

# Chapter 7

## Mole Features Extraction for a Melanoma Recognition System

Henrique C. Siqueira and Bruno J.T. Fernandes

The cancer is a painful disease that kill too many people. Skin cancer is among the most frequent types of tumors in the world, and melanoma is the most worrying type of skin cancer due to its high metastasis chances. Its global occurrence index is close to 133.000 people per year. Irresponsible exposure to the sun causes 40% out of the total. Melanoma is fatal when not diagnosed it its initial stages. The most common diagnosis method is done visually based on five features: asymmetry, border, color, diameter and elevation, also kwon as ABCDE method. We propose three algorithms to extract features of skin moles based on dermatological studies, using digital image processing techniques existing in the lecture. The first feature measures the asymmetry level of the mole; the second one calculates irregularity of the edges, and the third one computes the color variance of the mole. We also evaluate these features as input to classifiers creating a melanoma recognition model that indicates whether is melanoma or normal mole. The analysis of results are shown through ROC curve and 10-fold cross-validation from two dermatological datasets: Atlas of Clinical Dermatology and DermNet NZ.

---

H.C. Siqueira (✉) · B.J.T. Fernandes  
Escola Politécnica, University of Pernambuco, Recife, Brazil  
e-mail: hcs@ecomp.poli.br

B.J.T. Fernandes  
e-mail: bjtf@ecomp.poli.br

## 7.1 Introduction

Nowadays, skin cancer is among the most common cancers in the world, especially in tropical countries because of the high incidence of UV rays. In Brazil is the most common tumor corresponding to 25% of all malignant tumors already registered according to research conducted by INCA (National Cancer Institute of José Alencar Gomes da Silva) [11]. Its incidence has rapidly increased approximately 3–7% rates for people with light skin.

There are three main types of skin cancer: basal cell carcinoma (BCC), squamous cell carcinoma (SCC) and melanoma. The last one represents 4% of disease diagnostics and its major incidence is in adults with light skins. However, it is the worst case due to its high metastasis chances, the dissemination of the cancer lesion for the other organs.

The successful treatment of this cancer increases considerably if the tumor is identified in the early stages. As evidence, there was a great improvement in survival of melanoma patients that had early diagnosis in the last years [11], and about 90% of the cases have been completely cured when the tumors was found less than one millimeter of diameter [15].

Scientific works have been developed with the objective to create support systems for melanoma diagnosis since 1980s [3]. Nevertheless, there is not a definition about the most precise model to diagnose this pathology. One reason is the difficulty to compare the models, because are applied different statistical methods to validate, and from different databases, some of these data sets created by their authors, like in Manousaki et al. [12].

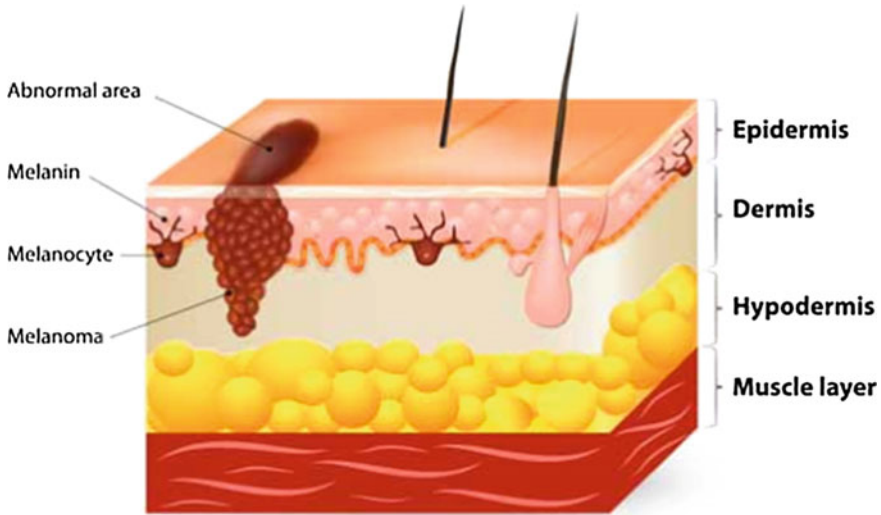
In this work, we propose three new algorithms capable to extract mole features from the human skin based on real features, defined by the medical community, using digital image processing techniques. We also propose a recognition model able to distinguish whether a mole image is a melanoma or not based on such features, using one of the following classifiers for decision-making: artificial neural network (ANN), logistic regression (LR) and support vector machine (SVM).

The chapter is organized as follows. Firstly, we general explain the melanoma disease. Thus, the features extraction algorithms are introduced, as well as the model to classify melanoma using our features as input. Finally, the experimental study is presented followed by some concluding remarks.

## 7.2 Melanoma

The melanoma is a type of skin cancer that affects the melanocytes cells, located in the bottom of the skin's epidermis as shown in Fig. 7.1. The melanocytes produce the melanin, the pigment responsible for the skin color.

The melanoma may begin like a mole that grows over time, may appear in almost any color (including red, blue, brown, black, gray, and tan), usually has irregular



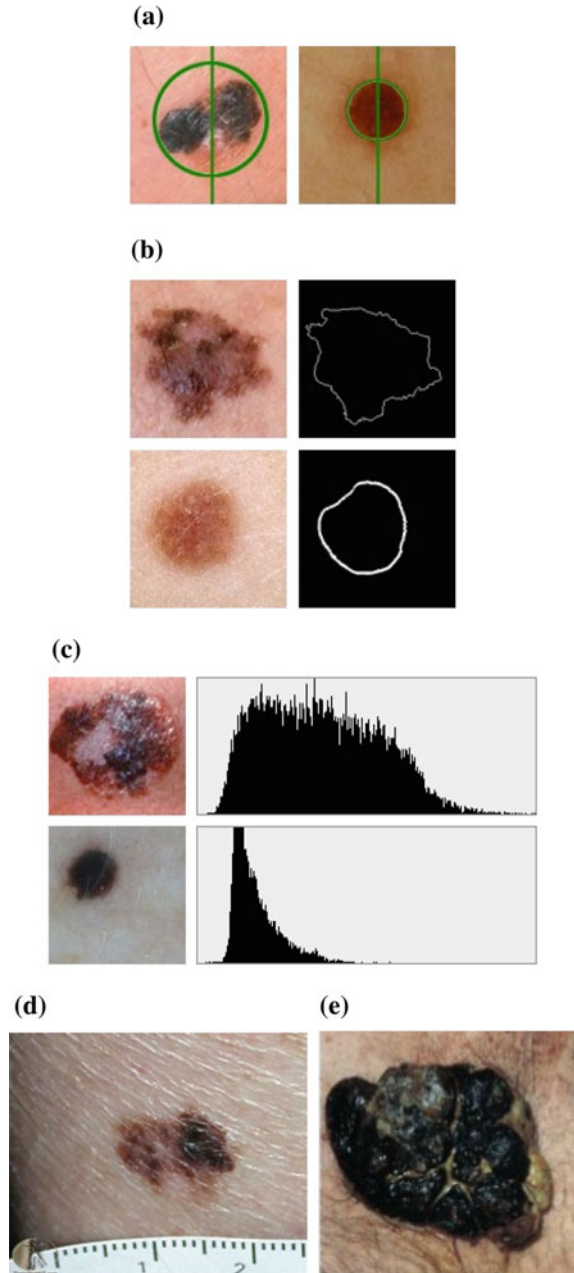
**Fig. 7.1** Skin's layers with a melanocyte cell and a melanoma. Adapted from The Skin and Cancer Foundation Inc. 2016, Retrieved from <https://www.skincancer.asn.au/page/2149/learn-about-melanoma>

edges, may be flat or raised on skin, may be painless or form wound. His appearance is independent of anywhere of the body, but it is more common in areas exposed to the sun, such as shoulders, head, arms and legs [5].

The most common procedure for the melanoma identification is made by dermatoscopy, an examination that usually performed through the dermatoscope, hand-held microscope that magnifies the skin ten times. Thus, the analysis identifies five main features, also known as ABCDE method (**A**symmetry, **B**order, **C**olor, **D**iameter and **E**levation) [5], as illustrated in Fig. 7.2.

- **Asymmetry:** indicates the level of similarity between the two halves of the mole. The Fig. 7.2a shows an asymmetrical melanoma on left and a symmetrical mole on right;
- **Border:** melanoma presents irregularity on the edges. In Fig. 7.2b, the top pictures show the border irregularity present in the melanoma, in opposite, the bottom pictures shows the mole with smooth transition on the edges;
- **Color:** melanoma has more than one color in the same mole. The Fig. 7.2c presents the histogram of the moles. In the top, melanoma case, the histogram shows the wide range of intensities, while in the bottom, has a small range of intensities.
- **Diameter:** melanoma is usually larger than 6 mm. Example of the melanoma case with approximately 2 cm of diameter in Fig. 7.2d.
- **Elevation:** It is more common to find melanoma which create a raised surface on the skin Fig. 7.2e.

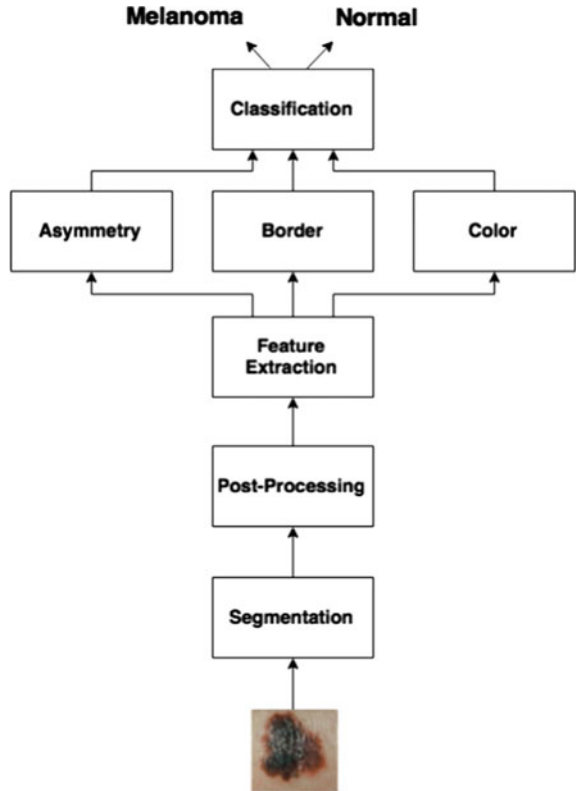
**Fig. 7.2** Illustration of mole features: asymmetry (a), border (b), color variance (c), diameter (d) and elevation (e)

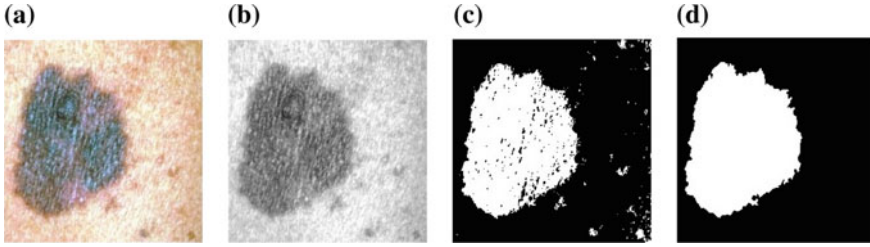


### 7.3 Melanoma Recognition Model

Figure 7.3 presents the complete workflow of the melanoma recognition model. The input of the model is a color image with any dimension, as illustrated in Fig. 7.4a. Next, it is necessary to identify the mole in the picture. So, we assume that image contains two classes: the first one is the mole defined as an object, and the other is the skin defined as background. The choice of the segmentation algorithm is an important decision because it contributes to the effectiveness of next steps. According to Bhuiyan et al. [2], which compares segmentation methods applied to binary images of a mole in the skin, the segmentation method by Otsu [14] achieves the best results. This method segments the grayscale image previous converted, as shown in Fig. 7.4b, in two classes as in Fig. 7.4c, based on the calculation of the optimal threshold in the image histogram that minimizes intra-class variance and maximize inter-class variance. Furthermore, does not require to set any parameter for different skin colors and moles.

Fig. 7.3 Workflow of proposed model for melanoma classification





**Fig. 7.4** Original image in (a). Converted grayscale image (b). The segmented image in (c). Image of segmented mole without fails in (d)

After segmentation, we make post processing to correct any fails in the binary image in Fig. 7.4d that has resulted of Otsu. Failures are skin areas that have been identified as a mole or holes present in the mole. A hole is a background region surrounded by pixels that represent the object. We used the algorithm of Suzuki et al. [16] to find the contour of all the object regions, and adopted as a mole the region with the highest number of connected components. With possession of contour points we perform the method *fillpoly()* of OpenCV [13] to fill the region, resulting in Fig. 7.4d, a segmented mole image without fails, after used as mask. Finally, the features are extracted and used as input to the supervised classifier. The output of the classifier indicates the presence of melanoma or not.

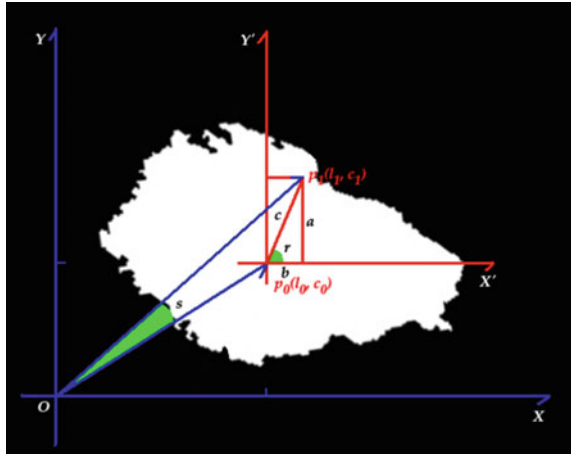
### 7.3.1 Feature Extraction

Three methods to extract features of the mole have been created, where we believe that they can be a good representation for the normal and the abnormal mole. The algorithms measure three melanoma characteristics previously covered: the asymmetry level, the border shape, and the color shades in the mole. The other features, diameter and elevation, are disregarded because we have not possible to get the real dimensions of the mole in the two-dimensional picture since the distance at which the image was captured is unknown.

### 7.3.2 Asymmetry

The asymmetry is calculated with the alignment mole by rotating the segmented image by an angle  $r$  between the mole orientation axis  $o$  and the  $x$  axis of the picture. The Fig. 7.5 illustrates the mole, the axes and the relation between angles. To obtain the  $r$  value, it is necessary to calculate the coordinates of the centroid point  $p_c(l_c, c_c)$  of the mole, given by,

**Fig. 7.5** Relation between angles in the mole



$$p_c(l_c, c_c) = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \tag{7.1}$$

where  $m_{xy}$  are the spatial moments of order  $xy$ .

According to Horn et al. [10], to trace the orientation axis through the centroid we have to get two points:  $p_0(l_0, c_0)$  and  $p_1(l_1, c_1)$  based on angle  $s$ , obtained by:

$$s = 0.5 * \arctan \left( \frac{2 * mu_{11}}{(mu_{20} - mu_{02})} \right) \tag{7.2}$$

$$p_0(l_0, c_0) = (l_c - (100 * \cos(s)), c_c - (100 * \sin(s))) \tag{7.3}$$

$$p_1(l_1, c_1) = (l_c + (100 * \cos(s)), c_c + (100 * \sin(s))) \tag{7.4}$$

where  $mu_{xy}$  are the central moments of order  $xy$ . Finally,  $r$  is obtained through the law of tangents, trigonometric formula given by,

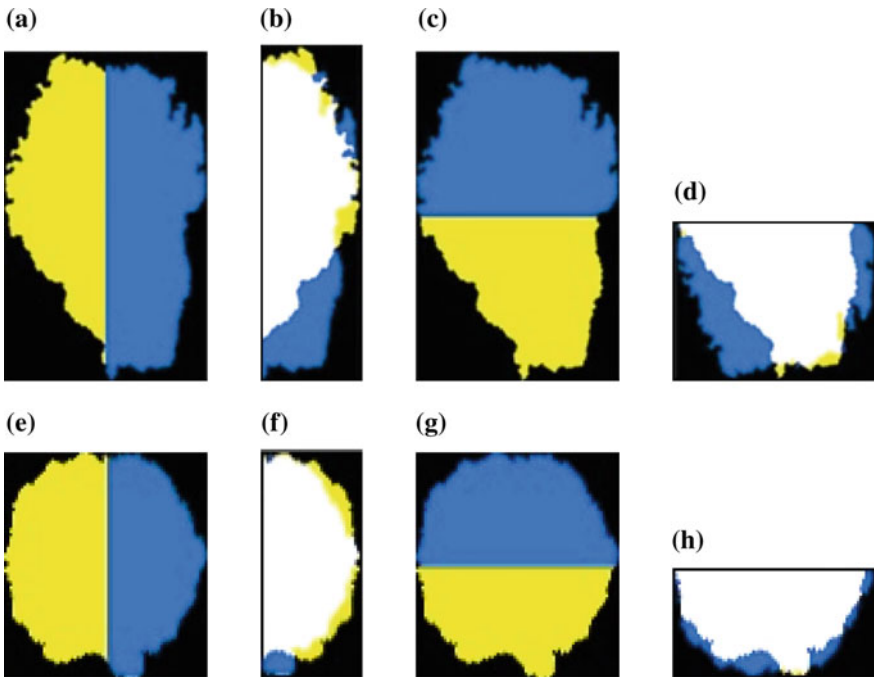
$$r = \arctan \left( \frac{a}{b} \right) \tag{7.5}$$

being  $a$  the difference between  $c_1$  and  $c_0$ , and  $b$  difference between  $l_1$  and  $l_0$ . Then, the mole is delimited with the smallest possible rectangle, resulting a new image containing only the mole (Fig. 7.6).

The Fig. 7.7 shows that the asymmetry can happen in relation with axis  $y$ , vertically, as well as in relation with axis  $x$ , horizontally. So, the both axis have to be considered to calculate this feature.



**Fig. 7.6** Image containing only the aligned mole



**Fig. 7.7** Asymmetry with y axis (a) and (e), overlapping halves on the y axis (b) and (f), asymmetric with x axis (c) and (g) overlapping halves on the x axis (d) and (h). Melanoma (a), (b), (c) and (d) and normal mole (e), (f), (g) and (h)

It is possible to identify the asymmetry difference between the melanoma, as illustrated in Figs. 7.7a–d and the normal mole, as shown in Figs. 7.7e–h by the analysis of the white region, which represents the overlapping parts of the image divided by the respective axis.



It can be observed in the non-melanotic case that the white region is occupying the largest relative area in the mole, whereas in the melanoma case this area is proportionally smaller. Lastly, we measure the mole asymmetry in percentage by the ratio between the number of pixels that do not coincide and the total number of pixels in the mole.

### 7.3.3 Border

To measure the irregularity of the border, the standard deviation  $\sigma_d$  of mole signature is calculated. Shape signature is a set of distances  $d_i$  of centroid point  $pc(l_c, c_c)$ , to each point  $p_i(l_i, c_i)$  of the mole contours. This distance is obtained applying the Pythagorean theorem,

$$d_i = \sqrt{(l_c - l_i)^2 + (c_c - c_i)^2}. \quad (7.6)$$

Finally, the standard deviation of the signature is

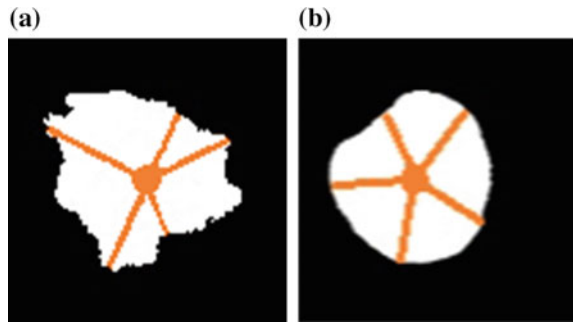
$$\sigma_d = \sqrt{\sum_{i=0}^{N_d} (d_i - \mu_d)^2}, \quad (7.7)$$

being  $N_d$  the number of calculated distances and  $\mu_d$  is

$$\mu_d = \frac{1}{N_d} * \sum_{i=0}^{N_d} d_i. \quad (7.8)$$

In melanoma cases this distance presents a large variance, as in Fig. 7.8a, whereas in normal moles, as in Fig. 7.8b it tend to remain constant.

**Fig. 7.8** Melanoma segmented image in (a) and the non-cancer mole in (b)



### 7.3.4 Color

To quantify the non-uniformity of the mole color, the variance  $\sigma^2$  of the mole's histogram is calculated. The grayscale image used in this process; besides reducing the computational cost, the non-uniformity of the pixels intensity from the mole is maintained.

Firstly, it is calculated the mean intensity  $\mu_i$  of the mole,

$$\mu_i = \frac{1}{L} * \sum_{i=0}^{L-1} i * p(i), \quad (7.9)$$

where  $i$  is the intensity value that can vary from 0 to  $L - 1$ . Being  $L$  the maximum number of the intensities which the pixel can represents; in case of grayscale image with eight bits, this value is 256. Finally,  $p(i)$  is the probability for the intensity  $i$  is included in the image. Thus, the variance is calculated by

$$\sigma_i^2 = \sum_{i=0}^{L-1} (i - \mu_i)^2. \quad (7.10)$$

In order to have the color variation rate of the mole  $c_{rate}$  in the range of zero and one, it was necessary normalize the result, applying the equation

$$c_{rate} = 1 - \frac{1}{1 + \left(\frac{\sigma_i^2}{L^2}\right)}, \quad (7.11)$$

according to Gonzalez et al. [9].

**Fig. 7.9** Histogram of the melanoma (a) and the normal mole (b)

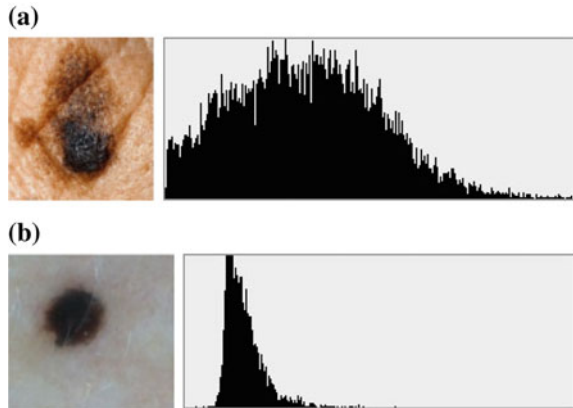
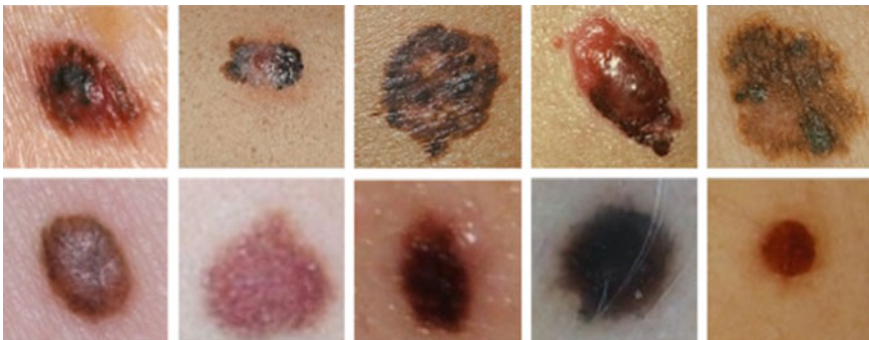


Figure 7.9a is a melanoma that presents the variance equal to 2141.9, while Fig. 7.9b has variance equal to 165.1. Therefore, it is observed that the variances of the histograms in the normal moles are usually small because the colors of the moles tend to be uniform while they are usually high in cases of melanoma.

## 7.4 Experimental Results

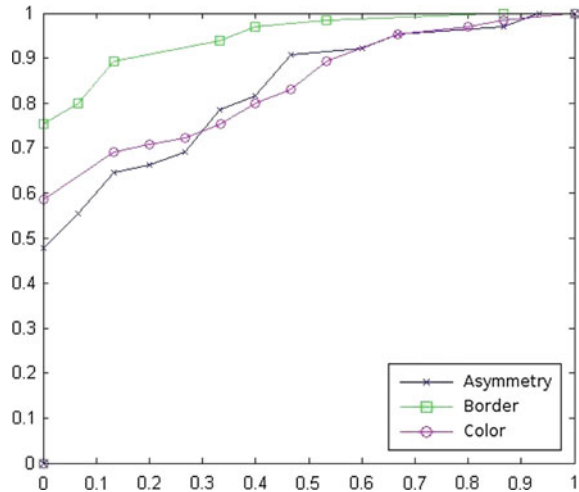
The proposed algorithms for features extraction and the models for melanoma classification are tested using images from two dermatological databases: Atlas of Clinical Dermatology and DermNet NZ. The first is a clinical dermatology atlas that has approximately 3000 images of dermatological diseases, all obtained by Niels K. Veien in his private dermatological clinic [1]. These images are intended for use in the study of dermatology area. The second, available since 1996 by New Zealand Dermatological Society Incorporated, has images and papers about skin. It is written and reviewed by health professionals and medical writers, with free access to the dataset via internet [6]. The Fig. 7.10 presents some image examples of these databases. We extract features from 139 images of moles in the skin, where 105 of these are cases of melanoma, and 34 are normal moles. All pictures are colors of 24-bit, 8-bit per channel in the RGB pattern (red, green and blue).

One way to demonstrate the antagonistic relationship between the melanoma and the normal moles for each feature is the analysis of the receiver operating characteristic (ROC) [8]. The ROC curve is a graph of true positive rates, that means the positive diagnosis with the presence of the pathology, against false positive rates, that means the negative diagnosis with the non existence of the pathology. In other words, the first rate is the ratio between the number of melanoma cases correctly classified over the total of melanoma images, while second rate is the relation between the number of normal moles misclassified as melanotic over the total cases of the normal mole



**Fig. 7.10** Examples of some images of these databases: DermNet NZ and Atlas of Clinical Dermatology

**Fig. 7.11** ROC curve of the extracted features: asymmetry, border and color



images. The quality of the result of the ROC curve is determined by the area under the curve (AUC) [4].

The Fig. 7.11 shows the comparison between the extracted features. The highest AUC obtained was 0.93 by the standard deviation of the edge of the mole. The asymmetry rate and the color rate obtained 0.82 and 0.83, respectively. So, the standard deviation of the edge can identify melanoma cases better than other features.

The results obtained using multilayer perceptron neural network (MLP), logistic regression with ridge estimator (LRR) and support vector machine (SVM) for the melanoma recognition problem are presented. We tested many different configurations for each classifier. As an evaluation approach of the models, we use the 10-fold cross-validation [7], where we divide the database into 10 sets mutually exclusive. At each iteration, one set is used for testing and the remaining sets are used for model training. The Tables 7.1, 7.2 and 7.3 present the evaluated configurations and accuracy obtained for each model, MLP, LRR and SVM, respectively. The accuracy is defined by the number of images classified correctly divided by the total number

**Table 7.1** Accuracy for Melanoma classification using MLP

MLP	Configuration	Accuracy (%)
MLP-1	$n_h = 3 \mid \alpha = 0.3 \mid m = 0.2$	81.3
MLP-2	$n_h = 10 \mid \alpha = 0.3 \mid m = 0.2$	84.9
MLP-3	$n_h = 20 \mid \alpha = 0.3 \mid m = 0.2$	83.5
MLP-4	$n_h = 10 \mid \alpha = 0.2 \mid m = 0.1$	79.9
MLP-5	$n_h = 10 \mid \alpha = 0.35 \mid m = 0.2$	80.6

**Table 7.2** Accuracy for Melanoma classification using logistic regression with ridge estimator

LRR	Configuration	Accuracy (%)
LRR-1	<i>ridge</i> = 0.1	84.2
LRR-2	<i>ridge</i> = 0.3	85.6
LRR-3	<i>ridge</i> = 0.5	86.3
LRR-4	<i>ridge</i> = 0.8	85.6
LRR-5	<i>ridge</i> = 1	84.9

**Table 7.3** Accuracy for Melanoma classification using SVM with sigmoid kernel

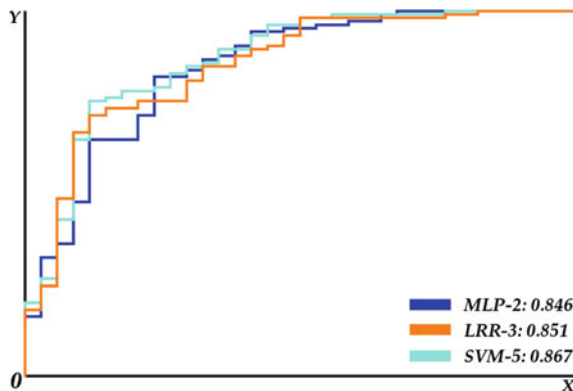
SVM	Configuration	Accuracy (%)
SVM-1	<i>gamma</i> = 2   <i>coef</i> = 2   <i>cost</i> = 1	85.6
SVM-2	<i>gamma</i> = 2.5   <i>coef</i> = 2   <i>cost</i> = 1	77.0
SVM-3	<i>gamma</i> = 2   <i>coef</i> = 1   <i>cost</i> = 1	84.9
SVM-4	<i>gamma</i> = 2   <i>coef</i> = 1   <i>cost</i> = 3	84.2
SVM-5	<i>gamma</i> = 2   <i>coef</i> = 3   <i>cost</i> = 3	86.3

of images. The best configuration of each model was selected for a more detailed study including the MLP-2 with 84.9% of accuracy, the LRR-3, and the SVM-5 with 86.3% of accuracy.

The Fig. 7.12 shows the ROC curves of the best configurations for each model. The SVM-5 had the best performance with 0.867 of AUC in comparison with MLP-2, which has 0.846 of AUC and e LRR-3 with 0.851 of AUC.

The Tables 7.4, 7.5 and 7.6 represent the confusion matrix of the models. The lines correspond to the real values (target) of classes and columns correspond to the values of the output of the model (predicted). The analysis of the confusion matrix in medical diagnostic systems is important to the detriment of the comparison

**Fig. 7.12** Comparison of MLP-2, LRR-3 and SVM-5 classifiers



**Table 7.4** Confusion matrix of Melanoma classification using MLP-2

	Normal	Melanoma
Normal	18	16
Melanoma	5	100

**Table 7.5** Confusion matrix of Melanoma classification using LRR-3

	Normal	Melanoma
Normal	17	17
Melanoma	2	103

**Table 7.6** Confusion matrix of Melanoma classification using SVM-5

	Normal	Melanoma
Normal	19	15
Melanoma	4	101

between false positive and false negative rates. Considering the problem of skin cancer classification, we can say that the minimization of false negative rates is crucial because it represents the reduction of the error where a skin cancer was classified as a normal mole, not masking the presence of a malignancy in the patient. This type of error must be avoided since the time is a critical factor in the success of the treatment. Rather, a consider number of false positive is not considered a serious mistake, since for the patient would generate only a warning about the presence of disease.

Thus, despite not having the largest area under the curve, the model LRR-3 has the lowest number of false negative, only 2 cases. While the SVM-5 has 4 cases and the MLP-2, with the worst performance, 5 cases of cancer signs classified as benign.

## 7.5 Conclusion

This work introduced three algorithms for features extraction from images of moles in human skin, capable to measure the asymmetry level, the border irregularity and the non-uniformity of the color. It is important to notice that they are invariants for scale, rotation and translation, important properties in features extraction tasks.

This work also presented models to classify the melanoma. The first was the artificial neural network multilayer perceptron, the second was the logistic regression with ridge estimator, and the last was the support vector machine. All models used as input the results of the algorithms developed here. The proposed models performed well, especially the LRR-3, with 86.4 % accuracy and only 2 instances of false negative. It was observed with the experiments that the extracted features can create a good representation of classes: melanoma and normal mole.

Moreover, good rates obtained in the experiments motivate the creation a system with user iteration, for that the diameter of features and elevation are taken into account, and may improve the rates obtained in the experiments.

## References

1. Atlas of Clinical Dermatology Homepage (2015). <http://www.danderm-pdv.is.kkh.dk/atlas/index.html>. Accessed May 2015
2. Bhuiyan AH, Azad I, Uddin K (2013) Image processing for skin cancer features extraction. *Int J Sci Eng Res*
3. Blum A, Zalaudek I, Argenziano G (2008) Digital image analysis for diagnosis of skin tumors. *Semin Cutan Med Surg*
4. Bradley AP (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn*
5. Denise MA (2004) Melanoma. nursing: patient education series
6. DermNet NZ Homepage (2015). <http://www.dermnetnz.org/sitemap.html>. Accessed May 2015
7. Duda RO, Hart PE, Stork DG (2000) *Pattern classification*, 2nd edn. Wiley-Interscience, New York
8. Fawcett T (2006) *An introduction to ROC analysis*. Elsevier Science Inc, Pattern Recogn. Lett
9. Gonzalez RC, Woods RE (2006) *Digital image processing*, 3rd edn. Prentice-Hall, Inc, Upper Saddle river
10. Horn BK (1986) *Robot vision*, 1st edn. McGraw-Hill Higher Education, New York
11. INCA Homepage (2015). <http://www.inca.gov.br>. Accessed Jun 2015
12. Manousaki AG, Manios AG, Tsompanaki EA, Evgenia I (2006). A simple digital image processing system to aid in melanoma diagnosis in an everyday melanocytic skin lesion unit. *Int J Dermatol*
13. OpenCV Homepage (2015). <http://opencv.org>. Accessed May 2015
14. Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern*. doi:[10.1109/TSMC.1979.4310076](https://doi.org/10.1109/TSMC.1979.4310076)
15. Raikar A, Sangani S, Hanabaratti K (2013) Diagnosis of melanomas by check-list method. In: *Fourth international conference communications and networking technologies (ICCCNT)*
16. Suzuki S, Be K (1985) Topological structural analysis of digitized binary images by border following. *Comput Vis Gr Image Process*. doi:[10.1016/0734-189X\(85\)90016-7](https://doi.org/10.1016/0734-189X(85)90016-7)