Skin Lesions Classification using Multichannel Dermoscopic Images

Conference Paper · September 2019

DOI: 10.5281/zenodo.3460912

3 authors:

Luís Vinícius de Moura
Pontifícia Universidade Católica do Rio Grande do Sul
7 PUBLICATIONS **1** CITATION
SEE PROFILE

Caroline Dartora
Pontifícia Universidade Católica do Rio Grande do Sul
12 PUBLICATIONS **6** CITATIONS
SEE PROFILE

Ana Maria Marques Da Silva
PUCRS - Pontifícia Universidade Católica do Rio Grande do Sul
93 PUBLICATIONS **215** CITATIONS
SEE PROFILE

 DOI: 10.5281/zenodo.3460912

Skin Lesions Classification using Multichannel Dermoscopic Images

Luís Vinícius de Moura
Escola Politécnica
PUCRS
Porto Alegre, Brazil
ORCID: 0000-0003-3429-3289

Caroline Machado Dartora
Escola de Medicina
PUCRS
Porto Alegre, Brazil
ORCID: 0000-0003-1357-3230

Ana Maria Marques da Silva
Escola Politécnica
Instituto do Cérebro
PUCRS
Porto Alegre, Brazil
ORCID: 0000-0002-5924-6852

***Abstract*— Visual inspection of visible light dermoscopic images can be challenging due to the similarity between melanoma and benign skin lesions, especially in the early stages. This paper aims to investigate the influence of independent use of color channels in feature extraction for skin lesions classification in dermoscopic images. Visible dermoscopic RGB images from 1850 benign and malignant (melanoma) skin lesions from ISIC database were used. RGB channels were split and analyzed independently to identify the relevance of each color channel information. Radiomics features were extracted, normalized, and selected by their scores for an unsupervised feature-ranking algorithm. Selected features were used to train a random forest classifier, which was cross-validated with K-fold method and optimized using exhaustive grid search. The results showed no significant differences in skin lesions classification using each color channels, although R channel produced the most relevant information, with an accuracy of 0.86 and an AUC of 0.93.**

***Keywords*—** *classification, medical imaging, multichannel images, skin lesion.*

I. Introduction

Melanoma accounts for about 1% of all types of skin cancer in the world but represents the majority in cases of death [1]. Melanoma diagnosis can be made by biopsy or imaging, such as the digital dermoscopy, by an expert dermatologist. Dermoscopy is a noninvasive visible light image technique that allows visualizing the subsurface morphology of cutaneous lesions up to dermis, revealing colors, textures, and structures not visible to the naked eye, improving precision and level of confidence of medical diagnosis [2]. Skin lesions appear as patches of darker color on the skin. Dermatologists apply the ABCD criteria to diagnose melanoma in visible dermoscopic images, using four visual features: asymmetry (A), border irregularity (B), color variation (C) and diameter (D) [3]. However, discriminating benign skin lesions and melanoma at their early stages is quite complicated, because they both exhibit similar features at early evolutional phase [4].

Dermoscopy image analysis is still a challenge for expert dermatologists [5]. Medical images have information that goes beyond visual aspects. Mathematical and computational methods applied to images allow the extraction of a variety of quantitative attributes, which can be used as tools for supporting clinical decisions [6-7].

The use of commercially available mobile cameras is quite common in skin lesion inspection systems, particularly for telemedicine purposes. However, the poor resolution and the variable illumination conditions are not easily handled. Nonuniform illumination or shadows can affect skin lesions classification systems based on features extracted from color images. The requirement for constant image colors for reproducibility remains unsatisfied, as it requires adjustments and corrections to operate within the camera dynamic range and measuring the same color regardless lighting conditions [4, 8].

Radiomics is a field of study that is growing due to a large amount of data, images, and outcomes from clinical studies and greater availability of tools and computational power. Studies have shown that descriptive models based on biomedical data can provide relevant predictive, prognostic, and or diagnostic information [9]. For skin lesion and melanoma diagnosis, several methods are used, mainly extracting and analyzing color, shape and texture features, mimicking the ABCD criteria.

Akram et al. (2018) [4] developed a classification method for skin lesion segmentation and recognition using features related to ABCD criteria. After extraction and features reduction, classification with support vector machine (SVM) using public datasets (ISIC, ISIC UDA, and ISIC MSK) was applied. Mean results were 97.5% for accuracy, 96.3% for sensitivity, 99.5% for specificity and 96.3% for precision. Gilmore, Hofmann-Wellenhof and Soyer (2010) [10], and Gautan et al (2018) [11] used SVM techniques to classify skin lesions using geometric, color and edge features, as well as GLCM features. [10] and [11] used less than 300 images, and accuracies of 0.80 and 0.75 were found, respectively.

Focusing in multichannel color analysis, Pathan et al (2019) [12] recently used 200 dermoscopic images to classify lesion with a set of absolute and relative colors features and eight different classification methods, finding the maximum accuracy of 0.88. Celebi and Zornberusing (2014) [13] applied K-mean to reduce the number of color features in 617 dermoscopic images. Symbolic regression was applied for classification, and sensitivity was 0.62, specificity 0.76 and overall accuracy 0.72.

Visually, multichannel color images are indispensable for benign and malignant skin lesion diagnosis [2], as well as in classification by artificial intelligence techniques. We propose to investigate the influence of independent use of color channels in feature extraction for skin lesions classification in dermoscopic images.

II. Materials and Methods

A. Image Dataset and Pre-processing

The dataset was collected from 1850 dermoscopic color images (RGB) of skin lesions, which were randomly selected, with their segmentation and clinical information, from *International Skin Imaging Collaboration* (ISIC). The groups were composed of 950 benign skin lesions images and 900 melanoma images. Eight images were excluded due to the

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finace Code 001 and PUCRS.

 DOI: 10.5281/zenodo.3460912

absence of segmentation. ISIC dataset has two types of segmentation: performed by expert doctors and beginners. When both were available, segmentations made by experts were prioritized. Fig. 1 shows an example of RGB and segmentation image from our ISIC dataset sample. All RGB images were converted to 32-bit unsigned and segmentation was applied as a binary mask. RGB images were split into each channel: red (R), green (G) and blue (B). Fig. 2 illustrates one image example with RGB channels separation.

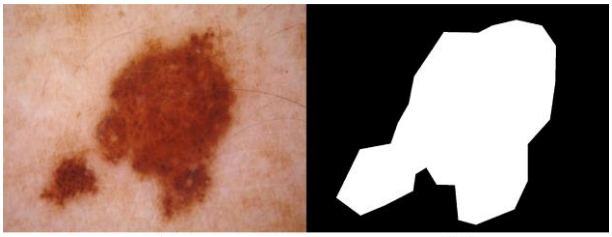

Fig. 1. Image example. (a) Original image (b) Segmentation.

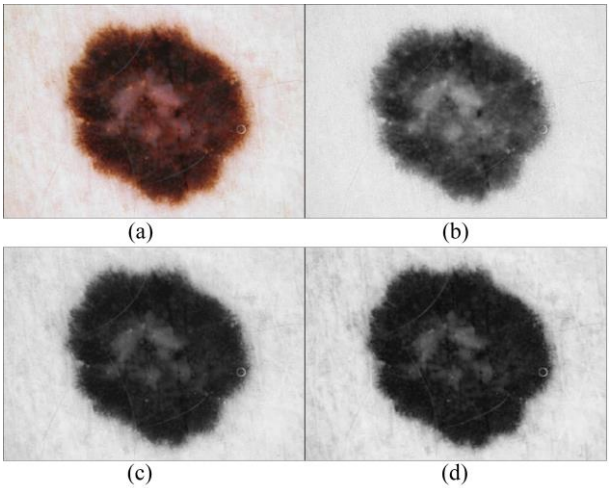

Fig. 2. Illustration of channels of RGB. (a) Original image (b) Red channel (c) Green channel (d) Blue channel.

B. *Radiomic Feature Extraction and Reduction*

Pyradiomcs library was used for the extraction of 6 groups of radiomic features [14]:

- 2D shape descriptors (9 features): related to the shape and size of the region of interest (ROI).
- First-order features (18 features): related to the ROI pixel intensity distribution, using basic metrics.
- Gray-level co-occurrence matrix (GLCM – 22 features): related to the ROI second-order probability function, based on 8 angles of pixel neighborhood.
- Gray-Level Run Length Matrix (GLRLM – 16 features): similar to GLCM, uses pixels intensities co-occurrence using 4 angles in the image length, giving the homogeneity of pixel intensity run.
- Gray-level size zone matrix (GLSZM – 16 features): used for texture characterization, provides a statistical representation by the estimation of a bivariate conditional probability density function of the image distribution values.
- Gray-level dependence matrix (GLDM – 14 features): quantify the dependence of image gray-level, calculating how much connected they are at a certain distance when their difference on pixel intensity is less than 1.

For feature selection, all features were normalized by scaling them between 0 and 1, independently. Features that had a better representation of the sample were selected using random forest classifier (RFC), defining weights for each feature, with feature ranking with recursive feature elimination and cross-validated (RFECV). RFC algorithm uses different models for forest classifiers [15], with several sub-samples of the original dataset and the mean prediction to get the best accuracy and to control data overfitting. RFECV method uses feature selection based on RFC classifier, adjusting the sample and removing features with lower weights until the maximum accuracy.

C. *Classification Model and Evaluation Metrics*

RFC was used with a K-fold cross-validation model (RFC-KF). It was implemented using Sickit-learn library on Python version 3.6.5 [16]. The K-fold method divides the original sample into K parts, using K-1 parts to train the classifier model and the other part of the sample to test it. The workflow of the proposed model is shown in Fig. 3.

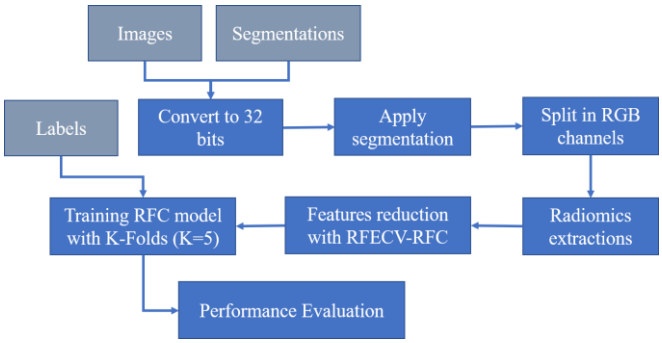

Fig. 3. The workflow of the proposed model.

For evaluation, receiver operating characteristic (ROC) and area under the ROC curve (AUC) were used. Accuracy (Eq. 1), sensitivity (Eq. 2), specificity (Eq. 3), precision (Eq. 4) and F1 score (Eq. 5) metrics were calculated, as follow:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$Sensitivity = TP / (TP + FN) \quad (2)$$

$$Specificity = TN / (TN + FP) \quad (3)$$

$$Precision = TP / (TP + FP) \quad (4)$$

$$F1\ Score = 2 / (Precision^{-1} + Sensitivity^{-1}) \quad (5),$$

where *TP*, *TN*, *FP*, and *FN* are counts of true positive, true negative, false positive and false negative, respectively [16].

Exhaustive grid search (EGS) technique was used for parameters optimization in the classification model. The optimization focus was given by the accuracy metric.

 DOI: 10.5281/zenodo.3460912

III. Results

For each RGB channel, 95 features were extracted. Using RFC, each channel features were independently reduced and 11, 62, and 66 features were selected for R, G and B channels, respectively.

Our model was trained using RFC-KF with K = 5. For each cross-validation step, performance metrics were calculated, and arithmetic means are presented (Table I).

TABLE I. METRICS FOR RGB CHANNELS

Metrics	Channel		
	R	G	B
# Feature	11	62	66
Accuracy	0.86	0.86	0.87
AUC	0.93	0.94	0.94
Precision	0.88	0.88	0.88
Sensitivity	0.83	0.83	0.85
Specificity	0.88	0.88	0.88
F1 Score	0.85	0.86	0.87

Figures 4, 5, and 6 show the ROC curve for each RGB color channel, respectively. Mean ROC for RFC-KF with K = 5 is shown as a straight blue line. The standard deviation of mean ROC is given as gray shade, and AUC for each fold is given in the text captions.

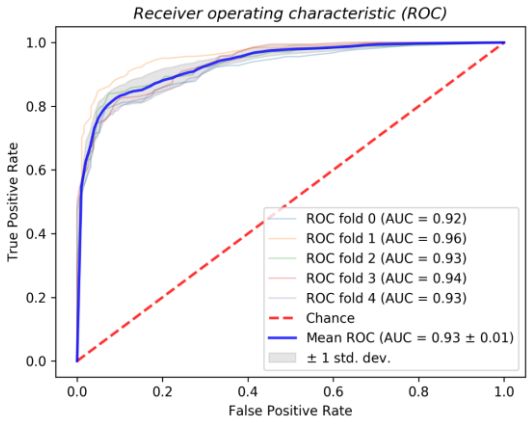

Fig. 4. ROC for the red channel.

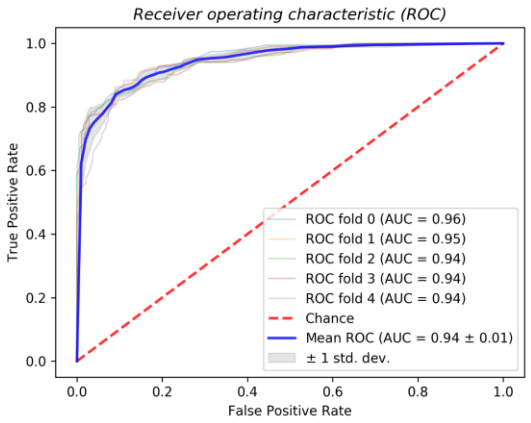

Fig. 5. ROC for the green channel.

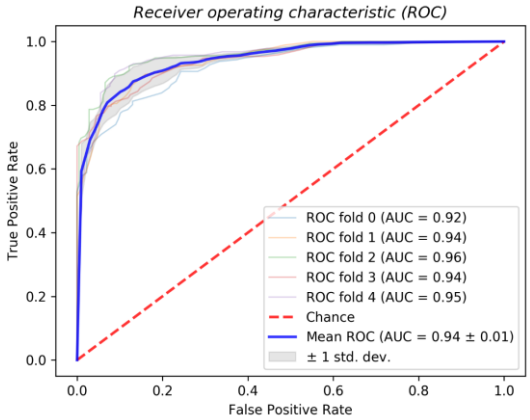

Fig. 6. ROC for the blue channel.

IV. Discussion

A wide variety of approaches for classification, selection, and reduction of colors and textures features are found in the literature. Usually, for color features, different channels and combinations (RGB, HSV, HIS, etc.) are used, and features related to mean values, kurtosis, standard deviation, and entropy are extracted. However, color channels do not use to be analyzed independently, with more than first-order feature extraction.

In our approach, segmented dermoscopic images were separated in RGB channels, and features were extracted, reduced in quantity and used to classify each channel independently. Model parameters were optimized using EGS and values of accuracy were 0.86, 0.87, and 0.87 for R, G, and B channels, respectively.

Another common approach is the analysis of the features by principal components. This method turns possible the use of an extensive range of features with a small dataset, avoiding overfitting. However, this technique does not allow the differentiation of which features are needed for the classification, usually increasing the computational time.

In this work, for all RGB channels analyzed independently, similar results were acquired. However, R channel showed higher optimization, using a smaller number of features (11 for R, and more than 60 for G and B channels). From 11 selected features for R channel, four were related to 2D shape, one to first-order features, two to GLRLM, three to GLSZM and one to GLDM.

The higher performance of the R channel can be attributed to the fact that, as shown in Fig.2, skin lesions contrast and non-uniformity is visibly distinguishable R channel. This can also be associated with the blood vessel dilation increase in the lesion, trauma, or neovascularization [13]. According to Ramezani et al. (2018) [8], some ethnic groups have the healthy skin color of reddish and the skin lesions are regions with altered color. Therefore, the red channel is more representative in the difference between benign and malignant regions of skin.

Related to features of greater importance, GLSZM showed higher weight in the R channel than other features. For G and B channels, 2D shape features had the highest weight. Soyer et al. (2004) [17], through visual evaluation in dermoscopic

 DOI: 10.5281/zenodo.3460912

images, showed the presence of blue and or white structures had some consideration in the malignant classification of skin lesion.

It is important to emphasize that the channels had almost the same performance. Thus, the use of only one channel, as R channel, can reduce the computational power and time needed for training and classification.

Maximum AUC obtained with the proposed method in this work was 0.94 and had values compatible with the literature (0.85 to 0.97) [18–21]. Differences can be attributed to the use of deep learning techniques, as well as different image features and dimensionality reduction techniques.

In our knowledge, until this moment, we have not encountered papers that differentiate classification results for each color channel. The use of principal component analysis for feature reduction is commonly used, but not shows to be optimized.

V. Conclusion

This paper presents the investigation of radiomic features applied to the classification of multichannel dermoscopic images of benign and malignant lesions. The independent analysis of RGB channels in the Random Forest Classifier model shown no statistical differences. However, between RGB channels, the red channel showed the most relevant information compared to G and B channels, using the smallest number of features to achieve similar performance.

Further studies are required to increase the number and diversity of skin lesions image samples. A possible reason for Red channel as more representative is the fact that almost all images come from ethnic groups with reddish-white healthy skin color and skin lesions with altered color. Application of image filters to reduce noise and techniques to reduce artifacts will be required to evaluate different skin colors, including the use of deep learning techniques.

References

[1] C. Facts, “Cancer facts and Figuers 2019,” 2019.

[2] S. Duma, “Dermoscopy of pigmented skin lesions,” *Eur. Handb. Dermatological Treat. Third Ed.*, pp. 1167–1177, 2015.

[3] R. J. Friedman, D. S. Rigel, and A. W. Kopf, “Early Detection of Malignant Melanoma: The Role of Physician Examination and Self-Examination of the Skin,” *CA. Cancer J. Clin.*, vol. 35, no. 3, pp. 130–151, 1985.

[4] T. Akram, M. A. Khan, M. Sharif, and M. Yasmin, “Skin lesion segmentation and recognition using multichannel saliency estimation and M-SVM on selected serially fused features,” *J. Ambient Intell. Humaniz. Comput.*, vol. 0, no. 0, p. 0, 2018.

[5] M. Binder *et al.*, “Epiluminescence Microscopy: A Useful Tool for the Diagnosis of Pigmented Skin Lesions for Formally Trained Dermatologists,” *Arch. Dermatol.*, vol. 131, no. 3, pp. 286–291, 1995.

[6] P. Lambin *et al.*, “Radiomics: Extracting more information from medical images using advanced feature analysis,” *Eur. J. Cancer*, vol. 48, no. 4, pp. 441–446, Mar. 2012.

[7] T. A. Data, R. J. Gillies, P. E. Kinahan, and H. Hricak, “Radiol.2015151169,” vol. 278, no. 2, 2016.

[8] M. Ramezani, A. Karimian, and P. Moallem, “Automatic Detection of Malignant Melanoma using Macroscopic Images.,” *J. Med. Signals Sens.*, vol. 4, no. 4, pp. 281–90, Oct. 2014.

[9] P. Afshar, A. Mohammadi, K. N. Plataniotis, A. Oikonomou, and H. Benali, “From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and opportunities,” *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 132–160, 2019.

[10] S. Gilmore, R. Hofmann-Wellenhof, and H. P. Soyer, “A support vector machine for decision support in melanoma recognition,” *Exp. Dermatol.*, vol. 19, no. 9, pp. 830–835, 2010.

[11] D. Gautam, M. Ahmed, Y. K. Meena, and A. Ul Haq, “Machine learning–based diagnosis of melanoma using macro images,” *Int. j. numer. method. biomed. eng.*, vol. 34, no. 5, pp. 1–21, 2018.

[12] S. Pathan, V. Aggarwal, K. G. Prabhu, and P. C. Siddalingaswamy, “Melanoma Detection in Dermoscopic Images using Color Features,” *Biomed. Pharmacol. J.*, vol. 11, no. 1, pp. 107–115, 2019.

[13] M. E. Celebi and A. Zornberg, “Automated quantification of clinically significant colors in dermoscopy images and its application to skin lesion classification,” *IEEE Syst. J.*, vol. 8, no. 3, pp. 980–984, 2014.

[14] J. J. M. Van Griethuysen *et al.*, “Computational radiomics system to decode the radiographic phenotype,” *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, 2017.

[15] L. Breiman, “(impo)Random forests(book),” *Mach. Learn.*, pp. 5–32, 2001.

[16] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[17] H. P. Soyer *et al.*, “Three-point checklist of dermoscopy: A new screening method for early detection of melanoma,” *Dermatology*, vol. 208, no. 1, pp. 27–31, 2004.

[18] S. Dreiseitl, L. Ohno-Machado, H. Kittler, S. Vinterbo, H. Billhardt, and M. Binder, “A comparison of machine learning methods for the diagnosis of pigmented skin lesions,” *J. Biomed. Inform.*, vol. 34, no. 1, pp. 28–36, 2001.

[19] M. Rastgoo *et al.*, “Classification of melanoma lesions using sparse coded features and random forests,” *Med. Imaging 2016 Comput. Diagnosis*, vol. 9785, no. March 2016, p. 97850C, 2016.

[20] S. Tomatis *et al.*, “Automated melanoma detection: Multispectral imaging and neural network approach for classification,” *Med. Phys.*, vol. 30, no. 2, pp. 212–221, 2003.

[21] K. Matsunaga, A. Hamada, A. Minagawa, and H. Koga, “Image Classification of Melanoma, Nevus and Seborrheic Keratosis by Deep Neural Network Ensemble,” pp. 2–5, 2017.