CWNN-Net: A New Convolution Wavelet Neural Network for Gender Classification using Palm Print

Elaraby A. Elgallad[1]
Deanship of Information Technology
Tabuk University, KSA

Wael Ouarda[2], Adel M. Alimi[3]
Research Groups in Intelligent Machines
ENIS, BP 1173, Sfax, 3038, Tunisia

Abstract—**The human hand is one of the body parts with special characteristics that are unique to every individual. The distinctive features can give some information about an individual, thus, making it a suitable body part that can be relied upon for biometric identification and, specifically, gender recognition. Several studies have suggested that the hand has unique traits that help in gender classification. Human hands form part of soft biometrics as they have distinctive features that can give information about a person. Nevertheless, the information retrieved from the soft biometrics can be used to identify an individual's gender. Furthermore, the soft biometrics can be combined with the main biometrics characteristics that can improve the quality of biometric detection. Gender classification using hand features, such as palm contributes significantly to the biometric identification domain and, hence, presents itself as a valuable research topic. This study explores the use of Discrete Wavelet Transform (DWT) in gender identification, with SqueezeNet acting as a tool for unsheathing features, and Support Vector Machine (SVM) operating as discriminative classifier. Inference is made using mode voting approach. Notably, the two datasets that were crucial for the fulfillment of the study were the 11k database and CASIA. The outcome of the tests substantiated the use of voting technique for gender recognition.**

Keywords—Deep learning; feature extraction; gender; voting

I. INTRODUCTION

Biometrics is a verification technique that incorporates science (biology) as well as technology. It is highly applicable in information assurance. The method uses human biological data, including DNA and fingerprints to ensure safe entry, data access, and protection. The systems normally have high-tech elements that are connected to produce powerful performance. Biometric systems are popular among institutions that have security systems as well as replacement systems including PIN, and ID replacement. The main distinction between biometric system and other conventional systems is the fact that the former requires the physical presence of the individual who is using the system. This physical presence creates an extra level of security and makes it difficult for identity thieves to use a false ID card or any other stolen mode of identification.

Biometric systems are principally divided into two categories on the basis of physiological traits and behavioral traits. Biometric systems with the physiological identifiers focus on authenticating the physical features of the individual being verified such as the face, fingerprints, structure of the fingers, iris, DNA, retina, and other human physical characteristics that are uniquely different. On the other hand, behavioral identifiers refer to the inimitable traits of individuals such as their typing patterns, manner of walking, use of gestures, and so on. The system can rely on the behavioral traits to offer uninterrupted verification rather than a solitary one-time check.

Accordingly, this paper focuses on gender recognition in biometric identification and explains how it can be used to enhance the accuracy of biometric identification processes. The paper is organized into five sections. The first section provides an introduction to this topic. The second section comprehensively explores the existing literature on gender identification. The third section details the manner in which the suggested technique operates. The fourth section outlines the trials and outcomes of the proposed method. Lastly, the fifth and sixth sections discuss the findings of these results and eventually provide the conclusion of this paper.

II. RELATED WORK

There literature on gender recognition remains scarce when compared to person identification that has attracted several academicians and experts in the field. The scarcity of related literature can be explained by the fact that person identification, as opposed to gender identification, has a practical security interest for the government and law enforcement agencies. Nevertheless, gender recognition has the potential to become the next hot topic when it comes to human-computer communication. Furthermore, gender identification can also be a valuable tool for various organizations that need intelligent advertising based on accurate gender identification. One of the pioneers is [1] who has explored the concept of using computers to classify gender based on the distinctive features of the hand. He came up with a system that splits the hand delineation into various fragments; it aligns with the fingers as well as the palms, and defines the features of every section using different techniques comprising boundary descriptors, file descriptors, region descriptors and ZMs.

The respective gender was allocated based on the consolidated evidence from the hand profile with the match score being determined using the score-level fusion and file descriptors. The best performance was a match score of 98%. Notably, the evidence provided by ZMs after utilizing the score-level fusion was almost similar to that of file descriptors. Afifi [2] shared a logical record of hand images that can be used in biometric detection and gender identification. Authors presented a collection of valuable metadata drawn from

Mahmoud's logical record. The researchers also proceeded to offer top-notch approaches for using the data record to in gender detection.

There is also a convolutional neural network (CNN) based on a two-steam data that indicates the gender identification challenge. CNN helps to disclose features that are then conveyed to a support-vector-machine that analyzes data and classifies the biometric recognition issue. There is a strong belief that the suggested data record will significantly contribute towards the creation of better gender detection and biometric systems that rely on hand images. Another researcher, Ming and Yubo [3] developed a palm geometric-oriented method for identifying whether a person is a male or a female with the support-vector machine. This method is less complicated and has been shown to be effective with regards to gender classification. They note that this method guarantees a high-level of accuracy without using any intricate computations, processes, and procedures that have been used less successfully in other biometric techniques.

Instead, it is characterized by few features, including the ability to record hand features without the users having to peg their hands on the device. Its efficiency, simplicity, and convenience make it appropriate for real-world execution. Meanwhile, Font-Aragonesin and Faundez-Zanuy [4] proposed a technique that uses anthropometric hand data to identify the respective genders of the sample being tested. Using their method, researchers collected data on unvaried number of several men and women in their visual hand database. Majority of the information was collected on men.

A simple approach was then devised to get the measurements of the users' hands. The information was transmitted through a Biometric Dispersion Matcher (BDM) to retrieve the suitable data. BDM functions as a quadratic discriminant classifier. This discriminant classifier begins with filtering out data that will not assist in disclosing the user's gender. It then proceeds to display a vector of the key computations. The technique had a performance rate of 95% where the ratio of men to women was 2:1, with a projection that the accuracy could become higher when data records increase.

Gender identification through the use of facial features remains a problematic issue in the field of computer vision. Previous efforts in the face image gender recognition concentrated on enhancing understanding the neural network. Key efforts included the double-layered neural network known as SEXNET that was the brainchild of Gollomb and Lawrence [5]. Another development in face image gender recognition was that of Yen [6], which sought to eliminate the inaccuracies that are caused by less important data such as changes in face lighting as well as adjustments in facial expressions. The authors developed a novel algorithm that could help to reduce the impact of the latter disturbances. The 2-D Gabor transform was effective in spotting the facial points, with the SVM discriminative classifier picking out distinctive features to determine the respective gender of the user.

Meanwhile, it is apparent that many researchers have focused on Local Binary Patterns (LBP) and the alternatives in computer vision that can be used in gender classification. One of the suggestions shared by researchers, in this case [7], is Binary Robust Independent Elementary Features (BRIEF), Oriented FAST & Rotated BRIEF (ORB) as well as Binary Robust Invariant Scalable Keypoints (BRISK), which have been reported to be reliable and fast in gender classification. When compared to LBP, the latter variants provide speedy detection while still maintaining high performance rates as LBP. This makes them better than LBP when used in systems that require constant gender recognition. Meanwhile, there is no sufficient literature to validate that these options are more suitable than LBP.

Notably, in another study [8], they sought to determine if face images with deep learning could help in gender classification. Specifically, they used Local Receptive Field-Extreme Learning Machine (LRF-ELM) as well as Convolutional Neural Networks (CNN). The technique was tested using data records that were meant to reveal the age and gender of the sample. The outcome of the experiment revealed an accuracy score of 80% for LRF-ELM, and 87.13% for CNN.

Another approach that captured the interest of researchers is gender recognition through boxing action. Author in [9] tested this technique whereby a period detection procedure is applied and thereafter, an averaged profile is used to denote a boxing sequence of a period. The human classification is done using Nearest Neighbor Classifier (NNC), which basically compares the recorded data with the data set that is being tested. The NNC was based on Euclidian metric. The tests were done on the KTH-Dataset, which has a good performance rate of at least 80%. PCA was used to extract features, and finally SVM was used to allocate the respective genders. In [10], they have also recommended Global Local Feature Fusion (GLFF) as a reliable method for gender recognition. The study relied on findings from psychophysics as well as neurophysiology observations, which asserted that universal and local data is important in image perception.

The initial step in the GLFF technique is mining universal and local traits through Active Appearance Model (AAM) and LBP tool respectively. One then proceeds to merge the universal and local features through sequent selection. Eventually, the chosen traits are used to determine the receptive genders in the dataset with the help of SVM. The test was made on a sample of 20 men and 20 women with the results having an average accuracy rate of 80% and an indication that the accuracy levels could go up. Based on these test results, it was apparent that blending local and global features could significantly improve the precision of gender recognition and the performance of gender classification tools.

Furthermore, the study presented the significance of different body features on gender classification. The numerical analysis revealed that the head makes a substantial contribution in gender recognition with the buttocks and leg making little contribution towards accurate gender identification.

Conversely, in [11], they suggested the use of facial features to classify gender by combining the findings of different SVM classifiers. Typically, three SVM classifiers are used to achieve this goal. The first descriptor, Histogram of Oriented Grardients (HOG), analyzes the features that pertain

to shape with the help of histogram intersection kernels. The second visual descriptor, LBP, focuses on the texture of the facial features, also with the help of histogram intersection kernels. The third descriptor analyses the raw pixel values using a linear kernel. The researchers used this approach to examine data that had been mined from FERET. The good performance rate of this test stood at 92.6%, which surpasses the accuracy of some commercial tools such as Face++.

Ardakany and Louis [12] dealt with the challenges of gender recognition through genetic algorithms. The researchers mined relevant data on facial feature records available in FERET database using LBP and PCA. The algorithm allows extracting the features that enhance the capability of the SVM classifiers to accurately recognize whether one is male or female. The genetic algorithm condenses the features being analyzed from 142 to about 71, and thereby increasing level of accuracy to about 98.5, making the results completely reliable.

Meanwhile, in the study [13], a discriminatively-trained CNN was found to be effective in classifying the gender of pedestrians. The CNN has a complex tier of neural networks that fuse feature mining and grouping into a solitary framework. The researchers recorded an accuracy of 80.4% when they used simple architecture and nominal prepossessing of dataset that provided complete body features of the pedestrians that were sampled. The accuracy of the results is equivalent to other high-tech approaches that do not use hand-manufactured trait miners.

In [14], they sought to identify gender through analyzing the different walking style of individuals. The researchers were inspired by previous studies, which have shown that gait can be used in behavior recognition tests. SVM was used to deduce the respective genders of the sample data in the analysis of the sparse spatio temporal features. The method attained a decent performance rate of 87%.

Another popular proposal in the current literature is the one made by Haitpoglu and Kose [15], which suggests combining Speed-Up Robust Features (SURF) of bags-of-visual-words (BOW) and SVM. This proposal was tested on different parts of 3560 face samples retrieved from the FERET database to determine the degree of its efficiency and reliability. The results showed that the approach is efficient in classifying gender for records retrieved from FERET database.

However, it provided more accurate recognition on the frontal face images than those on the left and right side. Meanwhile, in [16], author proposed an automated gender classification technique that relies on CNN. This method begins with network training, which is accomplished through combining numerous face datasets retrieved from different databases including ~70000 facial images from the World Wide Web. Once the data for the networks had been recorded, they were assessed, and compared with various network architectures that displayed better performance.

From their experiment, inception-v4 network had the highest accuracy level at 98.2% with Audience dataset coming second at 84%. Crime and Pedrini [17] used a pre-defined face silhouette prototype to create geometric descriptor for classifying gender. The researchers used this method to analyze four sets of face data records and reported better performance than other techniques that use geometric descriptors.

Another study [18], focusing on how the unique manner of walking could help in gender recognition shared a more advanced gait energy image (GEI) titled D-GEI. This was a step in the right direction in terms of building the literature on gait analysis and its place in gender recognition. Most of the existing studies have not extensively dealt with gait analysis, which has resulted in poor performance of gait evaluation in human identification.

The procedure in the D-GEI method begins with creating a dynamic region, followed by establishing the dynamic region of frame, and then computing the weighted average of the dynamic region. These steps help in determining the D-GEI. Once the latter has been established, HOG is used in gradient computation, while SVM classifier helps to determine the respective genders of the sample. Author in [19] presented another method that could be used in gender recognition based on the facial features of the user. In this proposal, the domain-specific as well as trainable traits are merged to help in gender classification. Fifty-two facial features linked to the eyes, nose, as well as the mouth were mined to act as SURF descriptors.

Likewise, COSFIRE acted as the trainable traits. This method responds strongly to some of the notorious challenges associated with face profiles such as the expression variations, light adjustments, and different poses. It attains a high accuracy rate on some of the top datasets in gender classification: GENDER-FERET and LFW. The technique was also highly reliable when real-world data records were used. The datasets comprised 206 training (144 men and 62 women) images, as well as 200 test (139 men and 61 women) images that were taken when the sample population on the normal walking motion. The purpose of assessing the performance of this algorithm on real scenarios was to predict its accuracy on analyzing images retrieved from videos, that are likely to be more problematic that standard data records.

Despite the problematic issues in real scenarios, the algorithm was 91.5% accurate in classifying the genders. The researchers further observe that the COSFIRE is more reliable than the SURF descriptors and therefore, recommends that it can be used solve several types of visual pattern classification problem.

III. METHODOLOGY

Fig. 1 indicates the CWNN-Net system which this paper seeks to propose. The system works together with Squeezenet to extract features while SVM remains as the discriminative classifier. The final interpretation is made using the mode voting method. There are two databases that are relied upon in this system are 11k database [2], and CASIA database [20] each containing sufficient hand images and palmprint images respectively. The images were changed to RGB and minimized to 227×227 pixels to enhance CNN feature extraction.

A. Features Extraction

Feature extraction discloses the nature of shape in a given pattern and thereby, simplifying the process of sorting the pattern using a formal method. It normally entails minimizing

the number of random variables being analyzed until one is left with the main variables. In pattern recognition and in image processing, feature extraction is a particular type of reducing dimensionality. Its key objective is to mine the pertinent material from the original sample and present in a way that makes it easy for image processing and pattern classification, Kumar and Bhatia [21].

1) Discrete Wavelet Transform (DWT): The Continuous Wavelet Transform (CWT) changes an uninterrupted signal into one that has two uninterrupted variables: translation and scale. The subsequent signal after the modifications made by CWT has a less complicated interpretation and beneficial when it comes to time-frequency analysis. The continuous wavelet transform of continuous function, x(t) relative to real-valued wavelet, ψ(t) is described by:

$$W(a,b) = \frac{1}{\sqrt{a}} \int \psi\left(\frac{t-b}{a}\right) s(t) dt \qquad (1)$$

where ψ is the analyzing wavelet, a represents a time dilation, b a time translation, and the bar stands for complex conjugate.

DWT is a robust signal processing tool. Although DWT has some similarities with CWT, the difference is more apparent when it comes to scale and position values: the former scales and position values apply powers of two.

The values of s and t are: s=2^j, τ= k*2^j and (j, k) $\in Z^2$ as shown in (2):

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-k*2^j}{2^j}\right) \qquad (2)$$

The focus in DWT as well as inverse DWT is disintegration and restoration. The disintegration and restoration are achieved through LPF and HPF. The effect of wavelet disintegration is a structured decomposition that occurs in tiers. The level of disintegration is selected as per the preferred cutoff frequency. Fig. 2 illustrates DWT undergoing a process of LPF and HPF [22].

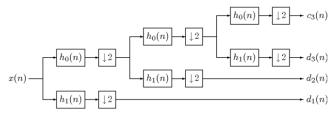Fig 1. CWNN-Net: A New Convolution Wavelet Neural Network for Gender Classification using Palm Print.

Fig 2. A Three-Level forward DWT via a Two-Channel Iterative Filter Bank [22].

2) Discrete haar wavelet transform

The original Haar definition is as follows:

$haar(0,t) = 1, for\ t \in [0,1);$

$$haar(1,t) = \begin{cases} 1, for\ t \in \left[0,\frac{1}{2}\right), \\ -1, for\ t \in \left[\frac{1}{2},1\right) \end{cases} \qquad (3)$$

One of the key characteristics of the Haar functions other than the Haar (0, t), the i–th Haar function can occur through the restriction of the (j − 1)–th function to be 50 percent of the interval where it is different from zero, through multiplying it with p2 and scaling over the interval [0, 1]. These traits make Haar function makes share a degree of link with the wavelet concept and can be used to define wavelet. In this context, two Haar functions represent universal functions, while the remaining functions represent the local functions. As such, the odd rectangular pulse pair, a Haar function in this setting, denotes the oldest and the most basic wavelet. The purpose of using DWT is to get data that is highly differentiated by offering distinct solution at separate segments of the time-frequency domain. Through wavelet transforms, one can create unequal segments of time-frequency plane in line with the time-spectral aspects of the signal. The wavelet technique has robust links to the traditional standard basis of the Haar functions. Scaling as well as shrinking a standard wavelet produces classical Haar functions.

Let ψ: R → R, the Harr wavelet function is defined by the formula:

$$\psi(t) = \begin{cases} 1, for\ t \in \left[0,\frac{1}{2}\right), \\ -1, for\ t \in \left[\frac{1}{2},1\right), \\ 0, otherwise. \end{cases} \qquad (3*)$$

for any Haar function (except function haar (0, t))

From basis (3*) may be generated by means of the formulas:

$$\psi_i^j(t) = \sqrt{2^j}\ \psi(2^j t - i)$$

$$i = 0,1,\dots,2^j - 1.\ and\ j = 0,1,\dots,log_2\ N - 1 \qquad (4)$$

The constant $\sqrt{2^j}$ is chosen so that:

the scalar produc $< \psi_i^j, \psi_i^j \geq 1, \psi_i^j(t) \in L^2(R)$

Let Φ: R → R, the Harr scaling function is defined by the formula:

$$\Phi(t) = \begin{cases} 1, & for \ t \ \in [0,1), \\ 0, & for \ t \ \notin [0,1). \end{cases} \qquad (5)$$

Similarly, to the properties of the wavelet function, for scaling function one can define the family of functions:

$$\Phi_i^j(t) = \sqrt{2^j} \ \Phi(2^j t - i)$$

$$i = 0,1, \dots, 2^j - 1. \ and \ j = 0,1, \dots, log_2 N - 1 \qquad (6)$$

The constant $\sqrt{2^j}$ is chosen so that:

the scalar product $< \Phi_i^j, \Phi_i^j \geq 1, \Phi_i^j(t) \ \in \ L^2(R)$

In the two-dimensional case, three sets of detail coefficients residing in the horizontal, vertical, and diagonal directions. The subbands LH_j, HL_j, and HH_j, j = 1, 2... J are the detail coefficients, as noted above, where j is the scale and J denotes the largest or coarsest scale in the decomposition. Fig. 3 shows a representation of the multilevel wavelet decomposition at level 3.
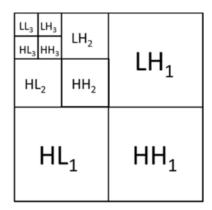

Fig 3.   Level 3 Wavelet Decomposition [23].

3) Squeeze net: The highlight of most researchers exploring CNN has been how the field can achieve create precision on machine vision datasets. Several models have been built to achieve different levels of accuracy. Reference [24] have reviewed different CNN techniques to establish which model would require simpler parameter but attain a degree of precision that is similar to some other popular techniques. From his analysis, he concluded Squeeze Net has minimal parameters but still achieves a performance level that is comparable to common models. Fig. 4 and Fig. 5 shows the organization of convolution filters in the Fire module.

Approach 1: Resizing the network 3x3 filters to 1x1 filters

This approach focused on decreasing the number of parameters resizing 3x3 filters to 1x1 filters. At the onset, this approach was confusing as one would assume that replacing 33x3 filters with 1x1 filters would provide less information and thereby, lower the accuracy of the model. Nevertheless, this was not the case. Naturally, 3x3 filters takes spatial data of pixels that are in close proximity. On the other hand, in 1x1 filters solely concentrates on one pixel and extracts the connection of channels within the pixel without focusing the adjacent one.
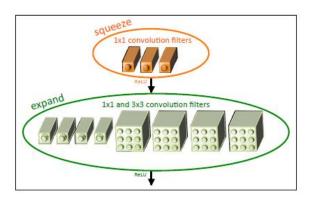

Fig 4.   Microarchitectural view: Organization of Convolution Filters in the Fire Module. In this Example, s1x1 = 3, e1x1 = 4, and e3x3 = 4. We Illustrate the Convolution Filters but not the Activations [24].
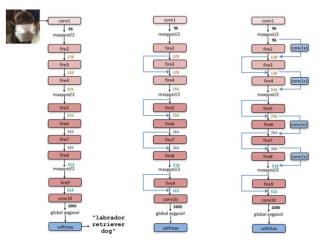

Fig 5.   Microarchitectural view: Organization of Convolution Filters in the Fire Module. s1x1 = 3, e1x1 = 4, and e3x3 = 4 [24].

Forrest used three main approaches when evaluating the different CNN architectures:

Approach 2: Minimizing the number of inputs in the residual 3x3 filters

The approach focuses decreasing the parameter through cropping filters. This methodical strategy is actualized through feeding "squeeze" stratums into "expand" layers. Squeeze layers consist of 1x1 filters while "expand" layers have both 3x3 filters and 1x1 filters. Therefore, aggregate number of parameters are minimized through the lowering the "squeeze" stratum filters. This process is referred to as the "fire module" is used as the foundational block of creating the Squeezenet design.

Approach 3: Downsample the end part of the network to create large feature maps in the convolution strata

Author in [24] propose that reducing the stride with the later convolution strata results in larger feature map, which enhances the accuracy of recognition. The manner in which the activation maps occur at the end of the network differentiates this design from well-known networks such as the VGG which is characterized with feature maps that shrink as one gets to end parts of the network.

The building block of Squeezenet is referred to as fire module. The fire module has two strata: squeeze layer and an expand layer. A Squeezenet has a pile of fire modules that are accompanied with minute pooling layers. Notably, both squeeze layer and the expand layer parallel activation map size. Nevertheless, the depth decreases in squeeze layer, while it surges the expand layer. Meanwhile, the neural models have a bottleneck layer and a trait of increasing. Besides, they have a tendency to generate a pattern of a surging depth, while minimizing activation map size with the intention of obtaining an elevated level abstract.

As illustrated in the chart above, the squeeze segment only has 1x1 filters. The latter implies that it operates as completely-attached layer dealing with feature points on the same area. The spatial abstract cannot be derived from the system. Notably, it helps in decreasing the depth of feature map, which simplifies and increases the speed of computation in the expand layer.

Reducing depth enables to achieve an effect when the number of computations performed by 3x3 filters in the expand layer decreases. The speed increases due to the fact that 3x3 filters required only as much computation as 1x1 filters need. Logically speaking, too much squeezing will decrease the exchange of data, while limited number of 3x3 filters is bound to inevitable limit space resolution. The SqueezeNet architecture enabled a 50X decrease in the size of the model when compared to AlexNet; it is worth noting that top-1 as well as top-5 accuracy of AlexNet was exceeded.

B. Cross Validation

The suggested system uses SVM for classifying the different genders on the basis of train-test split approach that is one of techniques for cross-validation. The predictive patterns are computed through the division of the basic model to create a training set to train the system, and set test for evaluation.

SVM was previously proposed By Vapnik [25]. The support vector machine is among the maximum margin classifiers and based on the Structural Risk Minimization. SVM plan input silhouette to a top dimensional domain where the uppermost separating hyperplane is gathered. Linear support vector machine is primarily pronounced for binary classification.

Take for instance that the training data set as well as the labels (x_n, y_n), $n=1,...,N$, $x_n \in \mathbb{R}^D$, $t_n \in \{-1, +1\}$, SVMs learning has the future controlled optimization:

$$\min_{w, \xi_n} = \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_n$$
$$s.t. \quad w^T x_n t_n \geq 1 - \xi_n \quad \forall_n \quad (7)$$
$$\xi_n \geq 0 \quad \forall_n$$

where ξ_n are the slack variables, w is the vector of coefficients, and C is the capacity constant.

The unconstrained optimization problem as in (8) which is recognized as the primal form problem of L1-SVM:

$$min_w = \frac{1}{2} w^T w + C \sum_{n=1}^{N} \max(1 - w^T x_n t_n, 0) \quad (8)$$

Meanwhile L1-SVM is not differentiable, the L2-SVM is used which minimizes the squared hinge loss as in (9):

$$min_w = \frac{1}{2} w^T w + C \sum_{n=1}^{N} \max(1 - w^T x_n t_n, 0)^2 \quad (9)$$

To expect the class label of a test data x:

$$arg_t \max(w^T x) \, t \quad (10)$$

To extend SVMs for multiclass problem, one-vs-rest approach is used. Representing the output of the *k-th* SVM as in (11):

$$a_k(x) = W^T x \quad (11)$$

the forecast class is

$$arg_k \max a_k(x) \quad (12)$$

C. Score Fusion

A key question that should be addressed when handling information on combining systems is the nature of data that needs to be merged using the fusion module. There are several approaches that have been devised to help in the latter at different levels including the sensor level, feature level, rank level, as well as the decision level. In line with our recommended system, we will analyze combining systems considering it is the principal level of fusion at the match score level [26].

At the match score level, the relationship between input and prototype biometric feature silhouettes. Synthesis at the core level entails the successful combination of various match scores output to obtain a biometric recognition decision [26]. Some of the strategies used in consolidating the different biometric matchers include Majority Voting, and Weighted Majority Voting. Our proposed system recommends the use of Mode Voting Technique (MVT) when consolidating information at the decision level.

The method utilizes the standard class label values that are retrieved from the predicted label array obtained through the SVM discriminate classifier. MVT is used to establish the common non-repeated values in the predict label array X for the purpose of biometric identification.

$$Z = \text{mode}(X_{k,i}) \quad (13)$$

where *Z* is the class label of the test image, *k* is the index of the test image, and *i* is the index of the descriptor.

IV. EXPERIMENTAL RESULTS

As apparent in the methodology section, two databases (11k Database and CASIA Database) were used to obtain the relevant data records. The 11k database had hand images taken from individuals whose photos were stored in jpg formats.

The hand images displayed the palm/dorsal, and left and right images of the individuals. In total, images of 190 subjects, aged between 18 and 75, were retrieved from the database. As for the CASIA database, it had 5,503 palmprint images belonging to 312 persons and representing both the left and right palms.

The images of JPEG format each 8 bit gray-level. Slight adjustments were made to the images to facilitate CNN feature extraction. Specifically, they were modified to RGB and minimized to 227×227 pixels.

SVM was applied to discriminately sort the features. The features were retrieved from single-level 2D DWT through the application of Haar wavelet filter. Through this process, DWT provided an approximation coefficient matrix cA as well as detail coefficients matrices cH, cV, and cD (horizontal, vertical, and diagonal, respectively).

Two experiments are carried out over 11k database, one on all dataset, and the other on the dataset after extracting hand images with accessories. The predicted label arrays obtained from SVM as a classifier for the obtained features are fused using mode voting technique. Table I and Table II summarize the results obtained for both experiments.

The recognition rate for all dorsal dataset ranged from 97.57% to 99.71%, with mode voting technique on the resulting 8 predicted label arrays, the recognition rate reached 99.86% with processing time 0.27 sec for each image.

For all palmer dataset ranged from 96.86% to 98.57%, with mode voting technique on the resulting 8 predicted label arrays, the recognition rate reached 100% with processing time 0.28 sec for each image.

By Applying mode voting technique on the resulting 16 predicted label arrays for previous both dorsal and palmar images, the recognition rate reached 100% with processing time 0.55 sec for each image.

After excluding images with accessories, the recognition rate for dorsal dataset ranged from 97.27% to 99.64%, and for palmer dataset ranged from 97.27% to 99.27%. Also, after applying mode voting technique on the resulting 16 predicted label arrays, the recognition rate reached 100% with processing time 0.57 sec for each image.

For CASIA database, the experiment is carried out over the database which contains only palmprint images. Table III summarize the obtained results.

The recognition rate for palmer dataset ranged from 81% to 97.75%. After applying mode voting technique on the resulting 8 predicted label arrays (as the database only contain palmar side), the recognition rate reached 98% with processing time 0.25 sec for each image as in Table III.

TABLE I. RESULTS OF 11K DATABASE FOR ALL IMAGES

Descriptor	dwt2	dorsal L	dorsal R	palmer L	palmer R
squeezenet fire9-concat	cA	99.43	99.71	97.29	98.57
	cH	98.71	97.57	97.71	98.29
	cV	98.75	98.29	97.43	98.00
	cD	98.14	97.86	96.86	98.43
Voting		99.86		100	
		100.00			
Sys. Time/image in sec.		0.27		0.28	
		0.55			

TABLE II. RESULTS OF 11K DATABASE EXCLUDING ACC. IMAGES

Descriptor	dwt2	dorsal L	dorsal R	palmar L	palmar R
squeezenet fire9-concat	cA	98.73	99.64	99.27	98.73
	cH	98.00	97.82	98.00	97.27
	cV	98.18	98.00	98.00	98.00
	cD	97.45	97.27	97.27	97.82
Voting		100		100	
		100.00			
Sys. Time/image in sec.		0.57			

TABLE III. RESULTS OF CASIA DATABASE

Descriptor	dwt2	palmar L	palmar R
squeezenet fire9-concat	cA	97.75	95.25
	cH	92.25	82.50
	cV	87.75	88.50
	cD	82.00	81.00
Voting		98.00	
Sys. Time/image in sec.		0.25	

It is noted that, the high recognition rate is nearly obtained from the approximation coefficient matrix cA in both databases.

V. DISCUSSION

From the results, it is clear that the fusion at decision level using the mode voting technique guarantees an excellent recognition rate regardless of low recognition rate of some datasets. The mode voting technique ranks top of the list of SVM classifiers used for each database.

Table IV shows the performance comparison of proposed system CWNN using 11k and CASIA databases. It is clear that the performance is stable for both databases, the dorsal and palmar results in 11k database are nearly the same, so one of them can be used alone to reduce the system processing time. Also, the processing time for two databases is nearly the same for each image.

TABLE IV. PERFORMANCE COMPARISON OF PROPOSED SYSTEM CWNN USING 11K AND CASIA DATABASES

Database	11k			CASIA
Hand	Dorsal L,R	Palmar L,R	All	Palmar L,R
No of Images	5680	5396	11076	5502
Feature extraction	Squeezenet & dwt2			
Classifier	SVM			
No of predicted arrays	8	8	16	8
Result by voting	99.86	100	100	98
Sys. Time/image in sec.	0.27	0.28	0.55	0.25

TABLE V.   PERFORMANCE COMPARISON OF PROPOSED SYSTEM CWNN
VS. EXISTING SYSTEMS USING 11K

11k	Afifi 2017		CWNN	
	Dorsal	Palmar	Dorsal	Palmar
All Images			99.86	100
no accessories	97.30	94.20	100	100

Table V shows the performance comparison of the proposed systems CWNN vs. existing systems in the previous work that use 11k database.

VI. CONCLUSION

This work reports a CWNN-Net: a new convolution wavelet neural network for gender classification using palm print system based on the mode voting technique, and compares the performance of the system using two datasets.

The novelty comes from using mode voting technique at decision level. Our experimental results show the efficiency of the suggested system. Using SqueezeNet and DWT in this system shows a promising result due to the advantage of SqueezeNet and wavelet decomposition.

We believe that the mode voting technique can serve as a step towards the construction of more accurate gender recognition and biometric identification systems. In the future, we will adopt this deep learning method in real-time palmprint recognition system and develop a more intelligent machine learning algorithm for feature extraction in palmprint recognition.

REFERENCES

[1] G. Amayeh, G. Bebis and M. Nicolescu, "Gender classi_cation from hand shape", In: Computer Vision and Pattern Recognition Workshops. CVPRW'08. IEEE Computer Society Conference on, IEEE, 2008, pp. 1-7.

[2] M. Afifi, "11K Hands: Gender recognition and biometric identification using a large dataset of hand images." arXiv preprint arXiv:1711.04322, 2017.

[3] W. Ming and Y. Yubo, "Gender Classification Based on Geometry Features of Palm Image", the Scientific World Journal, vol. 2014, Article ID 734564, 2014.

[4] X. Font-Aragones, and, M. Faundez-Zanuy, "Hand-Based Gender Recognition Using Biometric Dispersion Matcher", in: Apolloni, B., Bassis, S., Esposito, A., & Morabito, F., (eds) Neural Nets and Surroundings. Smart Innovation, Systems and Technologies, vol 19. Springer, Berlin, Heidelberg, 2013.

[5] B. A. Golombn and D. T. Lawrence, "SEXNET: A neural network identifies sex from human faces", Advances in Neural Information Processing Systems, 1991, pp. 572–577.

[6] C. Yan, "Face Image Gender Recognition Based on Gabor Transform and SVM", In: Shen, G., & Huang, X., Advanced Research on Electronic Commerce, Web Application, and Communication. Communications in Computer and Information Science, vol 144. Springer, Berlin, Heidelberg, 2011.

[7] F. S. Iglesias, M. E. Buemi, D. Acevedo and J. Jacobo-Berlles, " Evaluation of Keypoint Descriptors for Gender Recognition", In: Bayro-Corrochano, E., & Hancock, E., (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Lecture Notes in Computer Science, vol. 8827. Springer, Cham, 2014.

[8] Y. Akbulut, A. Şengür, and S. Ekici, "Gender recognition from face images with deep learning", International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, 2017, pp. 1-4.

[9] J. Wang, W. Hu, Z. Wang and Z. Chen, "Human Identification and Gender Recognition from Boxing", in: Sun, Z., Lai, J., Chen, X., & Tan, T., (eds) Biometric Recognition. Lecture Notes in Computer Science, vol 7098. Springer, Berlin, Heidelberg, 2011.

[10] W. Yang, C. Chen, K. Ricanek and C, Sun, "Gender Classification via Global-Local Features Fusion", In: Sun, Z., Lai, J., Chen, X., & Tan, T. (eds) Biometric Recognition. Lecture Notes in Computer Science, vol. 7098. Springer, Berlin, Heidelberg, 2011.

[11] G. Azzopardi, A. Greco and M. Vento, "Gender Recognition from Face Images Using a Fusion of SVM Classifiers", In: Campilho A., Karray F. (eds) Image Analysis and Recognition. Lecture Notes in Computer Science, vol 9730. Springer, Cham, 2016.

[12] A. R. Ardakany and S. J. Louis, "Improving Gender Recognition Using Genetic Algorithms", In: Bui L.T., Ong Y.S., Hoai N.X., Ishibuchi H., Suganthan P.N. (eds) Simulated Evolution and Learning. Lecture Notes in Computer Science, vol 7673. Springer, Berlin, Heidelberg, 2012.

[13] C. Ng, Y. Tay and B. Goi, "A Convolutional Neural Network for Pedestrian Gender Recognition", In: Guo, C., Hou, ZG., & Zeng, Z., (eds) Advances in Neural Networks. Lecture Notes in Computer Science, vol 7951. Springer, Berlin, Heidelberg, 2013.

[14] M. Collins, P. Miller and J. Zhang, "Gait Based Gender Recognition Using Sparse Spatio Temporal Features", In: Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H., & O'Connor, N., (eds) MultiMedia Modeling. MMM 2014. Lecture Notes in Computer Science, vol 8326. Springer, Cham, 2014.

[15] B. Hatipoglu and C. Kose, "A gender recognition system from facial images using SURF based BoW method ", International Conference on Computer Science and Engineering (UBMK), Antalya, 2017, pp. 989-993.

[16] C. Nistor, A. C. Marina, A. S. Darabant and D. Borza , "Automatic gender recognition for "in the wild" facial images using convolutional neural networks", 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, 2017, pp. 287-291.

[17] M.V.M Cirne and H. Pedrini, "Gender recognition from face images using a geometric descriptor", IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, 2017, pp. 2006-2011.

[18] T. Liu, B. Sun, M. Chi and X. Zeng, "Gender recognition using dynamic gait energy image", IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, 2017, pp. 1078-1081.

[19] G. Azzopardi, A. Greco, A. Saggese and M. Vento, "Fusion of domain-specific and trainable features for gender recognition from face images", IEEE Access, vol. PP, no. 99, 2018.

[20] CASIA Palmprint Database, http://biometrics.idealtest.org/

[21] G. Kumar and P. K. Bhatia, "A detailed review of feature extraction in image processing systems", In Advanced Computing & Communication Technologies (ACCT), Fourth International Conference, 2014, pp. 5-12.

[22] R. Haddadi, E. Abdelmounim, M. El Hanine and A. Belaguid, "Discrete Wavelet Transform Based Algorithm for Recognition of QRS Complexes", World of Computer Science & Information Technology Journal, 4(9), 2014.

[23] T. Williams and R. Li, "December. Advanced image classification using wavelets and convolutional neural networks", In Machine Learning and Applications (ICMLA), 15th IEEE International Conference, 2016), pp. 233-239.

[24] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally and, K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size", 2016.

[25] V. Vapnik, "The Nature of Statistical Learning Theory ". NY: Springer-Verlag, 1995.

[26] A. A. Ross, K. Nandakumar and A. Jain, "Handbook of multibiometrics", Vol. 6, Springer Science & Business Media, 2006, pp. 73-82.