

A machine learning approach for the recognition of melanoma skin cancer on macroscopic images

Jairo Hurtado, Francisco Reales

Pontificia Universidad Javeriana, Electronics Department. School of Engineering, Bogotá, Colombia

Article Info

Article history:

Received Aug 20, 2020

Revised Jan 02, 2021

Accepted Jan 20, 2021

Keywords:

Artificial intelligence

Image processing

Machine learning

Melanoma

Skin cancer

ABSTRACT

In the last years, computer vision systems for the detection of skin cancer have been proposed, especially using machine learning techniques for the classification of the disease and features based on the ABCD dermatology criterion, which gives information on the status of the skin lesion based on static properties such as geometry, color, and texture, making it an appropriate criterion for medical diagnosis systems that work through images. This paper proposes a novel skin cancer classification system that works on images taken from a standard camera and studies the impact on the results of the smoothed bootstrapping, which was used to augment the original dataset. Eight classifiers with different topologies (KNN, ANN, and SVM) were compared, with and without data augmentation, showing that the classifier with the highest performance as well as the most balanced one was the ANN with data augmentation, achieving an AUC of 87.1%, which saw an improvement from an AUC of 84.3% of the ANN trained with the original dataset.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Jairo Hurtado

Electronics Department

Pontificia Universidad Javeriana. Bogotá, Colombia

Email: jhurtado@javeriana.edu.co

1. INTRODUCTION

The high level of sun exposure, low use of sunscreens by people, and some environmental factors lead to an increase in the number of disorders and skin diseases, including cancer. There are three main types of skin cancer: basal cell carcinoma, squamous cell carcinoma, and melanoma, being melanoma the most lethal one [1]. Rural populations in the tropics, especially in mountainous areas are particularly affected by this disease due to the exposition to solar radiation products of their lifestyles, skin color, and geographical location, since the UV radiation increases between 10% and 12% for each kilometer of altitude [2]. Also, as the COVID-19 pandemic has caused limited physical access to health-care providers, this can generate further delay treatment of melanoma producing devastating consequences for the patients [3]. For this reason, the adoption of computational tools in medicine is arising [4].

Melanoma has been an illness of public concern due to the rapid increase of 25.9 % between 2006 and 2016 [2] and the World Health Organization predicts that in the next two decades, the number of people diagnosed with skin cancer will be double [5]. So that, it can be appreciated the usefulness of an algorithm that identifies malignant lesion patterns and suggests that the person go immediately to a specialist, because, if it is diagnosed early, the chance of surviving is about 95% [6]. Besides, automatic diagnosis has shown to overcome dermatologists when recognizing either malignant and benignant lesions or a particular type of

lesion [7]. Marco Albrecht *et al.* studied different computer methods for diagnosis and modeling of melanoma showing the helpfulness of the melanoma pattern recognition systems in order to start early treatment.

In the recent years, systems oriented to the automated diagnosis of skin cancer through images have been proposed [1], [7]-[17]. Variations depends especially on the type of image that is used as input and the architecture of the system. Firstly, there are overall three types of images used for this purpose. Macroscopic images that are lesions taken from standard cameras, dermatoscopic images where the images are taken using a device called dermatoscope which magnifies the skin lesion making malignant patterns more visible for the dermatologist [5] and finally the least used Histo-pathological images, which are photos of the disease using microscopic examination of a biopsy [18]. So that, while a system that works on macroscopic images may be more useful for common people, the amount of images in datasets of macroscopic lesions is very limited. The opposite happens with dermatoscopic images where there are many publicly available datasets with an amount of samples of the order of thousand images. With regards to the architecture, it is used traditional machine learning with hand-crafted features or deep learning where the features are calculated automatically.

In this paper an algorithm for detecting malignant patterns in a skin mole using traditional machine learning and hand-crafted features is proposed, counting with a pre-processing which reduces the shadows in the image produced by the circularity of some parts of the body. Secondly, the skin mole is segmented using the algorithm of unsupervised learning: Gaussian Mixture Model. After that, 70 features based on a dermatological criterion, which is used to diagnose melanoma skin cancer, are calculated, and finally, a classification is performed. The main contributions of this paper are: (i) The implementation of a novel malignant pattern recognition system that works on macroscopic skin lesions images. (ii) The comparison of the performance of the Gaussian mixture model to segment different types of skin lesions. (iii) A study on the impact of the Smoothed Bootstrap data augmentation method on the performance of different topologies of classifiers. (iv) A comparison of various state-of-the-art systems with different architectures and type of input images.

2. RESEARCH METHOD

To detect malignant patterns on skin lesions, the system is based on a medical criterion called the ABCD rule, this is one of the most used methods whose acronym refers to the four parameters used in the clinical dermatological diagnosis. These are Asymmetry, Border, Color, and structural differences [6]. **Asymmetry (A):** It is generated by the uncontrolled growth of the lesion, because of higher levels of melanin in different regions and tends to have an irregular shape. **Borders (B):** Melanocytic lesions have irregular borders. In contrast, benign lesions tend to have borders that fade smoothly and are symmetric. **Color (C):** It is related to the excess melanin under the surface of the lesion, causing a different pigmentation in a specific region. **Dermoscopic structures (D):** It refers to the generation of holes, points, cells, and inhomogeneity (texture) that indicates more melanin in a given region. The ABCD rule has been tested in multiple studies, which have documented its successful diagnostic accuracy in clinical practice. Also, has been confirmed with digital image analysis [19]. However, it is a medical criterion that only can be applied to pigmented lesions, which are lesions that look like spots. So that, the ABCD rule can not be applied to basal cell carcinoma nor squamous cell carcinoma [1]. For this reason, the system uses the ABCD rule aiming to recognize only benignant lesions and melanoma.

In the implementation, it was used the Dermatology Education atlas [20], which contains 173 images of macroscopic skin lesions of two types, melanoma (84) and benignant (89), with sizes from 154 by 186 to 1129 by 1241. This dataset was used to train the system. However, it is clear that it is not as large to ensure statistical significance, however, since datasets of macroscopic images tend to be small, previous works have had to deal with these situations with methods such as data augmentation [8]-[11]. In Figure 1 the block diagram of the proposed system is shown. The entire system was implemented in Python, using the OpenCV library for the pre-processing step as well as for the feature extraction. On the other hand, for the classification, it was used Scikit-learn and Tensorflow 2 libraries. Each block of the diagram of Figure 1 is explained below.

2.1. Pre-processing

In this block, the shadows, caused by the curvature of some body parts, that can affect the system performance are attenuated. Since there could be shadows that resemble the color of the skin mole, making a nonrecognition between shades and mole, Figure 2. To correct this problem, another image, obtained from the regression of the values near the corners of the original image, is created.

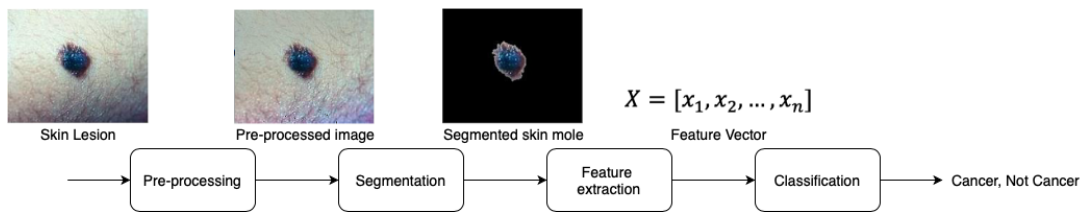


Figure 1. Block diagram of the system



Figure 2. Inaccurate segmentation caused by shadows

To use this method of attenuation of shadows, there are two assumptions. The shadows change smoothly and the mole is located in the center of the image. The first assumption ensures that a two polynomial degree adequately fits the shadows of the image. On the other hand, the second assumption ensures that samples can be taken at the corners of the image without touching the mole, samples that are subsequently used in the regression. The data taken for the regression is obtained from the channel V (Value) of the HSV color space and the two-degree polynomial is presented as shown in (1):

$$z(x, y) = p_1x^2 + p_2y^2 + p_3xy + p_4x + p_5y + p_6 \tag{1}$$

The six constants that minimize the squared error given as shown in (2) have to be found, where (i, j) are the indexes of all samples in the corners 'S'; Shown in Figure 3, V is the value channel of the space HSV and Z is the quadratic function that will be found. The result is shown in Figure 4.

$$l\epsilon = \sum_{j \in S} \sum_{i \in S} [V(j, i) - z(x_j, y_i)]^2 \tag{2}$$



Figure 3. Samples are taken from the corners used in the regression

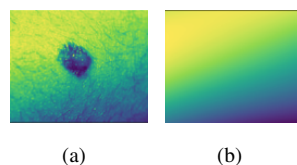


Figure 4. These figures are; (a) value component of the HSV space; (b) image after regression

To attenuate the shadows, the value component of the HSV space obtained above is divided by the quadratic polynomial found and is multiplied by the ratio of the average of the V values with the average of the values of V / z function as shown in (3). Finally, after changing the channel V of the original image in the HSV space for the new one found, it is passed to the RGB color space, Figure 5.

$$V_{new}(x, y) = \frac{\mu_V}{\mu_{V/Z}} \frac{V(x, y)}{z(x, y)} \quad (3)$$

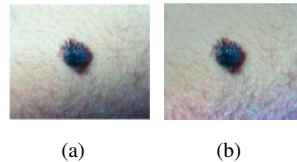


Figure 5. These figures are; (a) original image; (b) pre-processed image

2.2. Segmentation

The purpose of this step is to detect the skin mole automatically based on the color distribution of the image. Because images of skin lesions have near two clusters, light and dark colors, equivalent to background and skin lesion respectively, the Gaussians mixture model (GMM) can adequately describe the color distribution of the image and get the parameters of each of the two clusters in order to perform a pixel-wise classification. Also, GMM has shown to be capable of recognizing skin diseases with satisfactory efficiency [21].

Another reason to choose the Gaussian Mixture Model is that color-based clustering has been compared to other methods such as Graph-Cut Segmentation and Otsu, showing the best classification accuracy [2]. Besides, Pedro Pereira *et al.* [3] has compared the performance of 39 segmentation methods across three different large datasets concluding that these methods: Local Binary Patterns Clustering, Wu Quantifier, and Color Based Clustering had the best overall performance.

After finding the two clusters, the one with the largest area is classified as background and the other one as the lesion. So that, the pixels of each cluster are labeled as 0 and 1 respectively, generating the segmentation mask, shown in Figure 6 (a), which after filling blackheads and dilate the image show in Figure 6 (b), Figure 6 (c) is generated, which gives shape information, and Figure 6 (d) is obtained by multiplying the mask by the original image. The first is useful for evaluating the asymmetry of the mole and the second it can be evaluated, borders, variation in color, and texture presence.

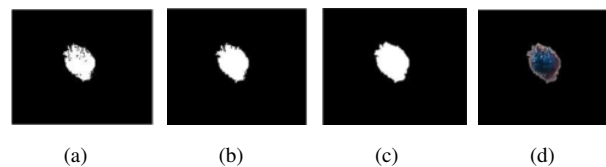


Figure 6. These figures are; (a) GMM Clustering, (b) Filling holes, (c) Dilated mask, (d) Skin mole segmented

To measure the accuracy of the proposed segmentation method on a different type of skin lesions that apparently would be difficult for the system to recognize, such as lesions with high color variation (nevus spilus, nevus repigmented, and some melanomas), Café-au-lait macule, which tends to have blurred shape, and lesions containing hair, see Figure 7, the Border Error (BE), shown in (4), which has been used in previous works to compare the segmentation efficiency [22], is calculated. Where SM is the segmentation mask, calculated automatically through GMM, and GT is the ground-truth which was hand-labeled from the dataset. So that, the BE measures the percentage of the non-overlapping area between the segmentation mask and the ground-truth.

$$lBE(SM, GT) = \frac{Area(SM \oplus GT)}{Area(GT)} \quad (4)$$

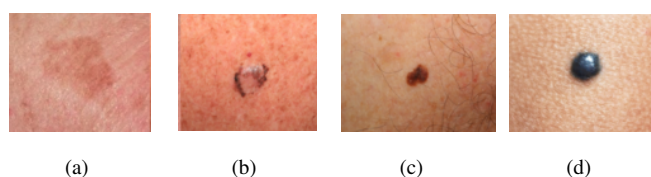


Figure 7. These figures are; (a) Café-au-lait macule; (b) Lesion with color variation; (c) Containing hair; (d) Lesion with color uniformity and sharp borders

Table 1 shows the average BE for the four different lesions of Figure 7, where the lower the Border Error the more accurate the segmentation is. Therefore, it suggests that the Café-au-lait macule is the type of lesion that the segmentation system works better with. On the other hand, images that contain hairs are not correctly segmented, due to the similarity in color between the skin lesion and hair. For this reason, nine haired samples were removed from the original dataset. So that, the classifiers were trained with 164 samples, 85 benign lesions and 79 malignant.

Table 1. Border Error (BE) for different type of lesions

Measure	Café-au-lait macule	Containing hair	With color variation	With color uniformity
BE(%)	31.1	770.2	67.1	49.0

2.3. Features based on the ABCD criterion

The skin lesion characterization is made using the criterion of the ABCD, as it gives information on the state of pigmented skin lesions using static parameters. The process to obtain these features is shown below.

2.3.1. Features based on asymmetry

To quantify the mole asymmetry, it is considered Figure 6 (c), and compared to other geometric figures, Figure 8, as the ellipse; in purple (proposed in this paper), as if the contour or shape of the mole resembles an ellipse is less likely to be malignant. A comparison with the bounding box is also made; in red, to have the dimensions of the lesion, and the Convex Hull; blue dotted. Besides, the area of the quadrants of the mole must be the same if the mole is completely symmetrical.

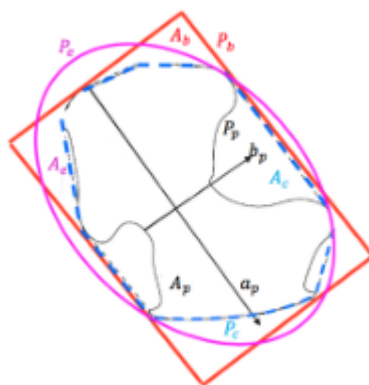


Figure 8. Geometric parameters of the skin lesion

The parameters used are, b_p and a_p ; minor and major axes. A_p , A_c , A_b and A_e ; areas of the lesion, convex hull, bounding box and ellipse, respectively, and P_p , P_c , A_b and P_e ; perimeter of the lesion, convex hull, Bounding Box, and ellipse respectively. On the other hand, A_1 and A_2 represent the areas of each division of the axis a_p . Similar to B_1 and B_2 for the axis b_p . The asymmetry features are presented in Table 2. For the features not to depend on the size and resolution of the image, those having units of area were divided by the area of the Bounding Box A_b and with length units were divided by its perimeter P_b .

Table 2. Features of asymmetry

Lesion area	Solidity	Equivalent diameter	Convex Hull Area
A_p	$\frac{A_p}{A_c}$	$\sqrt{\frac{4A_p}{\pi}}$	A_c
Circularity $\frac{4\pi A_p}{P_p^2}$	Lesion perimeter P_p	Aspect ratio $\frac{b_p}{a_p}$	Convex Hull perimeter P_c
Aspect ratio $\frac{b_b}{a_b}$	Elliptic area rate $\frac{A_p}{A_e}$	Elliptic rate $\frac{P_p}{P_e}$	Rate of areas b_p $\frac{(B1-B2)}{A_p}$
Rate of areas a_p $\frac{(A1-A2)}{A_p}$	Rate form b_p $\frac{B1}{B2}$	Rate form a_p $\frac{A1}{A2}$	

2.3.2. Texture variation, darkness and color information image

For features based on borders, color uniformity, and dermoscopic structures, a new image (I^N), with three channels, obtained from the original image is created 4. The first channel provides information on the texture variation, the second on the skin darkness, and the third, information on the color variation, I_i^N ($i = 1, 2, 3$). In order to calculate the texture variation channel (I_1^N), the brightness image L is obtained from as shown in (5), where the three channels of the original image I^C are averaged.

$$lL(x, y) = \frac{\sum_{i=1}^3 I_i^C}{3} \quad (5)$$

The texture $\tau(x, y, \sigma)$ is defined by as shown in (6) where $S(x, y, \sigma) = L(x, y) * G(\sigma)$ is the brightness smoothed by a Gaussian filter with standard deviation σ .

$$l\tau(x, y, \sigma) = \frac{L(x, y)}{S(x, y, \sigma)} - L(x, y) \quad (6)$$

The texture image $\tau(x, y, \sigma)$ is calculated for different values of $\sigma = (\sigma_1, \dots, \sigma_N)$, and it is selected for each pixel the highest texture among all scales, this is shown in (7). For this paper, the standard deviation was chosen as $\sigma = (1, \frac{11}{7}, \dots, \frac{43}{7})$ with a window of 7σ by 7σ . The values of this parameter were suggested by Cavalcanti *et al.* [10] due to the average size of the images in the dataset [5].

$$\tau(x, y) = \max_{\sigma} [\tau(x, y, \sigma)] \quad (7)$$

The texture variation channel will be obtained from the normalization of $\tau(x, y)$, shown in (8), where the minimum among all values is subtracted from the texture image, and then divided by the difference between the maximum and minimum, causing that all data is in the interval [0, 1]. The original image as shown in Figure 9 (a) and the result of I_1^N is shown in Figure 9 (b).

$$I_1^N = \frac{\tau(x, y) - \tau_{min}}{\tau_{max} - \tau_{min}} \quad (8)$$

For the darkness information image (I_2^N), since healthy skin tends to be reddish, when the red channel of the original image is brighter means that it is part of the fund and if it is darker injury. So that, the darkness is given by the complement of the red channel of the original image, as shown in (9). The result of I_2^N is shown in Figure 9 (c).

$$lI_2^N = 1 - I_1^C \quad (9)$$

In the color information channel (I_3^N), the three color channels of the original image I^C are represented in a single channel I_3^N using PCA (Principal Components Analysis), and the absolute value is taken creating the image $C(x, y)$. which is then normalized as shown in (10). The result of I_3^N is shown in Figure 9 (d).

$$I_3^N = \frac{C(x, y) - C_{min}}{C_{max} - C_{min}} \quad (10)$$

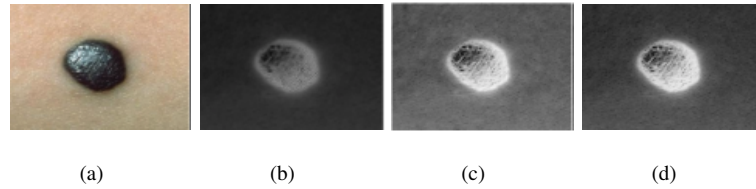


Figure 9. These figures are; (a) original image, (b) texture variation, (c) darkness information, (d) color information

2.3.3. Features based on borders

To quantify the variation of the intensity in the borders of the lesion, the magnitude of the gradient vector has to be calculated, as it measures if the border is sharp or soft. First, the border is obtained by subtracting the mask contracted from the mask dilated, it is important that the border is thick enough to ensure that it contains the change from lesion to the background. Then the gradient it is calculated for the values on the border B . So that, the features containing information of the intensity variation of the image in the borders are calculated from the mean and variance of the values of the gradient magnitude, as shown in (11) and (12).

$$l\mu = \frac{1}{N} \sum_{(x,y) \in B} |\nabla I(x, y)| \quad (11)$$

$$\sigma^2 = \frac{1}{N} \sum_{(x,y) \in B} (|\nabla I(x, y)| - \mu)^2 \quad (12)$$

Where $\nabla I(x, y)$ is the gradient vector for the scalar field $I(x, y)$. Due to the dependency of the gradient vector magnitude on the skin color of the original image, the value of the magnitude would be less if the skin color was darker. For this reason, the previously found channels I_1^N , I_2^N and I_3^N are used, due to the less dependency on this parameter. To approximate the gradient of the image, the Sobel operator is calculated, Figure 10. The color information channel is also divided into eight pieces, whose principal axes are oriented in the direction of the mole, this is ensured by using the eigenvectors, and the average and variance of the mean of the gradients are obtained in each fragment that belongs to the border of the lesion, thus obtaining two new features for each channel. This idea is similar to the proposed in Figure 11. The features based on the borders of the skin mole are shown in Table 3.

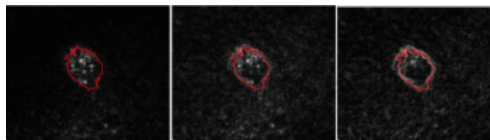


Figure 10. Borders and Sobel operator for I_1^N , I_2^N e I_3^N

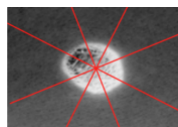


Figure 11. Pieces of the color information channel

Table 3. Features based on borders

Name	Notation	Quantity
Average gradient for each channel.	$\mu_{\nabla I_i^N}$	3
Gradient variance for each channel.	$\sigma_{\nabla I_i^N}$	3
Average mean gradient for the eight fragments per channel.	$avg(\mu_{\nabla I_i^N})$	3
Variance of average gradient for the eight fragments per channel.	$var(\mu_{\nabla I_i^N})$	3

2.3.4. Features based on color uniformity

Color features are obtained from the images of the segmented lesion, shown in Figure 6 (d), and the color information channel shown in Figure 9 (d). In order to remove color noise, both images are smoothed using a Gaussian filter. Table 4 is also used where the tones of interest are shown. These tones are the most common colors found in different types of lesions and allow the system to recognize particular color patterns. So that, six counters (C) that are increased depending on the Euclidean distance of a given pixel color to the tones in Table 4 are proposed. Thus, if the color of a given pixel is the closest to one of the tones of interest, the respective tone counter is incremented. Also, it is proposed in this paper, adding the mean and variance of a new image obtained by calculating the Euclidean distance between the color of the original image for each pixel and the tones of interest. These new parameters give information on how far the colors of interest are on average and how much they deviate from the values of the original image.

Table 4. Tones of interest in skin lesions. Alcon *et al.* [23]

Colour	Counter (C)	Red channel	Green channel	Blue channel
White (W)	c_W	1	1	1
Red (R)	c_R	0.8	0.2	0.2
Light brown (LB)	c_{LB}	0.6	0.4	0
Dark Brown (DB)	c_{DB}	0.2	0	0
blue-gray (BG)	c_{BG}	0.2	0.6	0.6
Black (BBL)	c_{BL}	0	0	0

In order to have information about the non-uniformity of the mole distribution color, features that depend on the location of the given color distribution in the image are proposed. For this, the channel of color information I_3^N is divided into eight pieces, shown in Figure 11, whose main axes are oriented in the mole direction (this is ensured by using the eigenvectors) and the average of the mean of the values of intensity in each fragment belonging to the lesion and its variance are calculated.

So that, in Table 5, the features that give information on the color distribution and uniformity of the skin mole are shown. Being R, G, and B the color channels of the original image whose pixels are part of the lesion. Likewise, for I_3^N , there will be taken only the values belonging to the lesion. On the other hand, the mean and variance of the data are represented with the symbols μ and σ respectively.

Table 5. Color distribution features

$max(R)$	$max(G)$	$max(B)$	$min(R)$	$min(G)$	$min(B)$
$\mu(R)$	$\mu(G)$	$\mu(W)$	$\sigma(R)$	$\sigma(G)$	$\sigma(B)$
$\mu(I^c - W)$	$\mu(I^c - R)$	$\mu(I^c - LB)$	<i>font-size</i> : $14px\mu(I^c - DB)$	$\mu(I^c - BG)$	$\mu(I^c - LB)$
$\sigma(I^c - W)$	$\sigma(I^c - R)$	$\sigma(I^c - LB)$	<i>font-size</i> : $14px\sigma(I^c - DB)$	$\sigma(I^c - BG)$	$\sigma(I^c - LB)$
$max(I_3^N)$	$\mu(I_3^N)$	$\sigma(I_3^N)$	$\mu(R)/\mu(G)$	$\mu(R)/\mu(B)$	$\mu(G)/\mu(B)$
c_{LB}	c_R	c_{DB}	c_{BG}	c_{BL}	c_W
$\mu(I_{3,\mu,8}^N)$	$\sigma(I_{3,\mu,8}^N)$				

2.3.5. Features based on dermoscopic structures

Although dermoscopic structures can only be measurable using a dermatoscope, which is a device that enables dermatologists to have a closer view of the skin lesion, differences between benign and malignant skin moles can be measured through macroscopic images using features based on the skin mole texture [4]. For this reason, the texture channel I_1^N , Figure 9 (b), which gives information of the mole rugosity (Holes, points and inhomogeneity), is used in order to obtain four more features which are the maximum, minimum, mean, and variance of the texture variation channel calculated inside the lesion. These features are shown in Table 6.

Table 6. Texture Features

Feature	Description
$\min(I_1^N)$	Minimum texture
$\max(I_1^N)$	Maximum texture
$\text{mean}(I_1^N)$	Average texture
$\text{var}(I_1^N)$	Variance of the texture channel

2.4. Data Augmentation

In the above processes, 70 features based on the medical criterion of the ABCD were calculated. With this, the system has enough information related to the status of the skin mole to perform a classification with labels (cancer, not cancer). However, as the dataset used for training and testing only has 164 samples, there is not supported statistical significance. For this reason, the smoothed bootstrap data cloning is used. This (re)sampling technique is based on the idea that new data samples can be added if these samples are distributed according to the same probability density as the real data set, resulting in a greater statistical significance [24]. In order to clone data through the smoothed bootstrapping method, a Gaussian distribution is a reasonable assumption [10]. This is adding to each sample a Gaussian error with zero mean and standard deviation ten times smaller than the deviation of each feature. Each new sample is given by as shown in (13), where $\vec{Y}_{k,n}, \vec{X}_n, \vec{G}_k \in R^{70}$ and \vec{G}_k is a random vector with normal distribution, $\vec{Y}_{k,n}$ is the new sample which is obtained from adding Gaussian noise to the current sample \vec{X}_n .

$$\vec{Y}_{k,n} = \vec{X}_n + \vec{G}_k \quad (13)$$

This data augmentation technique has been used previously by Cavalcanti *et al.* [4]. However, in their work the entire dataset was augmented five times, making that there was always in the test partition a similar sample in the training partition. This would make k-nearest neighbors the classifier with the best accuracy, nevertheless, those results would be biased due to the extended dataset. To avoid this situation, only the training partition is augmented two times and it is tested on the remaining samples. Also, the results of this method for different classifiers as well as the results without data augmentation are studied.

2.5. Classification system

For this research area, there are specially three classifiers used, K-nearest neighbors (KNN), artificial neural network (ANN), and support vectors machine (SVM). These classifiers have been used in previous works showing a dermatologist level accuracy [2], [4], [6], [7], for this reason, the performance of the three said classifiers, both with and without data augmentation, are compared in order to see how the behavior of these classifiers change with the Smoothed Bootstrapping Data Cloning.

3. RESULTS AND DISCUSSION

Performance metrics of the classification systems are presented in Table 7. These results are generated using a 10-fold cross-validation, showing the accuracy; which is the overall system performance, specificity; ability to recognize malignant lesions and sensitivity, ability to recognize benign lesions. Specificity becomes the most important parameter to consider because if the system does not suggest to visit a specialist, since the image is a malignant lesion, it can endanger the patient. Therefore, specificity should be the highest possible. On the other hand, it is better if the sensitivity has a high value, however, it does not represent an imminent danger to the patient.

Table 7. Comparison of the performance measures of different classification systems

Classifier	Specificity (%)	Sensitivity (%)	Accuracy (%)
SVM with kernel grade 5	15.9	100	70.6
SVM with kernel grade 5 augmented	15.9	100	70.6
KNN, Euclidean distance, k=2	37.6	83.1	66.7
KNN, Euclidean distance, k=2 augmented	56.3	76.7	66.7
KNN, Mahalanobis distance, k=2	47.4	90	72.5
KNN, Mahalanobis distance, k=2 augmented	71.7	87	80.4
ANN 200-180-150-100-50-20-1	80.2	94.4	88.2
ANN 200-180-150-100-50-20-1 augmented	86.9	87.8	86.3

Table 7 shows that the neural network has the best accuracy and with the original dataset the system achieves a high sensitivity level. In contrast, when the dataset is augmented the sensitivity decreases while the specificity increases making it a more balanced classifier. The problem with the measures taken in Table 7 is that while it gives an estimate of the classifier performance in one (specificity, sensitivity) point, it is better to compare the entire curve. This method to measure the performance is called the ROC curve, which stands for Receiver Operating Characteristic Curve and shows the dependence between sensitivity and specificity for each classifier, making possible a comparison not only for one point but for the entire spectrum of values of sensitivity and specificity. To compare the performance of the classifier using the ROC curve, the AUC is used, which is the area under the ROC curve, and measures the ability of the system to recognize benign lesions as benign and malignant lesions as malignant. So that, the closer the AUC to 100% the more accurate the system is.

The ROC curve and AUC for the three topologies compared (SVM, KNN, and ANN) are shown in Figure 12, which suggests that the overall performance of the neural network and SVM increases with data augmentation. However, being consistent with Table 7, the neural network has the best performance among the topologies compared. On the other hand, both Table 7 and Figure 12 show that the KNN tends to work better with the Mahalanobis distance than the Euclidean distance for this specific task. This is important because most approaches to skin cancer classification that use KNN are based on the Euclidean distance while they could increase their performance using a Mahalanobis distance-based KNN. On the other hand, Figure 12 makes clear that when the KNN with Mahalanobis distance is based on augmented data, the overall performance drops, the sensitivity increases and the specificity decreases.

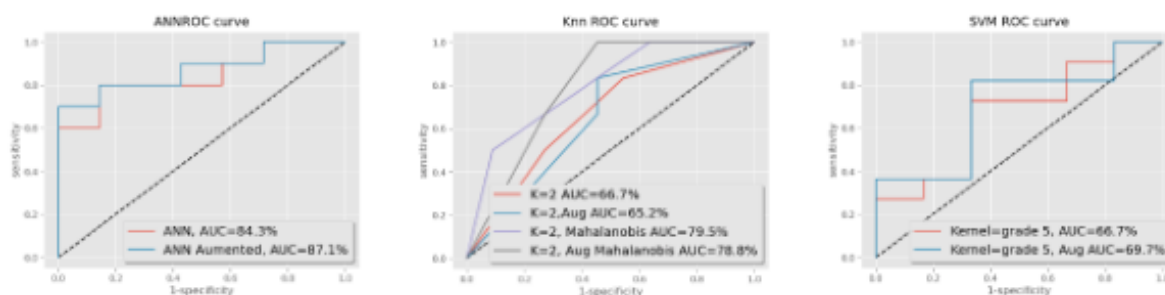


Figure 12. ROC Curves for the three topologies: ANN, KNN and, SVM

On the other hand, over the years traditional machine learning and deep learning systems have been proposed. However, while it is not possible to have a direct comparison of these systems' performance, since they are trained on different datasets, comparing the accuracy with previous systems gives information on whether the proposed approach is viable or not. Also, makes it possible to recognize patterns present among different systems, datasets, architectures, and types of images. So that, a comparison of different state-of-the-art systems is shown in Table 8. With regards to traditional machine learning usually counting with pre-processing, segmentation, feature extraction, and classification. Yuheng *et al.* [9] in 2019 proposed a system based on an SVM classifier and 143 macroscopic images using polarization speckle imaging which allowed the system to increase performance. Another interesting approach was made by Verosha *et al.* [8] in 2019. Their system used 170 macroscopic images of skin lesions for training and testing and implemented both the KNN with $k=5$ and SVM classifiers. Also, traditional machine learning has been used in histo-pathological images, for example, Takruri *et al.* [12] implemented a PSO-SVM hybrid system that optimizes SVM parameters and then performs a classification. On the other hand, deep learning systems working on the recognition of skin cancer have seen important improvements over the years. Pacheco *et al.* [1] proposed that the input of a convolutional neural network (CNN) was not only the image of the skin lesion but also clinical information such as age, location of the lesion and if it had bled, improving the average accuracy in over 7%. Maron *et al.* made a comparison between 112 dermatologists and a deep learning system, showing that the convolutional neural network was more accurate in both binary class problem (Melanoma, benign) and multiclass (Type of skin disease). Table 8 shows the results obtained by the 112 dermatologists in order to have a reference to compare the systems. Adegun *et al.* [15] proposed an encoder-decoder architecture which gave the system the possibility of extracting

strong and robust features. Finally, one very interesting approach is made by Almaraz-Damian *et al.* [17] who combined both hand-crafted features based on the ABCD rule and deep learning features.

Table 8. Comparison of results with the literature

System	Specificity (%)	Sensitivity (%)	Accuracy (%)	Type of image
(ANN) Proposed	86.9	87.8	{ 86.3	Macroscopic
(SVM) Verosha <i>et al.</i> (2019) [2]	76	74.3	75.3	Macroscopic
(KNN, k=5) Verosha <i>et al.</i> (2019)[2]	86	58.6	74.7	Macroscopic
(SVM) Yuheng <i>et al.</i> (2019) [4]	85.7	87.6	86.3	Macroscopic
(PSO-SVM) Mmahmoud <i>et al.</i> (2019) [6]	80.2	94.1	87.1	Histopathological
(CNN-SVM) Almaraz-Damian <i>et al.</i> (2020) [7]	90	86.4	89.7	Dermatoscopic
(CNN) Pacheco <i>et al.</i> (2020) [8]	76.4	77.9	76.4	Macroscopic
(CNN) Maron <i>et al.</i> (2019) [9]	84	90	86.9	Dermatoscopic
(CNN) Jianu <i>et al.</i> (2019) [10]	89	72	81	Dermatoscopic
(CNN) Adegun <i>et al.</i> (2020)[11]	96	97	95	Dermatoscopic
112 Average dermatologists accuracy [6]	74	60	67	Desmatoscopic
Average for macroscopic images	82.2	77.2	79.8	
Average for dematoscopic images	86.6	81.1	83.9	

From Table 8 it can be noticed that the proposed system has an overall accuracy above the average, proving the feasibility of the algorithm. Moreover, all the systems compared have an specificity, sensitivity and accuracy that outperforms the overall dermatologist accuracy obtained in [7] and the performance measure given in [25]. Besides, It is known that dermatologist accuracy may be reduced from (75-97)% for dermatoscopic images to (65-80)% for macroscopic images [12]. Table 8 suggests that the same applies to machine learning and deep learning systems where the average performance of the classifiers that work on dermatoscopic images is higher than the results with macroscopic images.

4. CONCLUSIONS

Firstly, it is deduced that the system is a reliable tool because it has an 86.3% accuracy which exceeds the average performance of an expert dermatologist. With regards to the data augmentation, it is concluded that this technique has a different impact depending on the classifier topology, decreasing the performance of the KNN classifier but increasing it in SVMs or ANNs. Nevertheless, it is shown that the data augmentation tends to balance the classifier specificity and sensitivity. Likewise, data cloning through Smoothed Bootstrap may be preferable to repeat images in the dataset with color and spatial transformations, since the computing cost of the first is relatively slow because it works directly on the feature vector. Another important aspect from the results is that even if the classifier with the best results in this paper is the ANN, the KNN classifier with Mahalanobis distance works fairly well for this task since it tends to recognize more effectively the clusters present in the data than the Euclidean distance and improves its results.

ACKNOWLEDGEMENTS

We would like to thank Pontificia Universidad Javeriana (www.javeriana.edu.co) and Minciencias (www.minciencias.gov.co), Bogotá, Colombia, for supporting this paper.

REFERENCES

- [1] Andre G. C. Pacheco, and Renato A. Krohling, "The impact of patient clinical information on automated skin cancer detection," *Computers in Biology and Medicine*, vol. 116, pp. 103545, Jan. 2020, doi: 10.1016/j.combiomed.2019.103545.
- [2] S. Núñez-González, E. Bedoya, D. Simancas-Racines, and C. Gault, "Spatial clusters and temporal trends of malignant melanoma mortality in Ecuador," *SAGE Open Medicine*, vol. 8, no. 6, May 2020, doi: 10.1177/2050312120918285.
- [3] A. M. Gomolin, T.Cline, "The danger of neglecting melanoma during the COVID-19 pandemic," *Journal of Dermatological Treatment*, 2020
- [4] C. Z. Tschandl, *et al.*, "Human-computer collaboration for skin cancer recognition," *Nature Medicine*, vol. 26, pp. 1229-1234, 2020, doi: 10.1038/s41591-020-0942-0.

- [5] M. A. Naeem, M. S. Farooq, A. Khelifi, A. Abid, "Malignant Melanoma Classification Using Deep Learning: Datasets, Performance Measurements, Challenges and Opportunities," *IEEE Access*, vol. 8, pp. 110575-110597, Jun. 2020, doi: 10.1109/ACCESS.2020.3001507.
- [6] American Cancer Society, "Melanoma: Statistics," 2020. [Online]. Available: <https://www.cancer.net/cancer-types/melanoma/statistics>
- [7] M. J. Maron, et al., "Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks," *European Journal of Cancer*, vol. 119, pp. 57-65, Sep. 2019, doi:10.1016/j.ejca.2019.06.013.
- [8] Verosha Pillay, and Serestina Viriri, "Skin cancer detection from macroscopic images," *Institute of Electrical and Electronics Engineers Inc.*, 2019.
- [9] Y. Wang, et al., "Classifying Melanoma and Seborrheic Keratosis Automatically with Polarization Speckle Imaging," *IEEE GlobalSIP 2019 : 7th IEEE Global Conference on Signal and Information Processing*, 2019, doi: 10.1109/GlobalSIP45357.2019.8969331.
- [10] P. G. Cavalcanti, J. Scharcanski, and G. V. G. Baranoski, "A two-stage approach for discriminating melanocytic skin lesions using standard cameras," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4054-4064, Aug. 2013, doi: 10.1016/j.eswa.2013.01.002.
- [11] B. Pallavi, and Keshvamurthy, "A Hybrid Diagnosis System for Malignant Melanoma Detection in Dermoscopic Images," *2019 4th IEEE International Conference on Recent Trends on Electronics, Information, Communication and Technology (RTEICT)*, May 2019, doi: 10.1109/RTEICT46194.2019.9016745.
- [12] M. Takruri, M. K. Abu Mahmoud, and A. Al-Jumaily, "PSO-SVM hybrid system for melanoma detection from histopathological images," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 4, pp. 2941-2949, Aug. 2019, doi: 10.11591/ijece.v9i4.pp2941-2949.
- [13] N. J. Sugiarti, Yuhandri, J. Naam, and Dolly Indra, "An artificial neural network approach for detecting skin cancer," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 2, pp. 788-793, Apr. 2019, doi: 10.12928/telkomnika.v17i1.9547.
- [14] S. R. S. Jianu, Loretta Ichim, and Dan Popescu, "Automatic Diagnosis of Skin Cancer Using Neural Networks," *11th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, Mar. 2019, doi: 10.1109/ATEE.2019.8724938.
- [15] A. A. Adegun, and S. Viriri, "Deep learning-based system for automatic melanoma detection," *IEEE Access*, vol. 8, Dec. 2020, doi: 10.1109/ACCESS.2019.2962812.
- [16] B. R. Esteva, et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115-118, 2017.
- [17] V. S. Almaraz-Damian, V. Ponomaryov, S. Sadovnychiy, and Heydy Castillejos-Fernandez, "Melanoma and nevus skin lesion classification using handcraft and deep learning feature fusion via mutual information measures," *Entropy*, vol. 22, no. 4, pp. 484, Apr. 2020, doi: 10.3390/e22040484.
- [18] K. C. Rashmi, K. Prasad, C. K. Udupa, and V. Shwetha, "A Comparative Evaluation of Texture Features for Semantic Segmentation of Breast Histopathological Images," *Journal of Dermatological Treatment*, vol. 8, pp. 64331-64346, Apr. 2020, doi: 10.1109/ACCESS.2020.2984522.
- [19] J. R. Rigel, Julie Russak, and Robert Friedman, "The Evolution of Melanoma Diagnosis: 25 Years Beyond the ABCDs," *CA: A Cancer Journal for Clinicians*, Jul. 2010, doi: 10.3322/caac.20074.
- [20] Dermnet Skin Disease Atlas, 2018. [Online]. Available: <http://www.dermnet.com>
- [21] C. Gupta, N. K. Gondhi, and P. K. Lehana, "Analysis and Identification of Dermatological Diseases Using Gaussian Mixture Modeling," *IEEE Access*, vol. 99, pp. 1-1, Jul. 2019, doi: 10.1109/ACCESS.2019.2929857.
- [22] C. Gupta, N. K. Gondhi, and P. K. Lehana, "Analysis and Identification of Dermatological Diseases Using Gaussian Mixture Modeling," *IEEE Access*, vol. 7, Jul. 2019, doi: 10.1109/ACCESS.2019.2929857.
- [23] J. F. Alcon, et al., "Automatic Imaging System With Decision Support for Inspection of Pigmented Skin Lesions and Melanoma Diagnosis," *IEEE Journal of Selected Topics In Signal Processing*, vol. 3, no. 1, pp. 14-25, Feb. 2009, doi: 10.1109/JSTSP.2008.2011156.
- [24] G. Shakhnarovich, R. El-Yaniv, Y. Baram, "Smoothed Bootstrap and Statistical Data Cloning for Classifier Evaluation," *Toyota Technological Institute at Chicago*, 2010.
- [25] C. Ocampo, "Herramienta Soporte al Diagnóstico del Melanoma usando Imágenes Dermatoscópicas," *Bogotá*, 2011.