

# Effective Features to Classify Skin Lesions in Dermoscopic images

Zhen Ma<sup>a</sup>, João Manuel R. S. Tavares<sup>b</sup>

*<sup>a</sup> Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, s/n 4200-465 Porto, Portugal; email: [zhen.ma@fe.up.pt](mailto:zhen.ma@fe.up.pt)*

*<sup>b</sup> Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal; email: [tavaresfe.up.pt](mailto:tavaresfe.up.pt)*

## **Corresponding Author:**

Prof. João Manuel R. S. Tavares

Faculdade de Engenharia da Universidade do Porto (FEUP)

Departamento de Engenharia Mecânica (DEMec)

Rua Dr. Roberto Frias, s/n, 4200-465 PORTO - PORTUGAL

Tel: +315 22 5081487, Fax: +315 22 5081445

Email: [tavares@fe.up.pt](mailto:tavares@fe.up.pt), Url: [www.fe.up.pt/~tavares](http://www.fe.up.pt/~tavares)

## **Effective Features to Classify Skin Lesions in Dermoscopic images**

### **Abstract**

Features such as shape and color are indispensable to determine whether a skin lesion is a melanoma or not. However, there are no fixed guidelines to define which features are effective and how to combine them for classification. This lack of definition impedes the development of the automatic analyses of dermoscopic images. In this work, a search for effective features was carried out using a support vector machine. Three image databases were used to verify the feasibility and sensitivity of the automatic classification used. The results showed which features had a major influence on the classification performance, and confirmed the need to use various types of features in this process.

**Keywords** *Skin lesion; melanoma; ABCD rule; shape features; color features; feature selection.*

# 1. Introduction

Dermoscopy has been widely used for the diagnosis of skin lesions. The additional details provided by a dermoscope, compared to the inspection by the naked-eye, contribute to a significant improvement in the detection rate of melanoma (Binder et al., 1995; Argenziano et al. 2002; Kittler et al., 2002; Heymann, 2005). However, the experience of the dermatologist can significantly affect the diagnostic accuracy (Binder et al., 1995; Kittler et al., 2002); therefore, an effective automatic computer-aided model to analyze dermoscopic images is urgently needed. A scheme for such a model includes three major steps: first, the border of the skin lesion is identified on the image under study (Ma & Tavares, 2016; Silveira et al., 2009); then, features are extracted based on the segmentation, and with these features the lesion is classified into an appropriate category. Many mature algorithms have been proposed for the classification step, such as support vector machines (SVMs) and artificial neural networks (ANNs) (Kulkarni et al., 1998; Hastie et al., 2009); however, their performances rely on the features from the second step. Unfortunately, there is no recommended set of features for detecting melanomas in the current literature. Although many novel features have been proposed to improve the classification accuracy of these algorithms, the validation is normally carried out using different methods with different feature sets. The use of redundant or irrelevant features for classification not only decreases the computational efficiency but also affects the performance. Moreover, in many cases, cross validation is used to test a classification scheme, which can present over-optimistic results since images in the same database are prone to have similar imaging conditions and alike structures of skin lesions. Feature selection algorithms have been proposed to find the significant features, especially in the

area of bioinformatics (Guyon et al., 2002; Saeys, Inza, & Larranaga, 2007). Nevertheless, the key benefit of using such an algorithm is to reduce the dimensionality of features without loss of significant information (Bermingham et al., 2015). Limitations such as the choice of the evaluation metric and image database can affect the decision of whether a feature should be kept or not.

In this work, effective features are explored without focusing on factors such as the accuracy of segmentation, the choice of the classification algorithm, and the representativeness of the training set. Three dermoscopic image databases were used in this study and the feasibility of automatic analysis on dermoscopic images was investigated based on the inter-database validation. This article is structured as follows: in the next section, the features selected for classification are introduced; then, in Section 3, the experiments are described and the results are presented and discussed; in the last section, the conclusions and perspectives for future work are pointed out.

## 2. Features

The ABCD rule or ABCDE rule (Friedman, Rigel, & Kopf, 1985) is a clinical guideline to determine whether a skin lesion is a melanoma or a common nevus (Abbasi, Shaw, & Rigel, 2004; Friedman et al., 1985). This evaluation criteria is composed of the asymmetry of the lesion region ( $A$ ), the geometric properties of the lesion border ( $B$ ), the color of the skin lesion ( $C$ ), the diameter of the lesion region ( $D$ ), and the elevation/evolution of the skin lesion ( $E$ ). The  $E$  rule requires follow-up inspections, and the other four criteria can be grouped into two categories: measures that reflect the geometric properties of a skin lesion ( $A$ ,  $B$ , and  $D$  rules), and measures that are related with the lightness and chromatic information ( $C$  rule). The ABCD rule requires a

subjective evaluation of the different aspects of a skin lesion; hence, in order to automatize the process in an expert system, these rules need to be quantified.

## 2.1 Shape features

To measure the shape of a skin lesion, the following six features were adopted to reflect the *B* and *D* rules: perimeter of the lesion; area of the lesion; perimeter to area ratio; aspect ratio - defined as the width divided by height of the minimum rectangle that bounds the lesion region; extent ratio - defined as the area of the lesion region divided by the area of the minimum bounding rectangle; and solidity ratio - defined as the area of skin lesion divided by the area of the convex hull of the lesion boundary (Haidekker, 2010; Olsen, 2011). The first two measures describe the size of the skin lesion, while the remaining ones indicate its regularity. Besides these commonly used features, the total variation of the lesion border is adopted as the seventh feature to represent the smoothness:

$$V = \int_C |\kappa| ds, \quad (1)$$

where the lesion border  $C$  is parameterized as  $u(s)$ , and  $\kappa = \nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right)$  is the mean curvature with  $\nabla$  denoting the divergence operator.

For the symmetry property - *A* rule, an ideal measure should be irrelevant to the position, size and orientation of the contour. Invariant image moments are well suited to these requirements (Flusser & Suk, 2006), and Hu's invariant set has been shown to be effective in various applications of image analysis (Hu, 1962; Flusser & Suk, 2006). Therefore, the seven image moments in Hu's invariant set were selected as the eighth to the fourteenth shape features for classification.

## 2.2 Color features

Although the appearance of a skin lesion can vary considerably, information concerning its color is always the main visual feature distinguishing it from the neighboring skin. The changes of pigment inside a skin lesion are important for both segmentation and classification steps. An ideal color feature should capture the differences between a melanoma and a non-melanoma nevus. Additionally, the imaging conditions can appreciably affect the appearance of a skin lesion on dermoscopic images; therefore, the ideal color features should be able to handle the diverse lighting conditions in different image databases.

The color information of a dermoscopic image is stored as a triplet  $(r, g, b)$  for each image pixel. The three color channels in the *RGB* color space are correlated and consequently they are hard to be used to evaluate the differences between colors. To overcome this problem, the *CIE L\*a\*b\** and *CIE L\*u\*v\** color spaces were adopted (Gonzalez and Woods, 2007). The benefit of separating lightness information from chromaticity can considerably decrease the distortions caused by lightness when defining color features. Accordingly, the first ten color features were chosen as the means and standard deviations of the  $L^*$ ,  $a^*$ ,  $b^*$ ,  $u^*$  and  $v^*$  values of the skin lesion. In addition, given that  $(a^*, b^*)$  and  $(u^*, v^*)$  are the coordinates representing the position of a chromaticity in the color coordination system, the difference between a chromaticity and the mean chromaticity of the skin lesion is calculated as the Euclidean distances:

$$d_1 = \sqrt{(a^* - a_1^*)^2 + (b^* - b_1^*)^2}, \quad (2)$$

$$d_2 = \sqrt{(u^* - u_1^*)^2 + (v^* - v_1^*)^2}, \quad (3)$$

where  $a_1^*$ ,  $b_1^*$ ,  $u_1^*$  and  $v_1^*$  are the means of  $a^*$ ,  $b^*$ ,  $u^*$  and  $v^*$  of the skin lesion and correspondingly  $(a_1^*, b_1^*)$  and  $(u_1^*, v_1^*)$  are the chromatic geometric centroids of the skin lesion. Then, the next four color features were defined as the mean and standard deviation of  $d_1$  and  $d_2$  inside the skin lesion:  $\bar{d}_1$ ,  $d_{1-\sigma}$ ,  $\bar{d}_2$ , and  $d_{2-\sigma}$ .

Moreover, the same 14 features of the neighboring healthy skin (the means and standard deviations of  $L^*$ ,  $a^*$ ,  $b^*$ ,  $u^*$ ,  $v^*$  and the means and standard deviations of the two Euclidean distances to the geometric color centroids of the healthy skin) next to the skin lesion were added to the color features; these features not only provide the information of color transitions from normal skin to skin lesion, but also compensate the possible inaccuracy of the lesion border. Also, the lightness difference and the Euclidean distance between the geometric centroids of healthy skin and skin lesion were adopted as three additional color features:

$$d_3 = \sqrt{(a_0^* - a_1^*)^2 + (b_0^* - b_1^*)^2}, \quad (4)$$

$$d_4 = \sqrt{(u_0^* - u_1^*)^2 + (v_0^* - v_1^*)^2}, \quad (5)$$

$$d_5 = |L_0 - L_1|, \quad (6)$$

where  $L_0$  and  $L_1$  are the mean lightness of the neighboring skin and skin lesion, and  $a_0^*$ ,  $b_0^*$ ,  $u_0^*$ , and  $v_0^*$  are the means of  $a^*$ ,  $b^*$ ,  $u^*$ , and  $v^*$  of the neighboring skin. Additionally, color saturation was used to correlate the lightness with chromaticity:

$$S = \begin{cases} 0 & \text{if } R + G + B = 0 \\ 1 - \frac{\min(R,G,B)}{(R+G+B)/3} & \text{otherwise} \end{cases}, \quad (7)$$

and the 32nd to 35th color features were defined as the mean and standard deviation of the saturation of the skin lesion ( $s_1$ ,  $s_{1-\sigma}$ ) and the neighboring skin ( $s_0$ ,  $s_{0-\sigma}$ ), plus the 36<sup>th</sup> feature as the ratio of the means of the two regions:

$$d_6 = s_1/s_0. \quad (8)$$

Table 1 lists all the shape and color features used for selection.

## 3. Experiments

### 3.1 Training and testing

Although an unbiased classification requires the skin lesions in the training set to be representative, different types of skin lesions can have large variations in shape and appearance, so it is difficult to define how typical a skin lesion is. Also, due to the size of the available image databases, a routine validation of a classifier or new feature is through the leave-one-out strategy. Consequently, the evaluation of the performance may be biased and the high sensitivity and specificity achieved with one image database may not hold true for another.

In order to make the evaluation objective, three databases of dermoscopic images were used. Details of these databases are given in Table 2, with samples of melanoma and non-melanoma skin lesions shown in Figure 1. As can be seen from the Figs 1a to f, the images were acquired under different conditions and from patients of diverse origins. In these databases, the skin lesions were actually classified into more accurate sub-categories; but for simplicity, this additional information was ignored and each skin lesion was classified as either melanoma or non-melanoma. The ground truth of lesion borders were provided in all the databases, which were manually segmented by qualified technicians. In order to obtain the statistics of neighboring healthy skin referred to in Section 2.2, a 10-pixel-wide band next to the lesion border was used; this outside band covers a moderate region that is sufficient to reflect the transition of healthy skin to skin lesion in the images.

Given the sizes of the databases, we adopted the strategy to choose one database as the testing set and combined the other two as the training set. However, if the training set was



formed by the 1<sup>st</sup> and 2<sup>nd</sup> image databases shown in Table 2, the total number of melanomas (69) would not be large enough to be used for extracting the differential information between melanoma and non-melanoma nevi. Consequently, the classification would not be able to achieve a good performance on the 3<sup>rd</sup> database, particularly with the fact that the 3<sup>rd</sup> database contains Spitz and Reed nevi (Lyon, 2010; Yoradjian et al., 2012). Thus, the other two cases were adopted: the combination (Comb 1) that used the 1<sup>st</sup> and 3<sup>rd</sup> databases as the training set and the 2<sup>nd</sup> database as the testing set; and the combination (Comb 2) that used the 2<sup>nd</sup> and 3<sup>rd</sup> databases as the training set and the 1<sup>st</sup> database as the testing set.

The next problem to solve was related to the different magnitudes of the 50 features chosen for classification; for example, the value of lightness ranges from 0 to 100, while saturation only varies from 0 to 1. A feature scaling step was necessary to balance such differences. Among the features to be selected, 31 were related to the L, a\*, b\*, u\*, and v\* channels, and since their region-based values were already at similar intervals. These features were kept unchanged and then the remaining 19 features were scaled as follows:

$$x = \begin{cases} 50 * \frac{x-x_{\min}}{x_{\max}-x_{\min}}, & \text{if } x_{\min} \neq x_{\max} \\ 0, & \text{if } x_{\min} = x_{\max} \end{cases}, \quad (9)$$

where  $x_{\min}$  and  $x_{\max}$  are the minimum and maximum values of feature  $x$  in the training set.

The support vector machine was chosen for this two-class classification, due to its effectiveness already shown in diverse areas of studies (Chu & Wang, 2005; Filho et al., 2015), and the radial basis function kernel was adopted for its flexibility, stability and general popularity (Celebi et al., 2007). Accordingly, three parameters had to be defined:  $C$ ,  $\varepsilon$ , and  $\gamma$ . Parameter  $C$  controls the trade-off between the error of a classifier on the

training set and the margin between classes; parameter  $\varepsilon$  determines the accuracy of the approximation; and parameter  $\gamma$  is related to the values of feature vectors and determines the behaviors of kernel functions for approximation. Values of these parameters can appreciably affect the performance of classification; therefore, to assure a stable evaluation of a feature combination, the values of  $\gamma$ ,  $C$  and  $\varepsilon$  of the SVM were fixed as  $1e - 3$ , 500 and 1 in the experiments.

### 3.2 Pre-selection criteria

In order to find the best combination among the 14 shape features and 36 color features with a training set and a testing set, an exhaustive search requiring  $2^{50}$  classifications would have to be made. This however could lead to an unfeasibly long computation time. To overcome this problem, the standard deviation of a feature would not be used if the mean of that feature was not used; for example, if the mean of  $a^*$  of the skin lesion ( $a_1^*$ ) was not used for classification then the standard deviation of  $a^*$  of the skin lesion ( $a_{1-\sigma}^*$ ) would not be used either. This strategy was adopted because there are 15 measures with both their means and standard deviations chosen as the features for selection. While the standard deviation indicates the variations of a measure, it can also be reflected by other features; for example, the variation of  $a^*$  channel inside the skin lesion is partly reflected by the mean of distance  $d_1$ . This strategy decreases the total number of classification to  $2^{50} \cdot \left(\frac{3}{4}\right)^{15}$ . In addition, since chromaticity is described by two-dimensional coordinates in the color spaces, the following strategies were implanted: features of  $a^*$  channel are bound with the features of  $b^*$  channel; for example, if  $a_1^*$  was used,  $b_1^*$  would be used; if  $a_{1-\sigma}^*$  was used, then  $b_{1-\sigma}^*$  would be. Likewise, features of  $u^*$  and  $v^*$  were bound for use; and features related to  $d_1$  and  $d_2$  were bound; and the same for  $d_3$  and  $d_4$ .

Furthermore, the following procedure was used to implement separate searches among the shape features and color features: In the first phase, all the color (shape) features were used for classification, and the combinations of shape (color) features were searched to find the optimal set; then, with the optimal set of shape (color) features, the combinations of color (shape) features were reversely searched to find the best match. With the color (shape) features selected from the second phase, we can again search the combinations of shape (color) features to find the best match. Hence, the two-phase procedure can be iterated, and the best performance of classification would be improved after each iteration. Thus, if the search is to find the best color features for a set of shape features, the number of classifications would be  $2^{23} \cdot \left(\frac{3}{4}\right)^{10}$ ; and if the searching is to find the best shape features for a set of color features, the number would be  $2^{14}$ .

In the experiments, the performance of classification was evaluated based on three measures: overall accuracy (OA), which is the percentage of the skin lesions that are correctly classified; sensitivity of a classification (S1) that is defined as the percentage of correctly classified melanomas among all the melanomas; and specificity (S2), which is defined as the percentage of correctly classified non-melanoma nevi among all the non-melanoma nevi. Higher values of these indices indicate a better classification.

### **3.3 First phase**

Following the procedure defined above, all the shape features were used for classification then the search was among the combinations of color features. For Comb 1, the highest overall accuracy was 83.5% and was achieved by two sets of color features. Both of them contained the mean lightness of skin lesion ( $L_1$ ), lightness difference ( $d_5$ ), standard deviation of lightness inside the skin lesion ( $L_1-\sigma$ ), mean saturation of the skin lesion ( $s_1$ ),

mean of  $d_1$  and mean of  $d_2$  inside the skin lesion ( $\bar{d}_1$  and  $\bar{d}_2$ ); while one set included two extra features: mean lightness of neighboring skin ( $L_0$ ) and ratio of mean saturations ( $d_6$ ). With all the shape features and the six mutual color features, the sensitivity was 87.5% and the specificity was 82.5%; while with the additional two color features, these two indices were 77.5% and 85.0%, respectively. The highest sensitivity achieved in this phase was 97.5%, but the corresponding specificity was only 40.0%, indicating that many non-melanoma nevi were wrongly classified as melanomas. If both sensitivity and specificity were considered, a ranking on the average of these two values gave the best combination that achieved 92.5% of sensitivity and 80.0% of specificity; the set of color features has 82.5% of overall accuracy and was composed of six elements:  $L_1$ ,  $s_1$ ,  $\bar{d}_1$ ,  $\bar{d}_2$ ,  $d_5$  and  $d_6$ . However, when we performed the classification for Comb 2 with all the shape features and the aforementioned three sets of color features, the overall accuracy of classification was only 55.0%, 54.0% and 61.0%, respectively. For Comb 2, the best overall accuracy of classification was 72.0% with 62.1% sensitivity and 76.1% specificity, achieved by a set of three color features: mean of  $a^*$  and mean of  $b^*$  inside the skin lesion ( $a_1^*$  and  $b_1^*$ ), and ratio of mean saturations ( $d_6$ ). Similarly, if this feature set is used for Comb 1, the overall accuracy was only 52.5% with 45.0% of sensitivity and 54.4% of specificity.

On the other hand, if all the color features were used for classification, the best results after searching among the combinations of shape features gave 84.5% of overall accuracy for Comb 1 with 60.0% of sensitivity and 90.6% of specificity, and 70% of overall accuracy for Comb 2 with 34.5% of sensitivity and 84.5% of specificity. Like the findings above, the set of shape features that achieved good results for Comb 1 did not perform well for Comb 2, and vice versa.

The results show that the sets of features that achieved the highest overall accuracy were not the ones with the highest sensitivity or specificity; consequently, a measure is needed to decide which set of features should be used in the second phase of the search. Given the composition of the training sets of Comb 1 and Comb 2, the total number of wrong classifications ( $WN$ ) in the two tests was chosen as the index to rank the features. Accordingly, seven sets of color features (found with all the shape features) and nine sets of shape features (found with all the color features) were chosen for the second phase of the search. Table 3 lists these 16 feature sets and the indices of the corresponding classification; the performance of these feature sets was moderate, but more balanced.

The minimum of  $WN$  (Wrong Number) equals 70 when all the shape features were used, the set of color features included: mean lightness of skin lesion ( $L_1$ ), lightness difference ( $d_5$ ), mean saturation of neighboring skin and skin lesion ( $s_0$  and  $s_1$ ), mean of  $d_1$  and mean of  $d_2$  inside the skin lesion ( $\bar{d}_1$  and  $\bar{d}_2$ ). The minimum of  $WN$  found with all the color features was 72, achieved by five sets of shape features. All of them contained the following three features: aspect ratio, smoothness of lesion border, and area of skin lesion; the difference between them was the choice of the Hu's seven invariant image moments. The frequency of occurrence (FO) of each feature appearing in the feature sets with  $WN \leq 80$  and  $WN \leq 90$  were calculated and are illustrated in Figure 2. For the color features, mean lightness of skin lesion ( $L_1$ ), mean lightness of neighboring skin ( $L_0$ ), lightness difference ( $d_5$ ), standard deviation of lightness inside the skin lesion ( $L_{1-\sigma}$ ), mean saturation of the neighboring skin and skin lesion ( $s_0$  and  $s_1$ ), mean of  $d_1$  and mean of  $d_2$  inside the skin lesion ( $\bar{d}_1$  and  $\bar{d}_2$ ), and means of  $a^*$ ,  $b^*$ ,  $u^*$  and  $v^*$  inside the skin lesion ( $a_1^*$ ,  $b_1^*$ ,  $u_1^*$ ,  $v_1^*$ ) were features that had high frequencies. For the shape features, aspect ratio, smoothness of lesion border, and area of skin lesion had the highest

frequencies. Additionally, the difference found among the frequencies of Hu's seven image moments was small, which implies that the effectiveness of these moments may vary when paired with other features, but generally they are of equal importance.

### **3.4 Second phase**

This phase included 16,384 classifications for each of the seven sets of color features indicated in Table 3, and 472,392 classifications for each of the nine sets of shape features. With the color features fixed, the optimal set of shape features found in this phase improved the performance for both Comb 1 and Comb 2. For example, for the first set of color features in Table 3, the optimal set of shape features achieved 78.0% of overall accuracy for Comb 2; and for the second set of color features referred to in Table 3, the highest overall accuracy for Comb 1 increased to 90.5% with 92.5% of sensitivity and 90.0% of specificity. Table 4 lists the best matches of shape and color features found in this phase.

Like in the first phase, the effective shape features were different for each of the seven sets of color features. Figure 3 shows the FO of each shape feature in the feature sets whose  $WN \leq 50, 60, 70, 80, 90$ , and Figure 4 illustrates the FO of each color feature based on the results for the nine sets of shape features. Aspect ratio, smoothness of lesion border, and area of the lesion region were the three shape features that always had high frequencies, especially among the feature sets with overall accuracy above 80.0%. This agrees with the finding of the first phase and indicates that these three shape features are effective and have a large influence on the performance of the classification. Figure 3 shows that the ranking of frequencies is generally the same, and when the  $WN$  increases, the difference between frequencies becomes smaller since more and more combinations

of shape features can achieve the same lower performance. For the color features, mean saturation of neighboring skin and skin lesion ( $s_0$  and  $s_1$ ), mean of  $d_1$  and mean of  $d_2$  inside the skin lesion ( $\bar{d}_1$  and  $\bar{d}_2$ ), and lightness difference ( $d_5$ ) were the ones with the highest frequencies, especially when the overall accuracy was greater than 80.0%. Similarly, the difference of frequencies among color features decreases when the  $WN$  increases. These five color features were also included in the ones with high frequencies of the first phase, which confirms their effectiveness; the missing features from the first phase, such as  $L_1$ ,  $a_1^*$  and  $b_1^*$ , are the ones that directly describe the color information; this phenomenon was probably caused by the large variations in the appearances of skin lesions.

### 3.5 Further selection

Since iterating the two-phase procedure generates a no-worse result, a further search was carried out based on the results of the second phase. The sets of features used in this phase are listed in Table 4. After this search, we confirmed that most of the combinations in Table 4 were already the optimal matches, however, we found two extra combinations that were able to achieve equal performance, and are listed in Table 5; the first one is a set of color features that achieved the same  $WN$  as the one presented in Table 4, and the second one is a set of shape features with which the classification can achieve an even better performance. Nevertheless, another search with these two sets of features confirmed that the matches of shape features and color features in Table 5 were already the best and no further sets of features would be able to achieve an equal or better performance. Hence, it is safe to conclude that with the search procedures performed the sets of features in Tables 4 and 5 were the optimal ones for classification.

Then, the frequencies of the shape features in the feature sets whose  $WN \leq 50, 60, 70, 80, 90$  were calculated. In the calculation we excluded the feature sets that had one of the following conditions:  $S1 \leq 60.0\%$  for Comb 1, or  $S1 \leq 40.0\%$  for Comb 2, or  $S2 \leq 70.0\%$  for either Comb 1 or Comb 2. These exclusions guaranteed that only the ones that had a good performance in all the indices of classification were taken into account. Similarly, the frequencies of color features were calculated based on the results of the second phase and the extra search; the results are shown in Figs. 5 and 6. The images show that the shape and color features with the highest frequencies in this phase are in line with the findings of the first and the second phase searches; hence, the effectiveness of these features was once again demonstrated.

### **3.6 Discussion**

Although the training set of Comb 2 contains more samples of skin lesions, the classification achieved a better performance for Comb 1. One possible reason is due to the low imaging resolution of the 1<sup>st</sup> database, which considerably affects the calculation of color features. However, the 2<sup>nd</sup> database contains images with incomplete profiles of skin lesions, and so the lesion borders on these images are inevitably inaccurate and can lead to wrong values of shape features. The inaccuracy of shape features and color features both have negative impacts on the classification; a pertinent question is which type of feature has the greater influence on the classification. Following up this idea, classification was carried out using features of just one type. The results showed that with only the color features, the best overall accuracy was around 83% for Comb 2, with sensitivity ranging from 48.3% to 72.4% and specificity from 87.3% to 97.2%, respectively. On the other hand, with only the shape features, the best overall accuracy



was about 59% with 31.0% sensitivity (also the highest value found for this index) and 70.4% specificity, respectively. Table 6 indicates the best sets of shape features and color features found according to the  $WN$ . The Table shows that the minimum of  $WN$  with features of only one type was not much different to the minimum found using both types. However, the sensitivity and the specificity with both types were higher, and the performance of classification was more stable for different image databases. The results also indicated that the color features are more robust for classification, which confirms the fact that these features are region-based.

The empirical search with Comb 1 and Comb 2 identified the features that are effective to achieve a satisfactory classification. However, it is worth pointing out that only using these features does not guarantee the best performance. In fact, classification using the five color features and three shape features that have the highest frequencies only achieved a moderate result with 83.0% of overall accuracy for Comb 1 and 73.0% of overall accuracy for Comb 2; remembering that there are features of low frequencies appearing in the set of features with the best performance. Nevertheless, despite this randomness, these eight features are most likely to achieve a good classification performance, and so should be prioritized in selection.

## 4. Conclusions

The features used in the automatic analysis of dermoscopic images have a critical influence on the performance of classification. However, a set of features that achieves good results on one database may not have an equal performance on another database, and a “redundant” feature may become indispensable to correctly detect a specific type

of skin lesion. In this work, we aimed to find the effective features for detecting melanomas experimentally. The three image databases used in the study were from different origins and acquired under diverse imaging conditions, which provides an objective basis for evaluation. The results obtained confirmed the effective use of both shape and color features and suggested the need to combine them to acquire high classification accuracy and robustness.

A comprehensive study on features for classification has many practical constraints, because image databases from diverse studies are always different and the diagnostic accuracy of dermoscopy itself is not 100% even under the optimal exam conditions (Kittler et al., 2002). Nonetheless, the findings in our work showed that the performance of the automatic analysis of skin lesions in dermoscopic images is comparable to experienced dermatologists. Furthermore, the ABCD rule may not always be able to detect a small-sized melanoma or a melanoma with a regular shape and homogeneous color (Grin et al., 1990). In addition, there are some specific features that can be effective in classifying a particular type of skin lesion; for example, ridges and furrows were shown to be effective in detecting acral lentiginous melanoma (Iyatomi et al., 2008; Bradford et al., 2009; Yang et al., 2017). These subjects are challenges to be explored, and future work will continue to focus on solving these issues and finding more effective features.

## **Acknowledgements**

This work was funded by European Regional Development Funds (ERDF), through the Operational Program ‘Thematic Factors of Competitiveness (COMPETE), and Portuguese Funds, through “Fundação para a Ciência e a Tecnologia” (FCT), under the

project: FCOMP-01-0124-FEDER-028160/PTDC/BBB-BMD/3088/2012. The first author also thanks FCT for the post-doc grant: SFRH/BPD/97844/2013.

Authors gratefully acknowledge the funding of Project NORTE-01-0145-FEDER-000022 - SciTech - Science and Technology for Competitive and Sustainable Industries, co-financed by “Programa Operacional Regional do Norte” (NORTE2020), through “Fundo Europeu de Desenvolvimento Regional” (FEDER).

## References

- Abbasi, N., Shaw, H., & Rigel, D. (2004). Early Diagnosis of Cutaneous Melanoma. *JAMA: The Journal of Medical Association*, 292, 2771–2776.
- Argenziano G., Soyer, H. P., De Giorgi, V., Piccolo, D., Carli, P., Delfino, M., Ferrari, A., Hofmann-Wellenhof, R., Massi, D., Mazzocchetti, G., Scalvenzi, M., & Wolf, I. H. (2002). *Dermoscopy: a tutorial*. EDRA Medical Publishing & New Media.
- Birmingham, M. L., Pong-Wong, R., Spiliopoulou, A., Hayward, C., Rudan, I., Campbell, H., Wright, A. F., Wilson, J. F., Agakov, F., Navarro, P., & Haley, C. S. (2015). Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Scientific Reports*, 5, 10312.
- Binder, M., Schwarz, M., Winkler, A., Steiner, A., Kaider, A., Wolff, K., & Pehamberger, H. (1995). Epiluminescence microscopy. A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Archives of Dermatology*, 131, 286–291.
- Bradford, P. T., Goldstein, A. M., McMaster, M. L., Tucker, M. A. (2009). Acral lentiginous melanomainscidence and survival patterns in the United States, 1986-2005. *JAMA Dermatology*, 145, 427-434.

- Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., Moss, R. H. (2007). A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, 31, 362-373.
- Celebi, M. E., Wen, Q., Hwang, S., Iyatomi, H., & Schaefer G. (2013) Lesion border detection in dermoscopy images using ensembles of thresholding methods. *Skin Research and Technology*, 19, e252-e258.
- Chu, F., & Wang, L. (2005). Applications of support vector machines to cancer classification with microarray data. *International Journal of Neural Systems*, 15, 475-484.
- Filho, M., Ma, Z. & Tavares, J. M. R. S. (2015). A review of the quantification and classification of pigmented skin lesions: from dedicated to hand-held devices. *Journal of Medical Systems*, 39: 177.
- Flusser, J., & Suk, T. (2006). Rotation moment invariants for recognition of symmetric objects. *IEEE Transactions on Image Processing*, 15, 3784–3790.
- Friedman, R. J., Rigel, D. S., & Kopf, A. W. (1985). Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. *CA: A Cancer Journal for Clinicians*, 35, 130–151.
- Gonzalez, R. C., & Woods, R. E. (2007). *Digital image processing* (3rd ed.). Prentice Hall.
- Grin, C. M., Kopf, A. W., Welkovich, B., Bart, R. S., & Levenstein, M. J. (1990). Accuracy in the clinical diagnosis of malignant melanoma, *Archives of Dermatology*, 126, 763-766.
- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46, 389-422.

- Haidekker, M. (2010). *Advanced biomedical image analysis*. (1st ed.) New Jersey: John Wiley & Sons, (Chapter 9).
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. (2nd ed.). Springer Science & Business Media.
- Heymann, W. R. (2005). Clinical and microscopic diagnosis of melanoma. *Journal of the American Academy of Dermatology*, 52, 133–134.
- Hu, M. K. (1962). Visual Pattern Recognition by Moment Invariants. *IRE Transactions on Information Theory*, 8, 179–187.
- Iyatomi, H., Oka, H., Celebi, M. E., Ogawa, K., Argenziano, G., Soyer, H. P., Koga, H., & Saida, T. (2008). Computer-based classification of dermoscopy images of melanocytic lesions on acral volar skin. *Journal of Investigative Dermatology*, 128, 2049-2054.
- Kittler, H., Pehamberger, H., Wolff, K., & Binder, M. (2002). Diagnostic accuracy of dermoscopy. *Lancet Oncology*, 3, 159–165.
- Kulkarni, S. R., Lugosi, G., & Venkatesh, S. S. (1998). Learning pattern classification-A survey. *IEEE Transactions on Information Theory*, 44, 2178–2206.
- Lyon, V. B. (2010). The Spitz Nevus: Review and Update. *Clinics in Plastic Surgery*, 37, 21–33.
- Ma, Z., & Tavares, J. M. R. S. (2016). A Novel Approach to Segment Skin Lesions in Dermoscopic Images Based on a Deformable Model. *IEEE Journal of Biomedical and Health Informatics*, 20, 615–623.
- Mendonca, T., Ferreira, P. M., Marques, J. S., Marcal, & A. R., Rozeira, J. (2013). PH<sup>2</sup> - a dermoscopic image database for research and benchmarking. *Conf Proc IEEE Eng*

*Med Biol Soc.* 5437-5440.

Olson, E. (2011). Shape factors and their use in image analysis—part 1: theory. *J GXP*

*Compliance*, 15, 85–96.

Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in

bioinformatics. *Bioinformatics*, 23, 2507–2517.

Silveira, M., Nascimento, J. C., Marques, J. S., Marçal, A. R. S., Mendonça, T.,

Yamauchi, S., Maeda, & J., Rozeira, J. (2009). Comparison of segmentation

methods for melanoma diagnosis in dermoscopy images. *IEEE Journal on Selected*

*Topics in Signal Processing*, 3, 35–45.

Yang, S., Oh B., Hahm S., Chung, K. Y., & Lee, B. U. (2017). Ridge and furrow pattern

classification for acral lentiginous melanoma using dermoscopic images. *Biomedical*

*Signal Processing and Control*, 32, 90-96.

Yoradjian, A., Simoes, M. M., Enokihara, S., & Paschoal, F. M. (2012). Nevo de spitz e

nevo de reed. *Anais Brasileiros de Dermatologia*, 87, 349–359.

## FIGURE CAPTIONS

**Fig. 1** Examples of dermoscopic images from the three databases overlapped with the lesion borders (blue contours): (a) non-melanoma in the 1<sup>st</sup> database, (b) melanoma in the 1<sup>st</sup> database, (c) non-melanoma in the 2<sup>nd</sup> database, (d) melanoma in the 2<sup>nd</sup> database, (e) non-melanoma in the 3<sup>rd</sup> database and (f) melanoma in the 3<sup>rd</sup> database.

**Fig. 2** Illustration of the frequencies of features based on the first search phase: (a) color features and (b) shape features.

**Fig. 3** Illustration of the frequencies of shape features based on the second search phase.

**Fig. 4** Illustration of the frequencies of color features based on the second search phase.

**Fig. 5** Illustration of the frequencies of shape features based on the second search phase and the extra searches.

**Fig. 6** Illustration of the frequencies of color features based on the second search phase of searching and the extra searches.

## TABLES

**Table 1** Features for selection

| Type           | Number | Details  |
|----------------|--------|--|
| Shape features | 14     | Ratio of perimeter to area of the skin lesion; smoothness of the lesion border; the seven Hu's invariant image moments; perimeter of the lesion border; area of the skin lesion; aspect ratio; extent ratio; solidity ratio.   |
| Color features | 36     | $L_1, a_1^*, b_1^*, u_1^*, v_1^*, L_0, a_0^*, b_0^*, u_0^*, v_0^*, d_5, d_3, d_4, s_1, s_0, d_6, L_{1-\sigma}, a_{1-\sigma}^*, b_{1-\sigma}^*, u_{1-\sigma}^*, v_{1-\sigma}^*, L_{0-\sigma}, a_{0-\sigma}^*, b_{0-\sigma}^*, u_{0-\sigma}^*, v_{0-\sigma}^*, s_{1-\sigma}, s_{0-\sigma}, \bar{d}_1, \bar{d}_2, \bar{d}_7$ (the mean distance to the geometric centroid $(a_0^*, b_0^*)$ of neighboring healthy skin), $\bar{d}_8$ (the mean distance to the geometric centroid $(u_0^*, v_0^*)$ of neighboring healthy skin), $d_{1-\sigma}, d_{2-\sigma}, d_{7-\sigma}, d_{8-\sigma}$ . |



**Table 2** Image databases used in the study

| Database | Size | Melanoma | Origin  |
|----------|------|----------|---|
| 1        | 100  | 29       | Private clinics in US and Australia (Ma and Tavares, 2016; Celebi et al., 2013) |
| 2        | 200  | 40       | PH2 image database (Mendonca et al., 2013)                                      |
| 3        | 404  | 105      | Interactive atlas of dermoscopy (Argenziano et al. 2002)                        |

**Table 3** Feature sets chosen for the second search phase

| Features   | WN <sup>3</sup> | Comb 1          |                 |                 | Comb 2          |                 |                 |
|--|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|  |                 | OA <sup>3</sup> | S1 <sup>3</sup> | S2 <sup>3</sup> | OA <sup>3</sup> | S1 <sup>3</sup> | S2 <sup>3</sup> |
| 100000000010011000000000000011000000 <sup>1</sup>  | 70              | 82.5%           | 70.0%           | 85.6%           | 65.0%           | 41.4%           | 74.6%           |
| 100000000010010100000000000011000000 <sup>1</sup>  | 74              | 82.5%           | 92.5%           | 80.0%           | 61.0%           | 44.8%           | 67.6%           |
| 000001000010011100000000000011000000 <sup>1</sup>  | 75              | 81.5%           | 70.0%           | 84.4%           | 62.0%           | 44.8%           | 69.0%           |
| 100001000010011010000000000011000000 <sup>1</sup>  | 75              | 81.5%           | 60.0%           | 86.9%           | 62.0%           | 51.7%           | 66.2%           |
| 1111100000000110100000000001011000000 <sup>1</sup> | 75              | 78.5%           | 25.0%           | 91.9%           | 68.0%           | 44.8%           | 77.5%           |
| 1000000000100110100000000001011000000 <sup>1</sup> | 76              | 80.0%           | 45.0%           | 88.8%           | 64.0%           | 51.7%           | 69.0%           |
| 100001000000011000000000000011000000 <sup>1</sup>  | 76              | 80.5%           | 60.0%           | 85.6%           | 63.0%           | 41.4%           | 71.8%           |
| 01010001101100 <sup>2</sup>                        | 72              | 84.5%           | 60.0%           | 90.6%           | 59.0%           | 34.5%           | 69.0%           |
| 01010010101100 <sup>2</sup>                        | 72              | 84.5%           | 60.0%           | 90.6%           | 59.0%           | 34.5%           | 69.0%           |
| 01010100101100 <sup>2</sup>                        | 72              | 84.5%           | 60.0%           | 90.6%           | 59.0%           | 34.5%           | 69.0%           |
| 01010101001100 <sup>2</sup>                        | 72              | 84.5%           | 60.0%           | 90.6%           | 59.0%           | 34.5%           | 69.0%           |
| 01010101101100 <sup>2</sup>                        | 72              | 84.5%           | 60.0%           | 90.6%           | 59.0%           | 34.5%           | 69.0%           |
| 11011110011100 <sup>2</sup>                        | 73              | 81.5%           | 57.5%           | 87.5%           | 64.0%           | 31.0%           | 77.5%           |
| 11011110111100 <sup>2</sup>                        | 73              | 81.5%           | 57.5%           | 87.5%           | 64.0%           | 31.0%           | 77.5%           |
| 11011111011100 <sup>2</sup>                        | 73              | 81.5%           | 57.5%           | 87.5%           | 64.0%           | 31.0%           | 77.5%           |
| 11011111111100 <sup>2</sup>                        | 73              | 81.5%           | 57.5%           | 87.5%           | 64.0%           | 31.0%           | 77.5%           |

<sup>1</sup> Combination of color features represented by a binary string, '0' – unused and '1' – used. From left to right: the sequence of features is listed in Table 1.

<sup>2</sup> Combination of shape features represented by a binary strings, '0' – not used and '1' – used. From left to right: the sequence of features is listed in Table 1.

<sup>3</sup> WN – Wrong number; OA – Overall accuracy; S1 – Sensitivity; S2 – Specificity.

**Table 4** Feature sets selected for another search

| Features  |   | W  | Comb 1 |       |       | Comb 2 |       |       |
|---|---|----|--------|-------|-------|--------|-------|-------|
|   |   | N  | OA     | S1    | S2    | OA     | S1    | S2    |
| 1000000000100110000<br>00000000011000000 <sup>1</sup> | 01100100001110 <sup>2</sup>                           | 45 | 90.0%  | 75.0% | 93.8% | 75.0%  | 44.8% | 87.3% |
|   | 01100100101110 <sup>2</sup>                           |    | 90.0%  | 75.0% | 93.8% | 75.0%  | 44.8% | 87.3% |
|   | 11010001101100 <sup>2</sup>                           |    | 90.5%  | 72.5% | 95.0% | 74.0%  | 44.8% | 85.9% |
|   | 11110001101100 <sup>2</sup>                           | 46 | 90.0%  | 72.5% | 94.4% | 74.0%  | 44.8% | 85.9% |
|   | 11110110101100 <sup>2</sup>                           |    | 89.5%  | 72.5% | 93.8% | 75.0%  | 44.8% | 87.3% |
| 1000000000100101000<br>00000000011000000 <sup>1</sup> | 11010100001100 <sup>2</sup>                           | 46 | 90.0%  | 67.5% | 95.6% | 74.0%  | 44.8% | 85.9% |
| 1000000000100101000<br>00000000011000000 <sup>1</sup> | 11100100101100 <sup>2</sup>                           | 45 | 90.0%  | 95.0% | 88.8% | 75.0%  | 34.5% | 91.5% |
|   | 11100000101100 <sup>2</sup>                           | 46 | 89.0%  | 95.0% | 87.5% | 76.0%  | 37.9% | 91.5% |
| 01010001101100 <sup>2</sup>                           | 00000000000001010000<br>0000001000000000 <sup>1</sup> | 46 | 88.5%  | 67.5% | 93.8% | 77.0%  | 41.4% | 91.5% |
| 01010010101100 <sup>2</sup>                           | 00000000000000000000<br>0000000011001100 <sup>1</sup> | 45 | 89.5%  | 62.5% | 96.3% | 76.0%  | 65.5% | 80.3% |
| 01010101001100 <sup>2</sup>                           | 00000000000000000000<br>0000000011110000 <sup>1</sup> | 46 | 89.5%  | 62.5% | 96.3% | 75.0%  | 62.1% | 80.3% |

<sup>1</sup> Combination of color features with the sequence defined in Table 1.

<sup>2</sup> Combination of shape features with the sequence defined in Table 1.

**Table 5** Feature sets selected for an extra search

| Features   | WN | Comb 1 |       |       | Comb 2 |       |       |
|--|----|--------|-------|-------|--------|-------|-------|
|  |    | OA     | S1    | S2    | OA     | S1    | S2    |
| 11010100001100 <sup>2</sup> 1000000000100110000000000000011000000 <sup>1</sup> | 46 | 90.0%  | 67.5% | 95.6% | 74.0%  | 44.8% | 85.9% |
| 0000000000000001000000000000011000000 <sup>1</sup>                             |    | 88.5%  | 60.0% | 95.6% | 77.0%  | 65.5% | 81.7% |
| 000000000000010100000000001000000000 <sup>1</sup> 01000001001100 <sup>2</sup>  | 43 | 90.0%  | 72.5% | 94.4% | 77.0%  | 44.8% | 90.1% |
| 01010001101100 <sup>2</sup>  | 46 | 88.5%  | 67.5% | 93.8% | 77.0%  | 41.4% | 91.5% |

<sup>1</sup> Combination of color features with the sequence defined in Table 1.

<sup>2</sup> Combination of shape features with the sequence defined in Table 1.

**Table 6** Best feature sets based on the classification with features of one type

|   | Shape features / Color features                   | Comb 1 |       |       | Comb 2 |       |        |
|---|---|--------|-------|-------|--------|-------|--------|
|   |   | OA     | S1    | S2    | OA     | S1    | S2     |
| 1 | 00010100011010 <sup>2</sup>                       | 90.0%  | 65.0% | 96.3% | 72.0%  | 6.9%  | 98.6%  |
|   | 000001001100000100000100000000000000 <sup>1</sup> | 86.5%  | 57.5% | 93.8% | 80.0%  | 31.0% | 100.0% |
| 2 | 10001110011010 <sup>2</sup>                       | 89.5%  | 60.0% | 96.9% | 72.0%  | 3.4%  | 100.0% |
|   | 000001110000001100000100000000000000 <sup>1</sup> | 87.0%  | 57.5% | 94.4% | 79.0%  | 37.9% | 95.8%  |
| 3 | 00010011111010 <sup>2</sup>                       | 89.5%  | 70.0% | 94.4% | 71.0%  | 3.4%  | 98.6%  |
|   | 011001000000000100000100000000000000 <sup>1</sup> | 90.5%  | 70.0% | 95.6% | 72.0%  | 10.3% | 97.2%  |
| 4 | 00011001011010 <sup>2</sup>                       | 89.0%  | 70.0% | 93.8% | 72.0%  | 3.4%  | 100.0% |
|   | 000001110000001000000100000000000000 <sup>1</sup> | 87.5%  | 52.5% | 96.3% | 77.0%  | 27.6% | 97.2%  |
| 5 | 00011100101110 <sup>2</sup>                       | 89.5%  | 55.0% | 98.1% | 71.0%  | 6.9%  | 97.2%  |
|   | 000001001100000000000100000000000000 <sup>1</sup> | 87.0%  | 60.0% | 93.8% | 77.0%  | 20.7% | 100.0% |
| 6 | 00101101110110 <sup>2</sup>                       | 89.5%  | 55.0% | 98.1% | 71.0%  | 3.4%  | 98.6%  |
|   | 011001000010000000000100000011001100 <sup>1</sup> | 88.0%  | 52.5% | 96.9% | 74.0%  | 37.9% | 88.7%  |

<sup>1</sup> Combination of color features with the sequence defined in Table 1.

<sup>2</sup> Combination of shape features with the sequence defined in Table 1.

**FIGURES**

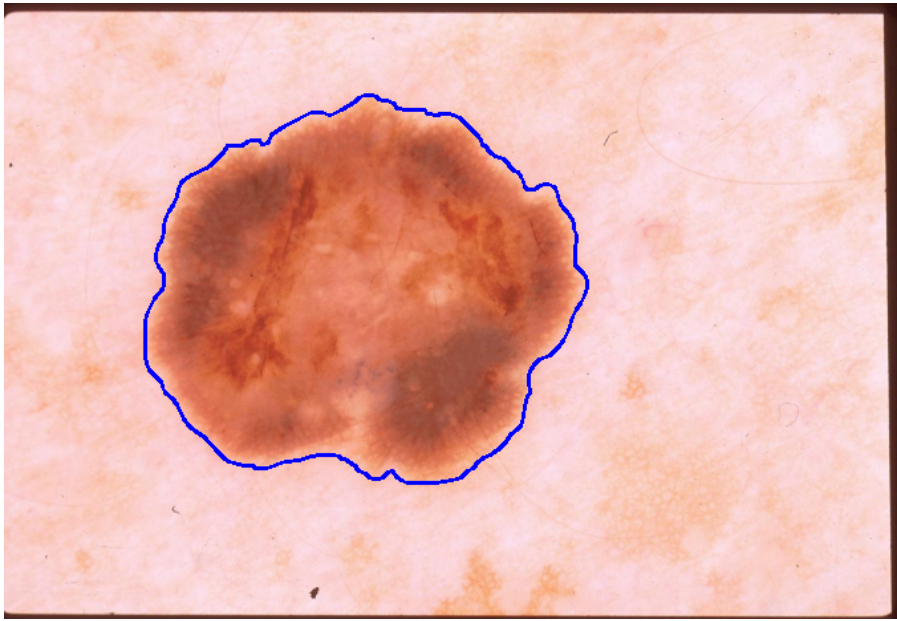


Figure 1a



Figure 1b



Figure 1c

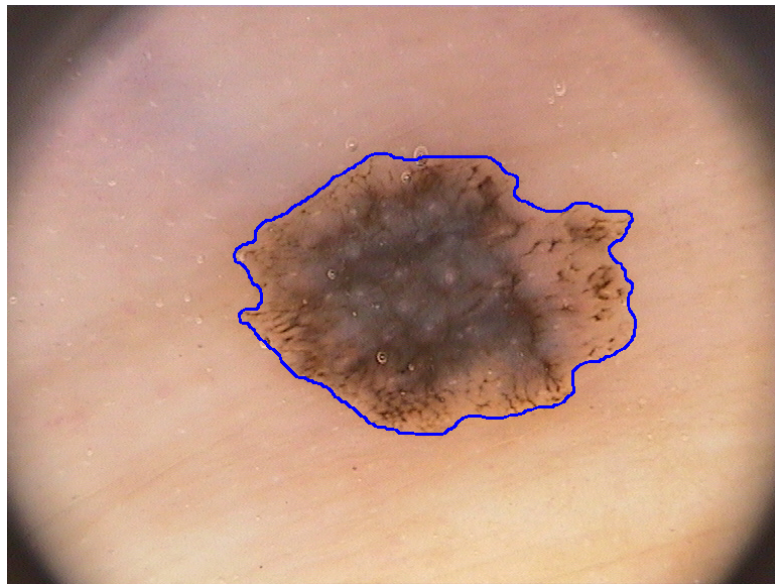


Figure 1d



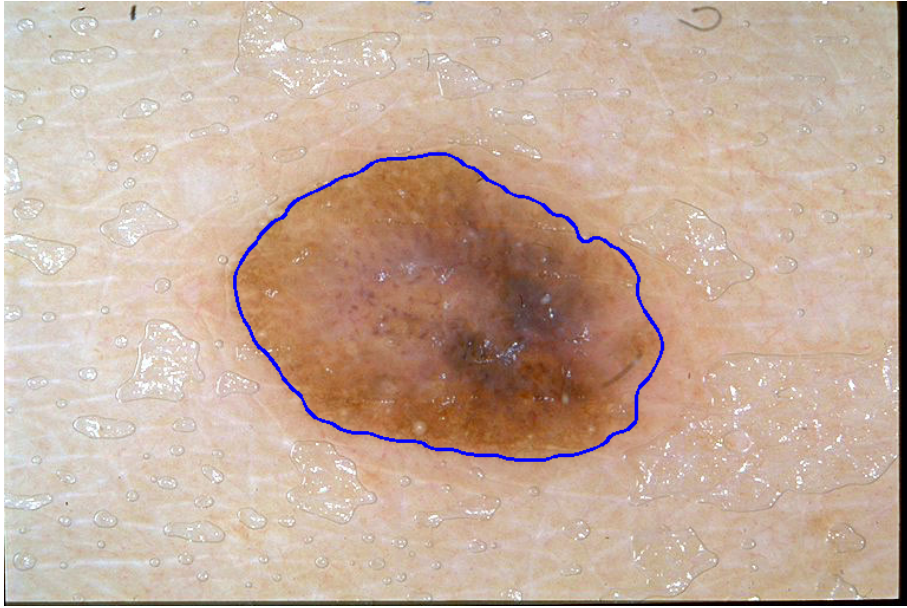


Figure 1e

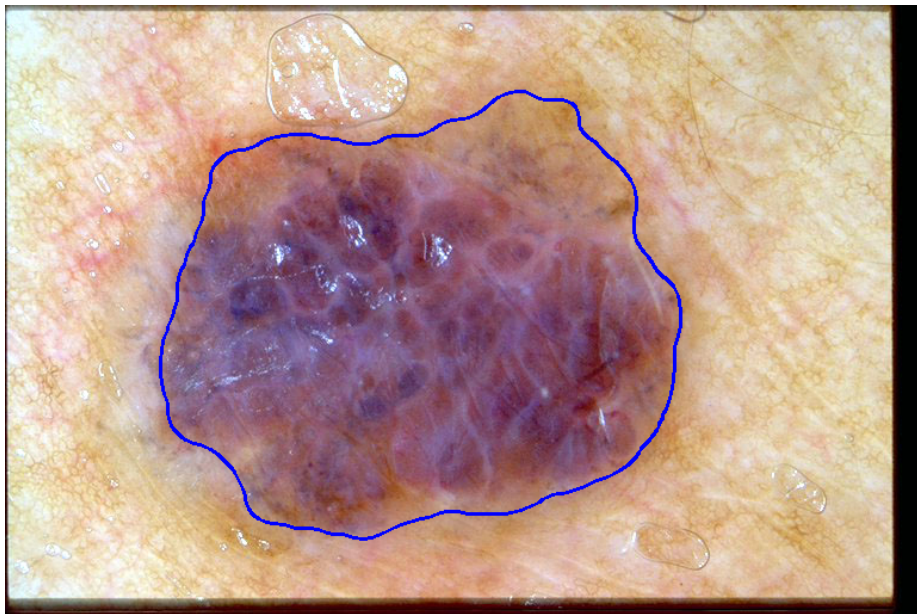


Figure 1f



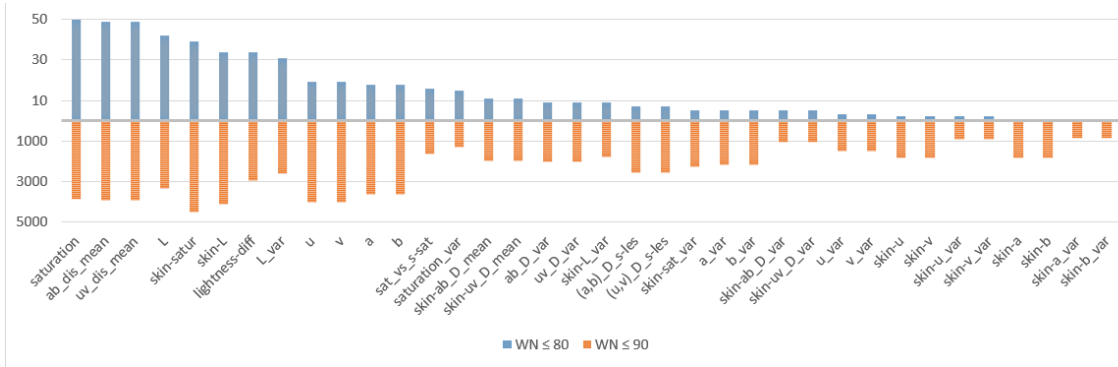


Figure 2a

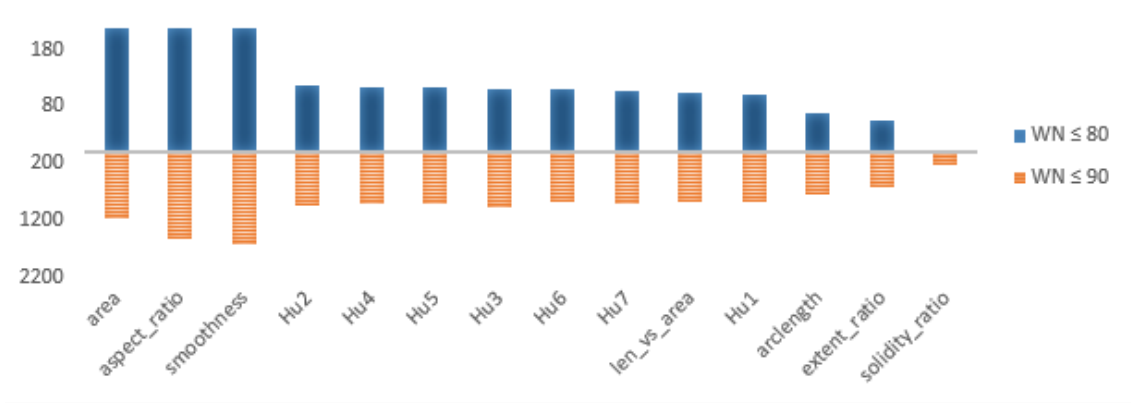


Figure 2b

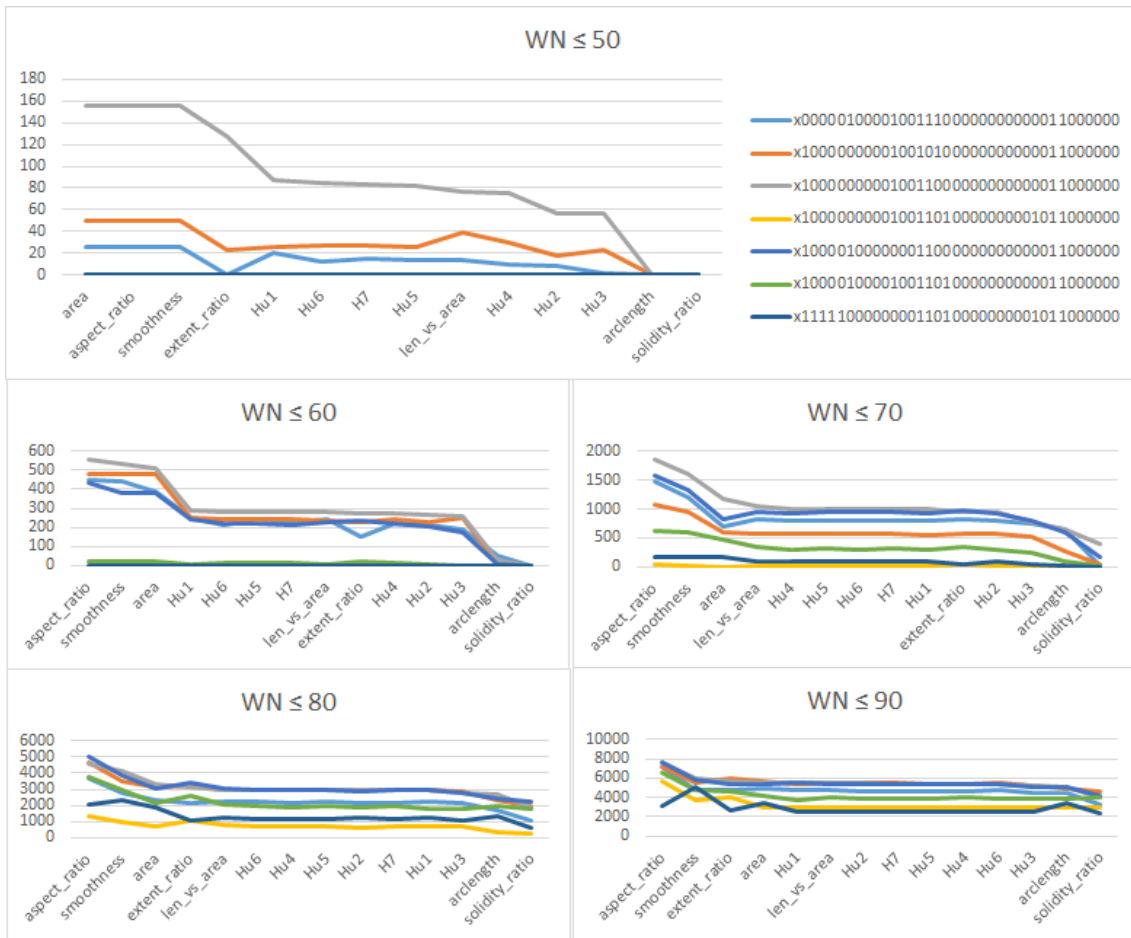


Figure 3

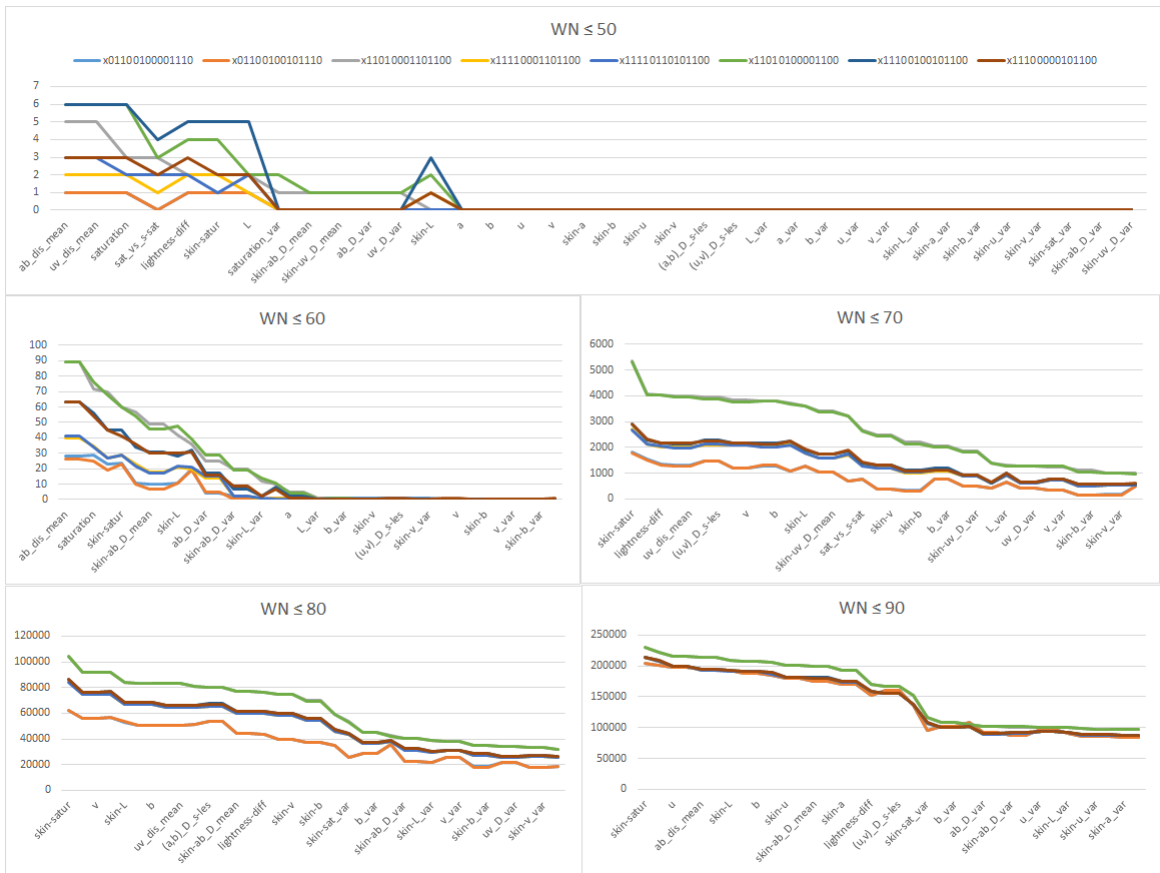


Figure 4

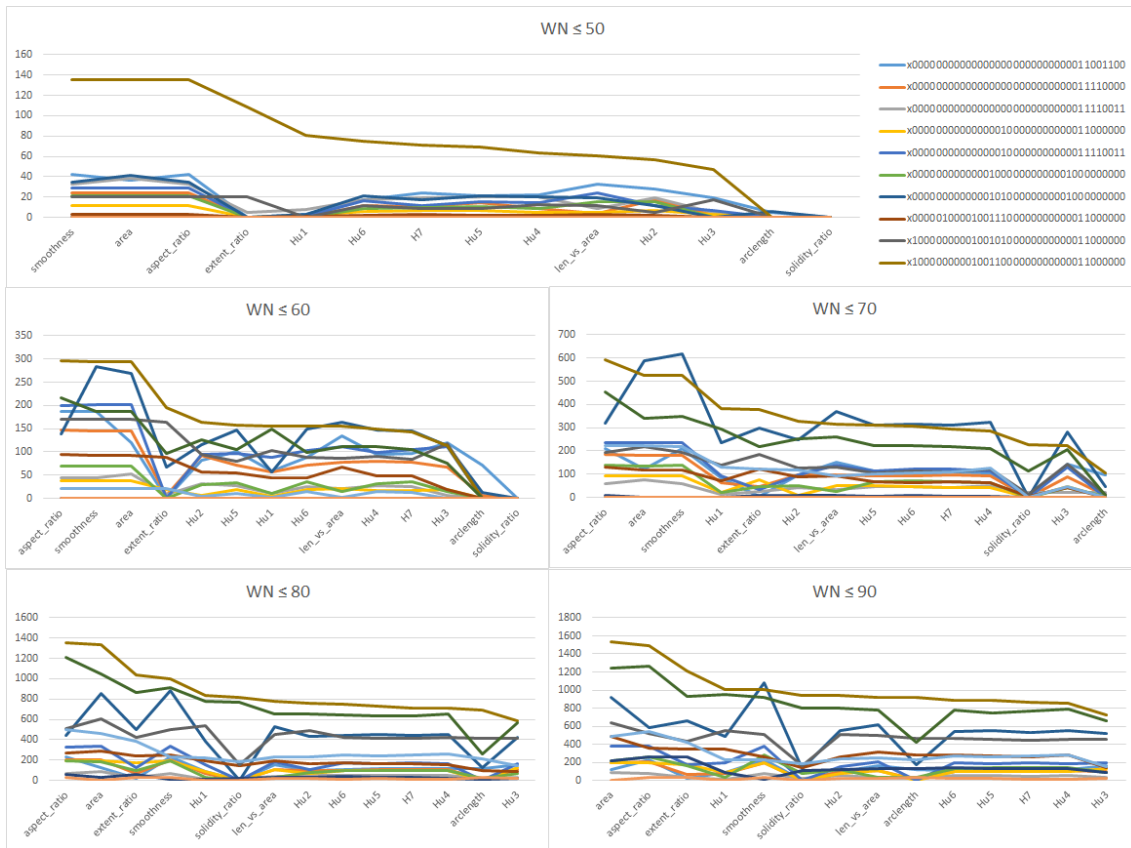


Figure 5

