

# Statistical Pattern Recognition for Thresholding between Human Skin and Background in Color Images

<sup>1,2</sup>Rafael Divino Ferreira Feitosa, <sup>2</sup>Vinicius Araújo Santos, <sup>2</sup>Leandro Luís Galdino de Oliveira, <sup>3</sup>Díbio Leandro Borges, <sup>2</sup>Lucas Calabrez Pereyra and <sup>1,2</sup>Adriano Soares de Oliveira Bailão

<sup>1</sup>Department of Informatics, Federal Institute Goiano - IF Goiano, Goiás, Brazil

<sup>2</sup>Institute of Informatics, Federal University of Goiás - UFG, Goiânia, Goiás, Brazil

<sup>3</sup>Department of Computer Science, University of Brasilia - UnB, Brasilia, Distrito Federal, Brazil

## Article history

Received: 07-03-2017

Revised: 24-03-2017

Accepted: 08-04-2017

## Corresponding Author:

Rafael Divino Ferreira Feitosa,  
Department of Informatics,  
Federal Institute Goiano,  
Goiânia, Goiás, Brazil  
Email:  
rafael.feitosa@ifgoiano.edu.br

**Abstract:** Many research works based on the tone of human skin have been developed to locate and track the human body for the purpose of recognition in color images. With respect to other techniques, some advantages of face detection based on skin color are the smaller processing time, invariant angles of rotation and the performance in semi-occluded faces. In this study we present the results of a survey that investigated the performance of 4 supervised classifiers in skin detection. In order to maximize the generalization of the models, a training set containing samples of individuals of different ages and ethnicities was used. Experimental results showed that the best performance was achieved by using an ANN and the worst results were yielded by LDA. With the Naive Bayes, QDA and ANN algorithms, we showed that the white, black, yellow and brown tones of human skin are in a well-defined range of the RGB color spectrum determined by common characteristics. We also compiled 2798 skin samples for treatment and 305 images with their manually obtained labels as supplementary material, which was made available to help in the development of further research in human skin detection.

**Keywords:** Applied Computing, Machine Learning, Skin Detection

## Introduction

Research related to the location and tracking of parts of the human body using skin detection techniques as a pre-processing stage for posterior recognition has been conducted with the objective of reducing the computational cost of algorithms (Sun, 2010). In the last few years, these techniques have attracted the attention of the academic community due to their numerous applications such as the tracking and detection of the face, the identification of naked people and the identification of hand movements, among others (Khan *et al.*, 2012).

These researches reported the use of several methodologies with the same objective of detecting human skin in color images based on skin tone. They make use of different color spaces such as RGB, HSV, YCbCr, CIELAB and others, separately or by combining two or more color systems (Hawari *et al.*, 2002; Lin, 2007; Snehal and Chougule, 2013). With different reported true positive rates, these studies show, in a

general way, that human skin tones are normally grouped into well-defined bands of color spaces which might pose as a problem of binary classification (skin or non-skin) which can be solved (Yang *et al.*, 1998; Störring *et al.*, 2001; Jones and Rehg, 2002).

In comparison to other techniques, the color-based thresholding to separate human skin from the background can be mentioned as an advantageous approach when used as a pre-processing step to detect exposed parts of the human body. The advantages include a smaller processing time as well as an invariant rotation angle and surpassing problems related to semi-occlusion (Khan *et al.*, 2012). Nevertheless, there are factors that might influence the results negatively, like variances in lighting conditions, complex backgrounds or backgrounds similar to certain skin tones, the variety of capture devices and ethnic diversity (Kakumanu *et al.*, 2007).

For Lin (2007), the detection and automatic recognition of human faces is one of the most intriguing

and important problems in the management of image databases of faces, as well as in computer vision and cybernetics. It can be applied as a safety mechanism, thus replacing keys, cards, passwords or Personal Identification Numbers (PIN). Apart from the usage in security systems, research on face detection can also be applied in the areas of Facial Expression Analysis, Human-Computer Interfaces and Content-Based Image Retrieval (CBIR) (Tripathi *et al.*, 2011).

In this article, we present the results of a survey that investigated the performance of 4 supervised classifiers in skin detection. To maximize the capacity of model generalization, we used a training set that contained samples of individuals of several ages and ethnicities. The Naive Bayes, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Artificial Neural Network (ANN) algorithms were used to perform the classification of skin and non-skin tones of the pixels in color images using the RGB color system. The article is organized as follows: In section II we present the classifiers and discuss the latest works which investigated skin detection in images as their main objective or as a pre-processing stage. In Section III we describe the methodology of the experiments, in section IV we present and discuss the results and in section V the conclusions and possibilities of future works are exposed.

### Related Works

The ethnic diversity of individuals may pose as one of the biggest challenges in the classification of pixels in skin and non-skin. In order to tackle this problem, it is fundamental to prioritize two aspects: The generalization capability of the classifier in the context of the problem at hand and the representativeness of the classifier's sample training set to contemplate, in a comprehensive manner, the variability of human skin tones.

According to Osorio (2013), the Brazilian Institute of Geography and Statistics (IBGE), which currently adopts the racial classification used in research carried out in national territory, distinguishes varieties by the skin color characteristic, namely white, black, yellow and brown. In the year of 2008 the Research of Ethno-racial Features of the Brazilian Population - PCERP 2008 - as carried out with the objective of raising an empirical database in order to support studies and analyses as an attempt to improve ethno-racial systems used in several statistical population surveys by IBGE (2011).

In an attempt to characterize the pattern distribution of the colors of human skin, Sun (2010) defines the technical predecessors as simple, because they use few computer resources with a reduced set of rules and a lack of flexibility by means of strictly fixed rules. In the training phase, a set of skin and non-skin images are

processed independently to determine the average histograms for each model. The RGB color space was used for standardization purposes and comparison of the results achieved by Jones and Rehg (2002).

Approaching the YCbCr color space, Tripathi *et al.* (2011) investigates the detection of faces in skin regions and validates the occurrences by implementing the method of similarity search between input images and previously trained image models. Initially, an input image is converted into the YCbCr color space, then the skin regions are segmented by thresholding and finally the corresponding model is applied for the selection of the faces.

In combination, 4 color spaces, namely the RGB, YCbCr, HSV and CIELAB, are used in Raghuvanshi and Agrawal (2012) to detect faces in the stages of skin segmentation, binarization of the image, rejection of non-face areas based on geometric properties of the human face and the determination of the facial area in the image.

In Khan *et al.* (2012), the effect of transformation of color spaces in regards of the efficacy of skin detection is investigated and validated to find the most suitable space for the application. The role of the luminance component of the color space and the most appropriate technique for the selection of pixels are also looked into. Nine skin-modeling approaches in color spaces IHLS, HSI, RGB, normalized RGB, YCbCr and CIELAB are considered.

With the same purpose of face detection, Hawari *et al.* (2002) suggest a new algorithm for the skin detection images, based on the YCgCr and HSV color spaces combined. Color balance was performed on the input image to correct the variations of light and then it is converted simultaneously into the color spaces HSV and YCgCr. After segmentation, morphological operations are performed in the skin region and the proportion of the face is calculated to discard areas of skin that do not belong to faces.

After converting RGB into HSV images, Snehal and Chougule (2013) uses the channel H (hue) to determine the color band for face detection, thus determining experimentally that the hue of the human skin is in a well-defined band of the color space. The method consists in the conversion of color spaces followed by checking, pixel by pixel in the image, to assign them as skin or non-skin and the usage of morphological operations to determine the region of interest.

In pattern recognition, it is preferable to have the original data extracted directly from the image as the input, in order to maximize efficiency in classification and reduce execution time. A limitation of the aforementioned works resides in the usage of several different color spaces and the consequent need to convert the data to RGB color space, which is the original pattern for color image representation. These operations require computer processing to an extent

which could be beyond the necessary for the training and classification of samples.

Using a set of samples of human skin tones, several works that aimed to construct mathematical models with a mechanistic approach can also be found in the literature. Those models attempt to describe how the variables of each color pixel relate to each other. Thus, one can determine rules that explain the behavior model from the observation of these variables, that is, the criteria that a certain color has to fulfill, in order to be classified as a skin tone. To outline the model and understand system response, the training samples are divided into components, such as the values of the RGB channels and the behavior of each of these components and the interactions that occurred between them are observed.

By using the mechanistic approach, Kovac *et al.* (2003) applies a set of fixed rules in an interactive art installation and utilizing the filter described in (1) to detect human skin in the RGB color space.

The color spaces RGB, YCbCr and HSI are used in combination by Zangana and Shaikhli (2013), suggesting a color filter to human skin detection as shown in (2), (3) and (4), respectively:

$$if \left\{ \begin{array}{l} R > 95, G > 40, B > 20 \\ \max\{R, G, B\} - \min\{R, G, B\} > 15 \\ |R - G| > 15 \\ R > G, R > B \\ or \\ R > 220, G > 210, B > 170 \\ |R - G| \leq 15 \\ R > B, G > B \end{array} \right. \quad then \ skin \quad (1)$$

$$if \{0.836 * G - 14 < B < 0.836 * G + 44\} \quad then \ skin \quad (2)$$

$$if \{102 < Cb < 128\} \quad then \ skin \quad (3)$$

$$if \{19 < H < 240\} \quad then \ non - skin \quad (4)$$

With the objective of determining a common set of characteristics shared by human skin tones in the RGB color space, Powar *et al.* (2013) proposes the model shown in (5) for skin detection applied for the selection of possible pictures with pornographic content in combination with the YCbCr color space shown in (6):

$$if \left\{ \begin{array}{l} R > 95, G > 40, B > 20 \\ |R - G| > 15 \\ R > G, R > B \end{array} \right. \quad then \ skin \quad (5)$$

$$if \left\{ \begin{array}{l} 76 < Cb < 128 \\ 132 < Cr < 174 \end{array} \right. \quad then \ skin \quad (6)$$

In Feitosa *et al.* (2014) the authors establish the model described in (7) from a set of skin samples for training obtained from Dass (2015). The reported results present a high specificity rate and a well-defined band of probable human skin tones inside the RGB color space:

$$if \left\{ \begin{array}{l} 9 \leq |R - G| \leq 71 \\ 8 \leq |R - I| \leq 57 \\ 4 \leq |B - I| \leq 47 \\ 0.227451 \leq \frac{R}{255} \leq 0.976471 \quad then \ skin \\ R > I > G > B \\ |R - G| \leq B \end{array} \right. \quad (7)$$

Although these proposed models yield reasonably satisfactory hit rates, they tend to establish the probable region of human skin in the RGB spectrum strictly. Thus, they are limited, once they don't consider the Probability Density Function (PDF) obtained through the average and covariance matrix provided by processing the samples during the training phase, as well as a determined degree of uncertainty as a limit for class decision making.

The described mechanistic models were trained to perform the classification of the images strictly into skin or non-skin. Thus, luminosity variations can interfere in a region, making it brighter or darker and thus not fitting the classification rules. The mechanistic classifier response is 1 (true) for skin or 0 (false) for non-skin. They reject likely tones that could be correctly classified in reason of minimum variations, e.g., a 0.99's response for skin. Instead, the classifiers utilized in this study were statistically trained, in a context of decision-making for binary classification, to classify a test sample as skin or non-skin using the class probability.

The statistical classifiers are based on considerations on probability because they minimize the average classification error. However, to make the optimization of previous techniques feasible, it is necessary that the Probability Density Function (PDF) and the probability of occurrence of each problem class are well-known a priori. In practice, the application of these classifiers in problems with few samples for each class or when the PDFs don't follow a uniform distribution is impracticable. In these cases, we assume that the data follows a theoretical distribution, accepting approximations in the dataset initial modeling, which impacts directly the classifier hit rate.

## Materials and Methods

In this study we propose a comparative study of 4 classifiers based on supervised learning for human skin detection in images stored using the RGB color space

from training samples containing tones of skin of people from different ethnicities. The RGB system was chosen as an object of study due to the advantages as presented by Feitosa *et al.* (2014):

- It is one of the most currently used systems for storage and representation of digital images
- The simplicity and intuitiveness of the model
- It does not need any computational transformation for the investigation

### *Utilized Classifiers*

This section presents the 4 classifiers utilized in the experiments - Naive Bayes, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Artificial Neural Network (ANN) - pointing their operating principles and characteristics.

#### *Naive Bayes*

The Naive Bayes classifier takes full independence of the object features, thus being considered a useful statistical classifier (Gonzalez and Woods, 2010). This is due to the fact that, on average, their use minimizes the probability of classification error or loss (Zhang, 2004). In pattern recognition applications, the uncertainties involved in real systems generate the need for conducting probabilistic considerations to respond better to situations not previously considered. In the case of the Bayesian classifier, a pattern is recognized from the probability of belonging to one of the existing classes. The main purpose of a Bayesian classifier is to minimize the total average loss. For that, it is necessary that the probabilities of occurrence of each class are known, as well as the probability of a continuous variable to belong to a certain interval, this being known as the Probability Density Function (PDF). In cases which the patterns to be classified belong to the n-dimensional domain, the calculation of PDF is not trivial, requiring the use of an analytical function for different density functions (Scholkopf and Mullert, 1999). Knowing that many natural physical observations follow a Gaussian distribution, the multivariate Gaussian probability density function is used. The implementation of this objective function defines the Bayesian decision function used in classification, also using the mean, standard deviation and covariance matrix of the classes as parameters to perform the classification.

#### *Linear Discriminant Analysis (LDA)*

The Linear Discriminant Analysis (LDA) is a supervised classification method derived from the generalization of Fisher's linear discriminant method (Fisher, 1936). One of its applications is to reduce the dimensionality of a problem as a preliminary stage of

classification similar to Principal Component Analysis (PCA). However, it does not preserve the variance of classes as an attempt to detect similarities (Martínez and Kak, 2001). Its main objective is to minimize the intraclass variance while maximizing interclass separation by means of the maximization of the average distance from each of the classes (Scholkopf and Mullert, 1999). It is a parametric method that uses the derivation of the same multivariate Gaussian function used in the Bayesian classifier. The derivation used is a simplification for classes with the same variance (Narsky and Porter, 2013), defining that the hyperplane that separates the classes is a linear combination of the attributes of these classes.

#### *Quadratic Discriminant Analysis (QDA)*

Once it is an extension of LDA, the Quadratic Discriminant Analysis (QDA) is also a parametric supervised classifier that uses, in its turn, a quadratic combination of attributes. Using a quadratic function to define the boundaries of decision between classes, the QDA differs from LDA in not assuming that the variances of each class are the same (Narsky and Porter, 2013). This characteristic allows new parameters to be inserted into the discriminant function, thus making it a curve, which is opposed to the LDA separation hyperplane, represented by a flat surface (Zhang, 1997).

#### *Artificial Neural Network (ANN)*

In several problems, the statistical properties of the classes are unknown or unidentifiable. For these cases, it is necessary to apply methods that obtain the decision functions necessary for classification by means of a training process (Gonzalez and Woods, 2010). For the modeling of non-linear decision hyperplanes, we can use the ANN, more specifically, the Multi-Layer Perceptron (MLP) type. It is composed of neurons, the perceptrons, interconnected and distributed in layers. The multilayer neural networks have an entrance, a hidden layer, which may contain a varying number of neurons and the output layer. Each entry is propagated through the neurons sequentially until the stimulus applied to the input reaches the output layer. By using multi-layer neural networks, the main problem lies in the adjustment of synaptic weights in the hidden layers. For this, the Scale Conjugate Gradient (SCG) was used as a back propagation method to update the bias and weight values.

#### *Metrics for Performance Evaluation*

The evaluation of a classifier is based on some performance metrics which are calculated during testing, providing their generalization capability. For a binary classification problem, these performance

measurements are based on the values of the confusion matrix: Number of True Positives (TP), number of True Negatives (TN), number of False Positives (FP) and number of False Negatives (FN) (Seliya *et al.*, 2009). In this work, we used six performance metrics - precision, recall, specificity, efficiency, *F-measure* and *G-measure* - as follows.

Precision (P) is defined as the rate between the number of pixels classified correctly (TP) and the total number of pixels defined as skin (TP + FP) by the classifier, as shown in (8):

$$P = \frac{TP}{TP + FP} \quad (8)$$

Recall or sensitivity is the rate of correctly classified pixels, i.e., the capability of the model to find the desired class (9):

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

Specificity is calculated in order to find the model's capability to distinguish the non-skin pixels, or the probability of a certain classification not be correct (10):

$$Specificity = \frac{TN}{TN + FP} \quad (10)$$

Efficiency is the arithmetic mean of the sensitivity and specificity (11). In practice, specificity and sensitivity are inversely proportional. Thus, this measure yields the level of balance between those:

$$Efficiency = \frac{recall + specificity}{2} \quad (11)$$

*F-measure* or *F-score* (F) relates to the performance of the classification model for the positive class (skin). It is calculated as the harmonic mean of the recall and precision, usually being  $\beta = 1$  (12).  $\beta$  is used to determine the relative importance of recall and precision (Castro and Braga, 2011):

$$F_{\beta} = (1 + \beta^2) * \frac{precision + recall}{\beta^2 * precision + recall} \quad (12)$$

*G-measure*(G) is also calculated using the precision and recall values (13). It normalizes the TP rate for the geometric mean of the predicted and actual positive classification values (Powers, 2011):

$$G = \sqrt{precision * recall} \quad (13)$$

## Experiments

For the execution of the experiments, we used the Matlab R2013a installed on an Ubuntu 14.04 operating system.

Our dataset is formed by skin samples of 2798 individuals and other 2798 samples extracted from images that knowingly do not contain human skin. The skin samples were taken from the face, in an area between the lines of the eyes and nose, in squares of 20x20 pixels. These were extracted from photos of individuals, men and women of different age groups (Fig. 1), captured in a controlled environment by Dass (2015) in the Human Project, which aims to store and catalog all possible human skin tones. The choice of base images for extraction of skin samples was made seeking to maximize the coverage of ethnic groups, ultimately ensuring that the training set contained egalitarian patterns with ethno-racial features of white, black, brown and yellow skin colors.

In skin samples, we extracted the average pixels from the chosen regions (Fig. 2) to reduce noise such as marks, blemishes, spots and/or scars from the skin samples. In the non-skin samples, we did not use the same methodology because the extracted cuts were composed by various objects as chairs, tables, doors, trees, animals, vehicles and others. For this non-skin set, we use all pixels from the cuts. Since our goal was also to present and provide human skin detection data using only the pixel colors, we trained the classifiers with 4 features: *R*, *G* and *B* channels and the *I* value (gray scale, obtained from average between RGB values to improve sample discrimination).

The skin tones of the resulting averages were subjected to the classifier together with the randomly selected non-skin samples. To verify the generalization capability in human skin detection in the RGB color space, we applied all 4 classifiers in a set of test images different from the training set, selected from the FDDB database of facial images (Jain and Miller, 2010).

Manually, we created the ground truth consisting of images with all pixels of non-skin removed from the 305 raw images used. With this methodology, it was possible to determine the accuracy rates and error in the classification of pixels of skin and non-skin in real images obtained from various sources and organized by Jain and Miller (2010), even in approximate terms, considering the loss at the edges. The accuracy measurements and the error, presented in the results section, considered the following rates:

- True positive for skin pixels classified as skin
- True negative for pixels of non-skin classified as non-skin
- False positive for pixel non-skin classified as skin
- False negative for skin pixels classified as non-skin

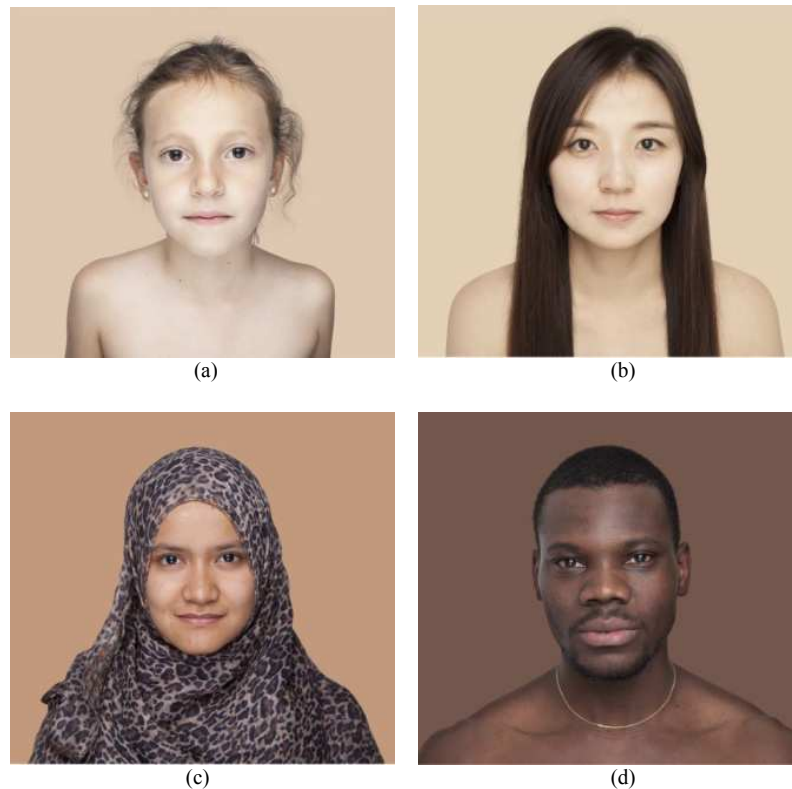


Fig. 1. Some individuals photographed by Dass (2015) representing the classification under their characteristic skin color adopted by the IBGE standards (Osorio, 2013)

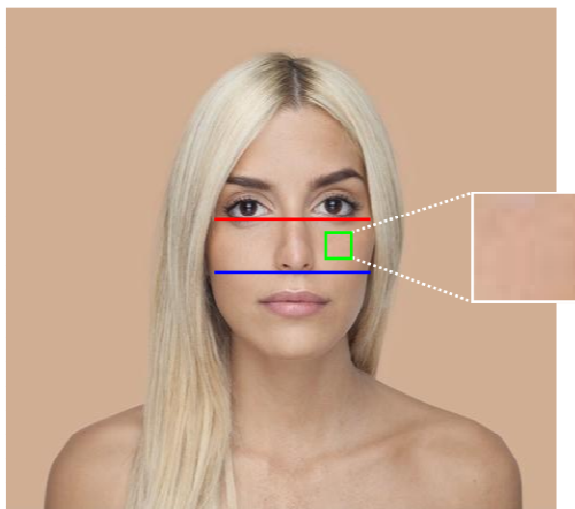


Fig. 2. Region chosen for the extraction of skin samples from the individuals

Each data set used on the training (skin and non-skin) was randomly divided in three groups: The first group, to be used during training, has 70% of the data, comprising 1958 samples; another group has 15% of the data (420 samples) for validation, to measure the generalization capacity of the ANN and in the case

when there is no further improvement, the process is stopped; the final 15% of the data, comprised of 420 samples, was used for testing and extraction of the ROC Curves and the Confusion Matrix of the trained network. In the training phase we used a feed forward network with 10 neurons on the hidden layer and only one neuron in the output layer. The maximum number of epochs was set to 1000.

## Results

For the purposes of comparison and evaluation of results, the classifiers' performance was assessed by using the confusion matrix and the Receiver Operating Characteristic (ROC) curve.

The confusion matrix allows the visualization of the hit and miss rates by means of the TP and TN rates, presented on the main diagonal and the FP and FN on the secondary diagonal. Classifier performances considered as perfect report a confusion matrix with the values of the main diagonal equal to 100% and the values of the secondary diagonal equal to 0%. The rates of trial and error measured for the utilized classifiers are shown in Fig. 3 and for the cited mechanistic models in Fig. 4.

Among the 4 classifiers, the ANN had the best performance in both sensitivity (56.7392%) and

specificity (96.7054%). The LDA had the worst performance in terms of the FP rate (22.0602%) and the worst FN rate (67.2487%). We observed that the ANN classifier improved the method's specificity proposed by Feitosa *et al.* (2014) in 1.2129% that, considering 39,615,571 pixels from test images - being 7,606,829 of skin and 32,008,742 of non-skin - which is equivalent to an increase of 388,234 pixels correctly classified as non-skin.

The ROC curve shows graphically the performance of the binary classifier as the discrimination threshold varies. Characteristically, the TP rate (sensitivity) is plotted as a function of the FP rate. Each point on the ROC curve is a pair [sensitivity, specificity]. A classifier with a capability to discriminate between classes regarded as perfect has a ROC curve that touches the upper left corner [100%, 100%]. The closer the ROC curve is to the upper left corner, the

higher the efficiency of the classifier, whereas curves near the diagonal graph indicates randomness in the classification and a curve below the diagonal is indicative of a performance lower than a random classifier's.

Figure 5 displays the ROC curves of the classifier performances. It is noticeable that the ANN classifier yielded the largest area under the ROC curve and therefore the greatest stability regarding sensitivity and specificity rates. The performance curves of the Naive Bayes and QDA achieved the same values for the area, but with different behaviors due to the fact that the classification criterion in the Naive Bayes classifier is the highest probability of a sample being skin class or non-skin class and in the QDA the lowest value representing the interclass distance. The worst performance was observed in the LDA with the smallest area under the curve.



Fig. 3. Confusion matrix of the Naive Bayes (a), LDA (b), QDA (c) and ANN (d) classifiers





Fig. 4. Confusion matrix of the mechanics models proposed by Kovac *et al.* (2003) (a), Zangana and Shaikhli (2013) (b), Powar *et al.* (2013) (c) and Feitosa *et al.* (2014) (d)

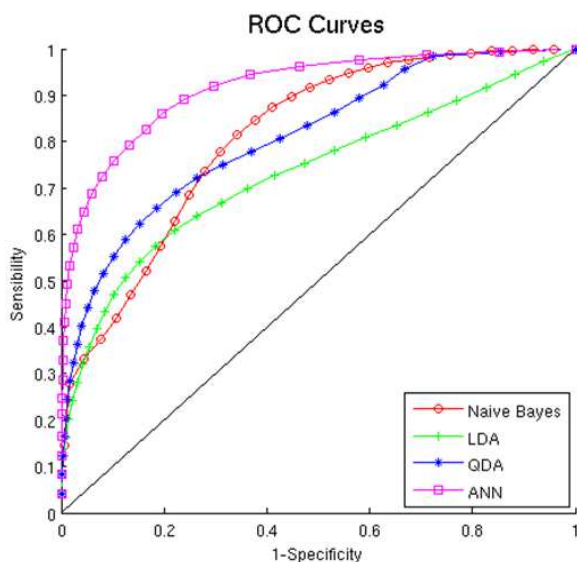


Fig. 5. ROC curves of the classifiers' performance applied to the test set

In Naive Bayes, a probabilistic classifier, the decision rule for the test sample is given by the largest value of the classes' predicted probability. This classifier calculates the similarity based on the covariance matrices of each class, obtained during training. The QDA, a statistical classifier, also uses the covariance matrices, but returns the distance between the test sample and the predicted classes, i.e., a dissimilarity is calculated. Similarly, the LDA, a statistical classifier, also returns a distance but based on the dissimilarity from the joint covariance matrix of the classes. So, in the QDA and LDA, the decision is performed by the smallest value returned by the classification between the classes. In the ANN, the network response parameter was configured for  $<0.3$  for skin, i.e., if the response of the activation function sigmoid of network is less than 0.3, the sample is classified as skin, otherwise ( $\geq 0.3$ ) as non-skin.

Figure 6 shows scatter plots of the probable human skin tones classified by the algorithms. The Naive Bayes



(Fig. 6a), QDA (Fig. 6c) and ANN classifiers (Fig. 6d) rated similar regions in the RGB color space. The results showed that the ANN was better in generalizing the correct classification of darker skin tones. Among these, the ones who had higher RGB spectrum reduction rates (98.8863%) were the Naive Bayes and QDA classifiers, selecting 186,841 possible tones, followed by the ANN (98.4040%), which selected 267,752 likely tones. The worst performance was presented by the LDA classifier, with a low spectrum reduction rate (73.2929%) considering 4,480,698 possible tones in the RGB space. The result of the LDA classifier (Fig. 6b) shows that the classes in the studied problem are not linearly separable, thus selecting a large amount of colors as possible skin tones.

We can observe that the Naive Bayes algorithm (Fig. 6a) and the QDA (Fig. 6c) rate the same color band colors as being skin tones. This behavior can be ascribed to the fact that both classifiers are based on Gaussian functions. One difference between these classifiers is the fact that the QDA considers that a given unknown pattern has the same probability of belonging to both classes (i.e., their PDFs are the same) and therefore will differ in their covariance matrices only. We can also compare the performance of the 4 classifiers utilized in this research in the examples of test images shown in Fig. 7. Figure 8 displays examples of human skin detection which suffered interference from lighting conditions and consequent loss of color-based characteristics.

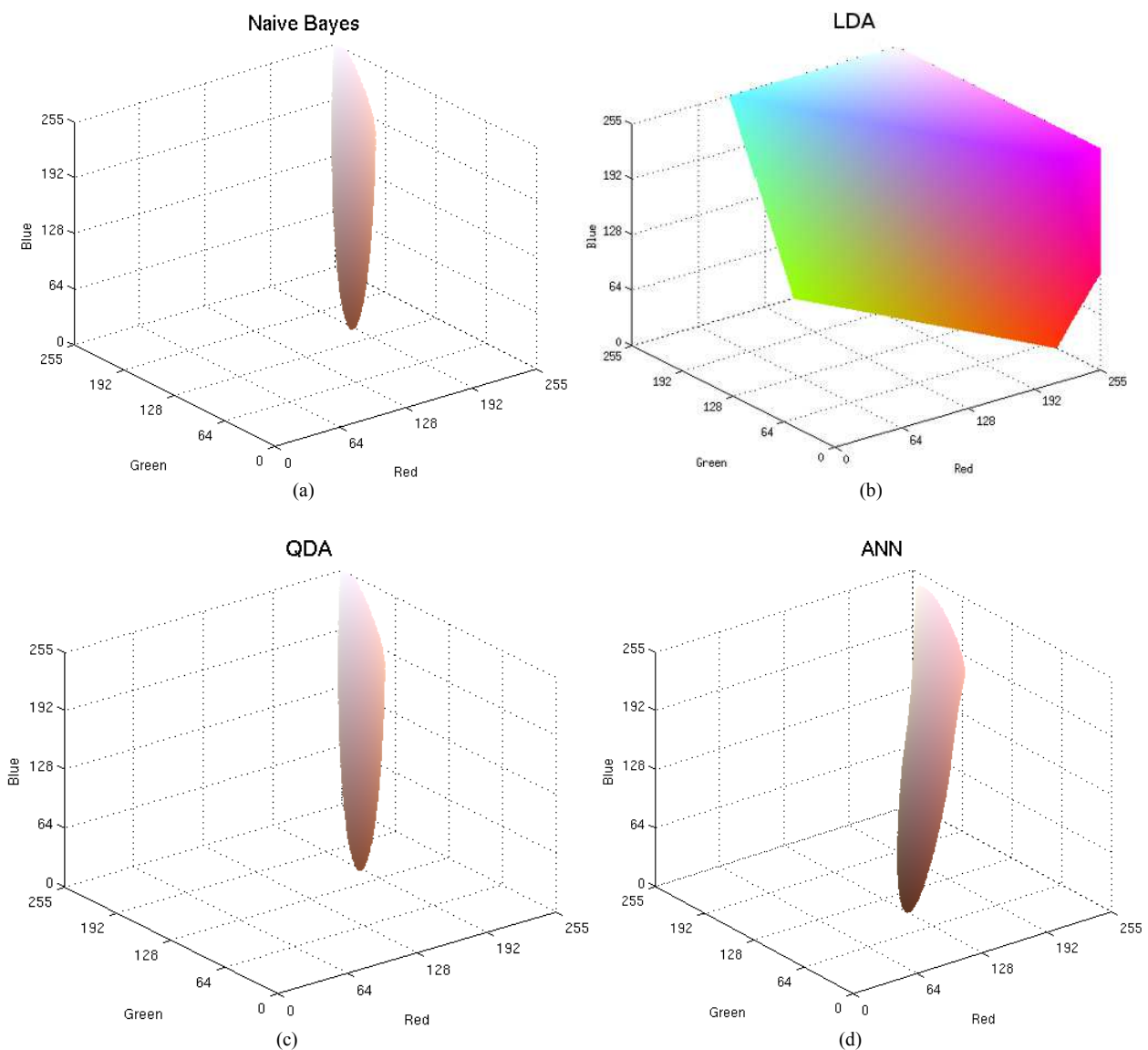


Fig. 6. Probable human skin tones selected in the RGB space by the Naive Bayes (a), LDA (b), QDA (c) and ANN (d) classifiers

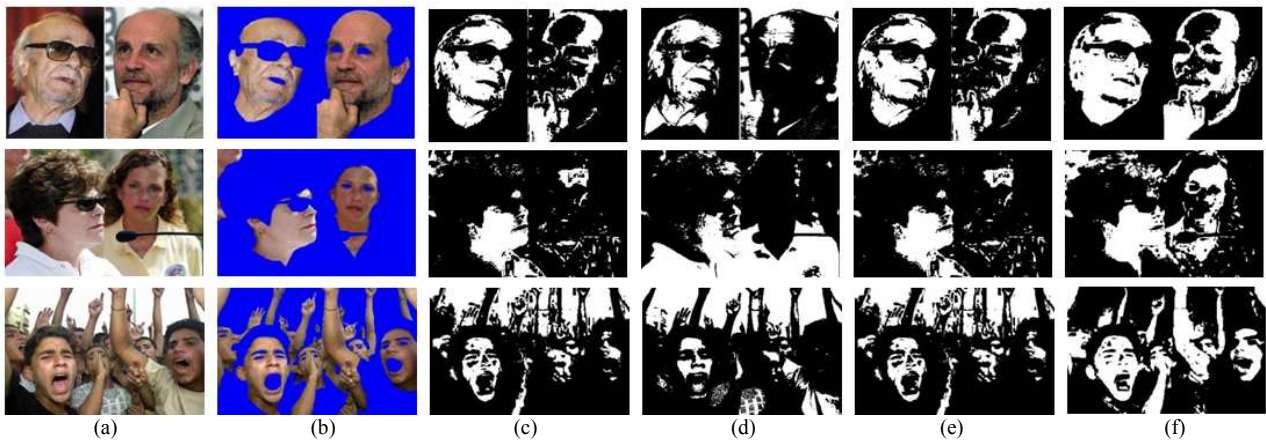


Fig. 7. Images tests with high hit rate (TP + TN) (a), ground truth (b) and the results of the Naive Bayes (c), of the LDA (d), the QDA (e) and a Neural Network (f)

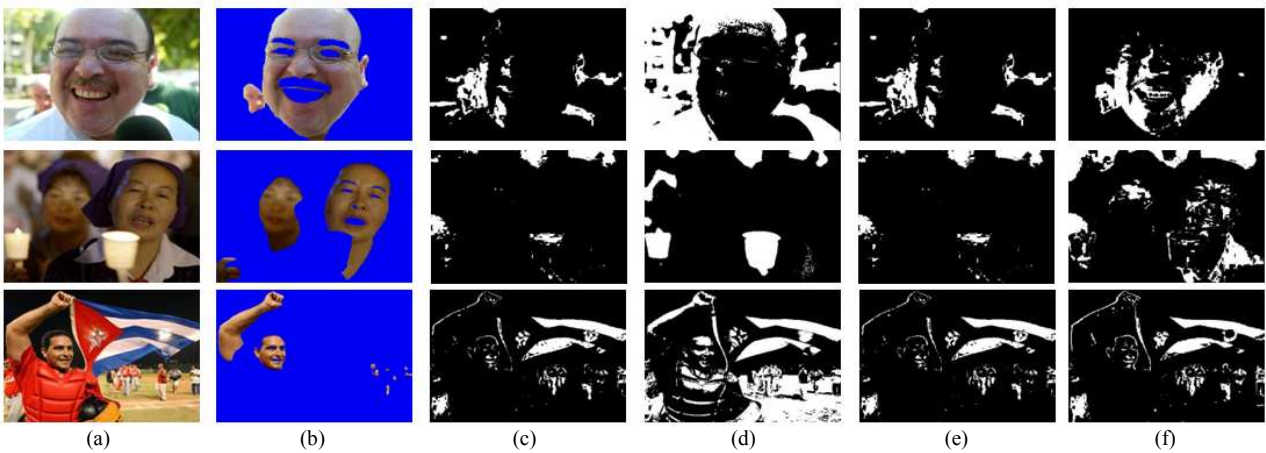


Fig. 8. Images tests with high error rate (FP + FN) (a), ground truth (b) and the results of the Naive Bayes (c), of the LDA (d), the QDA (e) and a Neural Network (f)

Table 1. Performance measurements for the classifiers and mechanistic models. For each metric: Top three results underlined and the best result identified by the character (\*)

Classifier	Precision	Recall		Efficiency	<i>F-measure</i>	<i>G-measure</i>
		Sensitivity	Specificity			
Naïve Bayes	0.6107	0.3504	<u>0.9469</u>	0.6486	0.4453	0.5760
LDA	0.2608	0.3275	0.7794	0.5535	0.2904	0.5052
QDA	0.6107	0.3504	<u>0.9469</u>	0.6486	0.4453	0.5760
ANN	<u>0.8036*</u>	0.5674	<u>0.9671*</u>	<u>0.7672</u>	<u>0.6652</u>	<u>0.7407</u>
Kovac <i>et al.</i> (2003)	0.5884	<u>0.8398*</u>	0.8604	<u>0.8501</u>	<u>0.6920</u>	<u>0.8501</u>
Zangana and Shaikhli (2013)	0.1870	<u>0.7712</u>	0.2034	0.4873	0.3011	0.3961
Powar <i>et al.</i> (2013)	<u>0.6480</u>	<u>0.8345</u>	0.8923	<u>0.8634*</u>	<u>0.7295*</u>	<u>0.8629*</u>
Feitosa <i>et al.</i> (2014)	<u>0.7502</u>	0.5697	<u>0.9549</u>	0.7623	0.6476	0.7376

We evaluated the predictive capability of the mechanistic models and the classification algorithms using widespread performance metrics for binary classification problems. According to Castro and Braga (2011), these criteria either focus on the detection of the minority class in unbalanced classification problems or consider the discrimination

of both classes as having the same relevance. All metrics used in this assessment yield values between 0 (poor performance) and 1 (high performance). The values obtained are reported on Table 1.

The ANN yielded at least one of the three best results in five performance measurements, except recall. As expected, due to the non-linear classification problem

previously mentioned, the LDA classifier yielded the worst results. However, it yielded some results which were superior than those of other models, such as in specificity, efficiency and *G-measure*.

Regarding precision and specificity values, the ANN obtained the best results, 0.8036 e 0.9671, respectively. The mechanistic model proposed by Kovac *et al.* (2003) yielded the best results in terms of recall/sensitivity (0.8398). Powar *et al.* (2013) yielded the best efficiency (0.8634), *F-measure* (0.7295) and *G-measure* (0.8629) results.

We also remark the results obtained by Zangana and Shaikhli (2013), which presented the lowest precision (0.1870), specificity (0.2034), efficiency (0.4873) and *G-measure* (0.3961). However, it yielded one of the 3 highest recall rates (0.7712).

In order to allow reproduction and encourage new research which could rely on the results of this study, our training and test sets were made available. This database comprises 2798 skin samples from areas between the eyes and nose lines which were cut from the faces of individuals pictured in the Humanae Project and used to train the classifiers. Other 305 images extracted from the Fddb database with their manually obtained labels were also made available.

## Conclusion

This work presents a study using 4 classifiers for human skin detection in color images. Experimental results show that the ANN classifier achieves the best performance whereas the LDA classifier achieves the worst. The ANN had the highest specificity rate, correctly classifying the pixel as non-skin (a TN rate of 96.7054%) and the greatest sensitivity rate (56.7392%).

With the Naive Bayes, LDA and QDA algorithms, we observed that the white, black, yellow and brown tones of human skin are in a well-defined range of the RGB color spectrum. Furthermore, they are determined by common characteristics and algorithms that trace a linear hyper plane of separation, such as the LDA, cannot solve the binary problem of classification into skin and non-skin patterns.

The ANN classifier improved the method's specificity proposed by Feitosa *et al.* (2014) in 1.2129% that, considering 39,615,571 pixels from test images - being 7,606,829 of skin and 32,008,742 of non-skin - which is equivalent to an increase of 388,234 pixels correctly classified as non-skin.

The ANN yielded at least one of the three best results in five performance measurements, except recall. As expected, due to the non-linear classification problem previously mentioned, the LDA classifier yielded the worst results. The proposed model by Kovac *et al.* (2003) yielded the best results in terms of recall/sensitivity (0.8398). Powar *et al.* (2013) yielded

the best efficiency (0.8634), *F-measure* (0.7295) and *G-measure* (0.8629) results.

The efficiency of supervised techniques is directly linked to the quality of the training patterns used when building the classification model. In this research, we presented a database of individuals photographed in a controlled environment with exposed upper torsos and faces and from which skin samples could be extracted to train classifiers or to perform detection tests.

Due to the fact that the characteristics used to train the classifiers are related to the colors of skin samples, we believe that lighting conditions are a limiting factor when applying these techniques. Thus, we intend to use lighting correction algorithms on the images in order to reduce or eliminate the negative influence of lighting conditions and improve hit rates in future studies.

We also intend to apply the classifiers that obtained the best performances - namely, Naive Bayes, QDA and ANN - in image tests with different pre-processing steps by color-correction algorithms and evaluate the effects of these procedures on the accuracy rates and error obtained by the classifiers.

## Acknowledgment

The authors would like to thank the Federal Institute Goiano and Federal University of Goiás for supporting the publication of this work and to Angelica Dass for providing the images of the Humanae Project.

## Authors Contributions

**Rafael Divino Ferreira Feitosa and Vinícius Araújo Santos:** Literature review, data set preparation, execution of the experiments, paper's writing, translation and reviewing.

**Leandro Luís Galdino de Oliveira and Dibio Leandro Borges:** Organization, conception, supervision and reviewing.

**Lucas Calabrez Pereyra and Adriano Soares de Oliveira Bailão:** Data analysis and interpretation, translation, paper's writing and reviewing.

## Ethics

This paper contains an original contribution and unpublished material of the research. The corresponding author declares the co-authors approved this manuscript. There are no ethical issues involved.

## References

- Dass, A., 2015. Humanae project.
- Castro, C.L. and A.P. Braga, 2011. Aprendizado supervisionado com conjuntos de dados desbalanceados. SBA Controle Automação, 22: 441-466. DOI: 10.1590/S0103-17592011000500002

- Feitosa, R.D., L.L. de Oliveira, D.L. Borges and M.M.N. Filho, 2014. A mathematical model for reducing the likely spectrum of human skin tones in the RGB color space. Proceedings of the IEEE International Conference on Imaging Systems and Techniques, Oct. 14-17, IEEE Xplore Press, pp: 329-334. DOI: 10.1109/IST.2014.6958499
- Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. *Annals Eugen.*, 7: 179-188. DOI: 10.1111/j.1469-1809.1936.tb02137.x
- Gonzalez, R.C. and R.C. Woods, 2010. *Processamento digital de imagens*, Pearson Prentice Hall, 3rd ed. ISBN: 978-85-7605-401-6.
- Hawari, K., B. Ghazali, J. Ma and R. Xiao, 2002. An innovative face detection based on skin color segmentation. *Int. J. Comput. Applic.*, 34: 6-10. DOI: 10.5120/4069-5244
- IBGE, 2011. *Características étnico-raciais da população: Classificações e identidades*. Instituto Brasileiro de Geografia e Estatística – IBGE, Tech. Rep.
- Jain, V. and E. Miller, 2010. *FDDB: A benchmark for face detection in unconstrained settings*. University of Massachusetts, Amherst, Tech. Rep., UM-CS-2010-009.
- Jones, M.J. and J.M. Rehg, 2002. Statistical color models with application to skin detection. *Int. J. Comput. Vis.*, 46: 81-96. DOI: 10.1023/A:1013200319198
- Khan, R., A. Hanbury, J. Stöttinger and A. Bais, 2012. Color based skin classification. *Patt. Recog. Lett.*, 33: 157-163. DOI: 10.1016/j.patrec.2011.09.032
- Kakumanu, P., S. Makrogiannis and N. Bourbakis, 2007. A survey of skin-color modeling and detection methods. *Patt. Recog.*, 40: 1106-1122. DOI: 10.1016/j.patcog.2006.06.010
- Kovac, J., P. Peer and F. Solina, 2003. Human skin color clustering for face detection. Proceedings of the IEEE Region 8 Computer as a Tool, Sept. 22-24, IEEE Xplore Press, pp: 144-148. DOI: 10.1109/EURCON.2003.1248169
- Lin, C., 2007. Face detection in complicated backgrounds and different illumination conditions by using YCbCr color space and neural network. *Patt. Recog. Lett.*, 28: 2190-2200. DOI: 10.1016/j.patrec.2007.07.003
- Martínez, A.M. and A.C. Kak, 2001. Pca versus LDA. *IEEE Trans. Patt. Anal. Machine Intell.*, 23: 228-233. DOI: 10.1109/34.908974
- Narsky, I. and F.C. Porter, 2013. *Statistical Analysis Techniques in Particle Physics: Fits, Density Estimation and Supervised Learning*. 1st Edn., John Wiley and Sons, ISBN-10: 3527410864, pp: 459.
- Osorio, R.G., 2013. *A Classificação de cor ou raça do IBGE revisitada, características étnico-raciais da população: Classificações e identidade*. Instituto Brasileiro de Geografia e Estatística - IBGE, Tech. Rep.
- Powar, V., A. Kulkarni, R. Lokare and A. Lonkar, 2013. Skin detection for forensic investigation. Proceedings of the International Conference on Computer Communication and Informatics, IEEE Xplore Press, pp: 1-4. DOI: 10.1109/ICCCI.2013.6466122
- Powers, D.M.W., 2011. Evaluation: From precision, recall and f-measure to ROC, informedness, markedness and correlation. *J. Machine Learn. Technol.*, 2: 37-63.
- Raghuvanshi, D.S. and D. Agrawal, 2012. Human face detection by using skin color segmentation, face features and regions properties. *Int. J. Comput. Applic.*, 38: 14-17.
- Scholkopf, B. and K.R. Mullert, 1999. Fisher discriminant analysis with kernels. Proceedings of the IEEE Signal Processing Society Workshop Neural Networks for Signal Processing, Aug. 25-25, IEEE Xplore Press, pp: 41-48. DOI: 10.1109/NNSP.1999.78812
- Seliya, N., T.M. Khoshgoftaar and J.V. Hulse, 2009. A study on the relationships of classifier performance metrics. Proceedings of the 21st IEEE International Conference on Tools with Artificial Intelligence, Nov. 2-4, IEEE Xplore Press, pp: 59-66. DOI: 10.1109/ICTAI.2009.25
- Snehal, M. and D.M.S.R. Chougule, 2013. Skin detection for face recognition based on HSV color space. *Int. J. Eng. Sci. Res. Technol.*, 2: 1883-1883.
- Störring, M., H.J. Andersen and E. Granum, 2001. Physics-based modelling of human skin colour under mixed illuminants. *Robot. Autonomous Syst.*, 35: 131-142. DOI: 10.1016/S0921-8890(01)00122-1
- Sun, H.M., 2010. Skin detection for single images using dynamic skin color modeling. *Patt. Recog.*, 43: 1413-1420. DOI: 10.1016/j.patcog.2009.09.022
- Tripathi, S., V. Sharma and S. Sharma, 2011. Face detection using combined skin color detector and template matching method. *Int. J. Comput. Applic.*, 26: 5-8. DOI: 10.5120/3119-4290
- Yang, J., W. Lu and A. Waibel, 1998. Skin-color modeling and adaptation. Proceedings of the 3rd Asian Conference on Computer Vision, Jan. 08-10, Springer, pp: 687-694.
- Zangana, H.M. and I.F. Shaikhli, 2013. A new algorithm for human face detection using skin color tone. *IOSR J. Comput. Eng.*, 11: 31-38.
- Zhang, H., 2004. The optimality of Naive Bayes. Proceedings of the FLAIRS Conference, AAAI Prima, pp: 3-3.
- Zhang, M.Q., 1997. Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proc. Nat. Acad. Sci.*, 94: 565-568. PMID: 9012824