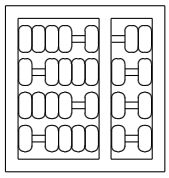


INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Skin Lesion Classification in Dermoscopy Images

William Tustumi *Helio Pedrini*

Relatório Técnico - IC-PFG-16-02 - Projeto Final de Graduação

December - 2016 - Dezembro

Skin Lesion Classification in Dermoscopy Images

William Tustumi* Helio Pedrini†

Abstract

Melanoma is one of the most aggressive types of skin-cancer. The success of the treatment is very reliant on early diagnosis. This project aims to study and develop an automatic tool to help physicians diagnose skin lesions. Initially, skin lesion images are segmented through well known image analysis algorithms, such as Otsu, Chan-Vese and Statistical Region Merging. Then, features are extracted by following the ABCD rule and second order image characteristics. From these features, a decision is made by means of a voting strategy combining Extra Tree, Decision Tree, AdaBoost, Linear Discrimination and Random Forest classifiers. Classification results achieved a success rate of 77% on the ISIC dataset, which is comparable to expert dermatologists that only had access to dermatology exams and patient history.

1 Introduction

Skin cancer is the most common type of cancer in Brazil, totaling 30% of the malignant tumors registered. From the different types of skin cancer, although melanoma corresponds to only 3% of the diagnosed cases, it is considered the most dangerous due to its high probability of metastasis. In Brazil alone, 1,547 people died in 2013 according to INCA (Instituto Nacional do Câncer), which estimates 5,670 new cases in 2016. Despite the high death ratio, the prognostic of this type of cancer is heavily improved with early diagnosis.

A common exam to diagnose skin cancer, used by specialized dermatologists, is the dermatoscopy. It consists in a non-invasive evaluation of the skin lesion surface that enhances and magnifies the skin features. The evaluation made with a dermatoscope and the skin covered with a fluid (oil, water, gel, or glycerin) to enhance the skin translucence, which allows the observation of hemoglobin, melanins and features from different levels of the skin [1].

The diagnosis success rate without a dermatoscope is about 75% when performed by an expert dermatologist; the percentage goes down when a general practitioner executes the diagnosis; however, the success rate reaches 90% when the dermatoscopy procedure is performed by a specialist [1]. Unfortunately, the number of specialists available is very limited and often not enough to cover those who need medical assistance, especially in poorer communities.

*Institute of Computing, University of Campinas, 13083-852, Campinas, SP
†Institute of Computing, University of Campinas, 13083-852, Campinas, SP

The main purpose of this project is to investigate methods for supporting automatic analysis of skin lesion images that can be used to improve diagnosis accuracy and speed, therefore, reducing the overhead of specialists and expanding the community service. Different classifiers are combined and the final performance is measured over the ISIC (International Skin Imaging Collaboration) dataset.

The remainder of the text is organized as follows. Section 2 briefly describes some techniques for segmenting the lesion images and extracting features from the segmented regions. Section 3 presents the proposed method. The experimental results are described and discussed in Section 4. Finally, some final remarks and directions for future work are outlined in Section 5.

2 Related Work

A subproblem of diagnosing a skin lesion is to segment the lesion part of the skin from the healthy part. The classification performance relies on segmentation quality, since the extracted features could be influenced by the skin area that is not part of the lesion. Many studies have used algorithms from computer vision and image analysis applied to medical image segmentation; however, skin lesion classification has its own special obstacles: the bruise can be gradually faded into the skin color, can present several different colors, or can be covered by noise hair or other objects. Subsections 2.1, 2.2 and 2.3 briefly describe the main segmentation methods used in this project.

The feature extraction follows the segmentation process. It is commonly based on characteristics used by expert dermatologists to diagnose skin cancer. A very standard and effective method is the ABCD rule, explained in Subsection 2.4.

Finally, the classification decision is produced by feeding a classification method, normally used in artificial intelligence, with the features extracted previously. Related works [2] and [3] have shown that the use of classifiers combined with the ABCD rule for feature extraction could achieve results above 70% accuracy when diagnosing skin cancer, close to expert dermatologists who had only access to medical history and dermatology inspection.

2.1 Otsu's Threshold

The threshold method proposed by Otsu [4] consists in finding a threshold value that minimizes the difference between pixels of the same segmentation group. This problem is equivalent to maximize the difference between pixels of different segmentation groups. Algorithm 1 shows the mains steps of the method.

A variation of the Otsu's threshold separates the image in sectors, such as a different threshold is calculated for each sector, which gives a better segmentation result on images where the illumination changes throughout them.

2.2 Chan-Vese Algorithm

A widely and successful segmentation method is the Chan-Vese algorithm [5], which utilizes a curve evolution that does not depend on the image gradient. Therefore, the method can

Algorithm 1: Otsu's threshold algorithm.

$var_{max} \leftarrow 0$;
$total_p \leftarrow height * weight$;
forall *intensities in* I **do**
    $total_w \leftarrow total_w + intensity * histogram[intensity]$;
forall *intensities in* I **do**
    $p_b \leftarrow histogram[intensity]$;
    $p_f \leftarrow total_p - p_b$;
    $w_b \leftarrow w_b + intensity * histogram[intensity]$;
    $m_b \leftarrow w_b / p_b$;
    $m_f \leftarrow (total_w - w_b)/p_f$;
    $var \leftarrow w_b * w_f * (m_b - m_f)^2$;
    if $var_{max} < var$ **then**
        $var_{max} \leftarrow var$;
        $threshold \leftarrow intensity$;

segment images where the edges are not well defined, which is helpful on lesion segmentation, since the contusion can gradually fade to the skin color. Zhao et al. [6] improved the Chan-Vese method of image segmentation by eliminating the necessity of one of its procedures. The improved version of the segmentation lowered the number of steps necessary for the algorithm when tested segmenting medical images.

The complete explanation on how the algorithm works is not covered in this project, because it is an algorithm based on a complex mathematical foundation that is not the focus of this work. However, it is very important to understand how the algorithm parameters influence its behavior.

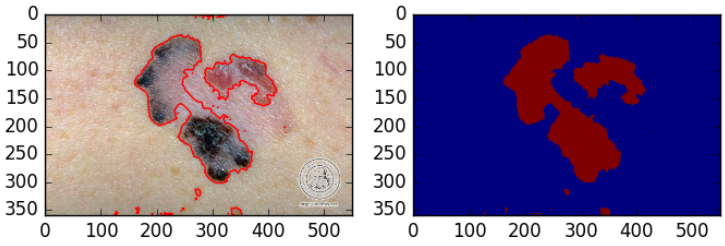
The parameter μ is the most influential on the curve behavior; it controls the penalty value correspondent to the border length, such that a high value of μ increases the penalty, encouraging merging groups or approximating segments instead of fitting the curve to all image details.

Figure 1 presents the segmentation results for different values of μ on the same lesion image. It is evident that the segmentation with $\mu = 1$ tries to follow all the lesion boundaries, hence it creates separated contours for each part of the lesion, whereas the segmentation with $\mu = 10$ approximates the contour and groups the main lesion together.

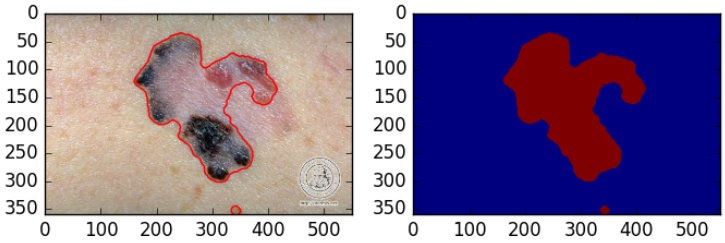
Another important parameter to the algorithm final result is the maximum number of iterations of the algorithm. This parameter is important because the algorithm is considerably costly, hence the cost-benefit has to be considered when developing the implementation. The results can suffer with the lack of iterations since every step, it tries to improve the segmentation with an approximation of the values to minimize the energy function that describes the segmentation contour. Getreuer [7] describes a method with an implicit gradient descent to find the result for minimizing the segmentation penalty.

Other factors besides the parameters can heavily interfere on the final result. For in-

Figure 1: Comparison between different values of μ on Chan-Vese segmentation algorithm.



(a) Segmentation using $\mu = 1$



(b) Segmentation using $\mu = 10$

stance, the starting point of the segmentation can influence the number of groups as well as the number of iterations that the algorithm takes to fit the curve to an object, which occurs since the penalty is complex and presents local minima. Therefore, the method might fall on a local minimum and return a bad segmentation, even if the parameter are well adjusted for the image; or it might take longer to converge to a minimum since it started too far from any acceptable segmentation.

2.3 Statistical Region Merging

The Statistical Region Merging (SRM) works in an opposite direction from the previous segmentation algorithms. It starts by considering that every pixel belongs to a different group and merges those groups depending on their pixel intensity disparity. Each group can only merge with its neighbors. The merging function is displayed in Equation 1, where two groups merge together if the function is evaluated true.

$$f(G, G') = \begin{cases} \text{true} & \text{if } \forall a \in \{R, G, B\}, |\bar{G}_a - \bar{G}'_a| \leq \sqrt{b^2(G) + b^2(G')} \\ \text{false} & \text{otherwise} \end{cases} \tag{1}$$

where $b(G)$ is a function that depends on the group size, average group size and the expected scene complexity. The main idea to understand is that the $b(G)$ function controls how the groups merge and it receives an expected scene complexity that essentially controls the number of groups the segmentation creates. The full explanation of the SRM method can be found in [8].

After executing the SRM algorithm, the segmented image has to be binarized because it can be divided into more than two different groups. The method used to group the regions into only two in our project was the Otsu's threshold, because the average color of the lesion is normally distant from the health skin color.

2.4 ABCD Rule

The main method for feature extraction addressed in this project is based on the ABCD rule proposed by Nachbar et al. [9], where the authors considered four main characteristics from a colored image lesion. Each letter stands for one of those characteristics: A for Asymmetric, B for Border, C for Color and D for Diameter.

The Asymmetric feature measures the difference between the largest diagonal of the lesion and the largest perpendicular segment that is entirely contained in the lesion. A large difference indicates that the lesion is a melanoma.

The Border feature measures how sharp the lesion border is; this measurement is calculated from the contusion perimeter length divided by its area; the perimeter has to be divided by the area because its length can be influenced by how far the photo was taken from the contusion. A sharper border indicates that the contusion is a melanoma.

The Color feature is extracted by segmenting the contusion into different groups formed by color and space proximity; after that, the group average color is calculated and compared to the colors commonly found in melanoma lesions. Table 1 presents such colors. Dermatologic studies also indicate that the presence of a large number of colors suggests that the lesion is a melanoma.

The Diameter feature measures the largest diameter of the contusion. A fifth characteristic, Evolution, is considered by dermatologists, but could not be a feature extracted in this project, since we do not have the images available of the same lesion from different views in time.

Black	**Red**	**White**
(0,0,0)	(255,0,0)	(255,255,255)
(10,10,10)	(255,50,0)	(245,245,245)
(20,20,20)	(200,0,0)	(235,235,235)
(30,30,30)	(200,50,50)	(225,255,255)
(50,40,40)	(150,0,0)	(215,215,215)
(50,50,50)	(150,50,50)	(205,205,205)
Light Brown	**Dark Brown**	**Blue and Gray**
(200,150,100)	(150,100,100)	(150,125,150)
(200,100,0)	(125,75,75)	(125,125,150)
(200,100,50)	(100,50,50)	(100,100,125)
(150,100,50)	(100,50,0)	(100,125,150)
(150,100,0)	(100,0,0)	(50,100,150)
(150,50,0)	(50,0,0)	(0,100,150)

Table 1: Colors commonly found in melanomas based on the table found on page 29 of [3].

The ABCD rule is fundamental for a large number of works in the field and is considered a solid ground to extract the image features, since it is very similar to methods used by dermatologists to diagnose lesions.

Soares [2] analyzed the ABCD rule to extract features, where an SVM (Support Vector Machine) is used to classify the images. The method achieved a classification success rate of 72%. Mahmudur et al. [10] used the same rule to extract features from lesion images, however, the authors combined different classifiers to improve the success rate, achieving accuracy of 75%.

3 Materials and Method

A diagram of the validation test is presented in Figure 2. The image set is divided into two groups the, training and testing. Both sets are filtered in order to eliminate noise and then are segmented through the same procedures. The training group of images has an additional step after the feature extraction, where the features used on the classification methods are selected. The training data is also used to calibrate the classifiers used on the validation test.

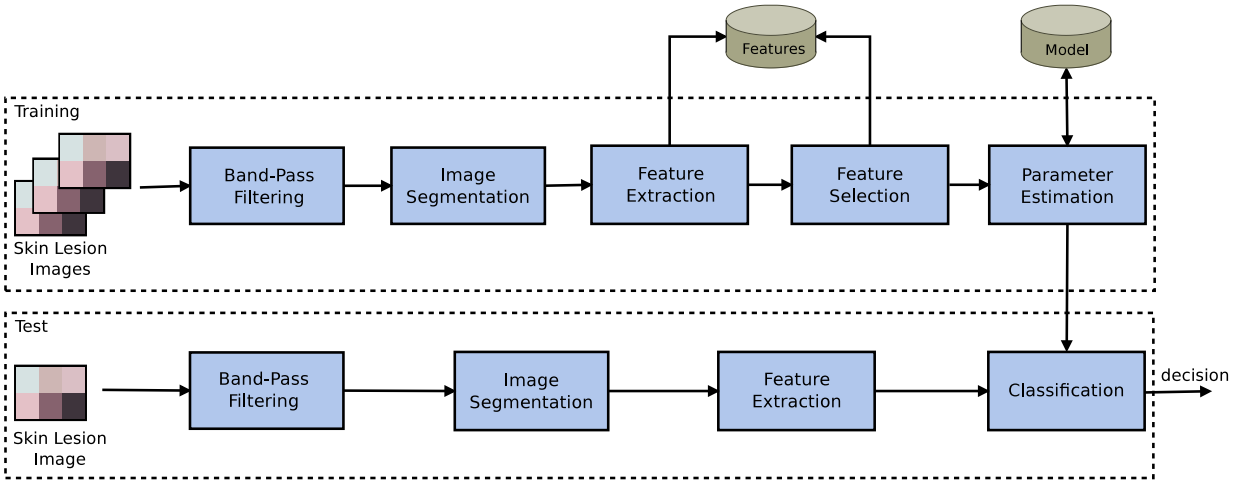


Figure 2: Main components of the developed method for skin lesion image analysis.

A software implementation of the methodology was developed using Python programming language version 2.7, along with the libraries skit-learn 1.8.0, skit-image and the modules srm [11] and chan-vese [12], respectively.

The application is divided into four main modules: image segmentation, feature extraction, lesion classification and decision result. The first module segments contusion from health skin. The second quantizes the characteristics from the lesion area. The third employes classifiers based on the extracted features. The forth stage combines all other modules. The source code can be found in [13].

The training and test data were download from [14] and [15], where 206 images were extracted from the ISDIS directory, such that 119 classified as melanoma images and 87 classified as non-melanoma images, whereas from the ISIC archive were extracted 189 images, 39 melanoma images and 150 non-melanoma images.

We applied a filter to suppress noise before performing the segmentation. The chosen filters were median of size 5×5 and Gaussian of the same window size. The median filter was chosen since it preserves edges while eliminating small noises such as hair and some skin steins. The Gaussian filter is widely used and also preserves some information of the edges, however, it blurs the image more than the median filter.

The segmentation module applies the Otsu's threshold algorithm, explained in Subsection 2.1, Chan-Vese method explained in Subsection 2.2, and SRM segmentation combined with Otsu's threshold explained in Subsection 2.3. Other algorithms were also evaluated, however, their performance were not relevant enough to be worthy mentioning in this section.

Most of the segmentation strategies mentioned previously only work on one image channel. Due to that, we chose to perform the segmentation on the Green channel of RGB color model, on gray scale images or the Hue channel of HSV color model. The performance of each segmentation method combined with a denoising filter is reported and discussed in Subsection 4.1.

Some of the segmentation methods used are heavily dependent on their calibration process since their parameters dictate how much weight they would give for different characteristics of the image. To tune their parameters, we perform an extended search on different combinations using a sample of the image pool.

A measurement of the segmentation performance is done by comparing the mask generated by the automatic segmentation algorithm with one performed by an expert dermatologist. The formula utilized to calculate the correspondence between the two masks is presented in Equation 2.

$$s = (p_b^b/(p_b^b + p_b^f) + p_f^f/(p_f^f + p_f^b))/2 \tag{2}$$

where

$$p_b^b = \text{pixels classified correctly as background}$$
$$p_b^f = \text{pixels classified wrongfully as background}$$
$$p_f^f = \text{pixels classified correctly as foreground}$$
$$p_f^b = \text{pixels classified wrongfully as foreground}$$

The score equation is composed of two parts: the first measures the relation between the background pixels that were labeled correctly (they belong to the background on the reference segmentation) and the ones that were labeled incorrectly (they belong to the foreground on the reference image), whereas the second part corresponds to the same measures but with the pixels labeled as foreground (belonging to the lesion area). The score is separated into two components, otherwise it would be influenced by the background/-foreground proportion of pixels (for instance, if an image is composed of 90% background

pixels, a segmentation that considers all the image background would achieve a score of 0.9). Thus, using Equation 2 to measure the segmentation quality bounds half the score to the image background and half to the foreground, which makes some pixel classification more important than others.

After the segmentation process is done, the program extracts features from the lesion area using the ABCD rule. Common second order features utilized on image analysis are also calculated to improve the classification method. The implementation used of the modules numpy, scipy, and scikit-image for feature extraction. The numpy module is used to calculate the mean value and the standard deviation of the color channels RGB and HSV.

The *stats* module from scipy was used to calculate the skewness value from each color channel. Finally, the modules *measure* and *feature* from scikit-image were employed. The first one was used to extract the area, perimeter, maximum diameter, eccentricity and thinness from the masks generated on the segmentation process. The second module was used to extract color (number of different color-groups) and second order features (angular second moment, energy, contrast, homogeneity and dissimilarity). The second order features are calculated from a gray level co-occurrence matrix (GLCM) generated using a function from the module *feature* from scikit-image. The matrix was created by considering the neighboring pixels upward, downward, leftward and rightward directions.

Before running the classification method, we performed a parameter refinement along with a feature selection. This eliminates redundant features and selects the best parameter combination for the classifiers. The first step is to execute a grid-search on all classifier parameters and select those that perform better when considering all the features. From these selected parameters, we select the features that influence results the most.

After the refinement process, five different supervised classifiers are executed using their selected parameters to provide a final decision. The classifiers used in this project are Extra Tree, Decision Tree, Random Forest, AdaBoost and Linear Discriminant Analysis. A combination of different learning methods was chosen since it achieves a better overall accuracy, as shown by Mahmudur et al. [10].

The combination method can use majority voting, that simply decides the result by following the decision given by the largest group of classifiers (hard voting system), or a weighted voting system that considers the certainty of each classifier response (soft voting system). The results of the different methods are discussed in Section 4.2.

4 Experimental Results

This section presents and discusses the experimental results for segmentation and classification stages through the proposed methodology.

4.1 Segmentation Results

The results reported in Table 2 show that the segmentation based on thresholding (Otsu's threshold and SRM), have improved when a filter is applied before the segmentation. However, the Chan-Vese algorithm presented a decrease in performance. The performance is heavily influenced by the color channel used in the segmentation. It is possible to observe a

distinct difference between the Otsu's threshold and Chan-Vese methods. The Hue channel presented the worst channel in any segmentation algorithm, however, its performance using Chan-Vese method was still close to the Green and Gray channels.

Table 2: Segmentation scores for Otsu's threshold, Chan-Vese and SRM.

	Otsu			**Chan-Vese**			**SRM**		
Channel	None	Gaussian	Median	None	Gaussian	Median	None	Gaussian	Median
Green	0.796	0.797	0.814	0.834	0.834	0.828	0.769	0.695	0.780
Hue	0.488	0.492	0.519	0.762	0.756	0.759	0.428	0.441	0.405
Gray	0.796	0.799	0.817	0.830	0.832	0.823	0.668	0.693	0.791

The images illustrated in Figure 3 show clearly the filters performance at eliminating hair and some small dots from the skin. Image 4(b) shows the mask generated from Image 4(a) without any filter; the white lines under the main body of the lesion correspond to the hair growing from it, whereas the small dots on the image border were created because of the skin texture. Image 4(c) shows the mask generated after applying Gaussian filter to the image; it eliminated some of the dots on the image border as well as some of the white lines from the hair. The median filter, however, was able to eliminate most of the white lines and most of the dots on the image border, as shown on Image 4(d). The influence of the filters on the Chan-Vese method was mostly negative in terms of performance or effectively non existent.

Even if the performance discrepancy was not much, the images in Figure 4 could help explain why the filters may generate a worse mask than without using any filter. Image 5(a) presents a lesion that has a fainted difference in color on its center. The correct segmentation unifies the three parts of the lesion in one because of that fainted coloration. Image 5(b) shows the mask without applying any filter; the three areas of the contusion are linked because of the proximity of those areas end the small differences of color. In Images 5(c) and 5(d), the areas are isolated since the coloration is smoothed and the algorithm cannot identify differences between the skin and the lesion center.

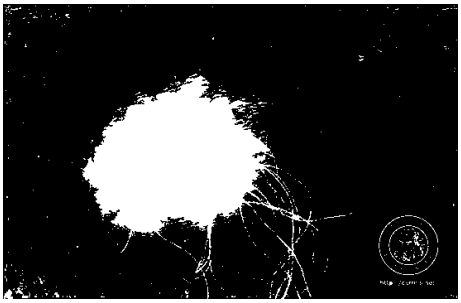
Although the average score of the segmentation was not bad, a critical problem is the difference in performance between different images. Figure 5 presents an example of a very good segmentation using Chan-Vese, whereas Figure 6 presents a poor one. The first image has sharper borders between the lesion area and the health skin. This provides a clear edge for most image segmentation schemes to delineate the lesion boundaries. Moreover, the background skin does not present much texture and different skin reflection. Additionally, the second image presents more noise on the background, a skin lesion with reflections and a health skin with a lot of texture and shadows. Therefore, the segmentation fails entirely. The score results for those images are 0.994 for the first one and 0.348 for the second one.

4.2 Classification Results

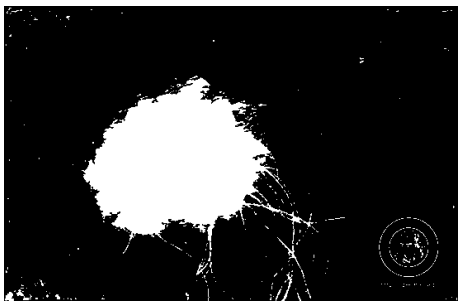
Table 3 presents the classification success percentage of the classifiers using different combinations of the database image and segmentation contours. The numbers inside the parentheses represent the highest and lowest values obtained when using these classifiers and

Figure 3: Image from ISDIS database and mask generated using Otsu's Threshold for image segmentation on gray scale images.

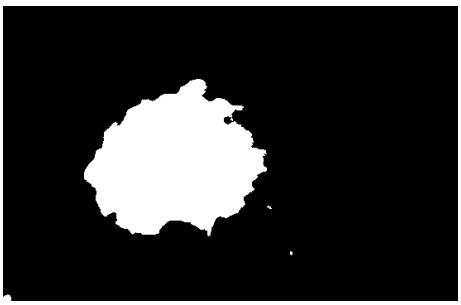
(a) Original image in RGB



(b) Mask generated without using filters



(c) Mask generated using Gaussian filter



(d) Mask generated using median filter

changing the partitioning of training and test data.

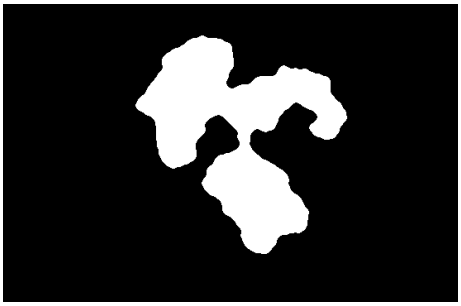
The results were improved when the image database was enlarged, which indicates the first set of images was not enough to train the classifiers properly. Furthermore, the classification process using a specialist segmentation performed better than using automatic segmentation, which indicates that the classification can be improved by better segmenting the lesion. The difference between the best and worst results decreased when using the specialist segmentation, which further corroborates the idea that improving the segmentation will improve all the steps of the classification.

Table 3: Success percentage of classifier voting.

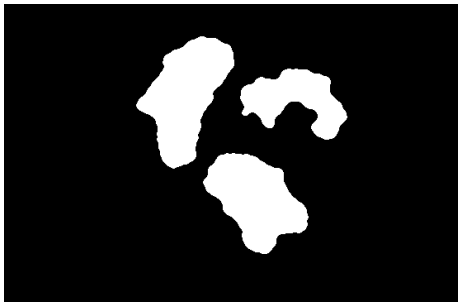
	ISDIS Archive		**ISDIS+ISIC Archive**	
Segmentation Method	voting (hard)	voting (soft)	voting (hard)	voting (soft)
SRM	59% (51-65)	58% (52-64)	74% (69 -78)	74% (72-77)
Chan-Vese	67% (58-77)	65% (58-76)	77% (71 -85)	75% (72-79)
Specialist	75% (72-77)	70% (66-78)	84% (81-85)	81% (79-85)

Figure 4: Image from ISDIS database and mask generated using Otsu's threshold for image segmentation on gray scale image.

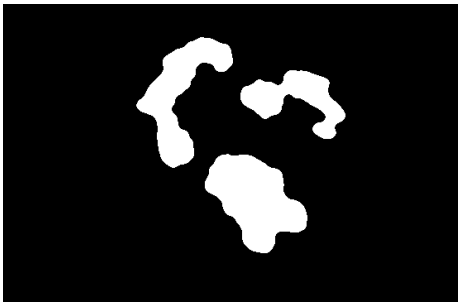
(a) Original Image in RGB



(b) Mask generated without using filters



(c) Mask generated using Gaussian filter



(d) Mask generated using median filter

5 Conclusions and Future Work

This project achieved its objective of implementing an automatic skin lesion analysis method using different segmentation and classification approaches. It also investigated the performance of the method in terms of its classification rate.

The classification success rate of 75% is comparable to those achieved by expert dermatologists without the use of dermatoscope. This result can be further improved by finding better segmentation methods for dermatology images and different features that depend less on the segmentation quality.

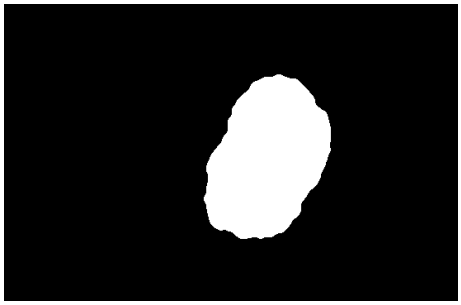
References

[1] Frange, V.M.N., Arruda, L.H.F., Daldon, P.E.C.: Dermatoscopia: Importância para a Prática Clínica. Revista de Ciências Médicas **18**(4) (2012)

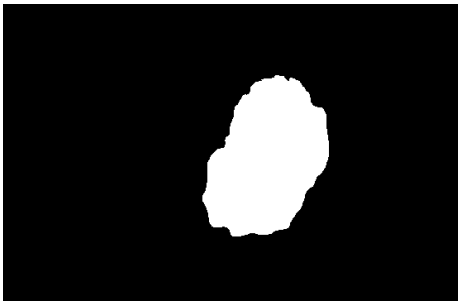
[2] Soares, H.B.: Análise e Classificação de Imagens de Lesões da Pele por Atributos de Cor, Forma e Textura utilizando Máquina de Vetor de Suporte. PhD thesis, Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Universidade Federal do Rio Grande do Norte, Natal, RN (2008)

Figure 5: Example of a proper segmentation performed automatically.

(a) Original image



(b) Mask generated using Chan-Vese



(c) Mask generated by a specialist

[3] Rosado, L.: Sistema Automático para Diagnóstico de Lesões Cutâneas baseado em Imagens Dermoscópicas. Master's thesis, Instituto Superior Técnico-Engenharia Biomédica, Universidade Técnica de Lisboa, Lisboa, Portugal (2009)

[4] Otsu, N.: A Threshold Selection Method from Gray-level Histograms. Automatica **11**(285-296) (1975) 23–27

[5] Chan, T.F., Vese, L.A.: Active Contours without Edges. IEEE Transactions on Image Processing **10**(2) (2001) 266–277

[6] Zhao, J., Shao, F., Xu, Y., Zhang, X., Huang, W.: An Improved Chan-Vese Model without Reinitialization for Medical Image Segmentation. In: 3rd International Congress on Image and Signal Processing. Volume 3., IEEE (2010) 1317–1321

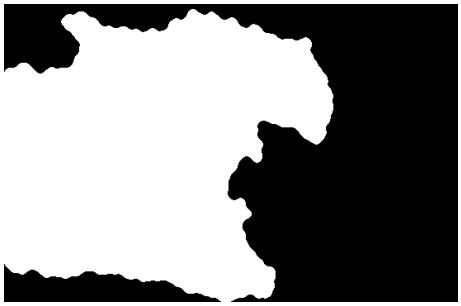
[7] Getreuer, P.: Chan-Vese Segmentation. Image Processing On Line **2** (2012) 214–224

[8] Nock, R., Nielsen, F.: Semi-Supervised Statistical Region Refinement for Color Image Segmentation. Pattern Recognition **38**(6) (2005) 835–846

[9] Nachbar, F., Stolz, W., Merkle, T., Cognetta, A.B., Vogt, T., Landthaler, M., Bilek, P., Braun-Falco, O., Plewig, G.: The ABCD Rule of Dermatoscopy: High Prospective Value in the Diagnosis of Doubtful Melanocytic Skin Lesions. Journal of the American Academy of Dermatology **30**(4) (1994) 551–559

(d) Original image

Figure 6: Example of inadequate segmentation performed automatically



(a) Mask generated using Chan-Vese

(b) Mask generated by a specialist

[10] Rahman, M.M., Bhattacharya, P., Desai, B.C.: A Multiple Expert-based Melanoma Recognition System for Dermoscopic Images of Pigmented Skin Lesions. In: 8th IEEE International Conference on BioInformatics and BioEngineering, IEEE (2008) 1–6

[11] SRM: (visited on 11/16/2016) `http://www.lix.polytechnique.fr/~schwander/darcs/python-srm/`.

[12] Morphological Snakes for Image Segmentation and Tracking: (visited on 11/16/2016) `https://github.com/pmneila/morphsnakes`.

[13] Code for Skin Lesion Classification: (2016) `https://bitbucket.org/whatust/tcc/overview`.

[14] ISIC: (visited on 11/16/2016) `https://isic-archive.com/`.

[15] ISDIS: (visited on 11/25/2016) `http://isdis.net/isic-project/`.