# Computer-aided Diagnosis of Melanoma Using Border and Wavelet-based Texture Analysis

Rahil Garnavi, Member, IEEE, Mohammad Aldeen, Senior Member, IEEE and James Bailey.

***Abstract*—This paper presents a novel computer-aided diagnosis system for melanoma. The novelty lies in the optimised selection and integration of features derived from textural, border-based and geometrical properties of the melanoma lesion. The texture features are derived from using wavelet-decomposition, the border features are derived from constructing a boundary-series model of the lesion border and analysing it in spatial and frequency domains, and the geometry features are derived from shape indexes. The optimised selection of features is achieved by using the Gain-Ratio method, which is shown to be computationally efficient for melanoma diagnosis application. Classification is done through the use of four classifiers; namely, Support Vector Machine, Random Forest, Logistic Model Tree and Hidden Naive Bayes. The proposed diagnostic system is applied on a set of 289 dermoscopy images (114 malignant, 175 benign) partitioned into train, validation and test image sets. The system achieves and accuracy of 91.26% and AUC value of 0.937, when 23 features are used. Other important findings include (i) the clear advantage gained in complementing texture with border and geometry features, compared to using texture information only, and (ii) higher contribution of texture features than border-based features in the optimised feature set.**

**Key words: Computer-aided diagnosis of Melanoma, Dermoscopy, Feature extraction, Classification, Wavelet.**

## I. Introduction

MALIGNANT melanoma is the deadliest type of skin cancer and the most common cancer in people aged 15–44 years in Australia [1]. Failure to diagnose a newly developed melanoma lesion may lead to lethal advanced melanoma, thus, early detection is of significant importance for the dermatology community.

Dermoscopy is a non-invasive *in vivo* imaging technique which provides a magnified and clear visualization of the morphological structures of the skin that are not visible to the naked eye. With the use of dermoscopy and dermoscopic clinical algorithms, such as pattern analysis [2], ABCD rule of dermoscopy [3], Menzies method [4], 7-point checklist [5], and the CASH algorithm [6], the diagnosis of melanoma has been improved compared with the simple naked-eye examination by 5–30%. This improvement rate depends on the type of skin lesion and the experience of the dermatologist [7]. However, clinical diagnosis of melanoma is inherently subjective and its accuracy has been an issue of concern, especially with equivocal pigmented lesions [7]. Despite the use of dermoscopy, the accuracy of expert dermatologists in diagnosing melanoma is still estimated to be about 75-84% [8].

Due to enhancements in skin imaging technology and image processing techniques in recent decades, there has been a significant increase in interest in the area of computer-aided diagnosis of melanoma, aiming to remove subjectivity and uncertainty from the diagnostic process, and provide a reliable second opinion to dermatologists. Some of these systems include[1]: SolarScan [10] developed by Polartechnics Ltd, in Australia, DermoGenius-Ultra [11] developed by LINOS Photonics Inc., DBDermo-MIPS [12], developed at the University of Siena in Italy, DANAOS expert system [13], Melafind [14] developed by Electro-Optical Sciences Inc., Irvington, etc. However, it is widely acknowledged that much higher accuracy is required for computer-based algorithms to be adopted routinely in the diagnostic process [9], [15], [16]. In this study, we have proposed a novel diagnostic system to facilitate accurate, fast and reliable computer-based diagnosis of melanoma.

A computer-aided diagnosis of melanoma generally comprises several components; image acquisition, segmentation, feature extraction, feature selection, and classification; the latter three are the main focus of this paper.

*Segmentation* or *border detection* is the process of separating the lesion from the surrounding skin in order to form the region of interest. Existing computer-based melanoma studies have either used manual, semi-automatic or fully automatic border detection methods. Various image features such as shape, colour, texture, and luminance have been employed to perform skin lesion segmentation. Numerous border detection methods have been reported in the literature [9], [17]. Recent methods include histogram thresholding followed by region growing [18], JSEG algorithm based on colour quantization and spatial segmentation [19], global thresholding on optimised colour channels followed by morphological operations [20], Hybrid thresholding [21]. In this study, we have applied our recently proposed Hybrid border detection method [21], which encompasses two stages; the first stage applies global thresholding to detect an initial boundary of the lesion, and the second stage applies adaptive histogram thresholding on optimized colour channels of X (from the CIE XYZ colour space) to refine the border.

*Feature extraction* is used to extract the features; similar to those visually detected by dermatologists, that accurately characterizes a melanoma lesion. The feature extraction method-

R. Garnavi has been with the Department of Electrical and Electronic Engineering, NICTA Victoria Research Laboratory, University of Melbourne, Australia, during the course of this research and now works at IBM Research Australia. e-mail: rahilgar@au.ibm.com

M. Aldeen is with the Department of Electrical and Electronic Engineering, University of Melbourne, Australia. e-mail: aldeen@unimelb.edu.au

J. Bailey is with the Department of Computer Science and Software Engineering, NICTA Victoria Research Laboratory, University of Melbourne, Australia. e-mail: baileyj@unimelb.edu.au

[1]A more comprehensive literature review can be found in [9].

ology of many computerised melanoma detection systems has been largely based on the conventional clinical algorithm of ABCD-rule of dermoscopy due to its effectiveness and simplicity of implementation. Its effectiveness stems from the fact that it incorporates the main features of a melanoma lesion such as asymmetry, border irregularity, colour and diameter (or differential structures), where quantitative measures can be computed. Different image processing techniques have been used to extract such features, see for example [9], [22], [23].

In this paper we propose a highly effective feature extraction method which combines different types of features; novel ones as well as a few adopted from existing studies. In our proposed method, texture and border (including geometry and boundary-series) features are extracted and combined. The texture-based feature extraction method employs tree-structured wavelet decomposition. The novelty here lies in (i) the use of four-level decompositions as opposed to three level used in the existing wavelet-based methods [24]–[26], (ii) the use of four colour channels of red, green, blue and luminance as opposed to only luminance in existing studies, (iii) application of various statistical measurement and ratios on all the nodes of the wavelet-tree.

The geometry-based feature extraction method uses a variety of statistical measures. These are derived via two distinctive processes: (i) various geometry-based features (standard and newly designed based on expert opinion of dermatologists) are defined to evaluate the properties of the overall shape of the lesion, and (ii) a boundary-series model of the border is constructed and analysed in both spatial and frequency domains to extract the main characteristics of the lesion boundary. To the best of our knowledge, the boundary-series analysis of process (ii) and some of the measures used in process (i) have not been previously applied on dermoscopy images.

*Feature selection* is an intermediate process that lies between the two steps of feature extraction and classification. In this process, irrelevant, redundant, or noisy features are excluded from the initial feature vector constructed in the feature extraction step. Some of the prominent feature selection method that have been used in existing computer-based melanoma recognition systems include sequential forward floating selection and sequential backward floating selection [27], genetic algorithm [28], ReliefF, mutual information and correlation-based methods used in [22], principal component analysis used in [25], statistical analysis used in [24], [26], etc.

However, in this study the Gain Ratio-based Feature Selection (GRFS) method is employed for the following reasons: (i) the method is computationally efficient, which is a vital factor because of the high dimensionality of the initial feature vector constructed in the feature extraction phase; (ii) it is highly comparable with a sample of well-known filter-based feature selectors as explained in Section V of this paper. To the best of our knowledge GRFS has not been applied in previous melanoma studies.

*Classification* is the final step in the diagnosis process, wherein the extracted features are utilized to ascertain whether the lesion is cancerous or normal. The most common classification methods that have been applied to computer-based melanoma recognition systems include [9], [23] Discriminate Analysis [24], Artificial Neural Network [25], [29], $k$-Nearest Neighbourhood [27], [29], Support Vector Machine [22], [23], [29], Decision Trees [23], [30], Bayes Networks [23] and Logistic Regression [29]. However, some enhanced versions of these classifiers have been reported in literature, including Random Forest (RF) [31], Logistic Model Tree (LMT) [32], and Hidden Naive Bayes (HNB) [33]. These three plus the very popular Support Vector Machine (SVM) were utilized in melanoma diagnosis in our preliminary study [34] for the first time, and will also be used in this study. All these classifiers and more particularly the Random Forest classifier are shown in our previous study [34] and in this paper to be highly effective in the classification of melanoma lesions.

TABLE I
LIST OF ABBREVIATIONS

| Term | Abbreviations |
|---|---|
| Area Under Curve of ROC | AUC |
| Receiver Operating Characteristic | ROC |
| Gain Ratio-based Feature Selection | GRFS |
| Hidden Naive Bayes | HNB |
| Logistic Model Tree | LMT |
| Random Forest | RF |
| Support Vector Machine | SVM |

To evaluate the accuracy of our diagnosis system, which is comprised of the three phases of feature extraction, feature selection and classification, we have conducted a set of four experiments, flow charts of which are shown in Figure 7. The proposed diagnosis system encompasses most of the existing features proposed by other studies, and blends them with novel ones carefully extracted to best reflect the property of the lesion.

In summary, the main contributions of this study are summarized as follows:

1) The use of four level of wavelet decomposition, incorporating four colour channels, applying various statistical measurement and ratios on all the nodes of the fully extended wavelet tree for texture-based feature extraction.
2) The use of boundary-series analysis in spatial and frequency domains and new geometrical measures for border feature extraction.
3) Establishing that the Gain Ratio feature selection method can deliver substantial accuracy for feature selector.
4) Combining different types of features in an optimised way to achieve higher performance for the diagnostic system.

The rest of the paper is organized as follows. An overview of the proposed texture feature extraction method is provided in Section II. The proposed geometry-based features and boundary-series features are explained in Sections III and IV, respectively. Section V details the proposed feature selection method, followed by discussing the applied classification methods. Experimental results and discussions are presented in Section VI. Section VII provides the summary and conclusion. The frequently used abbreviations in this study are listed in Table I.

## II. Proposed Texture-based Features

The visual characteristics of a lesion which constitutes the basis of clinical diagnostic approaches (e.g. ABCD rule of dermoscopy and pattern analysis) can be captured through texture analysis. The multi-scale properties of wavelets, makes the wavelet-based texture analysis a useful technique for analysing dermoscopy images, as these images are often taken under different circumstances, i.e. various image acquisition set ups (lighting, optical zooming, etc) and versatile skin colours. Among the two wavelet structures; Pyramid-structured [35] and Tree-structured wavelet analysis [36], the latter is more informative for classification of melanoma images, as it can capture information of both general (overall properties) and detail (detailed structure and internal texture) of the lesion [34].

In this paper, texture-based feature extraction is achieved by applying 4-level tree-structured wavelet decomposition on red, blue, green and luminance colour channel of dermoscopy images. Various statistical measures and ratios are proposed to extract useful information from the wavelet coefficients, and utilized for the classification. The measures include energy (E), mean (M), standard deviation (Std), skewness (S), kurtosis (K), norm (N), entropy (H), and average-energy (AvgE), expressed in Equations 1–8. Figure 1 displays a schematic illustration of wavelet tree with nodes marked by circles. For each colour channel, the measures are applied on the original image (called level 0) and also on wavelet coefficients of each node (sub-image) of the wavelet tree. The wavelet tree has 341 nodes in total (4 nodes in first level, 16 nodes in second level, 64 nodes in third level and 256 nodes in forth level, plus the actual image in level zero which is the parent node of the tree). The eight measures therefore yield a total of $8 \times 341$ features, per colour channel.

$$E(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} x_{jk}^2}{J \times K} \tag{1}$$

$$M(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} x_{jk}}{J \times K} \tag{2}$$

$$Std(n_i) = \sqrt{\frac{\sum_{j=1}^{J}\sum_{k=1}^{K} (x_{jk} - M(n_i))^2}{J \times K}} \tag{3}$$

$$S(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} (\frac{x_{jk} - M(n_i)}{Std(n_i)})^3}{J \times K} \tag{4}$$

$$K(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} (\frac{x_{jk} - M(n_i)}{Std(n_i)})^4}{J \times K} \tag{5}$$

$$N(n_i) = max(\sqrt{eig(X_i \times X_i')}) \tag{6}$$

$$H(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} \left(x_{jk}^2 \times log(x_{jk}^2)\right)}{J \times K} \tag{7}$$

$$AvgE(n_i) = \frac{\sum_{j=1}^{J}\sum_{k=1}^{K} |x_{jk}|}{J \times K} \tag{8}$$

where $i = 0, 1, ..., 341$ is an index of wavelet tree nodes starting sequentially from node $n_0$ of colour channel 1 and ending at node $n_{4.4.4.4}$ of colour channel 4 (colour channels: Red = 1, Green = 2, Blue = 3, Luminance = 4 ). $X_i$ is a $J_i \times K_i$ matrix of the $i^{th}$ node, $X_i'$ is its transpose, $x_{jk}$ is its $jk^{th}$ element, and $eig(X_i)$ are its eigenvalues. $J$ and $K$ are dimensions of the matrices over which wavelet decompositions take place. They start from dimension (resolution) of the original image and at each level of decomposition they are divided by a factor of two.

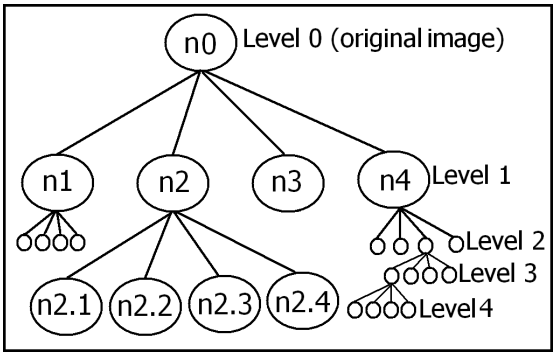


Fig. 1. Schematic 4-level wavelet tree with nodes in circle.

The ratios used in this study are maximum ratio ($R_m$), fractional ratio ($R_f$) and sub-level ratio ($R_s$), defined in Equations 9–11, respectively for a sample node $n_{2.3}$. As each decomposition sub-tree has four nodes (as illustrated in Figure 1), the maximum ratio divides each node by the maximum node among the four. The fractional ratio divides each node by the summation of the other three nodes. The sub-level ratio divides each node of the tree by its low frequency component (first child node from left hand side). These ratios are calculated for each of the above mentioned 8 measure, which will produce a total of $8 \times 765$ (340 maximum ratios, 340 fractional ratios, 85 sub-level ratios) features.

$$R_m(n_{2.3}) = \frac{m(n_{2.3})}{max(m(n_{2.1}), m(n_{2.2}), m(n_{2.3}), m(n_{2.4}))} \tag{9}$$

$$R_f(n_{2.3}) = \frac{m(n_{2.3})}{m(n_{2.1}) + m(n_{2.2}) + m(n_{2.4})} \tag{10}$$

$$R_s(n_{2.3}) = \frac{m(n_{2.3.1})}{m(n_{2.3})} \tag{11}$$

where $n_{2.3}$ refers to node 2.3 and $m(n_{2.1})$ refers to the measure (energy, entropy, etc.) applied on the node.

Moreover, the feature $ln(std{+}1)$ suggested by [24] and [25] has also been used. All of these features were calculated for four colour channels of red, green, blue and luminance (Equation 12) and a set of 35396 features is obtained.

$$luminance = (0.3 \times R) + (0.59 \times G) + (0.11 \times B) \tag{12}$$

where R, G and B represent red, green and blue colour channels, respectively.

As outlined in the introduction this is the first time four colour channel with four level of wavelet decomposition has been reported.

The features proposed in this section also account for colour information of the lesion (in addition to the textural properties). The colour features, which include the eight measures of energy, mean, standard deviation, entropy, average-energy, skewness, kurtosis, and norm of the four colour channels red, green, blue and luminance of the image, are measured by employing these eights measures on the node-0 (parent node) of the wavelet tree, i.e., the original image.

## III. Proposed Geometry-based Features

Border formation and geometrical properties of the shape of the lesion provide significant diagnostic information in the detection of melanoma. According to the ABCD rule of dermoscopy [3], asymmetry is given the highest weight among the four features of asymmetry, border irregularity, colour and differential structures. A number of studies have been carried out on quantifying asymmetry in skin lesions. In some approaches, the symmetry axis is determined in a certain manner (such as principal axis [37], [38], major axis of the best-fit ellipse [39], [40], Fourier transform [41], longest or shortest diameter [42]) and the two halves of the lesion along the axis are compared. In another approach, the symmetry feature is calculated based on geometrical measurements on the whole lesion, e.g. symmetric distance [42], circularity [42]. Other studies, consider the circularity index, also called thinness ratio, as a measure of irregularity of borders in dermoscopy images [40], [43]. Other features extracted from border and shape of the lesion include bulkiness, fractal dimension, irregularity indices [44], [45]. Here, we used some standard geometry features (Area, Perimeter, Greatest Diameter, Circularity Index, Irregularity Index A, Irregularity Index B, and Asymmetry Index) adopted from [23] complemented by new others (Shortest Diameter, Irregularity Index C and Irregularity Index D) that we drew from our discussions with expert dermatologists[2].

The main prerequisite for extracting the border features is the segmentation or border detection step, where the lesion is separated from the surrounding normal skin. The output of the segmentation step is the segmentation plane which is a black-white image. In this study, after forming the lesion pixels in a 2D matrix and the corresponding boundary pixels in a vector, the following set of 11 geometry-based features are extracted from each dermoscopy image:

**Area (A):** Number of pixels of the lesion.

**Perimeter (P):** Number of pixels along the detected boundary.

**Greatest Diameter (GD):** The length of the line which connects the two farthest boundary points and passes across the lesion centroid (C), which is given by:

$$(x_c, y_c) = (\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n}) \tag{13}$$

where $n$ is the number of pixels inside the lesion, and $(x_i, y_i)$ is the coordinates of the i-th lesion pixel.

**Shortest Diameter (SD):** The length of the line which connects the two nearest boundary points and passes across the lesion centroid.

**Circularity Index (CRC):** It explains the shape uniformity.

$$CRC = \frac{4A\pi}{P^2} \tag{14}$$

[2]Images used in this study have fairly similar spatial resolution, thus there has been no scale issue for features such as area and perimeter. For image sets where images are taken under different zooming condition and have various resolutions, a normalization procedure is required when measuring those features.

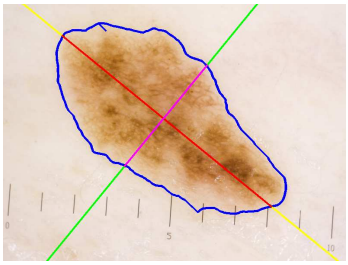

Fig. 2. Major and minor symmetry axes for sample dermoscopy image.

**Irregularity Index A (IrA):**

$$IrA = \frac{P}{A} \tag{15}$$

**Irregularity Index B (IrB):**

$$IrB = \frac{P}{GD} \tag{16}$$

**Irregularity Index C (IrC):**

$$IrC = P \times (\frac{1}{\text{SD}} - \frac{1}{\text{GD}}) \tag{17}$$

**Irregularity Index D (IrD):**

$$IrD = GD - SD \tag{18}$$

**Major and Minor Asymmetry Indices:** These indices are defined as the area difference between the two halves of the lesion, taken the principal axes (obtained by Equation 19) as the major symmetry axis, and its 90 degree rotation as the minor axes of the symmetry.

$$tan\ \ 2\theta = \frac{2\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i^2} \tag{19}$$

where $\theta$ is the orientation of the principal axis. Figure 2 shows a sample dermoscopy image with obtained principal axes (major and minor symmetry axes). After calculating the major and minor symmetry axes, the lesion is folded along the axes and the differences between the two halves of the lesion are calculated by applying the XOR operation on the binary segmentation plane. The asymmetry index is measured by:

$$\text{Asymmetry Index} = \frac{A_D}{A} \times 100 \tag{20}$$

where $A_D$ denotes the difference between the two halves. Figure 3 shows a dermoscopy image with the obtained major principal axes (symmetry axes), and the process of calculating the major asymmetry index.

## IV. Proposed Boundary-series Features

A time-series is an ordered sequence of values or observations $x$ that are measured and recorded at successive equally spaced time intervals $t$ [46]. Time-series have been widely applied in statistics, signal processing, data mining, mathematical finance and economy [47]. Inspired from this concept, in this study, we propose to build a boundary-series of the lesion by traversing the boundary of the lesion (starting from an arbitrary pixel on the lesion boundary[3]), and

[3]The choice of the starting point is not our concern because the measures applied on the resultant boundary-series, including histogram and wavelet analysis, are rotation invariant.

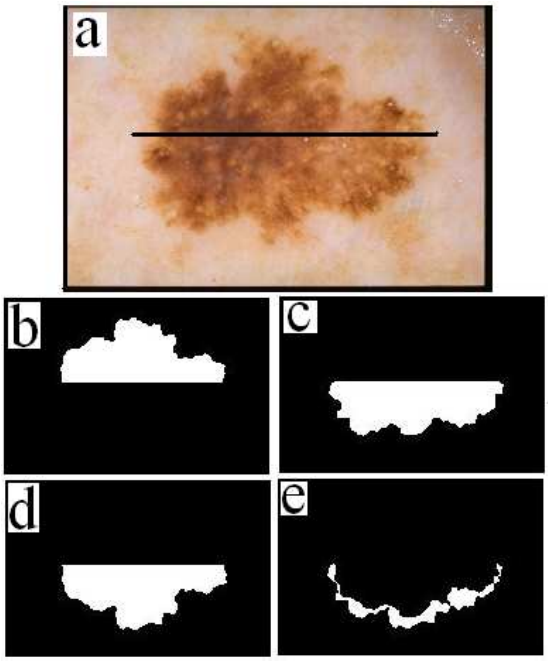


Fig. 3. Calculating the major symmetry index: (a) major symmetry axis, (b) upper half, (c) lower half, (d) folded upper half, (e) difference.

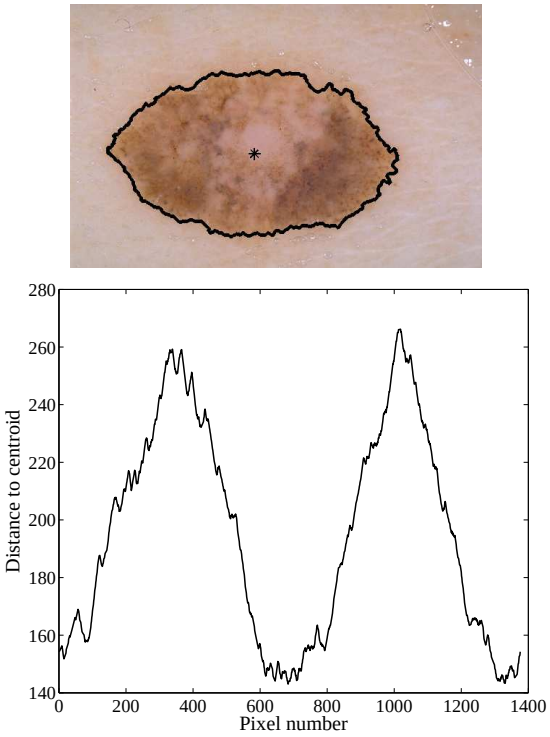


Fig. 4. Boundary-series for sample dermoscopy image.

calculating the distance between each border pixel and the centroid of the lesion, as expressed by Equation 21.

$$BS = d_1, d_2, ..., d_P \tag{21}$$

where $P$ is the number of pixels along the boundary (previously defined as perimeter of the lesion) and

$$d_j = \sqrt{(x_j - x_c)^2 + (y_j - y_c)^2} \tag{22}$$

where $(x_j, y_j)$ and $(x_c, y_c)$ are the coordinates of the j-th boundary pixel, and lesion centroid, respectively. Figure 4 shows a sample dermoscopy image and the corresponding boundary-series curve.

The obtained boundary-series is analysed in both the spatial and frequency domains. The former is done by calculating the histogram of the boundary-series (using 10 bins), and the latter is performed via applying a three-level wavelet transform, where the signal (the boundary-series) is decomposed into an approximate and a detail component. As in the pyramid-structured wavelet transform, the approximate component is used for the next level of decomposition. The wavelet decomposition was performed up to the third level, after which it is found that for the image set under study the boundary-series loses its time dependency and does not carry useful information regarding the border. A set of six statistical measures of Energy, Mean, Standard deviation, Skewness, Kurtosis and Norm (defined by Equations 1–6) are then applied on the resultant eight signals of boundary-series; histogram of boundary-series, and the three pairs of approximate and detail components of the wavelet transform, to produce a vector of 48 features.

## V. Feature Selection and Classification

### A. Feature Selection

Feature selection is the process of finding an optimised subset of features which provides the highest discriminating power when employed by the classifier. Exclusion of less significant features in the problem space can help the classifier by removing the noisy evidence in the training stage. Feature selection can also reduce the number of features to be considered in the problem space, thus lessens computation cost.

Feature selection algorithms can be categorized into three classes [48]: (1) the filter model, which uses the general characteristics of the data to evaluate the features, and selects the optimised feature subset(s) without using any classification/learning algorithm, (2) the wrapper model, which uses a predetermined learning algorithm and searches for features which improve the learning performance, (3) the hybrid model which combines the two approaches. The filter model feature selectors are less computationally expensive than wrapper-based approaches.

In this study, due to the high dimension of the feature vector, we have adopted the filter model by using the Gain Ratio-based feature selection method. GRFS (Equation 23) is an entropy-based feature valuator which assesses the worth of an attribute (feature) by measuring its Gain Ratio with respect to the class. The Gain Ratio is a measure of goodness of the feature and illustrates how well the feature predicts the class label. The Ranker search method is used to rank and sort features according to their evaluated Gain Ratio.

$$GainR(C, A) = \frac{H(C) - H(C|A)}{H(A)} \tag{23}$$

where A is the attribute or feature, C is the class label, and H specifies the entropy.

In our preliminary study [34], we proposed a two-stage feature selection scheme, which used the GRFS method in the first stage, and applied the Correlation-based Feature Selection method [49] on the subset obtained from the first stage,

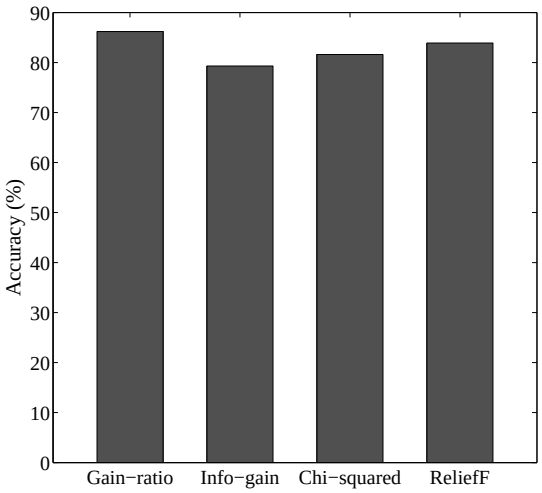


Fig. 5. Comparing various feature selection methods.

in the second stage. However, experimental results showed that adding the CFS method did not provide a noticeable improvement to the classification performance over that obtained by using the GRFS method only. For this reason we have elected to use GRFS only as the feature selection method in this study. Moreover, GRFS has been compared with other three measures; namely, Information Gain, Chi-squared and ReliefF [50]. In order to perform the comparison, the entire feature vector comprising 35455 features (including 35396 texture, 11 geometry, and 48 boundary-series features) is ranked individually by the four feature selectors. Then the first 100 highest ranked features by each selector are fed individually into the Random Forest classifier and the accuracy is computed. For all four feature selectors the optimal number of trees of the Random Forest is determined as 6. The results are shown in Figure 5, which shows GRFS is highly comparable with the other three.

In the experimental results (Section VI-D3) we have shown that applying GRFS method results in significant improvement in the system performance and a great reduction in the dimension of the feature vector, as well as the required time for classification.

### *B. Classification*

In this study four different classes of classifier are applied: Support Vector Machine [51] with SMO implementation [52], Random Forest [31], Logistic Model Tree [32] and Hidden Naive Bayes [33]. A brief introduction about each classifier can be found in [34] and for a more detailed explanation please refer to [9].

## VI. Experimental results

In our preliminary study [34], we investigated the effectiveness of using texture-based feature extraction method in the classification of melanoma lesions. The texture features were extracted from four-level Wavelet-decomposition in colour channels of red, green, blue and luminance. The feature selection was performed using the Gain Ratio and correlation-based feature selection methods on a train set of 103 dermoscopy images. Classification was conducted by four classifiers; namely SVM, RF, HNB, and LMT. When an optimised subset of texture features was extracted from a test set of 102 dermoscopy images, LMT classifier was able to successfully differentiate between melanoma and benign lesions with 88.24% accuracy. As shown in Figure 6, performing a comparative study showed that our wavelet-based textural feature extraction method is highly competitive with three other texture-based methods, proposed by Nimukar et al. [24], Patwardhan et al. [26] and Walvick et al. [25]. Table II also shows the 10-fold cross validation of classification accuracy, when applying different classifiers on test set, and using features suggested by [24], [26] and [25], and our proposed wavelet-based method. The statistical significance of the comparative result has been investigated using paired t-test.

TABLE II
Comparative study: 10-fold cross validation of classification accuracy of four classifiers, using wavelet-based features suggested by [24], [26] and [25], and our proposed wavelet-based feature extraction method. The symbol * next to measures indicates a statistically significant accuracy difference compared to our method, using paired t-test (p-value = 0.05).

| Method | SVM | RF | LMT | HNB |
|---|---|---|---|---|
| Our method | 86.27 | 86.27 | 88.24 | 86.27 |
| Nimukar [24] | 83.33* | 85.29 | 85.29 | 82.35 |
| Patwardhan [26] | 52.94* | 62.74* | 62.74* | 52.94* |
| Walvick [25] | 82.35* | 86.27 | 82.35 | 79.41 |

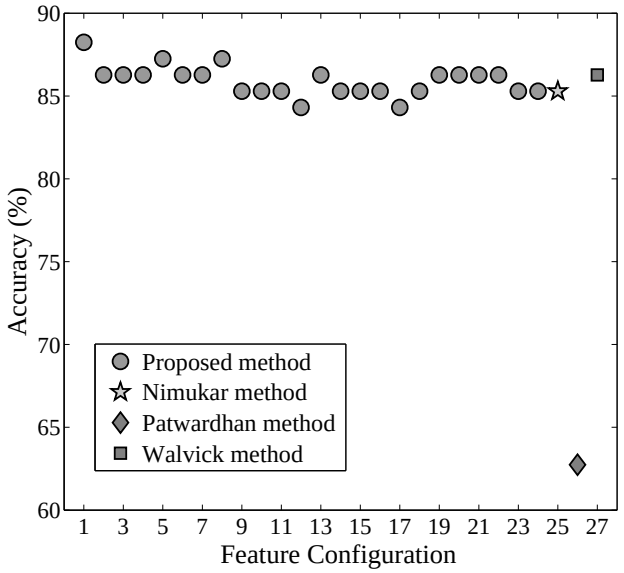


Fig. 6. Comparative study between our proposed wavelet-based method [34] and other three wavelet-based methods by [24], [25] and [26].

Figure 7(a) shows that the above-mentioned texture analysis piloted in our preliminary study [34], constitutes a foundation for the new experiments conducted in this study. Hereby, the purpose is to investigate the impact of incorporating border-based features in the classification of melanoma by appending them to the optimised texture features. To validate the comparison we repeat the texture-analysis procedure [34] in Experiment 1. This is followed by appending geometry and boundary-series features in Experiments 2 and 3 and

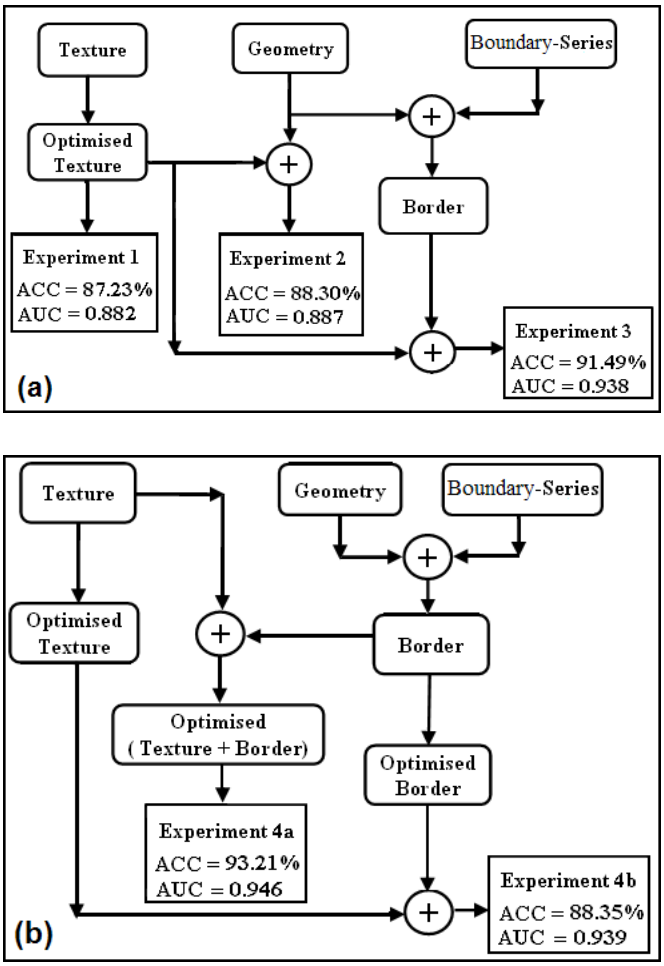


Fig. 7. Experiments flowchart: (a) Evaluating the effectiveness of three sets of features extracted from texture analysis (Experiment 1), texture and geometrical measurements (Experiment 2), and texture, geometrical and boundary-series (Experiment 3) and comparing the results,(b) Assessing the overall performance of the proposed diagnostic system.

performing a comparison between the three. Moreover, we determine the most effective feature optimisation approach and evaluate the performance of the proposed diagnostic system in an unbiased way in Experiment 4 (Figure 7(b)).

Two image sets have been used in this study; *Image-set1* is used in the first three experiments, and *Image-set2* is utilized in the last experiment. Table III shows the distribution of images for the two image sets.

TABLE III
IMAGE SETS USED IN THE STUDY; M : MALIGNANT, B : BENIGN.

| | Train set | Validation set | Test set |
|---|---|---|---|
| Image-set1: 88 M, 109 B | 48 M, 55 B | - | 40 M, 54 B |
| Image-set2: 114 M, 175 B | 40 M, 59 B | 30 M, 57 B | 44 M, 59 B |

### *A. Experiment 1: Optimised Texture Features*

As mentioned, in order to make a valid comparison with the other new experiments we repeat the experiment performed in [34] but using the new image set of 197 dermoscopy images (88 M + 109 B), labeled as *Image-set1* in Table III. The following initial steps are conducted:

(i) Dermoscopy images undergo pre-processing in order to minimise the impact of background skin in the analysis,

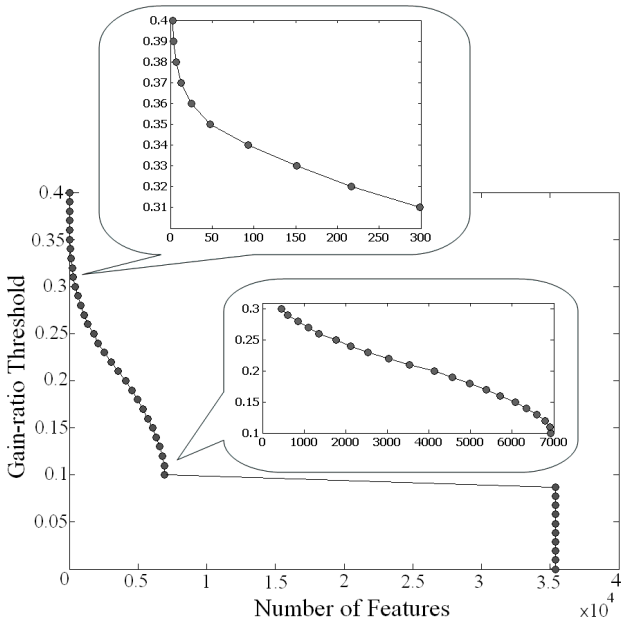


Fig. 8. Number of features selected by Gain Ratio feature selection method at different thresholds.

wherein the lesion borders are obtained manually by using dermatologists' expertise. Then, an enclosing rectangle frame is set around the lesion, the rest is cropped and the non-lesion pixel is set to black. This pre-processing step is used in all of the four experiments of this paper.

(ii) Using the MATLAB Daubechies-3 wavelet function, the texture-based features (explained in Section II of this paper) are extracted from half of the *Image-Set1* designated as *train image set* (See the first row of Table III).

(iii) The GRFS method is applied on the constructed feature vectors, resulting in various Gain Ratio values or evaluation rank. We consider these values as a cut off threshold; each corresponds to a certain subset of texture features. Figure 8 shows the number of features selected by the Gain-ratio feature selection method, at different feature evaluation rank. Figure 9 shows the obtained accuracy and the required time for classification at different feature evaluation ranks (set as threshold for the ranker search method) using the SVM classifier with RBF kernel (gamma=0.1). As illustrated in the figure, by reducing the number of features the accuracy is increased and then decreased after a certain threshold -the so-called peak phenomena [53]. Therefore the process of feature selection or optimisation is to find the optimised threshold. As expected and also shown in Figure 9 the time required for performing classification is proportional to the dimension of the optimised feature vector and thus decreases considerably as a result of the feature selection process.

(iv) By visual inspection, eleven of these thresholds are empirically chosen in range of [0.27 , 0.37] and the corresponding texture features are extracted from the second half of the dermoscopy images designated as a *test set* (See the first row of Table III). To test the effectiveness of the feature subsets selected in the previous step, we have applied the four classifiers of SVM, RF, MLT and HNB, using the corresponding parameter range shown in Table IV, on the *test set*; parameters outside the indicated regions did not produce

TABLE V
10-FOLD CROSS VALIDATION OF PERFORMANCE (ACCURACY AND AUC) OF DIFFERENT CLASSIFIERS ON THE TEST SET, USING THE 11 TEXTURE FEATURE CONFIGURATIONS RESULTING FROM THE GAIN RATIO FEATURE SELECTION.

| Configuration | Threshold | Feature No. | SVM (RBF) | RF | LMT | HNB |
|---|---|---|---|---|---|---|
| 1 | 0.27 | 1108 | 86.17 , 0.854 | 86.17 , 0.889 | 82.98 , 0.886 | 86.17 , 0.885 |
| 2 | 0.28 | 852 | 85.11 , 0.841 | 84.04 , 0.887 | 84.04 , 0.836 | 88.30 , 0.887 |
| 3 | 0.29 | 609 | 86.17 , 0.854 | 86.17 , 0.899 | 80.85 , 0.825 | 87.23 , 0.882 |
| 4 | 0.30 | 451 | 87.23 , 0.866 | 85.11 , 0.868 | 85.11 , 0.871 | 86.17 , 0.881 |
| 5 | 0.31 | 299 | 85.11 , 0.844 | 85.11 , 0.885 | 82.98 , 0.855 | 85.11 , 0.873 |
| 6 | 0.32 | 217 | 86.17 , 0.854 | 82.98 , 0.887 | 84.04 , 0.870 | 86.17 , 0.888 |
| 7 | 0.33 | 151 | 86.17 , 0.857 | 85.11 , 0.901 | 84.04 , 0.881 | 86.17 , 0.894 |
| 8 | 0.34 | 93 | 87.23 , 0.869 | 86.17 , 0.883 | 79.79 , 0.842 | 84.04 , 0.902 |
| 9 | 0.35 | 48 | 87.23 , 0.869 | **87.23 , 0.882** | 84.04 , 0.846 | 85.11 , 0.880 |
| 10 | 0.36 | 25 | 86.17 , 0.863 | 85.11 , 0.856 | 82.98 , 0.890 | 82.98 , 0.855 |
| 11 | 0.37 | 13 | 86.17 , 0.854 | 85.11 , 0.859 | 84.04 , 0.871 | 84.04 , 0.852 |

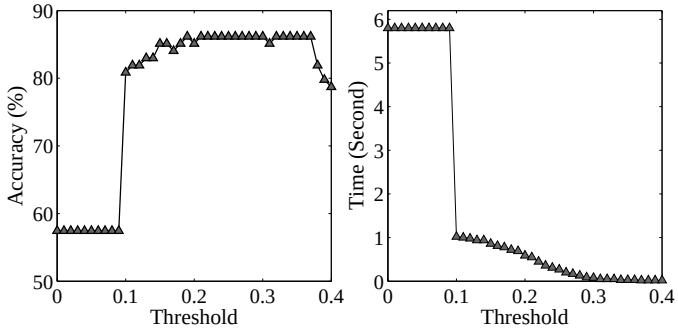


Fig. 9. Obtained accuracy and required time for classification for different thresholds using SVM classifier with RBF kernel.

TABLE IV
PARAMETERS SETTING FOR THE CLASSIFIERS

| Classifier | Parameter range |
|---|---|
| Support Vector Machine (RBF) | Gamma = [0.001, 0.5] |
| Random Forest | number of trees = [2, 200] |
| Logistic Model Tree | weightTrimBeta = [0, 0.9] |
| Hidden Naive Bayes | No parameters |

useful results. In the classification experiments, the highest accuracy obtained by each classifier is used as the measure of its performance. The Weka data mining toolkit [54] has been utilised in this study. This has resulted in the classification accuracy shown in Table V, which indicates that the best classification result (accuracy of 87.23% and AUC value of 0.882) is obtained by applying the Random Forest classifier using the 9-th configuration with 48 texture-based features. The obtained accuracy and AUC are calculated using a 10-fold cross-validation scheme [55].

In the next two experiments of this study, we expand the feature vector by augmenting the optimised texture features with geometry and boundary-series features. We then investigate the impact of such an expansion by comparing the classification results obtained from (1) texture, (2) texture and geometry, and (3) texture, geometry and boundary-series features. The test image set used in these experiments (40 melanoma and 54 benign, first row of Table III) is slightly different from what was used in our preliminary study [34]. The reason for excluding eight lesions is because the quality of these images is such that it is not feasible to perform border-based measurements. In other words, we have made this assumption that the input image contains the entire lesion[4].

### *B. Experiment 2: Optimised Texture and Geometry Features*

The purpose of this experiment as depicted in Figure 7(a) is to analyse the effect of integrating geometry with optimised texture features and assess the classification performance. The geometry features are extracted from the images in the *test set* of *Image-set1* (See the first row of Table III). The optimised texture feature sets are obtained from applying the GRFS method as per experiment 1. Eleven optimised features sets are chosen corresponding to the 11 thresholds shown in Table V.

The whole appended feature vector (texture plus geometry) is then fed into the same four classifiers to ascertain the status of the 103 lesions. Table VI shows the 10-fold cross-validation of classification performance where the highest accuracy for each of the classifier is quoted. Examination of Tables V and VI shows that there is an overall improvement in the classification results as a consequence of augmenting texture with geometry-based features; when using the RF classifier, the maximum classification accuracy has been increased from 87.23% (texture) to 88.30% (texture and geometry). Figure 10 illustrates this effect more clearly on the RF classifier, which shows that for most of the thresholds, the classifier accuracy increases when the geometry-based features are added. In the next experiment the feature vector is further extended to include the boundary-series features. Since the RF classifier exhibits the best overall results we proceed with this classifier in the remaining experiments.

### *C. Experiment 3: Optimised Texture, Geometry and Boundary-series Features*

As shown in flowchart of Figure 7(a), the purpose of this experiment is to analyse the effect of integrating border-based features (geometry and boundary-series) with texture features, and to investigate the classification performance. This is done by appending the 59 border features (the same 11 geometry features as per experiment 1 and 48 boundary-series features, Section IV) to the same 11 optimised texture feature subsets listed in Table V which were previously used in Experiments 1 and 2. The RF classifier is then applied

[4]It is worthwhile to mention that in some image acquisition setups dermoscopy images are taken such that the picture frame does only includes part of the lesion, and consequently the border can not be defined, thus any calculated border-based parameter would reflect an invalid value. The removed images in the experiment belong to these cases.

TABLE VI
10-FOLD CROSS VALIDATION OF PERFORMANCE (ACCURACY AND AUC) OF DIFFERENT CLASSIFIERS ON THE TEST SET, USING THE 11 TEXTURE FEATURE CONFIGURATIONS RESULTING FROM THE GAIN RATIO FEATURE SELECTION, COMBINED WITH GEOMETRY FEATURES.

| Configuration | Threshold | Feature No. | SVM (RBF) | RF | LMT | HNB |
|---|---|---|---|---|---|---|
| 12 | 0.27 | 1119 | 86.17 , 0.854 | 86.17 , 0.874 | 82.98 , 0.886 | 86.17 , 0.878 |
| 13 | 0.28 | 863 | 85.11 , 0.841 | 87.23 , 0.878 | 84.04 , 0.836 | 88.30 , 0.856 |
| 14 | 0.29 | 620 | 86.17 , 0.854 | 85.11 , 0.887 | 80.85 , 0.825 | 86.17 , 0.881 |
| 15 | 0.30 | 462 | 87.23 , 0.866 | 84.04 , 0.884 | 82.98 , 0.863 | 86.17 , 0.883 |
| 16 | 0.31 | 310 | 85.11 , 0.844 | 85.11 , 0.837 | 82.98 , 0.855 | 85.11 , 0.873 |
| 17 | 0.32 | 228 | 85.11 , 0.844 | 85.11 , 0.884 | 84.04 , 0.858 | 86.17 , 0.888 |
| 18 | 0.33 | 162 | 86.17 , 0.857 | 86.17 , 0.913 | 84.04 , 0.882 | 86.17 , 0.894 |
| 19 | 0.34 | 104 | 87.23 , 0.869 | 86.17 , 0.891 | 81.91 , 0.875 | 84.04 , 0.898 |
| 20 | 0.35 | 59 | 87.23 , 0.869 | 87.23 , 0.881 | 82.98 , 0.816 | 85.11 , 0.879 |
| 21 | 0.36 | 36 | 85.11 , 0.848 | **88.30 , 0.887** | 82.98 , 0.853 | 84.04 , 0.853 |
| 22 | 0.37 | 24 | 88.30 , 0.882 | 86.17 , 0.859 | 81.91 , 0.889 | 85.11 , 0.849 |

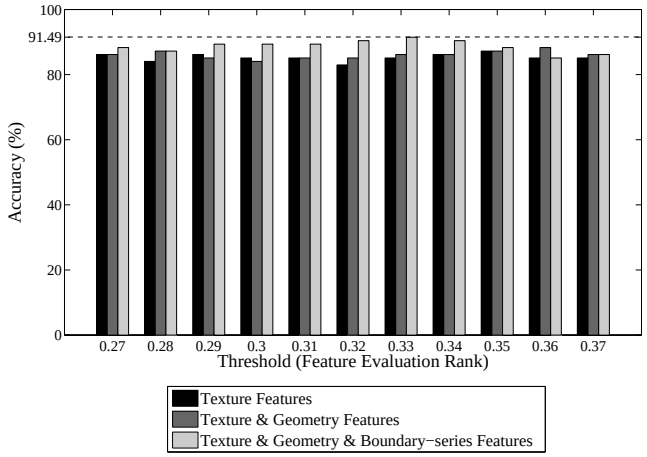


Fig. 10. 10-fold cross validation of accuracy of texture, geometry and boundary-series features at various thresholds of the Gain Ratio feature selection method, using the Random Forest classifier.

on test image set (48 malignant and 54 benign), resulting in the classification accuracies for each of the 11 feature vectors shown in Figure 10. The figure also shows the results of the highest accuracies obtained in Experiments 1 and 2 for easy comparison, where the highest accuracy (91.49%) is obtained by choosing the feature vector corresponding to the 0.33 threshold (the feature vector for this threshold comprises 151 texture, 11 geometry and 48 boundary-series features). The obtained AUC value for this feature configuration is 0.938.

### *D. Experiment 4: Performance Evaluation of the Proposed Computer-aided Diagnosis of Melanoma*

For this experiment the *Image-set2* from Table III is used. In order to to reduce bias in terms of evaluating prediction accuracy of the developed diagnostic system, we consider three independent image sets; a train, a validation and a test set of 99, 87 and 103 dermoscopy images, respectively. In other words, as our technique requires various parameters to be selected (such as what features to use and classifier specific parameters such as number of trees), these parameters need to be chosen without ever looking at the test set. The aim is to fix all the parameters prior to performing the final evaluation of the system on the test set (using 10 fold cross validation) in order to avoid any bias in selecting the features or in parameter selection[5].

As these images are in different conditions, the border detection step in two-third of them is done manually and the automated Hybrid border-detection method [21] is applied on the remainder to separate the lesion from the surrounding skin. This is followed by the pre-processing step explained in Section VI-A to prepare the images for performance evaluation of the diagnosis system.

As illustrated in Figure 7(b), we use two streams of optimisation. The first employs Global optimisation (labelled as experiment 4a in the flowchart) wherein the whole feature vector, consisting of texture and border (geometry and boundary-series) features is optimised. The second uses individual optimisations (labelled as experiment 4b in the flowchart) wherein the texture and border features are optimised individually and the resultant optimised features are appended, as further explained in the following.

*1) Global Optimisation:* In this experiment the whole feature vector, consisting of texture and border (geometry and boundary-series) is optimised, using the following three-step procedure.

(i) Feature selection: feature extraction is performed on images from the designated *train set* (the second row of Table III) and for each image a vector of 35455 features is extracted, including 35396 wavelet-based texture features and 59 border features (11 geometry and 48 boundary-series features). The extracted features are evaluated and ranked by using the Gain Ratio method. This results in various evaluation ranks in the range of [0, 0.6], each corresponding to a subset (a configuration) of features.

(ii) Parameter tuning: 26 various evaluation ranks, in range of [0.35 to 0.6] with 0.1 increment, are empirically chosen as thresholds and the corresponding feature subsets are extracted from the images of the designated *validation set* (the second row of Table III). Then, the RF classifier is applied on each feature configuration to classify the images of the validation set. In each classification process, various parameters of the classifier (number of trees) are tested and the parameter which resulted in highest classification accuracy is selected.

[5]One might attempt to merge all three subsets into a single dataset and then do 10 fold cross validation. However, in each fold of the validation, different features would be selected and different classifier parameters used, making it difficult to explain behaviour. Furthermore, 10 fold cross validation can be misused, if it is run repeatedly by the user on the same dataset, searching for best prediction accuracy by tweaking parameter behaviour.

TABLE VII
NUMBER OF FEATURES AND THE RESULTANT 10-FOLD CROSS-VALIDATION OF ACCURACY AND AUC VALUE USING RANDOM FOREST CLASSIFIER AT VARIOUS THRESHOLDS OF THE GAIN RATIO FEATURE SELECTION METHOD, APPLIED ON VALIDATION IMAGE SET.

| Threshold | Feature No. | No. of trees | Accuracy | AUC |
|---|---|---|---|---|
| 0.60 | 2 | 10 | 66.67% | 0.662 |
| 0.59 | 2 | 10 | 66.67% | 0.662 |
| 0.58 | 8 | 50 | 64.37% | 0.618 |
| 0.57 | 10 | 10 | 87.36% | 0.827 |
| 0.56 | 12 | 5 | 86.21% | 0.828 |
| 0.55 | 14 | 10 | 85.06% | 0.829 |
| 0.54 | 18 | 5 | 86.21% | 0.820 |
| 0.53 | 18 | 5 | 86.21% | 0.820 |
| **0.52** | **23** | **2** | **86.21%** | **0.832** |
| 0.51 | 23 | 2 | 86.21% | 0.832 |
| 0.50 | 29 | 10 | 85.06% | 0.843 |
| 0.49 | 44 | 2 | 86.21% | 0.83 |
| 0.48 | 51 | 3 | 86.21% | 0.848 |
| 0.47 | 62 | 10 | 86.21% | 0.872 |
| 0.46 | 67 | 90 | 85.06% | 0.887 |
| 0.45 | 78 | 70 | 85.06% | 0.823 |
| 0.44 | 91 | 5 | 87.36% | 0.829 |
| 0.43 | 99 | 125 | 86.21% | 0.891 |
| 0.42 | 122 | 55 | 83.91% | 0.847 |
| 0.41 | 149 | 200 | 85.05% | 0.900 |
| **0.40** | **166** | **15** | **87.36%** | **0.913** |
| 0.39 | 202 | 50 | 85.06% | 0.887 |
| 0.38 | 225 | 220 | 85.06% | 0.915 |
| 0.37 | 275 | 10 | 85.06% | 0.883 |
| 0.36 | 302 | 5 | 86.21% | 0.879 |
| 0.35 | 387 | 20 | 85.06% | 0.886 |

Table VII shows the number of features at various thresholds, the resultant 10-fold cross-validation of the classification performance (accuracy and AUC) and the corresponding classification parameter; number of trees of the RF classifier. As shown in the table the best AUC and accuracy results are obtained if 0.4 threshold, 166 features, and 15 trees in the random forest classifier are employed. For a more accurate evaluation of the system performance, and considering the number of images in the test set (103 images, according to the second row of Table III), we also consider threshold of 0.52 with 23 number of features, as explained in the following.

(iii) Classification: the parameters (Gain Ratio, threshold and number of trees of the RF) tuned in the previous step are then used to evaluate the overall performance of the proposed diagnostic system. The optimised feature configuration at threshold of 0.4 with 166 features is extracted from the images designated as the *test set* (the second row of Table III). The RF classifier with 15 trees is then applied and resulted in accuracy of 93.21% (using 10-fold cross-validation). As stated above the highest performance is obtained when 166 features are used on our image set. It might be noted that the ratio of features to image set used in this study could be considered to be high and may therefore introduce a degree of bias in the results due to problems associated with over-fitting. To obtain more reliable and robust results smaller ratios need to be used if somewhat lower accuracy and AUC could be tolerated. Therefore, we have also tested the system using threshold of 0.52, with 23 number of features, which results in an accuracy of 91.26%. Other performance measures of the proposed approach are listed in the first and second row of Table VIII.

TABLE VIII
10-FOLD CROSS-VALIDATION OF PERFORMANCE OF THE PROPOSED DIAGNOSTIC SYSTEM, USING TRAIN SET OF 99 IMAGES, VALIDATION SET OF 87 IMAGES AND TEST SET OF 103 DERMOSCOPY IMAGES, APPLYING THE TWO OPTIMISATION STREAMS.

| Stream | Feature No. | Accuracy | Sensitivity | Specificity | AUC |
|---|---|---|---|---|---|
| Global | 23 | 91.26% | 84.09% | 96.61% | 0.937 |
| Global | 166 | 93.21% | 90.90% | 94.92% | 0.946 |
| Individual | 387 | 88.35% | 88.64% | 89.06% | 0.939 |

*2) Individual Optimisation:* In this experiment the texture and border features are optimised individually and the resultant optimised features are combined. In order to do that, the first two steps of the procedure performed in the previous section (global Optimisation) is conducted independently for each of border and texture features. This involves performing feature selection and parameter tuning using *train* and *validation image sets* (the second row of Table III) to determine the optimised border features and optimised texture features. By visual inspection, thresholds for texture and border features are empirically chosen in range of [0.30, 0.59] and [0.11,0.57], respectively. The optimised texture feature vector includes 379 features, and the optimised border feature vector includes 8 features. The individual accuracy of texture and border features on evaluation image set is 78.16% and 90.81%, respectively. Then, the two optimised sets are combined yielding in total 387 optimised features, and the classifier parameter is tuned using the validation image set (number of trees = 35). Finally the RF classifier is applied on the test image set, using the optimised feature set and the classification parameter tuned in the previous step. This results in an accuracy of 88.35% (using 10-fold cross-validation). Other performance measures of the proposed method are listed in the second row of Table VIII.

*3) Comparison and Observations:* Table VIII shows the 10-fold cross-validation of performance (evaluated in terms of accuracy, sensitivity, specificity and the area under the ROC curve) of the proposed diagnostic system for the two optimisation streams; global and individual, which shows the overall performance of the global stream is higher than individual approach (higher accuracy, smaller feature space).

An interesting observation of this experiment is the distribution of features in the optimised feature vector; among 23 features (selected at the threshold of 0.52), four features are border-based and 19 are texture-based. Also, at the 0.4 threshold, out of 166 selected features only eight (one is geometry based and the rest are boundary-series based) are border (4.82%) and the majority are texture-based features. Table IX shows the ranking of these border-based features. As shown in the table, among the eight border features that contribute in the reported accuracy, the perimeter feature was adopted from previous works, and the other seven features are new and extracted from the boundary-series of the lesion.

Figure 11 shows the Gain Ratio values of the features nominated as optimised. We have estimated the contribution of the border features in the optimised feature set by calculating the accumulative border features' Gain Ratio divided by accumulative features' Gain Ratios (Equation 24). Considering the 35 highest-ranked features, which consists of six border features and 29 texture features, the approximate contribution of

TABLE IX
RANKINGS OF THE EIGHT OPTIMISED BORDER FEATURES APPEARING IN THE OPTIMISED FEATURE SET.

| Rank | feature | Description |
|---|---|---|
| 9 | Norm(BS) | BS: Boundary series |
| 10 | Norm(WA1(BS)) | WAn:Wavelet approximation component |
| 13 | Norm(WA3(BS)) | level n |
| 14 | Norm(WA2(BS)) | |
| 31 | Norm(Hist(BS)) | Hist: Histogram |
| 34 | Eng(Hist(BS)) | Eng: Energy |
| 139 | Perimeter | belongs to the Geometry category |
| 140 | Mean(Hist(BS)) | |

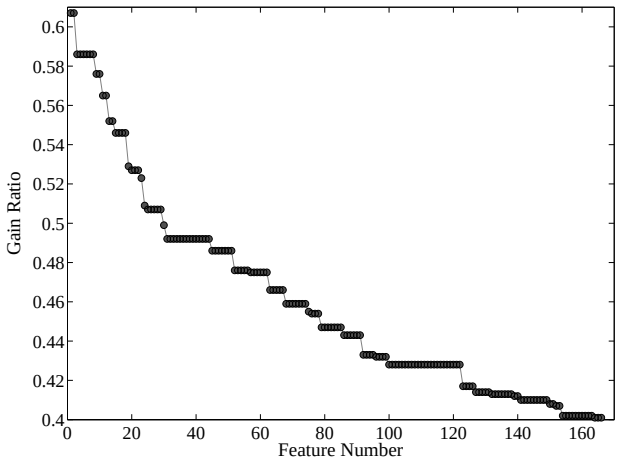


Fig. 11. The Gain Ratio values of the features in the optimised feature set.

border-based features is 17% , however its overall contribution in the optimised set of 166 features is only 5%.

$$\text{Approximate Contribution} = \frac{\sum GainRatio_{(Border)}}{\sum GainRatio_{(All)}} \times 100 \quad (24)$$

According to the estimated contribution, we suggest that texture-based features play the dominant role in the classification of melanoma. This finding is in fact fully aligned with the clinical knowledge we obtained from our meetings with experienced dermatologists at the Royal Melbourne Hospital, Melbourne, Australia. We are informed that, when they examine an image, they focus on the changes in the internal structure of the lesion while its external attributes such as border and shape are considered secondary.

The 20 first-ranked texture-based features are listed in Table X. Note the reason that ranking extends to 24 is that four features among the first 24 features are border-based (and thus not included in the table), and the remainder 20 features are texture-based. The following conclusions are derived from this table: (i) six out of 20 features belong to the nodes from the forth level, which vindicates our use of four-level wavelet decomposition as opposed to the three-level decomposition

TABLE X
THE 20 FIRST-RANKED TEXTURE-BASED FEATURES; L: LUMINANCE, R: RED, G: GREEN, B: BLUE.

| Rank | Measure | Channel | Level | Node |
|---|---|---|---|---|
| 1 | Average-energy | L | 3 | 4.1.3 |
| 2 | Average-energy | G | 3 | 4.1.1 |
| 3 | Kurtosis | B | 1 | 4 |
| 4 | Average-energy | G | 2 | 4.1 |
| 5 | Average-energy | L | 3 | 4.1.1 |
| 6 | Energy | L | 3 | 4.1.1 |
| 7 | Standard deviation | L | 3 | 4.1.1 |
| 8 | Kurtosis | G | 1 | 4 |
| 11 | Average-energy | L | 4 | 4.1.1.2 |
| 12 | Average-energy | L | 2 | 4.1 |
| 15 | Kurtosis | B | 2 | 4.2 |
| 16 | Standard deviation | L | 4 | 1.1.1.3 |
| 17 | Energy | L | 4 | 1.1.1.3 |
| 18 | Average-energy | L | 4 | 2.3.1.1 |
| 19 | Average-energy | L | 3 | 4.1.2 |
| 20 | Average-energy | L | 2 | 4.2 |
| 21 | Kurtosis | L | 1 | 4 |
| 22 | Sub-level ratio of Average-energy | B | 4 | 1.1.1.1 |
| 23 | Average-energy | B | 4 | 4.1.1.1 |
| 24 | Average-energy | L | 3 | 4.4.4 |

applied in the other studies[6], (ii) texture features extracted from colour channels of Green and Blue are among the first 20 optimised texture features, which justifies our decision to incorporate the independent colour channels of Red, Green, Blue along with the Luminance, as opposed to other studies where Luminance only is used[7], (iii) the optimised features are derived from both low and high frequencies which justifies our choice of tree-structured decomposition over the pyramid-structured wavelet, and (iv) particular measures we proposed to apply in this diagnostic system (*e.g.* Kurtosis) are prominent in the final optimised feature set. Table XI shows the advantages

TABLE XI
EFFECT OF GAIN RATIO-BASED FEATURE SELECTION METHOD.

| | Feature No. | Accuracy | Time (sec) |
|---|---|---|---|
| without feature selection | 35455 | 81.55% | 2.45 |
| with feature selection | 23 | 91.26% | 0.05 |

of applying the GRFS method in our system, which includes a significant reduction (by a factor of 1542) in the dimension feature space and in the time required for classification (50-times). Moreover, the classification accuracy is improved by 12%. Note that the figures listed in Table XI are obtained by applying the RF classifier on the test image set, using the classifier parameters tuned over the validation image set.

[6]To assess the statistical significance of 4th level of wavelet decomposition, we can consider the proportion of texture based features which are Level 4 in the original set of 35396 texture based features and compare it to the proportion of Level 4 features which get automatically selected using gain ratio for our optimised feature set (which has 158 texture based features). The optimised feature set contains 67 level 4 features and the original feature set contains 6656 level 4 features. Thus, using a hypergeometric distribution, we can compute the probability that the optimised feature contains at least 67 Level 4 features, under a model in which 158 features are randomly selected without replacement from 35396 features. This probability is 6.6E-12, indicating that there is a statistically significant overrepresentation of Level 4 features in the optimised feature set (using a significance p-value of 0.05).

[7]In the optimised feature set (166 features), the number of features extracted from R, G, B adds up to 97, i.e. more than 58%, and if we exclude the 8 border based features which are colour-independent, the ratio would be 62%.

## VII. Summary and Conclusion

In this paper we have presented a computer-aided diagnosis system for melanoma. The system covers the main diagnosis components of segmentation, feature extraction, feature selection and classification. We have elected to use both manual (under dermatologists' guidance) and automated (by applying the Hybrid border-detection method) borders in the segmentation step. This, we believe, is essential in accurately assessing the ability of diagnostic system in both semi and fully automated manners. It is also worthwhile to highlight that the system proposed in this paper is not designed to bring about complete autonomy in the diagnostic process or replace human judgment, but rather has potential as an assistive system that could be used to screen images and direct physicians attention to cases that have high risk.

The feature extraction process involved a guided generation of categorized features, extracted from (i) wavelet-based texture analysis, (ii) geometrical measurements, and (iii) boundary-series analysis in spatial and frequency domains. The texture features were extracted from application of four-level wavelet decomposition on red, green, blue and luminance colour channels in addition to employing a comprehensive set of statistical measures on all the nodes of the wavelet-tree. The geometry measures were extracted from indices that are commonly used in the literature as well as novel ones defined in this paper to reflect the expert opinion of dermatologists. Boundary-series analysis was applied on dermoscopy images for the first time in this paper, wherein features were extracted by constructing a boundary-series model of the lesion border and analysing it in spatial and frequency domains by applying histogram analysis and wavelet transform.

A computationally efficient feature selection method, based on Gain Ratio, is adopted and shown to be highly comparable with three other well-known feature selection methods, namely; Information Gain, Chi-squared and ReliefF. It has been shown that the use of the GRFS method provides significant improvement in the classification performance (10% increment) as well as a great reduction in the dimension of the feature vector and the required classification computational time. Furthermore, four advanced classes of classification techniques; namely, Support Vector Machine, Random Forest, Logistic Model Tree and Hidden Naive Bayes are employed, among which the Random Forest classifier was shown to be highly effective in the classification of melanoma lesions.

Comprehensive sets of experiments were conducted in this study to analyse the effectiveness of applying various combination of features in the classification of melanoma. Firstly, the texture-based features were analysed resulting in a classification accuracy of 87.23%. In the other two experiments the feature vector was extended, where the geometry-based features and then the boundary series features were appended to the optimised texture-based features. In both cases improvements in the classification accuracy (88.30% and 91.49%) were obtained. These results have demonstrated the superiority of integrating border with texture features over using texture features alone.

In the last experiment the overall performance of the proposed system was evaluated by conducting two streams of optimisation on a set of 289 dermoscopy images (114 malignant and 175 benign) subdivided into three image sets; train, validation and test. Two optimisation streams were studies: (1) the global optimisation, in which the whole feature vector, consisting of texture and border features is optimised, (2) Individual optimisation wherein the texture and border features are optimised individually and the resultant optimised features are integrated. The experimental results showed that by applying the Gain Ratio feature selection method and the Random Forest classifier, global optimisation outperforms individual optimisation. The proposed system achieves and accuracy of 91.26% and AUC value of 0.937, when 23 features are used. Although the results cited here were obtained through experiments conducted on a particular image set, they indicate that high diagnostic accuracy can be achieved when existing features are optimally blended with novel ones. The obtained performance of our system is in fact highly comparable with the state-of-the-art melanoma recognition systems reported in the recent survey by Maglogiannis et al. [23]. However, the authors acknowledge that due to lack of a standard benchmark for dermoscopy (melanoma) imaging, it is not easily feasible to provide a comprehensive and quantitative comparative study among the existing classification methods.

An important finding is the clear advantage gained in complementing texture with border and geometry features compared to using texture information only. Another interesting finding of this paper is that in the melanoma classification application texture features play the dominant role and make much higher contribution compared to border-based features. Among border features, which include geometry and boundary-series features, the latter has been shown to be more informative and obtain higher evaluation ranks (Gain Ratio values) than geometry in this application.

## Acknowledgment

The authors would like to thank Dr. Emre Celebi from Department of Computer Science, Louisiana State University, USA, for allowing access to the data used in this study, and Dr. George Varigos from Royal Melbourne Hospital, Australia, for providing us with an in-depth knowledge about clinical diagnosis of melanoma. This research is supported by NICTA Victoria Research Laboratory, Australia.

## References

[1] "Australia skin cancer facts and figures," Available at: http://www.cancer.org.au/, accessed September 2009.

[2] H. Pehamberger, A. Steiner, and K. Wolff, "In vivo epiluminescence microscopy of pigmented skin lesions. I: Pattern analysis of pigmented skin lesions," *Journal of the American Academy of Dermatology*, vol. 17, pp. 571–583, 1987.

[3] W. Stolz, A. Riemann, and A. Cognetta, "ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma," *European Journal of Dermatology*, vol. 4, pp. 521–527, 1994.

[4] S. Menzies, C. Ingvar, and W. McCarthy, "A sensitivity and specificity analysis of the surface microscopy features of invasive melanoma," *Melanoma Research*, vol. 6, no. 1, pp. 55–62, 1996.

[5] G. Argenziano, G. Fabbrocini, and P. Carli, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Arch. of Dermatology*, vol. 134, pp. 1563–1570, 1998.

[6] J. Henning, S. Dusza, S. Wang, and A. Marghoob, "The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy," *Journal of the American Academy of Dermatology*, vol. 56, no. 1, pp. 45–52, 2007.

[7] R. Braun, H. Rabinovitz, M. Oliviero, A. Kopf, and J. Saurat, "Dermoscopy of pigmented lesions," *Journal of the American Academy of Dermatology*, vol. 52, no. 1, pp. 109–121, 2005.

[8] G. Argenziano, H. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera, and M. Binder, "Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet," *Journal of the American Academy of Dermatology*, vol. 48, pp. 679–693, 2003.

[9] R. Garnavi, "Computer-aided diagnosis of melanoma," Ph.D. dissertation, University of Melbourne, Australia, 2011.

[10] V. Skladnev, A. Gutenev, S. W. Menzies, L. Bischof, G. Talbot, E. Breen, and M. Buckley, "Diagnostic feature extraction in dermatological examination," 2004, patent Number: 10/478078, United States.

[11] M. Jamora, B. Wainwright, S. Meehan, and J. Bystryn, "Improved identification of potentially dangerous pigmented skin lesions by computerised image analysis," *Arch. of Dermatology*, vol. 139, pp. 195–198, 2003.

[12] P. Rubegni, M. Burroni, G. Cevenini, R. Perotti, G. Dell'Eva, P. Barbini, M. Fimiani, and L. Andreassi, "Digital dermoscopy analysis and artificial neural network for the differentiation of clinically atypical pigmented skin lesions: A retrospective study." *The journal of investigative dermatology*, vol. 119, pp. 471–474, 2002.

[13] K. Hoffmann, T. Gambichler, and A. Rick, "Diagnostic and neural analysis of skin cancer. A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy," *British Journal of Dermatology*, vol. 149, pp. 801–809, 2003.

[14] M. Elbaum, "Computer-aided melanoma diagnosis," *Dermatologic clinics*, vol. 20, pp. 735–747, 2002.

[15] J. Boldrick, C. Layton, J. Nguyen, and S. Swetter, "Evaluation of digital dermoscopy in a pigmented lesion clinic: Clinician versus computer assessment of malignancy risk," *Journal of the American Academy of Dermatology*, vol. 56, no. 3, pp. 417–421, 2007.

[16] A. Perrinaud, O. Gaide, and L. French, "Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? A study comparing the results of three systems," *British Journal of Dermatology*, vol. 157, pp. 926–933, 2007.

[17] M. E. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, "Lesion border detection in dermoscopy images," *Computerised Medical Imaging and Graphics*, vol. 33, no. 2, pp. 148–153, 2009.

[18] H. Iyatomi, H. Oka, M. Saito, A. Miyake, M. Kimoto, J. Yamagami, S. Kobayashi, A. Tanikawa, M. Hagiwara, K. Ogawa, G. Argenziano, H. P. Soyer, and M. Tanaka, "Quantitative assessment of tumour extraction from dermoscopy images and evaluation of computer-based extraction methods for an automatic melanoma diagnostic system," *Melanoma Research*, vol. 16, no. 2, pp. 183–190, 2006.

[19] M. E. Celebi, Y. A. Aslandogan, W. V. Stoecker, H. Iyatomi, H. Oka, and X. Chen, "Unsupervised border detection in dermoscopy images," *Skin Research and Technology*, vol. 13, pp. 454–462, 2007.

[20] R. Garnavi, M. Aldeen, M. E. Celebi, A. Bhuiyan, C. Dolianitis, and G. Varigos, "Automatic segmentation of dermoscopy images using histogram thresholding on optimal color channels," *International Journal of Medicine and Medical Sciences*, vol. 1, no. 2, pp. 126–134, 2010.

[21] R. Garnavi, M. Aldeen, M. E. Celebi, S. Finch, and G. Varigos, "Border detection in dermoscopy images using hybrid thresholding on optimized color channels," *Computerized Medical Imaging and Graphics, Special Issue: Skin Cancer Imaging*, 2010, doi: 10.1016/j.compmedimag.2010.08.001.

[22] M. Celebi, H. Kingravi, B. Uddin, H. Iyatomi, Y. Aslandogan, W. Stoecker, and R. Moss, "A methodological approach to the classification of dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 31, pp. 362–373, 2007.

[23] I. Maglogiannis and C. Doukas, "Overview of advanced computer vision systems for skin lesions characterization," *IEEE Trans. on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 721–733, 2009.

[24] A. Nimunkar, A. Dhawan, P. Relue, and S. Patwardhan, "Wavelet and statistical analysis for melanoma," in *SPIE Medical Imaging: Image Processing*, vol. 4684, 2002, pp. 1346–1352.

[25] R. Walvick, K. Patel, S. Patwardhan, and A. Dhawan, "Classification of melanoma using wavelet-transform-based optimal feature set," in *SPIE Medical Imaging: Image Processing*, vol. 5370, 2004, pp. 944–951.

[26] S. Patwardhan, A. Dhawan, and P. Relue, "Classification of melanoma using tree structured wavelet transforms," *Computer Methods and Programs in Biomedicine*, vol. 72, pp. 223–239, 2003.

[27] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Trans. on Medical Imaging*, vol. 20, pp. 233–239, March 2001.

[28] H. Handels, T. Rob, J. Kreuschb, H. Wolffb, and S. Poppla, "Feature selection for optimized skin tumor recognition using genetic algorithms," *Artificial Intelligence in Medicine*, vol. 16, pp. 283–297, 1999.

[29] S. Dreiseitl, L. Ohno-Machado, and H. Kittler, "A comparison of machine learning methods for the diagnosis of pigmented skin lesions," *Journal of Biomedical Informatics*, vol. 34, pp. 28–36, 2001.

[30] M. Wiltgen, A. Gergerb, and J. Smolle, "Tissue counter analysis of benign common nevi and malignant melanoma," *International Journal of Medical Informatics*, vol. 69, pp. 17–28, 2003.

[31] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[32] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," *Machine Learning*, vol. 95, pp. 161–205, 2005.

[33] H. Zhang, L. Jiang, and J. Su, "Hidden naive bayes," in *Twentieth National Conference on Artificial Intelligence*, 2005, pp. 919–924.

[34] R. Garnavi, M. Aldeen, and J. Bailey, "Classification of melanoma lesions usingwavelet-based texture analysis," in *Digital Image Computing: Techniques and Applications (DICTA2010)*, Australia, 2010, pp. 75–81.

[35] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, 1989.

[36] T. Chang and C. J. Kuo, "Texture analysis and classification with tree structured wavelet transform," *IEEE Trans. on Information Technology in Biomedicine*, vol. 2, pp. 429–441, 1993.

[37] W. Stoecker, W. Li, and R. Moss, "Automatic detection of asymmetry in skin tumors," *Computerized Medical Imaging and Graphics*, vol. 16, no. 3, pp. 191–197, 1992.

[38] I. Maglogiannis and D. Kosmopoulos, "Computational vision systems for the detection of malignant melanoma," *Oncology Reports*, vol. 15, pp. 1027–1032, 2006.

[39] Y. Chang, R. Stanley, R. Moss, and W. Stoecker, "A systematic heuristic approach for feature selection for melanoma discrimination using clinical images," *Skin Research and Technology*, vol. 11, pp. 165–178, 2005.

[40] Z. She, Y. Liu, and A. Damatoa, "Combination of features from skin pattern and ABCD analysis for lesion classification," *Skin Research and Technology*, vol. 13, pp. 25–33, 2007.

[41] K. Clawson, P. Morrow, and B. Scotney, "Determination of optimal axes for skin lesion asymmetry quantification," in *IEEE International Conference on Image Processing*, vol. 2, 2007, pp. 453–456.

[42] V. Ng, B. Fung, and T. Lee, "Determining the asymmetry of skin lesion with fuzzy borders," *Computers in Biology and Med.*, vol. 35, pp. 103–120, 2005.

[43] A. Bono, S. Tomatis, and C. Bartoli, "The ABCD system of melanoma detection: A spectrophotometric analysis of the asymmetry, border, color, and dimension," *Cancer*, vol. 85, no. 1, pp. 72–77, January 1999.

[44] E. Claridge, P. Hall, M. Keefe, and J. Allen, "Shape analysis for classification of malignant melanoma," *Journal of Biomedical Engineering*, vol. 14, pp. 229–234, May 1992.

[45] I. Maglogiannis and E. Zafiropoulos, "Characterization of digital medical images utilizing support vector machines," *BMC Medical Informatics and Decision Making*, vol. 4, 2004.

[46] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. Springer, 2006.

[47] L. Ye and E. Keogh, "Time series shapelets: a novel technique that allows accurate, interpretable and fast classification," *Data Mining and Knowledge Discovery*, vol. 22, no. 1-2, pp. 149–182, 2011.

[48] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. on Knowledge and Data Engineering*, vol. 17, pp. 491–502, 2005.

[49] M. Hall, "Correlation-based feature subset selection for machine learning," Ph.D. dissertation, University of Waikato, New Zealand, 1998.

[50] I. Kononenko, "Estimating attributes: Analysis and extensions of relief," in *European Conference on Machine Learning*, 1994, pp. 171–182.

[51] V. N. Vapnik, *Statistical learning theory*. New York: Wiely, 1998.

[52] J. Platt, "Fast training of support vector machines using sequential minimal optimization," pp. 185–208, 1999.

[53] C. Sima and E. Dougherty, "The peaking phenomenon in the presence of feature-selection," *Pattern Recog. Let.*, vol. 29, pp. 1667–1674, 2008.

[54] I. H. Witten and E. Frank, *Practical machine learning tools and techniques*, 2nd ed. San Francisco: Morgan Kaufmann, 2005.

[55] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection." in *Proceedings of the 14-th International Joint Conference on Artificial Intelligence*, 1995, pp. 1137–1143.