



# Wavelet Entropy Based Probabilistic Neural Network for Classification

**Khaled Daqrouq<sup>1</sup>, Sheng Chen<sup>2</sup>, Emad Khalaf<sup>1\*</sup>, Ali Morfeqa<sup>1</sup>,  
Muntasir Sheikha<sup>1</sup>, Abdulrohman Qatawneh<sup>3</sup> and Abdulhameed AL-Khateeb<sup>1</sup>**

<sup>1</sup>Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah 21589,  
Saudi Arabia.

<sup>2</sup>Department of Electronics and Computer Sciences, University of Southampton,  
Southampton SO17 1BJ, UK.

<sup>3</sup>Higher Colleges of Technology, Sharjah, United Arab Emirates.

## Authors' contributions

*This work was carried out in collaboration among all authors. Author KD proposed the idea of the study, wrote the software and wrote the first draft of the manuscript. Author SC managed the writing. Author EK managed the editing and the whole procedure. Author AM did the editing and project management. Author MS did a scientific consultation for better results. Author AQ did a project management and software building. Author AAK did the result analysis and paper editing. All authors read and approved the final manuscript.*

## Article Information

DOI: 10.9734/CJAST/2019/v34i530145

### Editor(s):

(1) Dr. Wei Wu, Professor, Department of Applied Mathematics, Dalian University of Technology, China.

### Reviewers:

(1) Yahaya Shagaiya Daniel, Kaduna State University, Nigeria.

(2) Zlatin Zlatev, Trakia University, Bulgaria.

Complete Peer review History: <http://www.sdiarticle3.com/review-history/48124>

**Original Research Article**

**Received 18 January 2019**

**Accepted 26 March 2019**

**Published 11 April 2019**

## ABSTRACT

Recently, wavelet transform (WT) has been enormously effectual in various scientific fields. As a matter of fact, WT has overcome the FFT in the difficult nature data tackling. A wavelet entropy based probabilistic neural network (PNN) for classification applications is proposed. Specifically, wavelet transform is performed on the original input feature data, and the entropy values of the wavelet decomposition signals are then extracted to use as the input to the PNN classifier. Two benchmark data sets, Breast Cancer and Diabetes, are used to demonstrate the efficiency of our proposed wavelet entropy based PNN (WEPNN) classifier. The test classification rates of 80.3% and 77.0% are achieved respectively for the two data sets using the WEPNN with Shannon entropy. Other published methods are used for comparison. The method is promising. For results accuracy enhancement, large data set might be utilized in the future work.

\*Corresponding author: E-mail: [ekhalaf@kau.edu.sa](mailto:ekhalaf@kau.edu.sa);

**Keywords:** Wavelet transform; entropy; probabilistic neural network classifier.

## 1. INTRODUCTION

In the last two decades, wavelet transform has been broadly investigated. It has been used in many engineering areas [1-6]. The wavelet investigation is based on expansion and dilation of a function known as mother wavelet. Any signal of interest can be decomposed by wavelet transform to be implemented for different tasks. By adjusting these wavelet function, many signal processing algorithms may be created. In theory, the wavelet scale parameter should be with a positive real value and the translation process with an arbitrary real number [1,5] is taking into consideration. However, for computational efficiency, these parameters (the shift and scale parameter) are often constrained to some discrete lattices [7], [8]. The wavelet packet transform (WPT) decomposes the signal into a recursive form of a binary tree [9]. The difference between WPT and discrete wavelet transform (DWT) is that WPT decomposes into low pass frequency sub signals and high pass frequency sub signals unlike DWT that works at low pass part of frequency only. Therefore, WPT features have better recognition performance over DWT features [10]. Avci, Varol and Hanbay [11] suggested a method in digital modulation recognition task based on the entropy value of the wavelet norm. In [12] Wu Lin, proposes the energy indexes of WP for speaker identification. In another research, sure entropy is used for the signals' DWT sub signals for speaker identification task [13].

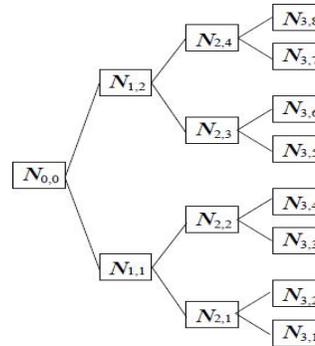
Since the derivation of the probabilistic neural network (PNN) in [14], to the original algorithm, many enhancements, modifications, and extensions have been suggested, that which aim to improving either the classification accuracy of PNNs, or the learning capability, or optimizing the network size [15]. Neural networks as an excellent classifiers, their performance depends, in general, on the quality of training samples and their size [16,17]. In many applications Fuzzy and wavelet theory has been used successfully [18], which demonstrates that the performance of neural networks can be improved incorporating neural fuzzy or wavelet techniques, especially through input matrix dimensionality reduction [19].

In this work, we develop a novel wavelet entropy based PNN (WEPNN) for data classification. We

first perform the wavelet transform on the original input feature data using WPT or DWT. Several types of entropy can be utilized to calculate the entropy values for the wavelet decomposed input feature data, which are then used as the feature input to feed into the PNN classifier. The motivation behind our choice of the WEPNN is two folds. Firstly, the wavelet and entropy techniques enable the creation of the uncorrelated features that are stable over a long period and are very differ from class to class, and this ability of working in a constructive unsupervised mode on the feature input will enhance the performance of the PNN classifier. Secondly, the PNN classifier makes the classification decision directly following the Bayes theorem [17]. Two data sets benchmark, Diabetes and Breast Cancer, are utilized to demonstrate the efficiency of our proposed WEPNN classifier.

## 2. WAVELET TRANSFORM FOR FEATURE EXTRACTION

We use wavelet transform for “feature extraction” by performing WPT or DWT on the original feature input data. In the DWT, each level of the decomposition process is calculated based only on the previous wavelet approximation coefficients while both the detail sub signals and approximation sub signals are generated to get the full image of the data [9].



**Fig. 1. Wavelet packet at depth 3**

### 2.1 Wavelet Packet Transform

The wavelet packet (WP) technique as a special case of the wavelet decomposition is explained in this section. The mother wavelet function as the basis function of the linear combination is defined by

$$\Psi_{a,b}(t) = \Psi\left(\frac{t-b}{a}\right), \quad (1)$$

Where  $a$  is the scale parameter and  $b$  is the shift parameter. Changing the magnitudes of parameters  $a$  and  $b$ , will scale and dilate the mother wavelet. Then, by the inner product of the signal ( $x(t)$  the data function) with  $\Psi(t)$  the mother wavelet the wavelet transform is calculated

$$W_{\Psi x}(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) * \Psi\left(\frac{t-b}{a}\right) dt, \quad (2)$$

Fig. 1 illustrates the WPT's recursive binary tree. A pair of filters, high-pass  $g[n]$  and low-pass  $h[n]$ , are utilized to create two sequences to illustrate the original signal into different frequency sub-band features. The wavelet orthogonal bases taken from a previous node are written as:

$$\Psi_{j+1}^{2p}(K) = \sum_{n=-\infty}^{\infty} h[n] \Psi_j^p(k - n/2) \quad (3)$$

$$\Psi_{j+1}^{2p+1}(K) = \sum_{n=-\infty}^{\infty} g[n] \Psi_j^p(k - n/2) \quad (4)$$

Where  $\Psi(n)$  is the wavelet function, while  $j$  is a wavelet transform level and  $p$  is the previous nodes level number [20].

## 2.2 Discrete Wavelet Transform

Taking the inner product of the wavelet functions with the signal, the DWT coefficients are defined. A simple algorithm known as Mallat's pyramid [21] is produced by forming the wavelet functions from a shifting and scaling of these wavelet functions. The DWT coefficients are expressed as [22]:

$$w_L(n, k) = \sum_i W_L(i, k-1) h(i-2n) \quad (5)$$

$$w_H(n, k) = \sum_i W_L(i, k-1) g(i-2n) \quad (6)$$

$W_L(p, j)$  is the  $j^{\text{th}}$  level where the scaling coefficient  $p^{\text{th}}$  is calculated, and  $W_H(p, j)$  is the  $j^{\text{th}}$  level where the wavelet coefficient  $p^{\text{th}}$  is calculated, while  $h(n)$  and  $g(n)$  are the dilation and scaled versions from the scaling and mother wavelet functions.

The multi-resolution decomposition is formed using the DWT, which generates an output sequence out of the input of a length  $N$ . The

output sequence at the highest frequency band width (level 1) has  $N/2$  values, at the next frequency bandwidth (level 2) has  $N/4$  values, and so on. Consider  $N = 2^m$ , where  $m$  is the number of the frequencies. Then  $k$  representing the frequency index will be vary in  $\{1, 2, \dots, m\}$  and corresponds to the scales  $2^1, 2^2, \dots, 2^m$ , as defined by the Mallat pyramid algorithm.

## 3. FEATURE EXTRACTION PROCEDURE

The entropy is excellent to describe the image, signal or any information data features. One of the examples of using entropy is to give information about the concentration of the image [20]. Another example is measuring the ordering of non-stationary processes. Our proposal is to use the wavelet entropy for data features extraction. Three properties are important to consider while thinking of the features [21]: 1) the features should be different from class to class; 2) repetitive for huge cases; and 3) should be uncorrelated with other features. In the following part the feature extraction will be explained. The data matrix maybe represented as

$$\Gamma = \begin{bmatrix} \delta_{1,1} & \delta_{1,2} & \dots & \delta_{1,Z} \\ \delta_{2,1} & \delta_{2,2} & \dots & \delta_{2,Z} \\ \vdots & \vdots & \vdots & \vdots \\ \delta_{N,1} & \delta_{N,2} & \dots & \delta_{N,Z} \end{bmatrix} \quad (7)$$

Where  $\delta_{ij}$  are the original data, and the columns of  $\Gamma$  contains the data for the cases. In this example, we have the  $Z$  cases. The feature extraction procedure by wavelet entropy is summarized as follows: DWT or WPT is calculated for each case (column). DWT and WPT will be tested separately.

When we calculate the DWT for each column vector in the matrix at a chosen level, the matrix  $D_{DWT}$  is obtained as

$$D_{DWT} = [cd_1 \quad cd_2 \quad \dots \quad cd_L \quad ca_L] \quad (8)$$

Where  $cd_l, 1 \leq l \leq L$ , are the detailed DWT subsignals, and  $ca_L$  is the approximation DWT subsignal. In this study,  $L = 5$ . Primarily, the approximation subsignals and the resulting detailed are at different lengths. Therefore, they are resized to have the same length in the matrix eq.8.

When the WPT is executed, the resulting WPT subsignals are represented as

$$D_{WPT} = [c_1 \quad c_2 \quad \dots \quad c_H] \quad (9)$$

Where  $c_h$ ,  $1 \leq h \leq H$ , denote the WPT nodes, while  $H$  is the number of the WPT nodes, and it rely on the levels used. In this study, WPT with level three is used. The next step is to evaluate the entropy for each column in the matrices  $D_{DWT}$  or  $D_{WPT}$  produced out of the WT by means of DWT or WPT. Different types of entropy may be employed, and they include:

*The Shannon entropy*

$$E(s) = - \sum_{\tau} s_{\tau}^2 \log(s_{\tau}^2) \quad (10)$$

*The log-energy entropy*

$$E(s) = - \sum_{\tau} \log(s_{\tau}^2) \quad (11)$$

*The threshold entropy*

$$E(s) = \text{the number of times that } |s_{\tau}| > P, \quad (12)$$

*And the sure entropy*

$$E(s) = N - \{ \text{the number of times that } |s_{\tau}| \leq \mathbb{P} \} + \sum_{\tau} \min\{s_{\tau}^2 \mathbb{P}^2\} \quad (13)$$

Where  $\mathbb{P}$  is the chosen threshold,  $s$  is the input signal, which is a column vector of  $D$ , and  $s_{\tau}$  are the elements of the input signal.

The entropy calculation results are arranged as the output feature vector

$$f_e = [f_{e_1} f_{e_2} \dots f_{e_{N_e}}] \quad (14)$$

Where  $N_e$  is the number of elements in  $f_e$ , which is the same the number of wavelet transform subsignals. Then, the generated  $f_e$  is the feature vector, which is the input to the classifier.

#### 4. PNN CLASSIFIER

The PNN classifier is used. Fig. 2 depicts the PNN scheme for classification for  $K$  classes. The layer number one of the PNN is the known as an input layer that receives  $x_p$  (input) of dimension  $d$  to be classified. Now, there is a node called the pattern layer in the second layer, are collected in the  $K$  groups related to the classes that they belong to. In the second layer, the nodes are connected as kernels

$$f_{i,j}(x_p; c_{i,j}, \sigma) = \frac{1}{(2\pi)^{d/2} \delta^d} \exp\left(-\frac{1}{2\delta^2} (x_p - c_{i,j})^T (x_p - c_{i,j})\right) \quad (15)$$

Where  $1 \leq i \leq K$ ,  $1 \leq j \leq M_i$  and  $M_i$  is the number of pattern units in the class  $i$ ,

while the superscript  $\top$  denotes the transpose operator  $c_{i,j} \in \mathbb{R}^d$  is the pattern unit's kernel center vector, and  $\sigma > 0$  is the smoothing or spreading parameter. The total number of the nodes in the second-layer is obviously given by

$$M = \sum_{i=1}^K M_i \quad (16)$$

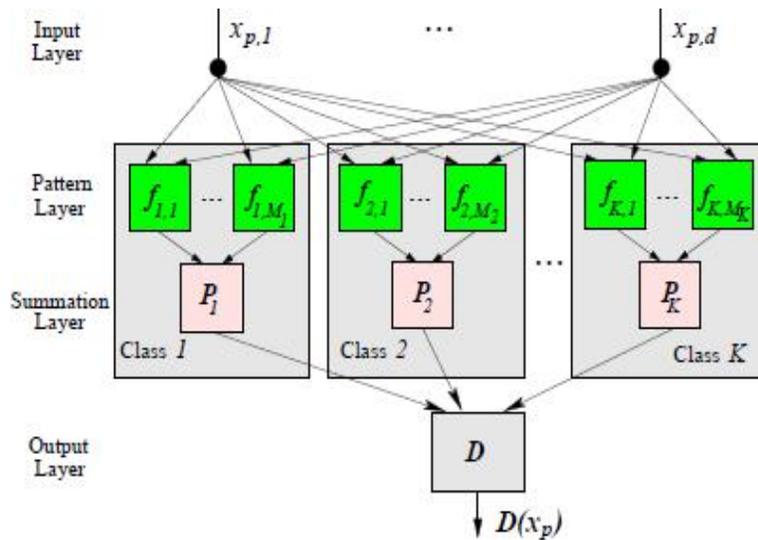


Fig. 2. The probabilistic neural network structure

Her outputs  $f_{ij}$  of the pattern units that belong to the group  $i$  are weighted by the coefficients  $\omega_{i,j} > 0$  and connected to the  $i$ th node of the third summation layer to form the class distribution for the class  $i$ . Generally, the positive weights associated with the  $i$ th node of the third layer need to fulfill

The condition

$$\sum_{j=1}^{M_i} \omega_{i,j} = 1, \quad 1 \leq i \leq K \quad (17)$$

And  $\omega_{i,j}$  is associated with the *previous* probability of  $f_{i,j}$ . The single node of the fourth output layer chooses the largest decision among the  $K$  class distributions and, therefore, decides the class of  $x_p$ .

We use the Matlab function, the NEWPNN [23], to train the PNN classifier.

### 5. RESULTS AND DISCUSSION

Two data sets, Breast Cancer and Diabetes, taken from [24,25], were utilized in our research. Each of them contains 100 realizations, and each realization has  $N_{tr}$  training patterns and  $N_{ts}$  test patterns, while each pattern is described by a vector in the  $m$  dimensional feature space. We used the first five realizations in our experiment and the results were averaged over these five realizations. In Table 6, 10 realizations and 15 realizations were tested. The two data sets are summarized in Table 1 [26].

**Table 1. Description of the two data sets [26]**

Data set	$m$	$N_{tr}$	$N_{ts}$
Breast Cancer	9	200	77
Diabetes	8	468	300

Firstly, The Breast Cancer data set is experimented. With Shannon entropy and the WPT of three levels, different wavelet functions were used in the proposed method, Table 2 shows the testing classification results. With Shannon entropy and the wavelet function Db1, the achieved test classification rate so for proposed system are summarized in Table 3 as the function of the WPT levels employed. Furthermore, the test classification rates achieved by our WEPNN classifier for three different entropy calculations are listed in Table 4, given the wavelet function Db1 and the WPT of three levels. In all these experiments, the spread  $\sigma$  was set to a value of 0.3. We now demonstrate that this choice was "optimal". Fig.3 illustrates the achievable test classification rate as the function of the spread parameter  $\sigma$ , given Shannon entropy, the wavelet function Db1, and the WPT of three levels. It can be seen from Fig.3 that the test classification rate is maximized at  $\sigma=0.3$ . The results achieved by our WEPNN classifier with Shannon entropy, and the WPT of three levels for Breast Cancer were compared with two other PNNs, and they are: a) WPID, which utilizes the Wavelet packet energy index distribution method [27]; b) GWPNN, which utilizes genetic wavelet packet neural network [27]; c) FFTPNN, which applies fast Fourier transform (FFT) directly on the original data columns to extract features [28] and then uses the resulting feature vectors to train the PNN classifier; and d) DWEPNN, which utilizes Shannon entropy and the DWT of five levels for the feature extraction and applies the resulting feature vectors to train the PNN classifier [21]. The results achieved by these:

Two methods are summarized in Table 5, in comparison with our proposed WEPNN.

**Table 2. Test classification rates achieved by the proposed system with Shannon entropy and the WPT of three levels for different wavelet functions. The data set is breast cancer**

Db1	Db2	Db5	Db10	Sym1	Sym8	Sym9	Sym10
80.33%	66.33%	63.33%	63.30%	80.33%	63.33%	61.33%	61.33%
Bior1.1	Bior2.2	Bior3.5	Bior3.7	Coif1	Coif2	Coif4	Coif5
80.33%	65.33%	63.34%	53.24%	65.00%	63.67%	63.00%	56.00%

**Table 3. Test classification rates achieved by the proposed system with Shannon entropy and the wavelet function Db1 for different WPT levels. The data set is breast cancer**

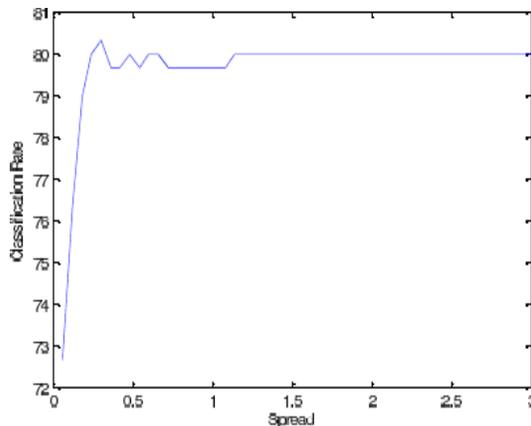
Level 2	level 3	level 4	Level 5	Level 7
78.63%	80.33%	72.00%	72.76%	77.786%

**Table 4. Test classification rates achieved by the proposed system with the wavelet function Db1 and the WPT of three levels for different entropy calculations. The data set is breast cancer**

Log energy	Threshold	Shannon	Sure
68.67%	72.00%	80.33%	44.33%

**Table 5. Comparison of test classification rates achieved by the five methods for the breast cancer data set**

Method	Feature dimension $d$	Test classification rate
WPID [24]	15	78.35%
GWPNN [25]	15	73.61%
FFTNN [23]	8	77.69%
DWENN [18]	5	74.67
Proposed WEPNN	15	80.33%



**Fig. 3. The test classification rate as the function of the spread parameter, given Shannon entropy, the wavelet function Db1, and the WPT of three levels. The data set is breast cancer**

**Table 6. The classification rates achieved by the proposed methods for 10 and 15 realizations of the breast cancer data set**

Method	No. Realizations $d$	Test classification rate
Proposed WEPNN	10	80.30%
Proposed WEPNN	15	79.1%

In Table 6, the first 10 or 15 realizations were tested and the results were averaged over these 10 or 15 realizations. The results were closed to

those achieved by five realizations. For results accuracy enhancement, large data set might be utilized in the future work.

We performed the similar experiment on the Diabetes data set. With Shannon entropy, the WPT of three levels, Db1 or Sym1 or Bior1.1 wavelet function as well as the spread parameter of  $\sigma = 0.3$ , we achieved the test classification rate of 77%.

## 6. CONCLUSIONS

In the proposed study, a wavelet entropy based probabilistic neural network (WEPNN) for classification applications, has been proposed. Wavelet transform in its two versions, discrete and packet have been tested. Several wavelet functions belong to different families with entropy formulations, over wavelet packet transform for feature extraction have been investigated. It has been shown that this feature extraction process aids the probabilistic neural network classifier. Initial testing results illustrate the promising potential of the proposed method. In particular, the WEPNN classifier using the WPT of three levels as well as Shannon entropy and certain types of wavelet functions achieves good test classification performance.

## ACKNOWLEDGEMENT

The authors acknowledge the financial support of King Abdulaziz University (GrantNo.1-4-1432/HiCi).

## COMPETING INTERESTS

Authors have declared that no competing interests exist.

## REFERENCES

1. Antonini M, Barlaud M, Mathieu P, Daubechies I. Image coding using wavelet transform. IEEE Trans. Image Processing. 1992;1(2):205–220.
2. Quiroga RQ. Quantitative analysis of EEG signals: Time-frequency methods and chaos theory. PhD Thesis, Medical University Lu"beck Germany; 1998.
3. Chen CC, Chen CT, Tsai CM. Hard-limited Karhunen-Loeve transform for text independent speaker recognition. Electronics Letters. 1997;33(24):2014–2016.
4. Lung SY, Chen CC. Further reduced form of Karhunen-Loeve transform for text

- independent speaker recognition. Electronics Letters. 1998;34(14):1380–1382.
5. Mallat S. A wavelet tour of signal processing: The sparse way. 3<sup>rd</sup> Edition. Academic Press, San Diego, CA; 1998.
  6. Vetterli M, Kovac'evic' J. Wavelets and sub-band coding. Prentice Hall, Englewood Cliffs, NJ; 1995.
  7. Mallat S, Zhong S. Characterization of signals from multi scale edges. IEEE Trans. Pattern Analysis and Machine Intelligence. 1992;14(7):710–732.
  8. Mallat S. Zero-crossings of a wavelet transform. IEEE Trans. Information Theory. 1991;37(4):1019–1033.
  9. Daubechies I. Ten lectures on wavelets. SIAM; 1992.
  10. Lung SY. Feature extracted from wavelet eigenfunction estimation for text-independent speaker recognition. Pattern Recognition. 2004;37(7):1543–1544.
  11. Avci E, Hanbay D, Varol A. An expert discrete wavelet adaptive network based fuzzy inference system for digital modulation recognition. Expert System with Applications. 2006;33:582-589.
  12. Wu JD, Lin BF. Speaker identification using discrete wavelet packet transform technique with irregular decomposition. Expert Systems with Applications. 2009;3(2):6313-63143.
  13. Avci D. An expert system for speaker identification using adaptive wavelet sure entropy. Expert Systems with Applications. 2009;36:6295-6300.
  14. Specht DF. Probabilistic neural networks. Neural Networks. 1990;3(1):109–118.
  15. Specht DF. Enhancements to probabilistic neural networks. In: Proc. IJCNN 1992 Baltimore, MD. 1992;761–768.
  16. Visser E, Otsuka M, Lee TW. A spatiotemporal speech enhancement scheme for robust speech recognition in noisy environments. Speech Communication. 2003;4:393–407.
  17. Kosko B. Neural networks and fuzzy systems. A dynamical approach to machine intelligence. Prentice Hall, Englewood Cliffs, NJ; 1992.
  18. Gowdy J, Tufekci Z. Mel-scaled discrete wavelet coefficients for speech recognition. In: Proc. ICASSP 2000, Istanbul, Turkey. 2000;1351–1354.
  19. Nava PA, Taylor JM. Speaker independent voice recognition with a fuzzy neural network. In: Proc. 5<sup>th</sup> IEEE Int. Conf. Fuzzy Systems, New Orleans, Louisiana. 1996;2049–2052.
  20. Daqrouq K, Al Azzawi KY. Average framing linear prediction coding with wavelet transform for text independent speaker identification system. Computers & Electrical Engineering. 2012;38(6):1467–1479.
  21. Daqrouq K. Wavelet entropy and neural network for text-independent speaker identification. Engineering Applications of Artificial Intelligence. 2011;24(5):796–802.
  22. Vishwanath M. The recursive pyramid algorithm for the discrete wavelet transform. IEEE Trans. Signal Process. 1994;42(3):673–676.
  23. Wasserman PD. Advanced methods in neural computing. Van Nos-trand Reinhold. 1993;35-55.
  24. Ra'tsch G, Onoda T, Mu'ller KR. Soft margins for Ada Boost. Machine Learning. 2001;42(3):287–320.
  25. Available:<http://www.raetschlab.org/Members/raetsch/benchmark>
  26. Xia Honga, Sheng Chen, Abdulrohman Qatawneh, Khaled Daqrouq, Muntasir Sheikhc, Ali Morfeq. A radial basis function network classifier to maximise leave-one-out mutual information. Applied Soft Computing. 2014;23:9–18.
  27. Engin A. A new optimum feature extraction and classification method for speaker recognition. GWPNN, Expert Systems with Applications. 2007;32:485-498.
  28. Quan ZH, Huang DS, Xia XL, Lyu MR, Lok TM. Spectrum analysis based on windows with variable widths for online signature verification. In: Proc. 18<sup>th</sup> ICPR, Aug. 20-24, Hong Kong, China. 2006;1122–1125.

© 2019 Daqrouq et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Peer-review history:*  
*The peer review history for this paper can be accessed here:*  
<http://www.sdiarticle3.com/review-history/48124>