

# WN-based approach to melanoma diagnosis from dermoscopy images

ISSN 1751-9659

Received on 3rd September 2016

Revised 27th January 2017

Accepted on 19th March 2017

E-First on 7th June 2017

doi: 10.1049/iet-ipr.2016.0681

www.ietdl.org

Amir Reza Sadri<sup>1</sup> ✉, Sepideh Azarianpour<sup>2</sup>, Maryam Zekri<sup>1</sup>, Mehmet Emre Celebi<sup>3</sup>, Saeid Sadri<sup>1</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran

<sup>2</sup>Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

<sup>3</sup>Department of Computer Science, University of Central Arkansas, Conway, AR, USA

✉ E-mail: ar.sadri@ec.iut.ac.ir

**Abstract:** A new computer-aided diagnosis (CAD) system for detecting malignant melanoma from dermoscopy images based on a fixed grid wavelet network (FGWN) is proposed. This novel approach is unique in at least three ways: (i) the FGWN is a fixed WN which does not require gradient-type algorithms for its construction, (ii) the construction of FGWN is based on a new regressor selection technique: D-optimality orthogonal matching pursuit (DOOMP), and (iii) the entire CAD system relies on the proposed FGWN. These characteristics enhance the integrity and reliability of the results obtained from different stages of automatic melanoma diagnosis. The DOOMP algorithm optimises the network model approximation ability rapidly while improving the model adequacy and robustness. This FGWN is then used to build a CAD system, which performs image enhancement, segmentation, and classification. To classify the images, in the first stage, 441 features with respect to colour, texture, and shape of each lesion are extracted. By means of feature selection, these 441 features are then reduced to 10. The proposed CAD system achieved an accuracy of 91.82%, sensitivity of 92.61%, specificity of 91%, and area under the curve value of 0.944 on a challenging set of 1039 dermoscopy images.

## 1 Introduction

The most serious type of skin cancer is malignant melanoma (MM) which is caused by irregular development of skin pigments. This type of cancer has increased dramatically over the past few decades [1–3]. Diagnosis of MM, especially in the early stages of the disease has a significant impact on its treatment [1]. There are several approaches for diagnosing melanoma. A compelling approach is the dermoscopy imaging (DI) method. In DI, a dermatoscope is used to acquire images of the lesion, and then these images are further analysed by a dermatologist. Although dermoscopy images provide a great potential to diagnose the MM immediately, their interpretation is quite time-consuming and demanding on the part of specialists. The recent advancements in computer technology have paved the way for interpreting and

analysing the dermoscopy images in a more expeditious and effective manner [4]. Consequently, there has been a tremendous tendency toward developing computer-aided diagnosis (CAD) systems that takes advantage of dermoscopy images. Standard CAD systems consist of five general stages (see Fig. 1) [5]: (i) image acquisition, (ii) image enhancement, (iii) image segmentation, (iv) feature extraction, and (v) lesion classification. Numerous approaches have been developed for each of these steps.

There is an extensive body of work on the determination of MM from dermoscopy images based on different algorithms such as: supervised learning and maximum a posteriori estimation [5], border and wavelet-based texture analysis [6], classical segmentation algorithms plus  $k$ -nearest neighbour ( $k$ -NN) [7], digital dermoscopy analyser plus artificial Multilayer Perceptron

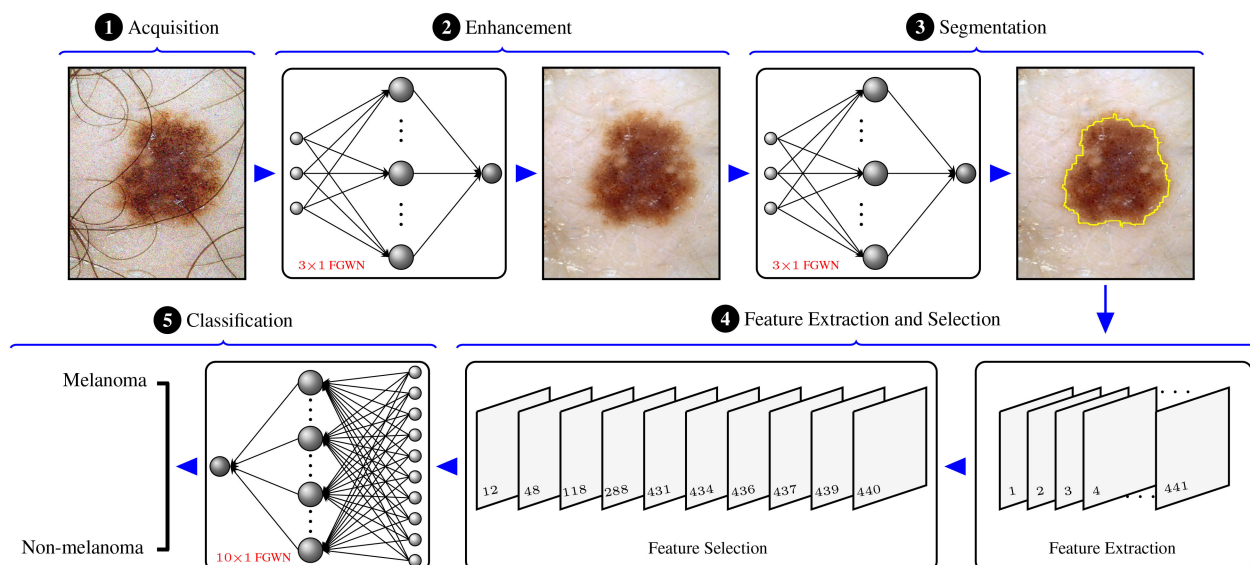


Fig. 1 Schematic representation of the proposed method for MM diagnosis from dermoscopy images

(MLP) [8], content-based image retrieval [9], colour mathematical morphology-based texture descriptors [3], and texture descriptors based on adaptive neighbourhoods [10]. This paper presents a novel system based on wavelet network (WN) for automatic MM diagnosis, which consists of dermoscopy images enhancement, segmentation, and finally melanoma classification.

As stated in [11], WNs are divided into two groups: adaptive WNs (AWNs) and fixed grid WNs (FGWNs). Owing to numerous advantages of the FGWNs such as no need of initial parameters and simpler calculations, they are much more suitable in many applications [4, 11, 12].

The current study proposes an algorithm for building a specific FGWN which is a WN with three layers, one hidden layer,  $d$  inputs, and  $c$  outputs (a  $d \times c$  FGWN). The structure of this algorithm is specified by the following stages. First, the input data is normalised using a linear map. Then an appropriate mother wavelet is chosen, based on which a wavelet lattice is formed. Wavelet lattice is a spatial arrangement of dilated and translated version of wavelets which is applied to the inputs. The very large dimensions of this arrangement should be reduced and effective wavelets should be selected. All these are accomplished through the proposed regressor selection algorithm, D-optimality orthogonal matching pursuit (DOOMP). Once the FGWN is constructed, the DOOMP algorithm is utilised to calculate the network weights and to optimise the network structure. As will be briefly discussed in Section 2, after the formation of wavelet lattice, the WN is a linear-in-the-parameters regression, which can then be solved using the conventional orthogonal least squares (OLS) algorithm. Although OLS is an efficient method, the computational cost of this technique is prohibitive. In this work, to overcome this issue, we use OMP (also known as residual-based selection), which gives the same results as OLS in less time [13]. Additionally, in order to optimise model adequacy and robustness, we use a different cost function based on the D-optimality criterion [14].

The CAD system block diagram which is proposed in this paper consists of five general stages which are illustrated in Fig. 1. After the dermoscopic image is acquired, if it is occluded with dark hair, a hair removal algorithm (based on the proposed  $3 \times 1$  FGWN) is used. Then, in order to segment the dermoscopic images, the R, G, and B pixel values of a dermoscopy image are considered as the inputs of a  $3 \times 1$  FGWN. The network structure is then determined by means of the proposed algorithm, and finally, the output of the network will specify the boundary of the skin lesion. After the segmentation stage, 441 features are extracted from the lesion and its boundary. These features characterise the colour, texture, and the shape of the lesion. To reduce computational time and improve classification accuracy, a feature selector is exploited. In this study, ReliefF feature selector [15] is selected which offers satisfactory results. The number of selected features after applying ReliefF drops to ten. Finally, these selected features are submitted to a  $10 \times 1$  FGWN. Classification is the final step in the diagnosis process, wherein the extracted features are utilised to decide whether the lesion is cancerous or normal. The proposed system is tested on 1039 digital dermoscopy images and the evaluation process is carried out using several metrics.

The main contributions of this paper are:

- To derive a WN approach of the MM diagnosis from dermoscopy images, which is applied to all detection stages (i.e. image enhancement, segmentation, and classification).
- To develop DOOMP algorithm for WN construction. The model class in this work covers a large range of medical image processing problems.
- To study the performance of a technique based on a single algorithm for all diagnosis stages.

## 2 Fixed GWNs

Let  $(\mathbf{x} = [x_1, \dots, x_d]^T, y)$  denote the input–output variable pairs of the network. The network output signal is the weighted sum of

multidimensional-wavelet families  $\psi_{m_i, n_i}$  (dilated and translated versions of the mother wavelet function  $\psi$ )

$$y = \sum_{i=1}^N \theta_i \psi_{m_i, n_i}(\mathbf{x}) = \sum_{i=1}^N \theta_i 2^{-m_i d/2} \psi(2^{m_i} \mathbf{x} - \mathbf{n}_i) \quad (1)$$

where  $\theta_i$  denotes the connection weights from hidden layer to the output layer,  $N$  is the number of wavelets (wavelet neurons) in the hidden layer, and  $m_i$  and  $\mathbf{n}_i$  are scale and shift parameters, respectively. The construction and implementation procedure of the proposed FGWN consist of two major phases: initialisation and optimisation. These phases are described next.

### 2.1 Initialisation of FGWN

In the first phase, after mapping the input data to a certain range (normalisation), the wavelet function, translations, and dilations of wavelet series (1) are selected. To this end, following [4, 11, 12, 16] we use multidimensional single-scaling radial-wavelet frame as a wavelet function. A prime example of these kinds of wavelets is Mexican hat which is expressed as

$$\psi_{m_i, n_i}(\mathbf{x}) = 2^{-m_i d/2} (d - \|\mathbf{x} - \mathbf{n}_i\|^2) \exp(-\|\mathbf{x} - \mathbf{n}_i\|^2 / 2) \quad (2)$$

By applying the selected wavelet function to the input vectors, a library of wavelets will be obtained. The number of these library members is determined by the wavelets scale levels parameter. Assume that the wavelet function support is  $[s_{\min}, s_{\max}]$ . It can be shown if all the inputs of FGWN are limited to be in the interval  $[x_{\min}, x_{\max}]$ , by considering the minimum and maximum scale scalar levels into the  $[m_{\min}, m_{\max}]$  the variation range for the shift parameter  $n$  is  $2^{m_{\min}} x_{\min} - s_{\max} \leq n \leq 2^{m_{\max}} x_{\max} - s_{\min}$  and the number of  $d$ -dimensional wavelets ( $L$ ) is as follows:

$$L = \text{range}(m) \times \text{range}(n) \quad (3)$$

In fact, in order to determine the wavelet library dimension, we only need the chosen range of the dilation index [16].

### 2.2 Optimisation of FGWN

Suppose we have  $M$  input–output pairs. The output vector is considered as  $\mathbf{y} = [y^{(1)} \dots y^{(M)}]^T$ . Now, we apply the  $L$  wavelets which we obtained in the previous phase to the  $M$  model samples and gather the results in a matrix form as follows:

$$\mathbf{W} = \begin{bmatrix} \psi_{m_{\min}, n_{\min}}(\mathbf{x}^{(1)}) & \dots & \psi_{m_{\min}, n_{\max}}(\mathbf{x}^{(1)}) & \dots & \psi_{m_{\max}, n_{\max}}(\mathbf{x}^{(1)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \psi_{m_{\min}, n_{\min}}(\mathbf{x}^{(M)}) & \dots & \psi_{m_{\min}, n_{\max}}(\mathbf{x}^{(M)}) & \dots & \psi_{m_{\max}, n_{\max}}(\mathbf{x}^{(M)}) \end{bmatrix} \quad (4)$$

$\mathbf{W} \in \mathbb{R}^{M \times L}$  is a large dimensional matrix and it is called *wavelet lattice*. The WN is now a linear-in-the-parameters regression and the output vector can then be expressed in terms of expanding the wavelet lattice members and constructed as

$$\mathbf{y} = \mathbf{W}\boldsymbol{\theta} \quad (5)$$

where *weight vector*  $\boldsymbol{\theta}_{L \times 1} = [\theta_1 \dots \theta_L]^T$  is composed of the weights between the wavelets of the hidden layer and output layer. The purpose of this phase is to determine the coefficient vector  $\boldsymbol{\theta}$  such that  $N \leq L$ .

Since a frame generally provides a redundant or over-complete basis, it is possible that some (or even many) elements of the

**Input:** Wavelet lattice  $\mathbf{W} = [\mathbf{w}_1 \dots \mathbf{w}_L]$ , output vector  $\mathbf{y}$   
**Output:** Weight vector  $\boldsymbol{\theta}$ , the number of hidden layer wavelons  $N$

*Initialisation :*  $I = \{1, \dots, L\}$ ,  $\mathbf{r}_0 = \mathbf{y}$  (residual error)

```

1: for  $j = 1$  to  $L$  do
2:   for  $i \in I$  do
3:      $J(i) = \underbrace{\frac{(\mathbf{w}_i^T \mathbf{r}_{j-1})^2}{\mathbf{w}_i^T \mathbf{w}_i}}_{\text{residual energy}} + \underbrace{\alpha \log(\mathbf{w}_i^T \mathbf{w}_i)}_{\text{D-optimality criteria}}$  (cost function)
4:   if  $(J(i) \leq 0)$  then
5:     set  $N = j$  and break the loop
6:   end if
7: end for
8:  $l_j = \underset{i \in I}{\operatorname{argmax}} J(i)$ 
9:  $\mathbf{q}_j = \mathbf{w}_{l_j} - \sum_{k=1}^{j-1} \frac{\mathbf{w}_{l_j}^T \mathbf{q}_k}{\mathbf{q}_k^T \mathbf{q}_k} \mathbf{q}_k$ 
10:  $\mathbf{r}_j = \mathbf{r}_{j-1} - \frac{\mathbf{r}_{j-1}^T \mathbf{q}_j}{\mathbf{q}_j^T \mathbf{q}_j} \mathbf{q}_j$ 
11:  $I = I - \{l_j\}$ 
12: end for
13:  $\mathbf{Q} = [\mathbf{q}_1 \dots \mathbf{q}_N]$  (with orthogonal columns)
14:  $\mathbf{A} = \mathbf{Q}^T \mathbf{W}$  (upper triangular)
15:  $\boldsymbol{\theta} = \mathbf{A}^{-1} \mathbf{y}$ 
16:  $N$ 

```

$\alpha$  is a small positive number (typically  $\alpha = 1 \times 10^{-6}$ ).

**Fig. 2** Algorithm 1: DOOMP subset selection

wavelet lattice do not contain any sample points in their supports. Therefore, the dimensionality of the wavelet lattice is too high. These redundant wavelets are useless for processing the available data, so they should be removed from the wavelet lattice and only effective wavelets should be selected. Moreover, because the output information is not taken into account in forming the wavelet lattice, some of the matrix elements are also redundant. Furthermore, in order to avoid overfitting problems that result from an over parameterisation of the model, it is important to select a reduced and effective subset of wavelets [16].

An efficient approach for model structure determination has been implemented using the OLS algorithm [4, 11, 16, 17]. In this algorithm, the regression matrix  $[\mathbf{W}$  in (5)] is decomposed into a product of an upper triangular matrix and a matrix with orthogonal columns.

Since, at every stage of the OLS algorithm, all non-selected regression matrix columns should be orthogonalised to the previously selected ones, the computational cost of the OLS method is very high [13]. In the case of WNs, to reduce the computational requirements of OLS, several approaches have been presented so far. For instance, in [16], before executing the OLS algorithm, by selecting effective wavelets in wavelet lattice, the dimension of the regression matrix is reduced. In addition, the researchers have proposed a technique based on two stages of screening for eliminating redundant wavelets in wavelet lattice in [4, 12]. Although, this procedure has decreased the OLS complexity significantly, because the structure of the OLS method has not changed yet by means of modification strategies, the applications of these approaches is restricted to low-dimensional problems.

As mentioned earlier, the OMP algorithm gives the same results as OLS, but the former requires less computational time than the latter [13]. To implement this algorithm, first, similar to the first stage of OLS algorithm, the most correlated column of the regression matrix to the output vector is selected and is considered as a new basis vector. Second, the output vector is perpendicular to the new basis and defined as the residual error. Third, most similar columns of the regression matrix to the residual error vector is selected and orthogonalised to the selected ones. Then, the residual error will be updated and so on. In the OMP algorithm, the residual energy from each basis vector can be considered as the cost function.

To optimise the WN model, taking advantage of the optimum experimental design technique is a proper strategy [14]. In this method, the quality of a model is assessed by a particular design

**Table 1** Hair removal methods comparison for 20 dermoscopy images

Method	DA = $\frac{TP}{TP + FP + FN}$ , %	PSNR = $20 \log_{10} \frac{MAX}{\sqrt{MSE}}$ , %
Dull-Razor	89.8	89.1
proposed	90.3	89.2
FGWN		

TP: true positives, FP: false positives, and FN: false negatives. MAX is the maximum possible pixel value of the image and MSE is the mean squared error.

criterion. The D-optimality measure is known to be one of the most effective among such criteria [14]. For this purpose, a logarithmic term related to the wavelet lattice is added to the cost function of OMP algorithm. Following [14], a subset selection procedure based on combination of D-optimality design and OMP, which is called DOOMP algorithm, is employed in the current study. The algorithm is summarised in Algorithm 1 (see Fig. 2).

The D-optimality measure in the cost function increases the condition number of the wavelet lattice which can be desirable for near ill-conditioned regression matrices [14]. According to Hong and Harris [14], in Algorithm 1 (Fig. 2) the selection procedure will terminate if the cost function  $J(i) \leq 0$ . This means that the error tolerance is no longer needed for the algorithm termination. So, the network parameters will be optimised in an automatic manner.

### 3 Acquisition, enhancement, and segmentation

#### 3.1 Database and ground truth

The dataset used in this paper includes 1039 dermoscopic true-colour images from different parts of the body, taken under the same imaging conditions with polarised light. All of these images are 24 bit red, green, and blue (RGB), with a spatial resolution of  $485 \times 716$  pixels, and are taken from patients suspected of having melanoma. The disease was confirmed by histopathological examination and biopsy and was diagnosed by the dermatologist. Lesion border segmentation was performed by an experienced team consisting of dermatologists, neuroradiologists, and pathologists. From these 1039 images, 528 were categorised as MMs and the remaining 511 were categorised as non-melanomas.

#### 3.2 Enhancement of dermoscopy images

Among the 1039 dermoscopy images in our database, only 54 images (5.2%) had lesions occluded with dark thick hairs that could cause problems in the segmentation and classification processes. For these images we used a new two-level hair removal procedure. In the first place, hair pixels are detected with the proposed FGWN and then replaced by the median value of the adjacent local non-hair pixels (in a  $3 \times 3$  square block with the centre of hair pixel). From all 54 images with dark hairs, 34 (about 63%) of them are randomly selected and used in the construction of a  $3 \times 1$  FGWN with R, G, and B matrices of an image pixels as the network inputs. The rest of them, 20 images, are then applied for the FGWN testing. The output layer included one node that represented the identified attribution of the pixel: hair or non-hair. The desired values of the output vector are considered as 1 for hair pixels and -1 for others. By applying the proposed Algorithm 1 (Fig. 2) to this network, 10 wavelons in the hidden layer were achieved. The derived mean squared error (MSE) over the training and the testing sets are 0.0024 and 0.0039, respectively. Increment in the number of wavelons has not significant changes in the MSE. The results of the proposed hair removal algorithm in comparison with the well-known Dull-Razor algorithm [18] are given in Table 1. For the implementation of the Dull-Razor algorithm, the software provided at [http://www.dermweb.com/dull\\_razor/](http://www.dermweb.com/dull_razor/) was used. The results were evaluated using the detection accuracy measure for hair detection part and peak signal-to-noise ratio criteria for the hair removal algorithm. From Table 1, it can be noted that the proposed hair removal algorithm achieves high performance.

**Table 2** Colour features extracted from dermoscopy images with their formulae and identifiers

Feature type	Feature name	Formula	Identifier							
			<i>R</i>	<i>G</i>	<i>B</i>	<i>H</i>	<i>S</i>	<i>V</i>	<i>L</i>	<i>a</i> <i>b</i>
colour histogram moments		$M_k^h = \left( \frac{1}{MN} \sum_{x=1}^N \sum_{y=1}^M (f_k(x, y) - M_k^1)^h \right)^{1/h}$	1, ..., 4 5, ..., 8 9, ..., <b>12</b> 13, ..., 16 17, ..., 20 21, ..., 24 25, ..., 28 29, ..., 32 33, ..., 36							
		$h = 1, 2, 3, 4, \quad k = R, G, B, H, S, V, L, a, b$								
	peak	$p_k = \max_{x=1 \dots N, y=1 \dots M} (f_k(x, y))$	37	38	39	40	41	42	43	44   45
	median	$m_k = \text{median}_{x=1 \dots N, y=1 \dots M} (f_k(x, y))$	46	47	<b>48</b>	49	50	51	52	53   54
autocorrelogram	—	—	55, ..., <b>118</b> , 119, ..., 246							

$f_k$  is the pixel value in the  $k$ th colour channel. The detail of shape features formula described in [19].

Bolded identifiers indicate selected features in feature selection stage.

**Table 3** Texture features extracted from dermoscopy images with their formulae and identifiers

Feature type	Feature name	Formula		<i>R</i>	<i>G</i>	<i>B</i>	<i>H</i>	<i>S</i>	<i>V</i>	<i>L</i>	<i>a</i>	<i>b</i>
texture wavelet	—	$E_l^t = \frac{1}{MN} \sum_{x=1}^N \sum_{y=1}^M  cD_l^{(t)} , \quad t = h, v, d$	db1	247, 248	249, 250	251, 252	253, 254	255, 256	257, 258	259, 260	261, 262	263, 264
			db4	265, 266	267, 268	269, 270	271, 272	273, 274	275, 276	277, 278	279, 280	281, 282
		$E_l = \frac{1}{3} \sum_t E_l^t, \quad l = 1, 2$	sym2	283, 284	285, 286	287, <b>288</b>	289, 290	291, 292	293, 294	295, 296	297, 398	399, 300
			sym8	301, 302	303, 304	305, 306	307, 308	309, 310	311, 312	313, 314	315, 316	317, 318
		coif1	319, 320	321, 322	323, 324	325, 326	327, 328	329, 330	331, 332	333, 334	335, 336	
		coif3	337, 338	339, 340	341, 342	343, 344	345, 346	347, 348	349, 350	351, 352	353, 354	
		bior1.3	355, 356	357, 358	359, 360	361, 362	363, 364	365, 366	367, 368	369, 370	371, 372	
		bior1.5	373, 374	375, 376	377, 378	379, 380	381, 382	383, 384	385,386	387, 388	389, 390	
		bior3.1	391, 392	393, 394	395, 396	397, 398	399, 400	401, 402	403, 404	405, 406	407, 408	
		bior5.5	409, 410	411, 412	413, 414	415, 416	417, 418	419, 420	421, 422	423, 424	425, 426	
	GLCM	energy	$Ene = \sum_i \sum_j C(i, j)^2$					427				
		entropy	$Ent = - \sum_i \sum_j C(i, j) \log C(i, j)$					428				
		inertia	$Ine = \sum_i \sum_j (i - j)^2 C(i, j)$					429				
		homogeneity	$Hom = \sum_i \sum_j (C(i, j)/(1 +  i - j ))$					430				
		correlation	$Cor = (\sum_i \sum_j (i - \mu_x)(j - \mu_y) C(i, j))/(\sigma_x \sigma_y)$					<b>431</b>				

$E_l^t$  is the  $L_1$ -norm of horizontal, vertical, and diagonal detail coefficient and  $E_l$  is the average of these three norms.  $C$  is a CM.

Bolded identifiers indicate selected features in feature selection stage.

### 3.3 Segmentation

The authors introduced a new approach for the skin lesions segmentation using an FGWN based on the OLS algorithm. Preliminary results of this work were published in [4]. This paper also took advantage of the proposed FGWN based on the DOOMP algorithm for segmenting the boundary of the lesion accordingly. Segmentation based on the DOOMP algorithm is faster than the OLS method, but the accuracy of the two implementations is identical. Therefore, for the sake of brevity, we do not report the experimental results of this stage. For more details we refer the readers to [4].

## 4 Melanoma diagnosis

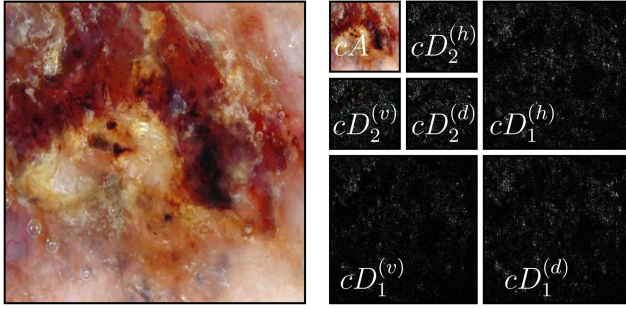
### 4.1 Feature extraction

After the segmentation stage, features characterising colour, texture, and shape of the lesion should be extracted. Tables 2–4 present these features along with their formulae and identifiers.

Here, in order to describe and represent the colour of image pixels which encompass the lesion, colour histogram, and

autocorrelogram are utilised. The extracted features of colour histogram include the first- to fourth-order colour moments, peak, and median. In addition, since histogram does not contain spatial information, in this paper, the autocorrelogram has also been used. Colour autocorrelogram characterises colour distributions of pixels and spatial correlation of colours.

A colour correlogram (henceforth correlogram) expresses how the spatial correlation of pairs of colours changes with distance. Informally, a colour correlogram of an image is a vector indexed by colour pairs, where the  $k$ th entry for  $(i, j)$  specifies the probability of finding a pixel of colour  $j$  at a distance  $k$  from a pixel of colour  $i$  in the image, where the colours  $i, j$  are quantised from  $m$  quantisation levels [20]. The autocorrelogram is a subset of the correlogram which captures spatial correlation between identical colour levels (i.e.  $i=j$ ). The computational cost of the autocorrelogram is lower than that of the correlogram and also since the autocorrelogram gives the occurrences of the similar colours instead of different colours, it may be useful as a proper feature for melanoma diagnosis. For these reasons, we take the advantages of the autocorrelogram instead of correlogram in this



**Fig. 3** Two-level 2D-wavelet decomposition of a dermoscopy image

work. For convenience, we use the  $L_\infty$  norm to measure the distance between pixels and the number of quantisation colours is considered as  $m=64$ . In addition, for the sake of simplicity the proximity of the pixels are applied as  $k=1, 2, 3$ . In this way,  $64 \times 3 = 192$  colour autocorrelation features are extracted from available dermoscopy images and they are numbered from 55 to 246 in Table 2. Moreover, we used the popular wavelet and grey-level co-occurrence matrix (GLCM) features to capture the texture characteristics of the lesions. The GLCM is created from a grey-scale dermoscopy intensity image.

As shown in Fig. 3, at each decomposition level, two-dimensional (2D)-wavelet coefficients are divided into four sub-blocks. The sub-block labelled  $cA$  corresponds to coefficients representing the *approximation* image. For *detail* coefficients, the superscripts  $h$ ,  $v$ , and  $d$  stand for horizontal, vertical, and diagonal, respectively, and the subscripts denote the decomposition level. For computing the texture descriptor, at each decomposition level, the  $L_1$ -norm of each detail coefficient was computed. The average of the three norms was calculated as it is shown in Table 3. As portrayed in this table, in the present study ten common and useful wavelets were utilised. In this way, overall  $10 \times 9 \times 2 = 180$  wavelet texture features were computed (feature number 247–426 in Table 3). In total, the number of features extracted from each lesion was 441 (246 colour-related, 185 texture-related, and 10 shape-related features).

## 4.2 Feature selection

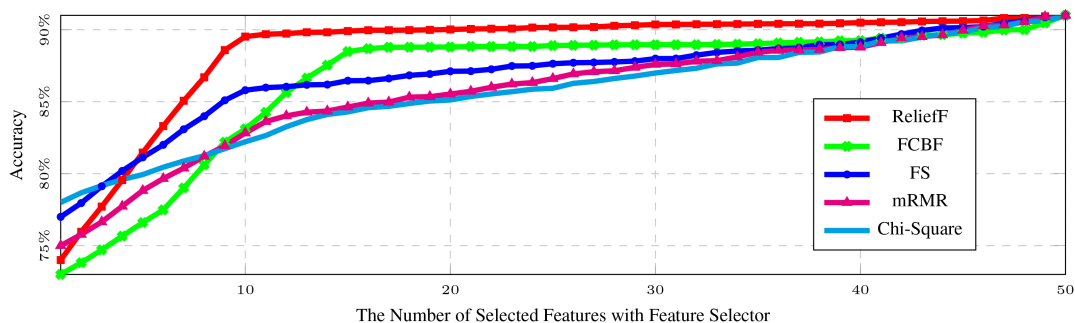
The extracted features were considered as the inputs of a feature selector in order to choose more effective and weakly correlated features and also to reduce the classifier complexity. To do this, we chose five conventional feature selectors and through a comparison procedure the best feature selector was opted. To perform the comparison, all the 441 features related to an independent set of 200 images (100 melanoma and 100 non-melanoma) from the database were submitted to the  $k$ -NN classifier and the *accuracy* measure was computed. In addition, the *redundancy rate* (RR) of each feature selector was calculated as well.

The feature selection algorithms utilised in our work are as follows: (i) ReliefF algorithm, which is noise-tolerant and robust to feature interactions [15], (ii) fast correlation-based filter (FCBF) which measures feature–class and feature–feature correlations and provides an effective way to handle feature redundancy in feature selection [21], (iii) Fisher score (FS) which selects features that assign similar values to the samples from the same class and different values to samples from different classes [22], (iv) minimum-redundancy-maximum-relevance (mRMR) algorithm which selects features that are mutually far away from each other, while they still have high correlation to the classification variable [23], and (v) chi-square which is used as a test of independence to assess whether the class label is independent of a particular feature [24]. The reason why  $k$ -NN classifier is applied in this section is that this classifier is shown to be highly effective in the classification of melanoma lesions and easy execution [7]. In extensive experiments,  $K=12$  neighbours delivered the most reliable results for several different selection strategies which are implemented in the MATLAB software. In the classification experiments presented in this section, the dataset of 200 lesions was split into two distinct train and test subsets. The training data consisted of 60% of 200 independent images and the remaining subset served as the test set to estimate classification performances for the selected features. The accuracy rate is obtained by running  $k$ -NN classifier on the first  $z$  ( $1 \leq z \leq 441$ ) highest ranked features selected by different feature selection algorithms. Fig. 4 shows the obtained accuracies of various feature selectors.

**Table 4** Shape features extracted from dermoscopy images with their formulae and identifiers

Feature type	feature name	Formula	Identifier
shape	border	perimeter	432
		eccentricity	433
		convexity	<b>434</b>
		elongation	435
		sphericity	<b>436</b>
	region	area	<b>437</b>
		irregularity	438
		compactness	<b>439</b>
		solidity	<b>440</b>
		extent	441

Bolded identifiers indicate selected features in feature selection stage.



**Fig. 4** Accuracy of various feature selectors for some features

**Table 5** Comparison of different feature selector results

Parameter	Feature selector				
	Relieff	FCBF	FS	mRMR	Chi-square
mean of accuracy, %	<b>87.14</b>	85.79	85.78	86.97	85.79
RR	<b>78.8</b>	82.7	80.2	80.1	81.1

**Table 6** Decomposition of features in the feature selection stage

Feature numbers	Colour		Texture		Shape	
	Histogram	Autocorrelogram	Wavelet	GLCM	Border	Region
441	54	192	180	5	5	5
50	16	5	14	5	5	5
48	14	5	14	5	5	5
42	12	4	13	4	5	4
35	10	3	10	3	5	4
20	4	2	5	2	4	3
17	3	1	4	2	4	3
15	2	1	4	2	3	3
12	2	1	3	1	2	3
10	2	1	1	1	2	3

Let  $F$  be the set of selected features. The following measurement is used for measuring RR of  $F$  [25]:

$$RR(F) = \frac{1}{m(m-1)} \sum_{f_i, f_j \in F, i > j} c_{i,j} \quad (6)$$

where  $m$  is the number of features and  $c_{i,j}$  is the correlation between the two features  $f_i$  and  $f_j$ . The measurement assesses the averaged correlation among all feature pairs, and a large value indicates that many selected features are strongly correlated and thus redundancy is expected to exist in  $F$ .

The accuracy and RR evaluation measurements are calculated for each of five feature selectors and these are shown in Table 5. As depicted in this table, the values of Relieff algorithm for all the evaluation criteria are better than the other methods; for this reason, we adopted the Relieff feature selector in this paper.

As illustrated in Fig. 4, when the number of selected features is more than ten, the accuracy rate for Relieff feature selector does not change very much. So it could be concluded that the best subset of features has a cardinality of ten, which is small enough to avoid overfitting. These ten selected features are made bolded in Tables 2–4. These features include: median and kurtosis of the blue channel, the probabilities of the 64th quantised colour occurrence in RGB image (autocorrelogram), level-2 decomposition of RGB colour space image with Symlet (sym2) wavelet, correlation, convexity, sphericity, area, compactness, and solidity.

In Table 6, we compare the number of selected features in the feature selection stage in terms of the type of features. For example, if the feature selector selects 50 features out of 441 extracted features, the decomposition of these 50 selected features will be as follows: 21 colour-related (16 histogram plus 5 autocorrelogram), 19 texture-related (14 wavelet plus 5 GLCM), and 10 shape-related (5 border plus 5 region). The rest of the feature numbers illustrated in Table 6 can be interpreted similarly.

It is interesting to point out that among the ten selected features in feature selection stage only one of them is related to wavelets. The reason is that we have decomposed the signal on a wavelet tree with fix structure and we have selected different wavelets such as db1, db4, sym2 etc. for each of ten leaves. Therefore, there are ten trees, one for each wavelet. Although the decomposed data on the leaves of the tree are different for various wavelets, there is some sort of similarity between ten wavelets. The decomposed data on the relevant ten tree leaves also have some sort of correlation. Furthermore, wavelet texture features are based on the 2D-wavelet decomposition level of each sub-block that has important discriminatory power, but are highly correlated [26]. Thus, they are

normally removed in feature selection stage and only the most effective ones of them are kept.

Then, selected features are considered as the inputs of classifier which is described in the following section.

#### 4.3 Classification

In the final step of MM diagnosis, the feature set selected by Relieff algorithm is submitted to the proposed FGWN classifier. To avoid overfitting and also reduce bias in terms of evaluating the prediction accuracy of the developed CAD system, we used a ten-fold cross-validation scheme [27]. To do this, the database is divided into ten random groups with no overlap as the configuration train and test data. For each train set, all previous steps are applied, then ten selected features described in the previous section are calculated from each image and considered as FGWN inputs for building a 10 1 FGWN (formation of the wavelet lattice, determination of the shift and scale parameters, and calculation of the network weights). In this way, the FGWN is formed. After that, the values of the ten features from the test set are considered as FGWN inputs (according to the best result of each of the ten configuration related to ten-fold cross-validation), whereas the output of FGWN is a number (1 or -1) that discriminates melanoma from non-melanoma in the images. In our experiments, 18–21 wavelons in hidden layer were large enough to achieve good results.

This algorithm is compared with four state-of-the-art techniques that have been successfully used in many classification problems: MLP,  $k$ -NN, AWN, and the Support Vector Machine (SVM). The evaluation process is carried out using four metrics including sensitivity, specificity, accuracy, and precision.

The MLP employed here is a three-layered multilayer perceptron with ten inputs, a hidden layer, and an output. A matrix of ten selected feature values is considered as the MLP inputs. The MLP was trained by means of the fastest training function available in MATLAB. Comparable results have been found for a number of neurons ranging from 24 to 27 for our MLP. The AWN was implemented following [28]. Comparable results have been found for a number of wavelons ranging from 18 to 22 for our WN.

SVMs perform classification by constructing a set of  $N$ -dimensional hyperplanes that optimally separate the given data into classes, using the largest possible margin. Margin is the distance between optimal hyperplane and the nearest training data points of any class, and the larger the margin the lower the generalisation error of the classifier [6]. An SVM classifier with a Gaussian radial basis function kernel was implemented using MATLAB.



**Table 7** Ten-fold cross-validation of classification performance on database

Evaluation formula	Sensitivity	Specificity	Accuracy	Precision
	TP TP + FN	TN TN + FP	TP + TN TP + TN + FP + FN	TP TP + FP
MLP	86.17	87.08	86.62	87.33
k-NN	88.07	87.28	87.68	87.74
AWN	90.91	89.24	90.09	89.72
SVM	92.22	90.41	91.34	90.86
FGWN	<b>92.61</b>	<b>91.00</b>	<b>91.82</b>	<b>91.40</b>

TP: true positives, TN: true negatives, FP: false positives, and FN: false negatives.  
In each algorithm the best configuration results are shown.

#### 4.4 Evaluation of results

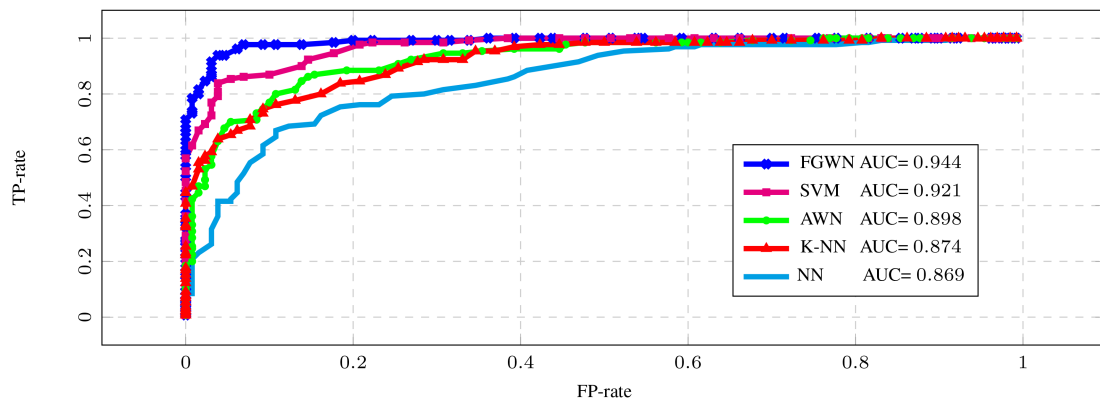
In our experimental evaluation, we used a personal computer with Intel(R) Core(TM)2 Duo central processing unit T9550 (2.66 GHz) and 4 GB random access memory. All algorithms were implemented in MATLAB 7.12.

In this paper, four criteria of standard evaluation which have been used in a large number of related research are employed as classification performance criteria. These criteria are presented in Table 7. As depicted in this table, the values of our proposed algorithm for all the evaluation measure are better than other methods.

Furthermore, when images with different qualities are taken (for example, under different zooming conditions and having various resolutions or different lightings), the inputs of the network in each stage of detection (e.g. feature values in classification stage) may be highly scattered. This problem is solved by the normalisation step in the algorithm of FGWN structure formation. The details of normalisation step are available in [4].

Since the proposed FGWN has satisfactory characteristics such as robustness and optimal approximation, it can be used as a good classifier in MM diagnosis. This results in the precision rate in Table 7 to be high. The point that is a failure in classification could be attributed to inaccurate lesion segmentation. Failure often happens when the lesion has multiple or unclear boundaries or when it exhibits regression (scar-like depigmentation) structures [19]. In these cases, segmentation algorithm and the following post-processing stage face problems and the lesion boundary may be detected incorrectly. In this way, it is likely that some important features are lost and classification algorithm fails.

The Receiver Operating Characteristic (ROC) curve as a metric to quantify the performance of a classifier is plotted in Fig. 5. An ROC diagram plots the true positive (TP) rate versus the false positive (FP) rate for different discrimination thresholds of the classifier. The point (0, 1) represents a perfect classification and the diagonal line represents random performance. A convenient way of comparing ROC curves is the area under the curve (AUC). A larger AUC means better classification performance. In other words, the superior classifier is located at the left uppermost while the poorer is closest to the diagonal. The AUC measures for all the evaluated classification algorithms are also computed and illustrated in



**Fig. 5** ROC curve for the classifier. For each ROC diagram, the AUC measures were also computed

**Table 8** Computational time (seconds) comparison of some WN-based method for triple stages of MM detection for each image on average

Stage	AWN [28]	FGWN [16]	FGWN [4]	Proposed FGWN
hair removal	181	102	87	<b>63</b>
segmentation	215	197	164	<b>118</b>
classification	288	242	201	<b>185</b>

Fig. 5. It is evident from this figure that the presented FGWN algorithm achieves a high value of AUC and is more discriminating than other classifiers. Our method is quite simple and is applicable for melanoma detection by means of a CAD system.

As may be recalled, in the proposed FGWN, a new construction algorithm (DOOMP) is employed for building the WN. This method is different from our earlier work which applied OLS technique with two stages of screening for the formation of WN. In comparative Table 8, the execution-time of some WN-based method for the middle three stages of MM diagnosis is presented.

In AWN [28], the network parameters are initialised with some random values and then optimised through a backpropagation procedure. Thus, such an iterative approach with tedious calculations can increase the computational time. In FGWN [16], a WN based on OLS algorithm is proposed. Since this WN does not need an iterative training procedure, it has lower complexity than AWN, but because of the OLS structure, the computational time of this technique is high. Moreover, in our previous FGWN [4], the network computation was reduced by adding two screening stages before OLS execution. Nevertheless, because of OLS restrictions, the computational time of that method is higher than that of the FGWN proposed in this paper.

According to similarities between FGWN and SVM, we compare their accuracies in Table 9, which includes a reduction in the feature space dimension. Since SVM is known for coping with large dimensionality feature spaces, it turned out that in higher dimensions SVM gets better results than FGWN. However, in our dataset, dimensional reduction via feature selection not only reduced the computational time of classification, but also it improved the performance of discrimination in terms of accuracy. Besides, applying a technique based on a single algorithm makes MM diagnosis become a uniform, entirely automatic, and fast process.

## 5 Conclusion

Building up on our previous work [4], and presenting a fully-automatic pre-screening method, in this work we proposed a novel CAD system for the diagnosis of MM to assist dermatologists. The most important advantage of the proposed algorithm was its uniformity and accuracy which are crucial in pathological diagnosis. All important standard stages for detecting MM were done through a single algorithm based on WNs. This type of WNs instead of using gradient-type algorithms converts the network into

**Table 9** Accuracy of the proposed FWGN classification method in comparison with SVM in the different dimension feature spaces

Classifier	Feature numbers	Accuracy
SVM		
without feature selection	441	82.67
	120	88.12
	50	91.95
with feature selection	12	91.39
	10	91.34
FGWN		
without feature selection	441	80.51
	120	86.77
	50	91.90
with feature selection	12	91.83
	10	91.82

a linear-in-the-parameters problem, which can then be solved using OMP plus D-optimality criterion. Uniform algorithm, simple structure, and quick running time make this method a good option for designing automatic skin cancer diagnosis software.

## 6 References

- [1] Kasmi, R., Mokrani, K.: 'Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule', *IET Image Process.*, 2016, **10**, (6), pp. 448–455
- [2] Xiang, L., Aldridge, B., Fisher, R., *et al.*: 'Estimating the ground truth from multiple individual segmentations incorporating prior pattern analysis with application to skin lesion segmentation'. Proc. Eighth Int. Symp. on Biomedical Imaging (ISBI.11), Electronic Proc., April 2011, pp. 1438–1441
- [3] Gonzalez-Castro, V., Debayle, J., Wazaefi, Y., *et al.*: 'Automatic classification of skin lesions using color mathematical morphology-based texture descriptors'. Proc. Int. Conf. Quality Control by Artificial Vision, Le Creusot, France, June 2015
- [4] Sadri, A.R., Zekri, M., Sadri, S., *et al.*: 'Segmentation of dermoscopy images using wavelet networks', *IEEE Trans. Biomed. Eng.*, 2013, **60**, (4), pp. 1134–1141
- [5] Wighton, P., Lee, T.K., Lui, H., *et al.*: 'Generalizing common tasks in automated skin lesion diagnosis', *IEEE Trans. Inf. Technol. Biomed.*, 2011, **15**, (4), pp. 622–629
- [6] Garnavi, R., Aldeen, M., Bailey, J.: 'Computer-aided diagnosis of melanoma using border and wavelet-based texture analysis', *IEEE Trans. Inf. Technol. Biomed.*, 2012, **16**, (6), pp. 1239–1252
- [7] Ganster, H., Pinz, A., Rohrer, R., *et al.*: 'Automated melanoma recognition', *IEEE Trans. Med. Imaging*, 2001, **20**, (3), pp. 233–239
- [8] Iyatomi, H., Oka, H., Saito, M., *et al.*: 'Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system', *Melanoma Res.*, 2006, **16**, (2), pp. 183–190
- [9] Ballerini, L., Xiang, L., Fisher, R.B., *et al.*: 'A query-by-example content-based image retrieval system of non-melanoma skin lesions'. Proc. MICCAI-09 Workshop MCBR-CDS 2009: Medical Content based Retrieval for Clinical Decision Support, 2010 (LNCS, **5853**), pp. 31–38
- [10] González-Castro, V., Debayle, J., Wazaefi, Y., *et al.*: 'Texture descriptors based on adaptive neighborhoods for classification of pigmented skin lesions', *J. Electron. Imaging*, 2015, **24**, (6), p. 061104
- [11] Billings, S.A., Hua-Liang, W.: 'A new class of wavelet networks for nonlinear system identification', *IEEE Trans. Neural Netw.*, 2005, **16**, (4), pp. 862–874
- [12] Hosseinabadi, H.Z., Nazari, B., Amirfatahi, R., *et al.*: 'Wavelet network approach for structural damage identification using guided ultrasonic waves', *IEEE Trans. Instrum. Meas.*, 2014, **63**, (7), pp. 1680–1692
- [13] Xu, J., Ho, D.W.C.: 'A basis selection algorithm for wavelet neural networks', *Neurocomputing*, 2002, **48**, (1), pp. 681–689
- [14] Hong, X., Harris, C.J.: 'Nonlinear model structure design and construction using orthogonal least squares and D-optimality design', *IEEE Trans. Neural Netw.*, 2002, **13**, (5), pp. 1245–1250
- [15] Kira, K., Rendell, L.A.: 'A practical approach to feature selection'. Proc. Ninth Int. Workshop Machine Learning, Aberdeen, Scotland, UK, July 1992, pp. 249–256
- [16] Zhang, Q.: 'Using wavelet network in nonparametric estimation', *IEEE Trans. Neural Netw.*, 1997, **8**, (2), pp. 227–236
- [17] Sadri, A.R., Zekri, M., Sadri, S., *et al.*: 'Impulse noise cancellation of medical images using wavelet networks and median filters', *J. Med. Signals Sens.*, 2012, **2**, (1), pp. 25–37
- [18] Lee, T., Ng, V., Gallagher, R., *et al.*: 'Dullrazor®: a software approach to hair removal from images', *Comput. Biol. Med.*, 1997, **27**, (6), pp. 533–543
- [19] Celebi, M.E., Kingravi, H.A., Uddin, B., *et al.*: 'A methodological approach to the classification of dermoscopy images', *Comput. Med. Imaging Graph.*, 2007, **31**, (6), pp. 362–373
- [20] Huang, J., Kumar, S.R., Mitra, M., *et al.*: 'Image indexing using color correlograms'. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition, San Juan, Puerto Rico, June 1997, pp. 762–768
- [21] Yu, L., Liu, H.: 'Feature selection for high-dimensional data: a fast correlation-based filter solution'. Proc. 20th Int. Conf. Machine Learning, Washington, DC, USA, August 2003, pp. 856–863
- [22] Duda, R., Hart, P., Stork, D.: '*Pattern classification*' (John Wiley & Sons, New York, 2001, 2nd edn.), 2012
- [23] Peng, H., Long, F., Ding, C.: 'Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (8), pp. 1226–1238
- [24] Huan, L., Setiono, R.: 'Chi2: feature selection and discretization of numeric attributes'. Proc. Seventh Int. Conf. Tools with Artificial Intelligence, Washington, DC, USA, November 1995, pp. 388–391
- [25] Arizona State University: '*Advancing feature selection research-ASU feature selection repository*' (Arizona State University, Glendale, AZ, USA, 2010), pp. 1–28
- [26] Huang, K., Aviyente, S.: 'Wavelet feature selection for image classification', *IEEE Trans. Image Process.*, 2008, **17**, (9), pp. 1709–1720
- [27] Kohavi, R.: 'A study of cross-validation and bootstrap for accuracy estimation and model selection'. Proc. 14th Int. Conf. Tools with Artificial Intelligence, Montreal, Quebec, Canada, August 1995, pp. 1137–1145
- [28] Zhang, Q., Benveniste, A.: 'Wavelet networks', *IEEE Trans. Neural Netw.*, 1992, **3**, (6), pp. 889–898