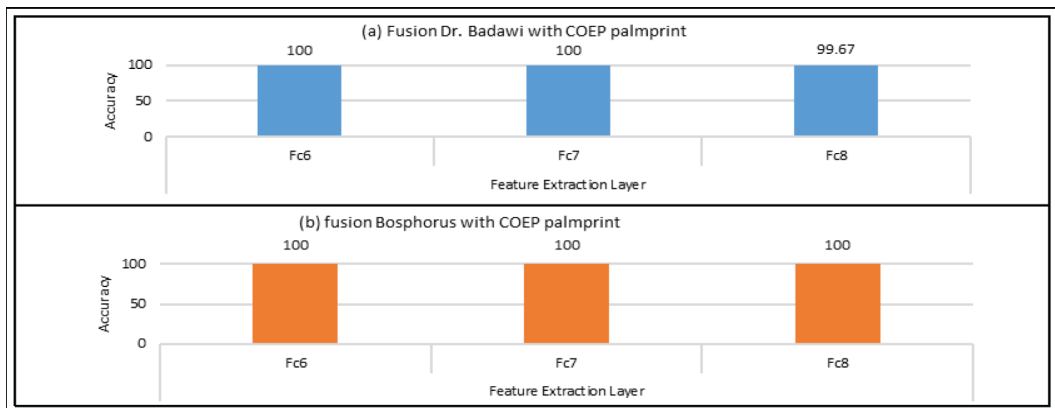


Figure 11. Experiment result of serial method by all model (AlexNet, VGG16 and VGG19)



5.4.2. Second Approach: Score Level Fusion (SLF) Experiment

Again, the multimodal system was tested two times by repeating PP database for each DHV system database: Dr. Badawi dataset with COEP and Bosphorus dataset with COEP. The Fusion of PP and DHV at SLF is done by applying two strategies:

5.4.2.1. Strategy 1: Transfer Learning With CNN

This strategy fine-tuning a pre-trained convolutional neural network (VGG16, VGG19 and AlexNet) for features extraction and classification then fusion with Sum, max and product methods. We keep the features from the early layers of the pretrained network and we tried numbers of epochs of 30 and mini-batch size of 10 with an initial learning rate of 0.0001.

Figure 12. Experiment result of parallel feature level fusion

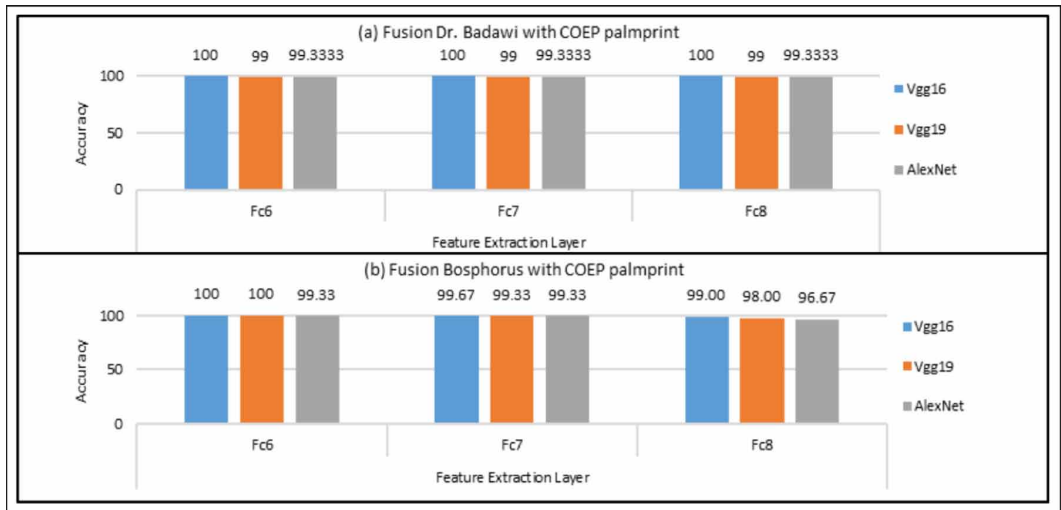


Figure 13. Experiment result fusion by score level with transfer learning

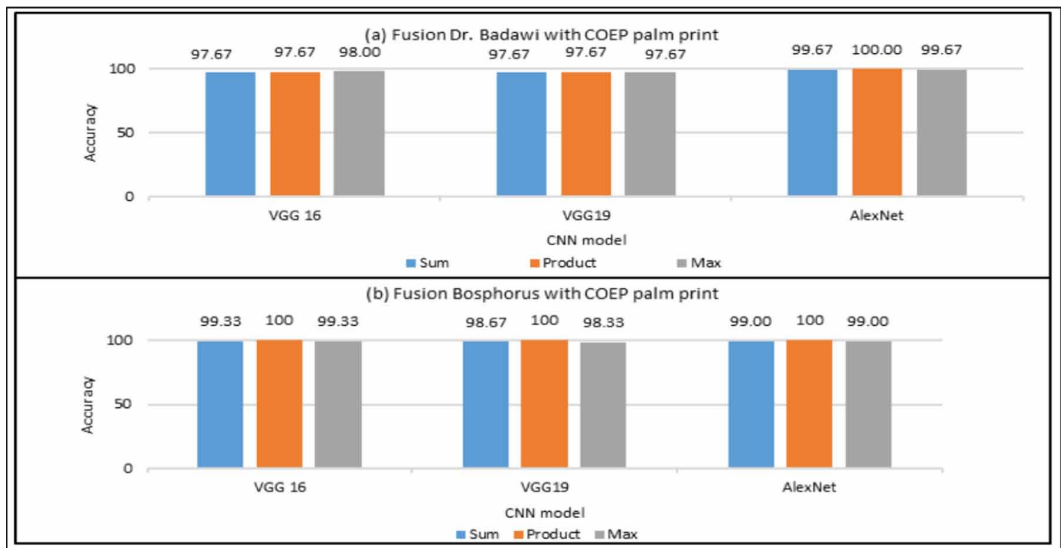
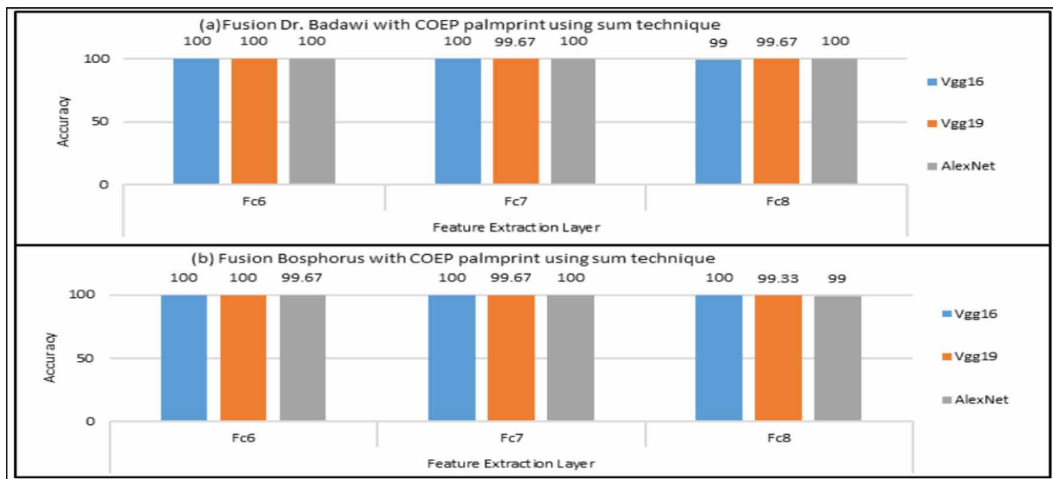


Figure 13 shows experimental result of SLF for the Bosphorus dataset with COEP dataset and Dr. Badawi dataset with COEP dataset where features for each dataset are extracted and classification result is normalized by the Min-max normalization techniques, then fused by the max, sum, or product.

As can be seen from Figure 13 (a), the sum technique achieved accuracy range from 99.67% to 97.67% as well as the highest accuracy in AlexNet model. While product technique achieved accuracy range from 100% to 97.67% as well as the highest accuracy in AlexNet model. and with max technique accuracy ranges from 99.67% to 97.67% as well as the highest accuracy in AlexNet model. Overall, the best results in AlexNet model.

Figure 14. Experiment fusion by SLF using sum technique



As can be seen from Figure 13 (b), the sum technique achieved accuracy range from 99.33% to 98.67% as well as the highest accuracy in VGG16 model. While product technique achieved accuracy of 100% in all models. Max technique accuracy ranges from 99.33% to 98.33% as well as the highest accuracy is in VGG16 model. Overall, that the best results when using product technique.

5.4.2.2. Strategy 2: Using Feature Extraction With CNN

This strategy using AlexNet, VGG16 and VGG19 models for feature extraction from ‘fc6’, ‘fc7’ or ‘fc8’ layers then classification by ECOC with SVM, we use linear kernel function. Classification result normalized by the min-max normalization techniques, then fused by the max, sum, or the product. The experimental result of fusing the Bosphorus dataset with COEP dataset and fusing Dr. Badawi dataset with COEP dataset by the max, sum, or the product are shown in Figures 14,15, and16.

As can be seen from Figure 14 (a and b) with sum technique. Figure 14 (a) shows that fusing Dr. Badawi dataset with COEP dataset achieved accuracy range from 100% to 99% as well as the highest accuracy in all model of 100%. While in Figure 14 (b), fusing Bosphorus dataset with COEP dataset achieved accuracy range from 100% to 99% and the highest accuracy in VGG16 and VGG19 model is about 100%.

Figure 15 (a and b) shows product technique results. In Figure 15 (a), fusing Dr. Badawi dataset with COEP dataset achieved accuracy range from 100% to 99% as well as the highest accuracy in all models is about 100%. While in Figure 15 (b), fusing Bosphorus dataset with COEP dataset achieved accuracy range from 100% to 99% and the highest accuracy in VGG16 model is about 100%.

Figure 16 (a and b) shows max technique results. In Figure 16 (a), fusing Dr. Badawi dataset with COEP dataset achieved range from 100% to 97.33% as well as the highest accuracy in AlexNet model is about 100%. While in Figure 16 (b), fusing Bosphorus dataset with COEP dataset achieved accuracy range from 100% to 96.67% and the highest accuracy in VGG16 model is about 100%. Overall, we note the common model between the two datasets that achieving the best results is VGG16 model.

5.5. Result Discussion

Table 1 shows a comparison between the proposed approaches results. In general, we note feature extract with CNN and classification by ECOC with SVM in general achieved the highest accuracy is about 100% in all approach. However, we can see that FLF is better than SLF for two tests. Also, we note the feature extract with CNN and classification by ECOC in FLF achieved the highest accuracy when we serial fused all model (VGG16, VGG19 and AlexNet) together. While the feature extract

Figure 15. Experiment fusion by SLF using product technique

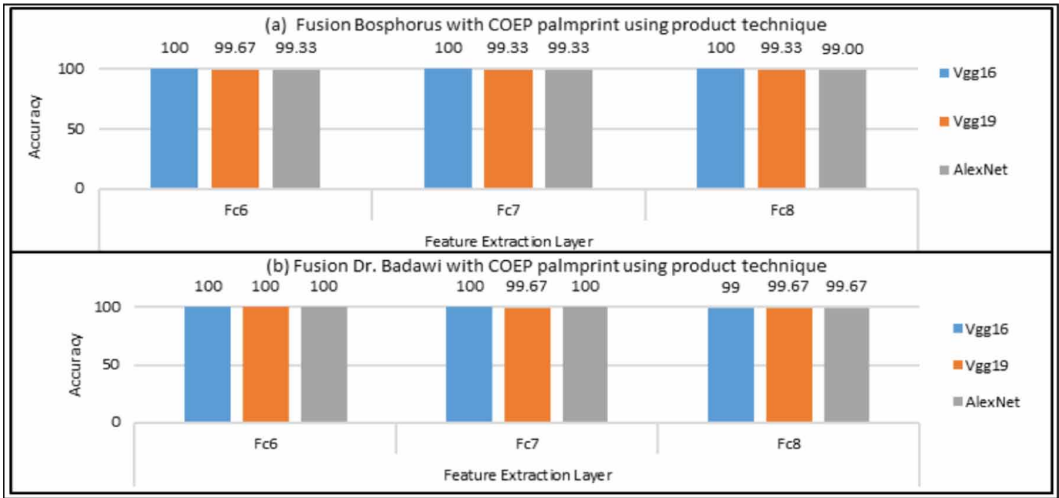
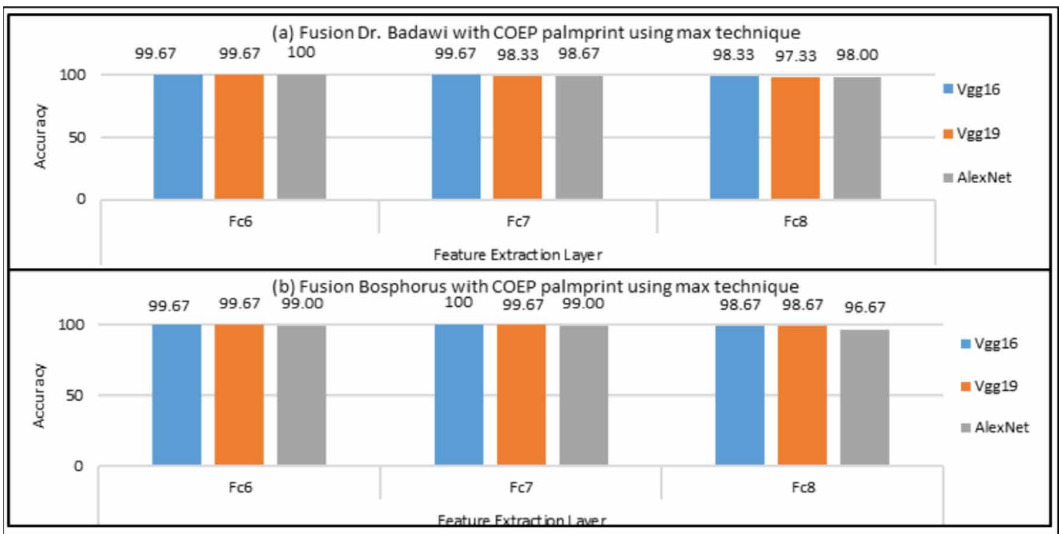


Figure 16. Experiment fusion by SLF using max technique



with CNN and classification by ECOC with SLF achieved the highest accuracy when we use sum method in all model (VGG16, VGG19 and AlexNet). Also, in transfer learning with CNN product technique giving best result.

5.6. Comparison With State-of-the-art

Comparison between the performance of the proposed multi-modal biometric fusing of DHV and PP with the state of the art is illustrated in Table 2. These systems use different feature extraction and classification methods. In (Shahin, Badawi, & Rasmy, 2008) used STFT, maxima and minima and patterns tree extraction feature extraction algorithms and point-matching and least square for classifier. On the other hand, in (Sanchit, Ramalho, Correia, & Soares, 2011) used Gabor filtering feature extraction algorithms and HD classifier. In (Wang, Yau, Suwandy, & Sung, 2008), used thinning

Table 1. Comparison between the results

Fusing Dataset	Approach	Feature Extract and Classification	Fusing Method	Model			Highest Result
				VGG 16	VGG19	AlexNet	
Dr.Badawi with COEP	FLF	Feature extract with CNN and ECOC with SVM	serial fusion one model	100	99.67	99.67	100
			serial fusion all model	100			100
			parallel fusion	100	99	99.33	100
	SLF	Transfer learning with CNN	Sum	97.67	97.97	99.67	99.67
			product	97.67	97.67	100	100
			max	98	97.67	99.67	99.67
		Feature extract with CNN and ECOC with SVM	Sum	100	100	100	100
			product	100	99.67	99.33	100
			Max	99.67	99.67	100	100
Bosphorus with COEP	FLF	Feature extract with CNN and ECOC with SVM	serial fusion one model	100	99.67	99.33	100
			serial fusion all model	100			100
			parallel fusion	100	100	99.33	100
	SLF	Transfer learning with CNN	Sum	99.33	98.67	99	99.33
			Product	100	100	100	100
			max	99.33	98.33	99	99.33
		Feature extract with CNN and ECOC with SVM	Sum	100	100	100	100
			product	100	100	100	100
			max	100	99.67	99	100

algorithm and minutiae points feature extraction algorithms with MHD classifier. In (Wang, Yau, & Suwandy, 2008) used LPP for feature extraction and KNN for classification. In (Bharathi, Sudhakar, & Balas, 2015) used Shearlet transform and Scale-invariant feature transform features and ED for classification. In (Ben Khalifa, Gazzah, & BenAmara, 2013) used DWT feature extraction algorithms similarity measure for classification. In (Nageshkumar, Mahesh, & Swamy, 2009) used eigen vectors of the image then Use the canonical for feature extraction and ED classifier. In (Noushath, Imran, Jetly, Rao, & Kumar, 2013) use LDA and LPQ for feature extraction and logical AND combine the output decisions by different matchers. In (Gopal, Srivastava, Bhardwaj, & Bhargava, 2016) used AHE, GMF and AAD for feature extraction and KNN, RF and SVM for classification.

However, the proposed system recognition is based on CNN depth models that select and express the depth feature of the image automatically by two approaches: using feature extracted with pre-trained CNN and transfer learning with pre-trained CNN that given the highest accuracy in two approaches.

6. CONCLUSION

Palmprint (PP) recognition has many advantages such stable line features, decreased distortion, that it is easier to manually position, and its ability to reach a higher recognition rate quicker [4]. Also, dorsal hand veins (DHV) are gaining popularity among recognition systems since their patterns possess the

following four biometric characteristics: universality, uniqueness, permanence, and circumvention. There are some applications that use the DHV and PP for recognition as ATM machines, bank transactions, computers, and entry systems. We noticed Most of researchers used hand-crafted methods for PP and DHV recognition systems. To overcome these issues, the researchers used feature-learning algorithms, especially the CCN.

In this paper, we proposed a multimodal biometric system based on PP and DHV by (VGG16, VGG19 and AlexNet) CNN models. In the proposed system, we used two fusing approaches: First Approach: fusion at feature level (FLF) where features are extracted with (VGG16, VGG19 or/and AlexNet) CNN models and used ECOC with SVM for classification in this approach we use serial and parallel methods for feature fused. Second Approach: fusion at score level (SLF) by two strategies: transfer learning that use (VGG16, VGG19, and AlexNet) CNN models for features extraction and classification then fused at score level. Second strategy feature extraction with (VGG16, VGG19 and AlexNet) models and train ECOC with SVM for classification then fusion. In this approach we use sum, max and product fusing methods.

In our model, the test is done two times by repeating PP database for each DHV databases: Dr. Badawi with COEP palmprint and Bosphorus with COEP palmprint. The experiment results feature extract with CNN and classification by ECOC with SVM in general achieved the highest accuracy. However, FLF is better than SLF for two tests. Also, the feature extract with CNN and classification by ECOC in FLF achieved the highest accuracy when we serial fused all model (VGG16, VGG19 and AlexNet) together. While the feature extract with CNN and classification by ECOC with SLF achieved the highest accuracy when we used sum method in all model (VGG16, VGG19 and AlexNet). Also, transfer learning with CNN product technique giving best result.

In the future, we will try the cross validation beside the spilt validation. Also, we will evaluate the performance of the proposed algorithm using the different pre-trained model on more DHV and PP datasets.

Table 2. Comparison with state-of-the-art

Author/ Ref # / Year	The Used Biometric	Feature Extraction Algorithm	Matching/ Classification	Results
(Shahin, Badawi, & Rasmy, 2008)	Fingerprints (FP), dorsal hand vein (HV), and hand geometry (HG) shape. Fusion at decision level	FP extract by STFT and Chain Code Contours algorithm. GH extract by maxima and minima points detection. VH extract by patterns tree extraction based on correlation between the two trees patterns	Point matching and least square algorithms.	Not Mentioned
(Sanchit, Ramalho, Correia, & Soares, 2011)	hand-palm vein and hand-DHV Fusion at the score level	Gabor filtering	HD	Accuracy= 100%
(Wang, Yau, Suwandy, & Sung, 2008)	Palmprint and palm vein. Fusion at the feature level	thinning algorithm, minutiae points	MHD algorithm	Accuracy= 99.4%
(Wang, Yau, & Suwandy, 2008)	palmprint and palm vein. Fusion at the feature level	LPP	KNN	Accuracy= 99.5%
(Bharathi, Sudhakar, & Balas, 2015)	Dorsal hand vein, finger vein and palm vein. Fusion at the score level	Shearlet transform and Scale-invariant feature transform	Euclidean distance (ED)	Accuracy of 94%,
(Ben Khalifa, Gazzah, & BenAmara, 2013)	Face, palmprint and the fingerprint. Fusion at score levels.	DWT	Similarity measure	EER= 0.47
(Nageshkumar, Mahesh, & Swamy, 2009)	Face and palmprint Fusion at matching score level	Determine the eigen values and eigen vectors of the image then Use the canonical	ED	Accuracy= 98%
(Noushath, Imran, Jetly, Rao, & Kumar, 2013)	Face and Palmprint Fusion at sensor level, feature level, decision level and score level	LDA and LPQ	adopted logical AND, OR to combine the output decisions by different matchers	Accuracy= 99.5%
(Gopal, Srivastava, Bhardwaj, & Bhargava, 2016)	Palmprint, palm-phalanges print and dorsal hand vein. Fusion at score level	AHE, GMF, AAD and Mean Features.	KNN, RF and SVM	Accuracy= 99%
Propose system	Palmprint and dorsal hand vein. Fusion at feature level and score level	Transfer learning (Feature extract and classification with VGG16, VGG19 and AlexNet models)		100%
		ECOC+ SVM	Feature extract with VGG16, VGG19 and AlexNet models	100%

REFERENCES

- Al-johania, N. & Elrefaei, L. A. (2019). Dorsal Hand Vein Recognition by Convolutional Neural Networks: Feature Learning and Transfer Learning Approaches. *International Journal of Intelligent Engineering and Systems*, 12(3), 178-191. doi: 10.22266/ijies2019.0630.19.
- Al-Waisy, A. S., Qahwaji, R., Ipson, S., & Al-Fahdawi, S. (2018). A multimodal deep learning framework using local feature representations for face recognition. *Machine Vision and Applications*, 29(1), 35–54. doi:10.1007/s00138-017-0870-2
- Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., . . . Awwal, A. (2018). The history began from alexnet: A comprehensive survey on deep learning approaches.
- Autonomous Institute of Government of Maharashtra. (2018). COEP Palm Print Database. Retrieved from www.coep.org.in/resources/coeppalmprintdatabase
- Belean, B., Streza, M., Crisan, S., & Emerich, S. (2017). Dorsal hand vein pattern analysis and neural networks for biometric authentication. *Studies in Informatics and Control*, 26(3), 305–314.
- Ben Khalifa, A., Gazzah, S., & BenAmara, N. E. (2013). Adaptive score normalization: A novel approach for multimodal biometric systems. *World Acad. Sci. Eng. Technol. Int. J. Comput. Sci. Eng.*, 7(3), 882–890.
- Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1), 1–127. doi:10.1561/2200000006
- Bharathi, S., Sudhakar, R., & Balas, V. E. (2015). Hand Vein-based Multimodal Biometric Recognition. *Acta Polytechnica Hungarica*, 213–229.
- Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.
- Bolle, R., & Pankanti, S. (1998). *Biometrics. Personal Identification in Networked Society: Personal Identification in Networked Society*.
- Bora, K., Chowdhury, M., Mahanta, L. B., Kundu, M. K., & Das, A. K. (2016). Pap smear image classification using convolutional neural network. In *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing. ACM*. doi:10.1145/3009977.3010068
- Byahatti, P., & Hatture, S. M. (2017). *A Fusion Model for Multimodal Biometric System. International Journal of Engineering Research & Technology. IJERT*.
- Chaki, J. (2018). An efficient two-stage Palmprint recognition using Frangi-filter and 2-component partition method. In *Proceedings of the Fifth International Conference on Emerging Applications of Information Technology (EAIT)*. Academic Press. doi:10.1109/EAIT.2018.8470444
- Chan, T.-H., Jia, K., Gao, S., Lu, J., Zeng, Z., & Ma, Y. (2015). Pcanet: A simple deep learning baseline for image classification. *IEEE Transactions on Image Processing*, 24(12), 5017–5032. doi:10.1109/TIP.2015.2475625 PMID:26340772
- Chuang, S.-J. (2018). Vein recognition based on minutiae features in the dorsal venous network of the hand. *Signal, Image and Video Processing*, 12(3), 573–581. doi:10.1007/s11760-017-1195-3
- Ding, Y., Zhuang, D., & Wang, K. (2005). A study of hand vein recognition method. In *Proceedings of the IEEE International Conference Mechatronics and Automation* (pp. 2106-2110). IEEE Press.
- Dubey, P., Kanumuri, T., & Vyas, R. (2017). Palmprint Recognition using Binary Wavelet Transform and LBP representation. In *Proceedings of the International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE)* (pp. 201-205). Academic Press. doi:10.1109/RISE.2017.8378154
- Gopal, S., Srivastava, S., Bhardwaj, S., & Bhargava, S. (2016). Fusion of palm-phalanges print with palmprint and dorsal hand vein. *Applied Soft Computing*, 47, 12–20. doi:10.1016/j.asoc.2016.05.039
- Haghighat, M., Zonouz, S., & Abdel-Mottaleb, M. (2015). CloudID: Trustworthy cloud-based and cross-enterprise biometric identification. *Expert Systems with Applications*, 42(21), 7905–7916. doi:10.1016/j.eswa.2015.06.025

Han, X. H., Lei, J., & Chen, Y. W. (2016). HEp-2 cell classification using K-support spatial pooling in deep CNNs. In *Deep Learning and Data Labeling for Medical Applications* (pp. 3–11). Springer.

Heenaye, M., & Khan, M. (2012). A multimodal hand vein biometric based on score level fusion. *Procedia Engineering*, 41, 897–903. doi:10.1016/j.proeng.2012.07.260

Hertel, L., Barth, E., Käster, T., & Martinetz, T. (2015). Deep convolutional neural networks as generic feature extractors. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*. Academic Press. doi:10.1109/IJCNN.2015.7280683

Hijazi, S., Kumar, R., & Rowen, C. (2015). *Using convolutional neural networks for image recognition*. San Jose, CA: Cadence Design Systems Inc.

Hu, Y.-P., Wang, Z.-Y., Yang, X.-P., & Xue, Y.-M. (2014). Hand vein recognition based on the connection lines of reference point and feature point. *Infrared Physics & Technology*, 110–114.

Huang, D., Zhu, X., Wang, Y., & Zhang, D. (2016). Dorsal hand vein recognition via hierarchical combination of texture and shape clues. *Neurocomputing*, 214, 815–828. doi:10.1016/j.neucom.2016.06.057

Huang, Z., Pan, Z., & Lei, B. (2017). Transfer Learning with Deep Convolutional Neural Network for SAR Target Classification with Limited Labeled Data. *Remote Sensing*, 9(9), 907. doi:10.3390/rs9090907

Jia, W., Hu, R.-X., Lei, Y.-K., Zhao, Y., & Gui, J. (2013). Histogram of Oriented Lines for Palmprint Recognition. *IEEE Transactions on Systems, Man, and Cybernetics. Systems*, 44(3), 385–395. doi:10.1109/TSMC.2013.2258010

Joutsijoki, H., Haponen, M., Rasku, J., Aalto-Setälä, K., & Juhola, M. (2016). Error-correcting output codes in classification of human induced pluripotent stem cell colony images. *BioMed Research International*, 1–13. doi:10.1155/2016/3025057 PMID:27847810

Kabir, W., & Ahmad, M. O. (2017). Palmprint Recognition Based on Histograms of Sparse Codes. In *Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)* (pp. 965–968). IEEE Press. doi:10.1109/MWSCAS.2017.8053086

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105). ACM.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.

Lee, J.-C., Lo, T.-M., & Chang, C.-P. (2016). Dorsal hand vein recognition based on directional filter bank. *Signal, Image and Video Processing*, 10(1), 145–152. doi:10.1007/s11760-014-0714-8

Li, W., Zhang, B., Zhang, L., & Yan, J. (2012). Principal Line-Based Alignment Refinement for Palmprint Recognition. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews*, 42(6), 1491–1499. doi:10.1109/TSMCC.2012.2195653

Liu, C., & Wechsler, H. (2001). A shape- and texture-based enhanced Fisher classifier for face recognition. *IEEE Transactions on Image Processing*, 10(4), 598–608. PMID:18249649

Marcelino, P. (2019). Transfer learning from pre-trained models. *Towards Data Science*. Retrieved from <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>

Mi, A., Wang, L., & Qi, J. (2016). A Multiple Classifier Fusion Algorithm Using Weighted Decision Templates. *Scientific Programming*. doi:10.1155/2016/3943859

Nageshkumar, M., Mahesh, P. K., & Swamy, M. N. (2009). An Efficient Secure Multimodal Biometric Fusion Using Palmprint and Face Image.

Noushath, S., Imran, M., Jetly, K., Rao, A., & Kumar, G. H. (2013). Multimodal biometric fusion of face and palmprint at various levels. In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1793–1798). Academic Press. doi:10.1109/ICACCI.2013.6637453

Orság, F., & Drahanský, M. (2003). Biometric Security Systems: Fingerprint and Speech Technology. In *IICAI* (pp. 703–711). Academic Press.

- Palma, D., Montessoro, P. L., Giordano, G., & Blanchini, F. (2015). A dynamic algorithm for palmprint recognition. In *Proceedings of the IEEE Conference on Communications and Network Security (CNS)* (pp. 659-662). IEEE Press. doi:10.1109/CNS.2015.7346883
- Ramvalho, M., Correia, P. L., & Soares, L. D. (2011, April). Biometric identification through palm and dorsal hand vein patterns. In *Proceedings of the 2011 IEEE EUROCON-International Conference on Computer as a Tool* (pp. 1-4). IEEE.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. doi:10.1016/j.neunet.2014.09.003 PMID:25462637
- Shahin, M. k., Badawi, A. M., & Rasmy, M. E. (2011). 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*, 1107–1122.
- Shahin, M. K., Badawi, A. M., & Rasmy, M. E. (2008). A Multimodal Hand Vein, Hand Geometry, and Fingerprint Prototype Design for High Security Biometrics. In *Proceedings of the Cairo International Biomedical Engineering Conference*. Academic Press. doi:10.1109/CIBEC.2008.4786038
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1), 1929–1958.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Erhan, D. (2015). Going deeper with convolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE Press.
- The Bosphorus Hand Database. (2010). Retrieved from <http://bosphorus.ee.boun.edu.tr/>
- Trabelsi, R. B., Masmoudi, A. D., & Masmoudi, D. S. (2016). Hand vein recognition system with circular difference and statistical directional patterns based on an artificial neural network. *Multimedia Tools and Applications*, 75(2), 687–707. doi:10.1007/s11042-014-2315-5
- Wan, H., Chen, L., Song, H., & Yang, J. (2017). Dorsal hand vein recognition based on convolutional neural networks. In *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1215-1221). IEEE Press. doi:10.1109/BIBM.2017.8217830
- Wang, J.-G., Yau, W.-Y., & Suwandy, A. (2008). Feature-level fusion of palmprint and palm vein for person identification based on “a Junction Point” representation. In *Proceedings of the 15th IEEE International Conference on Image Processing* (pp. 253-256). IEEE Press.
- Wang, J.-G., Yau, W.-Y., Suwandy, A., & Sung, E. (2008). Person recognition by fusing palmprint and palm vein images based on Laplacian palm representation. *Pattern Recognition*, 41(5), 1514–1527. doi:10.1016/j.patcog.2007.10.021
- Wild, P., Radu, P., Chen, L., & Ferryman, J. (2016). Robust multimodal face and fingerprint fusion in the presence of spoofing attacks. *Pattern Recognition*, 50, 17–25. doi:10.1016/j.patcog.2015.08.007
- Yang, J., & Yang, J.-y. (2002). Generalized K-L transform based combined feature extraction. *Pattern Recognition*, 35(1), 295–297. doi:10.1016/S0031-3203(01)00152-2
- Yang, J., Yang, J.-y., Zhang, D., & Lu, J. (2003). Feature fusion: Parallel strategy vs. serial strategy. *Pattern Recognition*, 36(6), 1369–1381. doi:10.1016/S0031-3203(02)00262-5
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in neural information processing systems* (pp. 3320-3328). Academic Press.
- Zhang, H., Gao, W., Chen, X., Shan, S., & Zhao, D. (2006). Robust Multi-view Face Detection Using Error Correcting Output Codes. In *Proceedings of the European Conference on Computer Vision*. Springer. doi:10.1007/11744085_1
- Zhu, X., & Huang, D. (2012). Hand dorsal vein recognition based on hierarchically structured texture and geometry features. In *Proceedings of the Chinese Conference on Biometric Recognition* (pp. 157-164). Springer. doi:10.1007/978-3-642-35136-5_20

Norah Abdullah Al-johani received her bachelor's degree with first honors in computer science in 2014 from Faculty of Computer Science and Engineering in Medina, Taibah University, Saudi Arabia. She Received her master's degree in 2019 at King Abdulaziz University, Jeddah, Saudi Arabia. Her research interests include programming languages, artificial intelligence, encryption and image processing.

Lamiaa A. Elrefaei received her B.Sc. degree with honors in Electrical Engineering (Electronics and Telecommunications) in 1997, her M.Sc. in 2003 and Ph.D. in 2008 in Electrical Engineering (Electronics) from faculty of Engineering at Shoubra, Benha University, Egypt. She held a number of faculty positions at Benha University, as Teaching Assistant from 1998 to 2003, as an Assistant Lecturer from 2003 to 2008, and has been a lecturer from 2008 to date. She is currently an Associate Professor at the faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. Her research interests include computational intelligence, biometrics, multimedia security, wireless networks, and nano networks. She is a senior member of the IEEE.