

Classification of Melanoma Lesions Using Wavelet-based Texture Analysis

Rahil Garnavi, Mohammad Aldeen
NICTA VRL, Dept. of Electrical & Electronic
Engineering, The University of Melbourne
rgarnavi@ee.unimelb.edu.au,
aldeen@unimelb.edu.au

James Bailey
Dept. of Computer Science & Software
Engineering, The University of Melbourne
baileyj@unimelb.edu.au

Abstract

This paper presents a wavelet-based texture analysis method for classification of melanoma. The method applies tree-structured wavelet transform on different color channels of red, green, blue and luminance of dermoscopy images, and employs various statistical measures and ratios on wavelet coefficients. Feature extraction and a two-stage feature selection method, based on entropy and correlation, were applied to a train set of 103 images. The resultant feature subsets were then fed into four different classifiers: support vector machine, random forest, logistic model tree and hidden naive bayes to classify melanoma in a test set of 102 images, which resulted in an accuracy of 88.24% and ROC area of 0.918. Comparative study carried out in this paper shows that the proposed feature extraction method outperforms three other wavelet-based approaches.

1. Introduction

Malignant melanoma, the most dangerous type of skin cancer and among the four most common cancers in Australia, is the most common cancer in people aged 15–44 years [1]. Melanoma, if detected at early stages, is curable. However, differentiation of melanoma from other pigmented skin lesions is not trivial even for experienced dermatologists. Failure to diagnose a newly developed melanoma lesion leads to lethal advanced melanoma.

Epiluminescence microscopy or dermoscopy is a non-invasive in vivo imaging technique which allows for 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 [2], such as pattern analysis, ABCD rule of dermoscopy, Menzies method, 7-point checklist, and the CASH algorithm, the diagnosis of melanoma has been improved compared with the simple naked-eye examination between 5% and 30% depending on the type of skin le-

sion and the experience of the dermatologist [5]. However, clinical diagnosis of melanoma is inherently subjective and its accuracy has been an issue of concern, especially with equivocal pigmented lesions [5]. Despite the use of dermoscopy, the accuracy of expert dermatologists in diagnosing melanoma is still estimated to be about 75–84% [2].

Due to the enhancements in skin imaging technology and image processing techniques in the recent years, there has been a significant increase in interest in the computer aided diagnosis of melanoma, which aims to remove subjectivity and uncertainty from the diagnostic process and provide a reliable second-hand opinion to dermatologists. However, it is widely acknowledged that much higher accuracy is required for computer-based algorithms to be adopted routinely in the diagnostic process [4, 17].

A computer aided diagnosis of melanoma generally comprises four components; image acquisition, border detection, feature extraction, and classification; the latter two are the main focus of this paper.

Feature extraction is used to extract the most important features that accurately characterize a lesion. These features are mainly similar to those visually detected by dermatologists. In computerized melanoma detection systems, feature extraction has been generally based on the conventional clinical algorithm of ABCD-rule of dermoscopy, where asymmetry, border irregularity, color and diameter (or differential structures) of the lesion are examined. Numerous feature extraction methods have been proposed and different image processing techniques have been employed [7, 14] to obtain color, border and texture information of the lesion. It has been shown that using wavelet coefficient along with ABCD features augment the classification accuracy [10]. Texture-based features have been explored in the literature, however, in this paper we propose a complementary feature extraction method which extracts the textural information of the lesion using a 4-level wavelet decomposition on red, green, blue and luminance color channels, and applies a variety of measures and ratios.

Classification is the final step of the diagnosis pro-

cess, where the extracted features are utilized to ascertain whether the lesion is malignant, benign or suspicious. The most popular classification methods that have been applied to computer-based melanoma recognition include discriminant analysis, artificial neural network, K-nearest neighbourhood, support vector machine and decision trees [14]. In this study we have used and compared various classifiers; namely, support vector machine, random forest, logistic model tree, and hidden naive bayes.

The rest of the paper is organized as follows. An overview of the existing wavelet-based texture analysis methods in dermoscopy images is provided in Section 2. The proposed feature extraction method is detailed in Section 3. Section 4 explains the proposed feature selection method. The proposed classification method is provided in Section 5. Experimental results and discussions are presented in Section 6. Section 7 provides the conclusion.

2. Wavelet-based texture analysis in dermoscopy images

In clinical diagnostic approaches (e.g. ABCD rule of dermoscopy and pattern analysis) dermatologists look into the visual differences within the lesion and also changes in the appearance of the lesion over the time. These visual characteristics can be captured through texture analysis. Wavelet-based texture analysis provides a multiresolution analytical platform which enable us to characterize a signal (an image) in multiple spatial/frequency spaces. The multi-scale characteristics of wavelet can be very useful since dermoscopy images are taken under different circumstances such as various image acquisition set up (lighting, optical zooming, etc) and versatile skin colors.

The 2D wavelet transform has been widely applied in image processing applications. There exists two wavelet structure; (1) Pyramid-structured wavelet transform which decomposes a signal into a set of frequency channels with narrower bandwidths in lower frequency channels, useful for signals which their important information lies in low frequency components [8], (2) Tree-structured wavelet analysis which provides low, middle and high frequency decomposition which is done by decomposing both approximate and detail coefficients as shown in Figure 1. In dermoscopy image analysis, the lower frequency components reveal information about the general properties (shape) of the lesion, which is clinically important, and the higher frequency decomposition provides information about the textural detail and internal patterns of the lesion which is also significant in the diagnosis. Thus the decomposition of all frequency channels are useful in this application. Therefore, the tree-structured wavelet analysis can be more informative for classification of skin lesions.

Some applications of wavelet transform to classify

melanoma has been reported in the literature; Nimunkar et al. [15] applied pyramid-structured wavelet transform to differentiate melanoma from dysplastic nevi on a set of 28 images. The luminance color channel was decomposed to three levels and different statistical ratios such as energy, entropy, etc. was used to initially build up a vector of 34 features. Using statistical analysis, this was ultimately narrowed down to five features.

Inspired by Chang and Ku [8]'s method on tree-structured wavelet transform, Patwardhan et al. [16] proposed an adaptive wavelet-based tree-structured method to classify melanoma on a set of 60 images. The average energy, maximum energy and fractional energy ratios were used, where the luminance image was decomposed into three levels, resulted in a vector comprising 231 features. Then, a statistical analysis based on population mode was conducted to select those features with bimodal distribution. As a result, the process yielded five optimal features.

The two feature sets proposed in [15] and [16], were later combined [20] and optimized using principal component analysis (over 33 images) to obtain an optimal set of four features. The classification was performed on a set of 27 images using back propagation neural network, leading to true positive of 100% and false positive of 23.53% which was shown to be higher than each individual feature set proposed by [15] and [16].

In this paper we introduce a feature extraction approach that yields higher classification accuracy than those reported in [15], [16] and [20], as outlined in the next section.

3. Proposed feature extraction method

In this study, we expand the feature extraction methods reported in [15], [16] and [20] by incorporating red, blue and green color channel, in addition to the luminance color channel used by the previous studies, because each color channel maintain certain information of the lesion which can be useful in characterizing and classification of the lesion. Furthermore, we decompose the wavelet-tree to the forth level for the fact that certain textural details, which are significant in the diagnosis process, might appear in deeper levels. Moreover, different statistical measures and ratios are proposed to extract useful information from the images and utilized for the classification.

In this paper we use eight measures of energy (E), mean (M), standard deviation (Std), entropy (H), average-energy (AvgE), skewness (S), kurtosis (K), and norm (N), expressed in Equations 1–8. Figure 2 displays a schematic illustration of wavelet tree with nodes marked by circles. 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 which has 340 nodes in total (4 nodes in first level, 16 nodes in second level, 64 nodes on third level and 256 nodes on forth level). The eight measure will

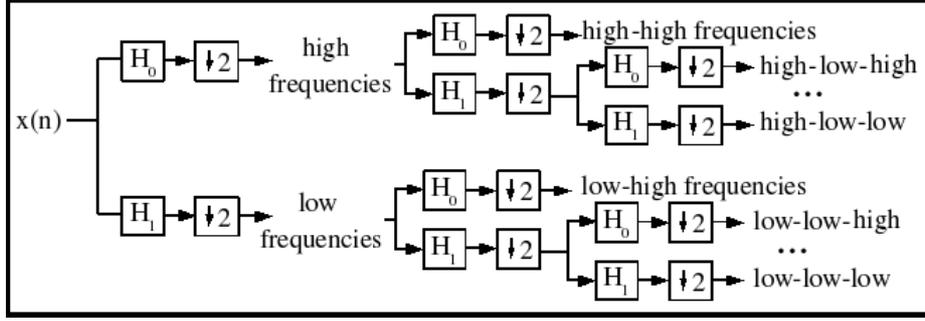


Figure 1. Tree-structured wavelet decomposition.

therefore yield a total of 8×341 features.

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

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

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

$$H(n_i) = \frac{\sum_{j=1}^J \sum_{k=1}^K (x_{jk}^2 \times \log(x_{jk}^2))}{J \times K} \quad (4)$$

$$AvgE(n_i) = \frac{\sum_{j=1}^J \sum_{k=1}^K |x_{jk}|}{J \times K} \quad (5)$$

$$S(n_i) = \frac{\sum_{j=1}^J \sum_{k=1}^K \left(\frac{x_{jk} - M(n_i)}{Std(n_i)}\right)^3}{J \times K} \quad (6)$$

$$K(n_i) = \frac{\sum_{j=1}^J \sum_{k=1}^K \left(\frac{x_{jk} - M(n_i)}{Std(n_i)}\right)^4}{J \times K} \quad (7)$$

$$N(n_i) = \max(\sqrt{\text{eig}(X \times X')}) \quad (8)$$

where n_i refers to node i , X is the matrix of the node (or sub-image) with dimension of $J \times K$, x_{jk} is the (j, k) -th element of the matrix, and X' is its transpose. $\text{eig}(X)$ is the eigenvalue of the matrix X .

The ratios used in this study are maximum ratio (R_m), fractional ratio (R_f) and sub-level ratio (R_s). As each decomposition sub-tree has four nodes (as illustrated in Figure 2), the maximum ratio divides each node by the maximum node among four. The fractional ratio divides each node by the summation of other three nodes. The sub-level ratio divides each node of the tree by its parent node. Equations 9–11 show the formula of these ratios for the sample node $n_{2.3}$. These ratios are calculated for each of the above mentioned 8 measure, which will produce a total of 8×765

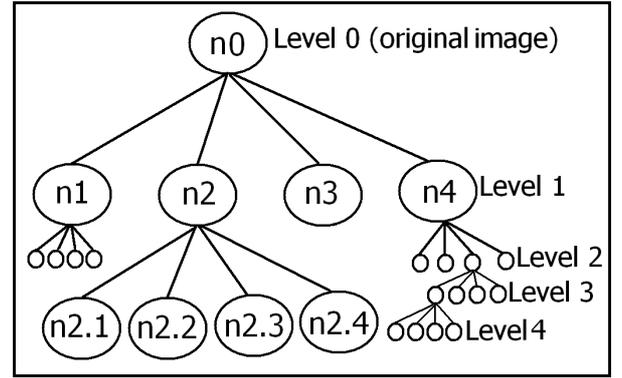


Figure 2. Schematic 4-level wavelet tree with nodes in circle.

(255 maximum ratio, 255 fractional ratio, 85 sub-level ratio) features.

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

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

$$R_s(n_{2.3}) = \frac{m(n_{2.3})}{m(n_2)} \quad (11)$$

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.

We have also calculated the feature $\ln(std+1)$ suggested by [15] and [20]. All of these features were calculated for four color channels of red, green, blue and luminance (Equation 12) and by including the area of the lesion, a set of 35397 features is obtained.

$$\text{luminance} = (0.29 \times R) + (0.59 \times G) + (0.11 \times B) \quad (12)$$

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

4. Proposed feature selection method

Feature selection is the process of finding an optimal 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 evidences in the training stage. Feature selection can also reduce the number of features to be considered in the problem space, thus reduces computational cost.

Numerous feature selection techniques have been proposed in the literature [9]. In this study, we first apply an entropy-based feature valuator called gain ratio (Equation 13) which evaluates the worth of an attribute (feature) by measuring its gain ratio with respect to the class. A ranker search method has been used along with gain ratio to rank features according to their evaluated gain ratio.

$$GainR(C, A) = \frac{H(C) - H(C|A)}{H(A)} \quad (13)$$

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

In order to optimally reduce the number of calculated features (35397 in total), we have also used correlation-based feature evaluator [11] which evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy between them. This feature evaluator selects the subset of features that are highly correlated with the class while having low inter-correlation. The Best-first method [19] has been applied to search for the features. The Best-first method searches the space of feature subsets using greedy hill-climbing algorithm, which can either start with the empty set of features (forward direction), or with the full set of features (backward direction), or at any point and search in both directions.

5. Proposed classification method

In this study we have used the following classifiers: support vector machine (SVM), random forest (RF), logistic model tree (LMT) and hidden naive bayes (HNB), briefly introduced in the following.

5.1. Support Vector Machine

Support Vector Machine performs classification by constructing an N-dimensional hyper-plane that optimally separates the given data into classes. In this study we apply sequential minimal optimization (SMO) algorithm for training a support vector classifier [18]. The kernel used in this study is a Gaussian radial basis function (RBF).

5.2. Random Forest

Random forest (forest of random trees) is an ensemble classifier which consists of many decision trees. The output of RF is the class label that is the mode of the class's output by individual trees. Each individual tree classifies the given feature vector - each tree votes for a class - and the forest outputs the classes with the highest vote [6].

5.3. Logistic Model Tree

Logistic model trees combine logistic regression and tree induction to perform the supervised learning. In another word, LMTs are classification trees with logistic regression functions at the leaves. Landwehr et al. [13] proposed this approach for building logistic models which uses the CART algorithm for pruning. It has been shown that [13] it is more accurate than C4.5 decision trees and standalone logistic regression on real-world datasets, and even competitive with boosted C4.5 trees.

5.4. Hidden Naive Bayes

Hidden Naive Bayes is a promising attempt to improve the performance of Naive Bayes method. In an HNB, attribute dependencies are represented by creating a hidden parent for each feature (attribute), which combines the influences of all other features [21].

6. Experimental results

An image set of 205 digital dermoscopy images obtained from the Interactive Atlas of Dermoscopy [3], is applied in this study. These were JPEG images with a resolution of 768×512 pixels. The diagnosis distribution include 96 melanoma and 109 benign (65 Clark nevi, 14 blue nevi, 30 Reed/Spitz nevi), where the histopathological results have been determined by biopsy or excision.

6.1. Preprocessing

In this study, the lesion borders were obtained manually, according to practice used by dermatologists. The motivation for using manual borders rather than automatic computer-detected borders was to avoid propagating the border detection error into the feature extraction and classification phase and reducing any possible bias.

Moreover, a rectangle frame enclosing the lesion with horizontal and vertical sides is set around the lesion and the rest is cropped. Also, within this frame, the pixels that do not belong to the lesion are set to black. This process is done to minimize the impact of background skin in the analysis and maximize the focus on lesion pixels.

| Conf. | Threshold | Feat no | SVM (RBF kernel) | RF | LMT | HNB |
|-------|-----------|---------|-----------------------------|----------------------------|----------------------------|----------------------|
| 1 | 0.30 | 451 | (0.01) 86.27 , 0.862 | (100) 86.27 , 0.917 | (0.4) 88.24 , 0.918 | 86.27 , 0.896 |
| 2 | 0.31 | 299 | (0.01) 86.27 , 0.862 | (120) 86.27 , 0.922 | (0.1) 82.35, 0.886 | 86.27 , 0.892 |
| 3 | 0.32 | 217 | (0.01) 86.27 , 0.862 | (130) 85.29, 0.923 | (0.2) 86.27 , 0.908 | 84.31, 0.91 |
| 4 | 0.33 | 151 | (0.01) 86.27 , 0.862 | (160) 85.29, 0.917 | (0.2) 83.33, 0.902 | 85.91, 0.907 |
| 5 | 0.34 | 93 | (0.01) 85.29, 0.853 | (150) 86.27 , 0.921 | (0.2) 87.25 , 0.917 | 86.27 , 0.921 |
| 6 | 0.35 | 48 | (0.01) 85.29, 0.852 | (180) 85.29, 0.901 | (0.3) 86.27 , 0.871 | 85.29, 0.885 |
| 7 | 0.36 | 25 | (0.01) 84.31, 0.843 | (100) 85.29, 0.897 | (0.3) 83.33, 0.865 | 86.27 , 0.878 |
| 8 | 0.37 | 13 | (0.01) 86.27 , 0.862 | (120) 87.25 , 0.893 | (0.4) 84.31, 0.891 | 82.35, 0.87 |

Table 1. 10-fold cross validation of performance (accuracy and ROC area) of different classifiers on the test set, using the 8 feature configurations resulting from the first stage of feature selection on the train set.

| Conf. | Threshold | Dir | Feat no | SVM (RBF kernel) | RF | LMT | HNB |
|-------|-----------|-----|---------|-----------------------------|----------------------------|----------------------------|----------------------|
| 9 | 0.30 | F | 28 | (0.02) 85.29, 0.852 | (220) 81.37, 0.894 | (0.2) 81.37, 0.88 | 85.29, 0.899 |
| 10 | 0.30 | B | 27 | (0.01) 85.29, 0.852 | (15) 82.35, 0.897 | (0.1) 81.37, 0.877 | 83.33, 0.888 |
| 11 | 0.31 | F | 33 | (0.01) 85.29, 0.852 | (150) 82.35, 0.893 | (0.25) 80.39, 0.843 | 85.29, 0.893 |
| 12 | 0.31 | B | 25 | (0.01) 84.31, 0.843 | (250) 80.39, 0.875 | (0.15) 82.35, 0.89 | 84.31, 0.883 |
| 13 | 0.32 | F | 29 | (0.04) 84.31, 0.844 | (170) 82.35, 0.896 | (0.05) 82.35, 0.877 | 86.27 , 0.896 |
| 14 | 0.32 | B | 29 | (0.03) 85.29, 0.854 | (250) 83.33, 0.902 | (0.05) 80.39, 0.878 | 84.31, 0.898 |
| 15 | 0.33 | F | 29 | (0.01) 84.31, 0.843 | (300) 85.29, 0.906 | (0.05) 84.31, 0.892 | 84.31, 0.9 |
| 16 | 0.33 | B | 29 | (0.01) 85.29, 0.853 | (150) 85.29, 0.898 | (0.1) 81.37, 0.88 | 84.31, 0.901 |
| 17 | 0.34 | F | 17 | (0.03) 82.35, 0.822 | (150) 84.31, 0.9 | (0.2) 82.35, 0.866 | 83.33, 0.907 |
| 18 | 0.34 | B | 23 | (0.03) 85.29, 0.853 | (190) 85.29, 0.914 | (0.25) 82.35, 0.868 | 83.33, 0.907 |
| 19 | 0.35 | F | 21 | (0.01) 84.31, 0.844 | (110) 86.27 , 0.895 | (0.1) 85.29, 0.907 | 82.35, 0.881 |
| 20 | 0.35 | B | 20 | (0.01) 85.29, 0.854 | (100) 85.29, 0.891 | (0.2) 86.27 , 0.906 | 81.37, 0.876 |
| 21 | 0.36 | F | 8 | (0.03) 86.27 , 0.863 | (50) 86.27 , 0.903 | (0.05) 83.33, 0.885 | 82.35, 0.874 |
| 22 | 0.36 | B | 8 | (0.03) 86.27 , 0.863 | (150) 86.27 , 0.894 | (0.15) 81.37, 0.87 | 82.35, 0.875 |
| 23 | 0.37 | F | 6 | (0.01) 80.39, 0.803 | (220) 83.33, 0.893 | (0.1) 85.29, 0.887 | 82.35, 0.87 |
| 24 | 0.37 | B | 6 | (0.01) 80.39, 0.803 | (150) 83.33, 0.902 | (0.1) 85.29, 0.881 | 82.35, 0.87 |

Table 2. 10-fold cross validation of performance (accuracy and ROC area) of different classifiers on the test set, using the 16 feature configurations resulting from the second stage of feature selection on train set.

6.2. Feature extraction and feature selection

We have designated half of the images (48 melanoma and 55 benign) as train set to apply the feature selection, in order to determine the most significant features contributing in the classification of melanoma. The feature extraction method explained in Section 4 is applied on the train set. MATLAB Daubechies-3 wavelet function is used for this application.

For selecting the optimal features we have used a two-stage approach; at the first stage, gain ratio feature selection method is applied on the feature vector. This resulted in 8 different feature subsets, depending on the threshold set for the ranker search method (shown in Table 1). At the second stage, the correlation-based feature evaluation along with best-first search method (in forward and backward directions) has been applied on the subsets resulted from the first stage of feature selection, resulting in 16 feature configurations, as shown in Table 2.

6.3. Classification

To test the effectiveness of feature subsets selected in the previous step (Section 6.2), we have applied the classification schemes described in Section 5 on a test set consisting of 48 melanoma and 54 benign dermoscopy images, using Weka data mining toolkit [12].

Table 1 shows 10-fold cross validation of performance (accuracy and ROC area) of different classifiers on the test set, using the 8 feature configurations resulting from the first stage of feature selection on the train set. The ‘‘Threshold’’ column shows the ranking threshold for including the features in the subset. The third column shows the number of features in the determined subset. The figures in the bracket in ‘‘SVM’’, ‘‘RF’’ and ‘‘LMT’’ columns are the corresponding optimal parameters of each classifier. Table 2 shows 10-fold cross validation of performance (accuracy and ROC area) of the for classifiers on the test set, using the 16 feature configurations resulting from the second stage of feature selection

| Measure | Color Channel | Tree Node |
|--------------------|---------------|-----------|
| Kurtosis | Red | 4.3.4.4 |
| Kurtosis | Red | 2.4.1 |
| Kurtosis | Luminance | 4.3.4.4 |
| Kurtosis | Green | 4.3.4.4 |
| Kurtosis | Blue | 4.1.4 |
| Average-energy | Luminance | 3.2.1.3 |
| Kurtosis | Green | 2.2.4 |
| Average-energy | Luminance | 4.1.4 |
| Average-energy | Luminance | 4.1.4.4 |
| Average-energy | Red | 4.1.4 |
| Average-energy | Blue | 4.1.4.4 |
| Standard deviation | Luminance | 4.2.4.1 |
| Energy | Luminance | 4.2.4.1 |

Table 3. The 13 optimal features selected by the 8-th feature configuration, resulted in the accuracy of 87.25%.

on train set. The “Dir” column refers to the direction of the best-first method, and F and B show the forward and backward directions, respectively.

According to Tables 1 and 2, the highest accuracy (88.24%) and ROC area (0.918) are obtained by first feature configuration, i.e. applying the gain ratio at the threshold of 0.3 which results in 451 features, when classified by LMT. This is followed by 5-th and 8-th configurations, where gain ratio was used with thresholds of 0.34 (with LMT classifier) and 0.37 (with RF classifier) resulting in 93 and 13 optimal features. Both feature sets obtain an accuracy of 87.25%. However, the ROC area for 93-element subset is 0.917, and 0.893 for 13-element feature set. Multiple feature configurations; e.g. 2, 3, 4, 6, 7, 19, 20, 21 and 22 exhibit an accuracy of 86.27%, among which 21 and 22 have the smallest number of features, i.e. eight optimal features. The decision of which configuration is best to use is a trade-off between having a more accurate classification decision and the expense of increased complexity and computational cost.

Table 3 shows the 13 optimal features selected by the 8-th feature configuration, which resulted in the accuracy of 87.25%. As indicated in Table 3, certain measures and nodes appear more frequently in the optimal feature subset suggested by this configuration, such as the kurtosis and average-energy measures, and 4.3.4.4 and 4.1.4 nodes of the wavelet tree. Moreover, 60% of the optimal features are happen to be in the forth level of the wavelet decomposition. In this feature configuration, the luminance color channel is the dominant one, following by red. Blue and green both have two occurrences.

6.4. Comparison with others methods

The proposed feature extraction method has been compared with three other wavelet-based melanoma classifica-

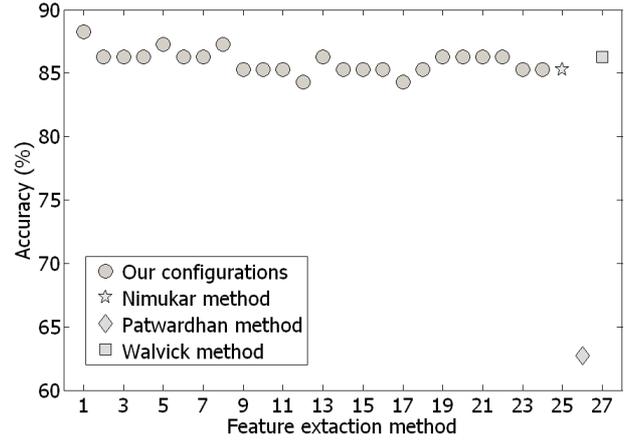


Figure 3. Accuracy of various feature configurations proposed by our method as opposed to three other feature extraction methods by [15], [16] and [20].

tion methods proposed by Nimukar et al. [15], Patwardhan et al. [16] and Walvick et al. [20]. In order to make this comparison we have applied the suggested feature sets by [15], [16] and [20] on our test set of 102 images (48 melanoma and 54 benign dermoscopy images). Table 4 shows the classification accuracy and obtained ROC area of our best three feature configurations along with the result obtained by the above-mentioned methods. The parameters of the classifiers have been tuned in a way that the highest performance for the given feature set is gained.

According to Table 4, our proposed method with different configuration outperforms the other three. However, the best performance among the three methods is obtained by Walvick et al. [20] with the accuracy of 86.27% and ROC area of 0.901 when RF classifier is used. It is followed by Nimukar et al. [15] with the highest accuracy of 85.29% and Patwardhan [16]’s methods come last. Figure 3 shows the obtained accuracies of various feature configurations proposed by our method as opposed to three other feature extraction methods by [15], [16] and [20]. These accuracies are the highest of all four classifiers for each feature selection method.

7. Conclusion

This paper presents a wavelet-based texture analysis method for differentiating malignant melanoma from benign nevi. We have applied tree-structured wavelet transform on different color channels of red, green, blue and luminance of skin image. Various statistical measures and ratios have been employed on wavelet coefficients. The proposed feature extraction method along with a two-stage feature selection method, based on entropy and correlation, has been applied on a train set of 103 images. Then, using the

| Method | Feat no | SVM (RBF kernel) | RF | LMT | HNB |
|-----------------|---------|-----------------------------|----------------------------|----------------------------|----------------------|
| configuration 1 | 451 | (0.01) 86.27 , 0.862 | (100) 86.27 , 0.917 | (0.4) 88.24 , 0.918 | 86.27 , 0.896 |
| configuration 5 | 93 | (0.01) 85.29, 0.853 | (150) 86.27 , 0.921 | (0.2) 87.25 , 0.917 | 86.27 , 0.921 |
| configuration 8 | 13 | (0.01) 86.27 , 0.862 | (120) 87.25 , 0.893 | (0.4) 84.31, 0.891 | 82.35, 0.87 |
| Nimukar [15] | 5 | (0.01) 83.33, 0.834 | (100) 85.29, 0.91 | (0.3) 85.29, 0.879 | 82.35, 0.81 |
| Patwardhan [16] | 5 | (0.01) 52.94, 0.5 | (90) 62.74, 0.623 | (0.1) 62.74, 0.692 | 52.94, 0.507 |
| Walvick [20] | 4 | (0.01) 82.35, 0.822 | (70) 86.27 , 0.901 | (0.15) 82.35, 0.904 | 79.41, 0.759 |

Table 4. Comparative study: 10-fold cross validation of performance (accuracy and ROC area) of different classifiers on test set, using features suggested by [15], [16] and [20], and our best three feature configurations.

resultant optimal feature subsets, four different classifiers of support vector machine, random forest, logistic model tree and hidden naive bayes, have been applied on a test set of 102 dermoscopy images. The best performance obtained by applying gain ratio at the threshold of 0.3 and LMT classifier, leading to accuracy of 88.24% and ROC area of 0.918. A comparative study has been also performed, which shows that the proposed feature extraction method outperforms three other wavelet-based approaches.

8. Acknowledgment

The authors would like to thank Dr. Emre Celebi from Department of Computer Science, Louisiana State University in Shreveport, USA, for assisting with the data. 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] G. Argenziano, H. Soyer, S. Chimenti, R. Talamini, and R. Corona. Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet. *Journal of the American Academy of Dermatology*, 48:679–693, 2003.
- [3] G. Argenziano, H. Soyer, and V. D. Giorgi. *Interactive Atlas of Dermoscopy*. EDRA Medical Publishing and New Media, Milan, Italy, 2002.
- [4] J. Boldrick, C. Layton, and J. Nguyen. Evaluation of digital dermoscopy in a pigmented lesion clinic: Clinician versus computer assessment of malignancy risk. *Journal of the American Academy of Dermatology*, 56(3):417–421, 2007.
- [5] R. Braun, H. Rabinovitz, M. Oliviero, H. Rabinovitz, and W. Stolz. Dermoscopy of pigmented lesions. *Journal of the American Academy of Dermatology*, 52(1):109–121, 2005.
- [6] L. Breiman. Random forests. *Machine learning*, 45:5–32, 2001.
- [7] M. Celebi, H. Kingravi, and B. Uddin. A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, 31(5):362–373, 2007.
- [8] T. Chang and C. J. Kuo. Texture analysis and classification with tree structured wavelet transform. *IEEE Trans. on Information Technology in Biomedicine*, 2:429–441, 1993.
- [9] M. Dash and H. Liu. Feature selection for classification. *Intelligent Data Analysis*, 1:679–693, 1997.
- [10] M. Elbaum, A. Kopf, H. Rabinovitz, R. Langley, and H. Kamino. Automatic differentiation of melanoma from melanocytic nevi with multispectral digital dermoscopy: A feasibility study. *Journal of the American Academy of Dermatology*, 44:207–218, 2001.
- [11] M. Hall. *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, University of Waikato, Hamilton, New Zealand, 1998.
- [12] M. Hall, E. Frank, and G. Holmes. The WEKA data mining software: An update. *SIGKDD Explorations*, 11:10–18, 2009.
- [13] N. Landwehr, M. Hall, and E. Frank. Logistic model trees. *Machine Learning*, 95:161–205, 2005.
- [14] I. Maglogiannis and C. Doukas. Overview of advanced computer vision systems for skin lesions characterization. *IEEE Trans. on Information Technology in Biomedicine*, 13(5):721–733, 2009.
- [15] A. Nimunkar, A. Dhawan, P. Relue, and S. Patwardhan. Wavelet and statistical analysis for melanoma. In *SPIE Medical Imaging 2002: Image Processing*, volume 4684, pages 1346–1352, 2002.
- [16] S. Patwardhan, A. Dhawan, and P. Relue. Classification of melanoma using tree structured wavelet transforms. *Computer Methods and Programs in Biomedicine*, 72:223–239, 2003.
- [17] A. Perrinaud, O. Gaide, L. French, J. Saurat, A. Marghoob, and R. Braun. Can automated dermoscopy image analysis instruments provide added benefit for the dermatologist? A study comparing the results of three systems. *British Journal of Dermatology*, 157:926–933, 2007.
- [18] J. Platt. Fast training of support vector machines using sequential minimal optimization. pages 185–208, 1999.
- [19] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1 edition, 2003.
- [20] R. Walvick, K. Patel, S. Patwardhan, and A. Dhawan. Classification of melanoma using wavelet-transform-based optimal feature set. In *SPIE Medical Imaging 2002: Image Processing*, volume 5370, pages 944–951, 2004.
- [21] H. Zhang, L. Jiang, and J. Su. Hidden naive bayes. In *Twentieth National Conference on Artificial Intelligence*, pages 919–924, 2005.