Classification of Melanoma Lesions Using Sparse Coded Features and Random Forests

Mojdeh Rastgoo[a,b] and Guillaume Lemaître[a,b] and Olivier Morel[a] and Johan Massich[a] and Frank Marzani[a] and Rafael Garcia[b] and Désiré Sidibé[a]

[a]LE2I UMR6306, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté, 12 rue de la Fonderie, 71200 Le Creusot, France;

[b]ViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona, Spain

ABSTRACT

Malignant melanoma is the most dangerous type of skin cancer, yet it is the most treatable kind of cancer, conditioned by its early diagnosis which is a challenging task for clinicians and dermatologists. In this regard, CAD systems based on machine learning and image processing techniques are developed to differentiate melanoma lesions from benign and dysplastic nevi using dermoscopic images. Generally, these frameworks are composed of sequential processes: pre-processing, segmentation, and classification. This architecture faces mainly two challenges: (i) each process is complex with the need to tune a set of parameters, and is specific to a given dataset; (ii) the performance of each process depends on the previous one, and the errors are accumulated throughout the framework. In this paper, we propose a framework for melanoma classification based on sparse coding which does not rely on any pre-processing or lesion segmentation. Our framework uses Random Forests classifier and sparse representation of three features: SIFT, Hue and Opponent angle histograms, and RGB intensities. The experiments are carried out on the public PH^2 dataset using a 10-fold cross-validation. The results show that SIFT sparse-coded feature achieves the highest performance with sensitivity and specificity of 100% and 90.3% respectively, with a dictionary size of 800 atoms and a sparsity level of 2. Furthermore, the descriptor based on RGB intensities achieves similar results with sensitivity and specificity of 100% and 71.3%, respectively for a smaller dictionary size of 100 atoms. In conclusion, dictionary learning techniques encode strong structures of dermoscopic images and provide discriminant descriptors.

Keywords: Melanoma, Classification, Sparse coding, Random forests, Dermoscopy

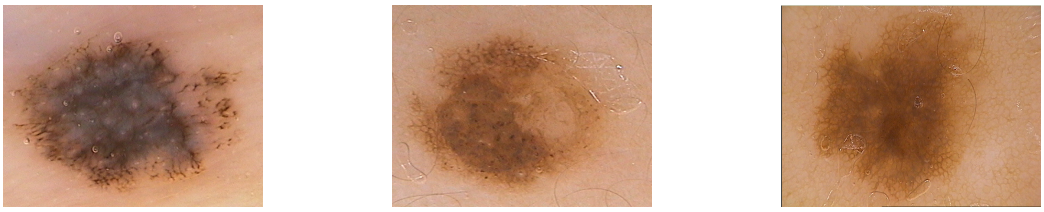
1. INTRODUCTION

Malignant melanoma is a type of skin cancer and although it accounts for almost 2% of all skin cancer cases, it is the deadliest type causing the vast majority of deaths. The incidence of melanoma has increased in the past decades to currently reach 132,000 melanoma cases, according to the World Health Organisation.[1] The American Cancer Society also reported the estimated deaths of melanoma in 2014 as 9710 individuals and new cases as 76,100 individuals.[2] Melanoma cancer is incurable in its advanced stages and the patient should go through surgery, possibly immunotherapy, chemotherapy, and/or radiation therapy. However, if it is diagnosed at its early stage, it is the most treatable kind of cancer.[2,3] In fact the patient survival rate has increased significantly, in the past decades, thanks to early diagnosis and treatment of melanoma in its early stages.

Add the ABCDE lexicon here

The clinical prognosis of early stage of melanoma is based on the "ABCDE" rule,[4] standing for: Asymmetry, irregular Borders, variegated Colors, Diameter greater than 6 mm and Evolving stages of the lesion over time. At each clinical routine, these criterion are used to visually inspect skin lesions which are acquired through different imaging techniques such as dermoscopy. The similarity between the lesions and the necessity to perform patients follow-up over years makes visual inspection difficult and more prone to errors. Thus, Computer-Aided

Further author information: (Send correspondence to M.R. or G.L)
M.R.: E-mail: mojdeh.rastgodast@udg.edu
G.L.: E-mail: guillaume.lemaitre@udg.edu



(a) Melanoma lesion (b) Dysplastic lesion (c) Benign lesion

Figure 1: Samples of PH^2 dataset, representing melanoma, dysplastic and benign lesions, respectively.

Diagnosis systems based on machine learning and image processing techniques have been proposed to assist the dermatologists and clinicians. The proposed algorithms generally attend to mimic the characteristics of the "ABCDE" rule and consist of common steps of pre-processing, segmentation and classification of extracted features.[5] This sequential architecture is complex and each process of the framework is data-driven.

In this paper, we propose a more general framework which does not need pre-processing and segmentation of the lesions and is based on sparse coded features and Random Forests (RF) classifier[6] to detect melanoma in dermoscopic images.

The rest of this paper is organized as follows: an overview of related work is presented in Sect. **??**. The proposed method is discuss in Sect. 2 while the validation and obtained results are presented in Sect. **??**. Finally the paper is concluded in Sect. **??**.

2. METHODOLOGY

As mentioned in the Sect. 1, the proposed framework do not rely on pre-processing an segmentation and focus only on feature detection, extraction and classification.

Add the classification framework graphic

2.1 Feature extraction

2.1.1 Low-level features

In clinical environment, the prognosis of early stage melanoma relies upon visual cues as stated by the "ABCDE" rule. As introduced in Sect. 1, this lexicon characterizes cancerous lesions based on their local textures, shapes, and colors from dermoscopic images. Thus, in order to mimic clinicians assessment, the choice of low-level features encoding similar contextual information is beneficial-crucial?. In that regard, three low-level features in line with the previous requirements are used and presented into details in the remainder of this section. Furthermore, these features are locally extracted by partitioning the dermoscopic images into patches.

Dense Scale-Invariant Feature Transform (SIFT) descriptors are used to encode the local gradient information using an histogram-based representation.[7] SIFT descriptors are extracted over a regular grid such that each grid point is fixed at the center of each image patch. Typically, a region around the center is divided into 4×4 sub-regions from which a 8-bins histogram of the gradient orientations weighted by the gradient magnitudes is computed. Finally, these histograms are concatenated to build the final descriptor with a final size of 128 dimensions. The SIFT implementation used in this work is provided by Vedaldi *et al.*[8]

The opponent color space angle and hue histogram (*C1*) have been first proposed by Van de Weijer and Schmidt as local color features.[9] These descriptors are robust to photometric variations (i.e., shadow, shading, specularities, and light source changes) as well as geometrical variations (i.e, viewpoint, zoom, and object orientation). The hue ($H_{\mathcal{O}}$) and angle ($\theta_{\mathcal{O}}$) of opponent color space ($\mathcal{O}_{1,2,3}$) are formulated as shown in Eq. 2 and Eq. 3, respectively. The opponent color space transformation is defined as in Eq. 1.

$$\begin{pmatrix}\mathcal{O}_1\\ \mathcal{O}_2\\ \mathcal{O}_3\end{pmatrix} = \begin{pmatrix}(R-G)/\sqrt{2}\\ (R+G-2B)/\sqrt{6}\\ (R+G+B)/\sqrt{3}\end{pmatrix} , \tag{1}$$

$$H^{\mathcal{O}} = \arctan\left(\frac{\sqrt{3}(R-G)}{R+G-2B}\right) , \tag{2}$$

$$\theta_d^{\mathcal{O}} = \arctan\left(\frac{\sqrt{3}(R'_d-G'_d)}{R'_d+G'_d-2B'_d}\right) , \tag{3}$$

where d denotes the spatial coordinates of (x,y) and R'_d, G'_d, B'_d denote the first order derivatives of RGB with respect to the coordinates.

This color descriptor is built by taking a 42 bins histogram for the opponent angle $\theta_d^{\mathcal{O}}$ and the hue channel $H^{\mathcal{O}}$, for a final descriptor size of 84 dimensions.

The color intensities (*C2*) represent the color information in a simplest form, their intensities. This descriptor concatenates the color intensities R, G and B to create the feature descriptor.

2.1.2 High-level features

Add the sparse coding framework graph

High-level descriptor is computed using sparse coding techniques. Sparse signal representation has become very popular in the past decades and lead to state-of-the-art results in various applications such as face recognition,[10] image denoising, image inpainting,[11] and image classification.[12] The main goal of sparse modeling is to efficiently represent the images as linear combination of a few typical patterns, called atoms, selected from the dictionary. Here, we intend to use sparse representation of the low-level extracted features for melanoma classification. Sparse coding consists of three main steps: (i) dictionary learning, (ii) low-level features projection, and (iii) feature pooling.[13]

Sparse approximation Given a dictionary $\mathbf{D} \in \mathbb{R}^{n\times K}$ composed of K atoms and an original signal $\mathbf{y} \in \mathbb{R}^n$ (i.e., one feature vector), the sparse approximation corresponds to find the sparest vector $\mathbf{x} \in \mathbb{R}^K$ such that:

$$\arg\min_{\mathbf{x}} \|\mathbf{y}-\mathbf{D}\mathbf{x}\|_2 \qquad \text{s.t. } \|\mathbf{x}\|_0 \leq \lambda \tag{4}$$

where λ is a specified sparsity level.

Solving the above optimization problem is an NP-hard problem.[14] However, approximate solutions are obtained using greedy algorithms such as Matching Pursuit[15] or Orthogonal Matching Pursuit (OMP).[16,17] We used the batch-OMP variant which offers a more efficient algorithm than the standard OMP for our specific problem.[13]

Dictionary learning As stated previously, the sparse approximation is computed given a specific dictionary $\mathbf{D}$, which involves a learning stage from a set of training data. This dictionary is learned using K-SVD which is a generalized version of K-means clustering and uses Singular Value Decomposition (SVD). The dictionary is built, by iteratively solving the optimization problem of Eq. 5, by alternatively computing the sparse approximation of $\mathbf{X}$ and the dictionary $\mathbf{D}$.

$$\arg\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_2 \qquad \text{s.t. } \|\mathbf{x}_i\|_1 \leq \lambda \tag{5}$$

where $\mathbf{Y}$ is a training set of low-level descriptors, $\mathbf{X}$ is the associated sparse coded matrix (i.e., set of high-level descriptors) with a sparsity level λ, and $\mathbf{D}$ is the dictionary with K atoms.

Given $\mathbf{D}$, $\mathbf{X}$ is computed using the batch-OMP algorithm, while given $\mathbf{X}$, $\mathbf{D}$ is sequentially updated, one atom at a time using SVD.

Low-level features projection Once the dictionary is learned, each set of low-level features $\mathbf{F}_I \in \mathbb{R}^{n \times p}$ extracted from p patches in an image is encoded using the dictionary $\mathbf{D}$, solving the optimization problem presented in Eq. 4 such that $\mathbf{F}_I \simeq \mathbf{D}\mathbf{X}_I$.

Feature pooling The sparse coded matrix $\mathbf{X}_I$ is max-pooled to build a final descriptor $\mathbf{f}$ characterizing the given image, such that:

$$\mathbf{f}_i = \max_j \left(|\mathbf{X}_I(i,j)| \right), \ \forall i = 1, \cdots, K \ . \tag{6}$$

2.2 Feature classification

2.2.1 Over-sampling from imbalanced dataset

Similarly to Barata *et al.*,[18] the imbalanced issue of the dataset is tackled by over-sampling the samples of the minority class. New samples are generated to get a balanced set by randomly repeating original samples of the minority class with an additional Gaussian noise $\mathcal{N}(0, 0.0001)$.

2.2.2 Random Forests (RF)

The classification is performed using a RF classifier. RF is an ensemble of decision trees[6] which generalizes the classification process by applying two types of randomization: at the tree level, each tree is fed by a bootstrap made of S' samples built from the original data of size S such that $S = S'$, and at the node level, a subset of feature dimensions m is randomly selected from the original dimension M such that $m = \sqrt{M}$. The trees in RF are grown to their maximum length without any pruning. Each tree in the ensemble casts a unit vote in the final prediction and the final prediction is based on combination of all the votes. RF is used with 1000 un-pruned trees using gini criterion and the original feature dimension of size TO CHECK $M = \{144, 84, 228, 26, 48, 74, 254, 276, 302\}$.

3. EXPERIMENTS

The experiments are conducted on the public PH^2 dataset which is acquired at *Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal*.[18] The dataset contains 200 dermoscopic images divided into two classes: (i) 160 benign and dysplastic, and (ii) 40 melanoma lesions. Seven images have been discarded due to artefacts such as hair occlusions; thus, our experiments are conducted on a subset of the dataset consisting of 39 melanoma, 78 benign, and 76 dysplastic lesions. The patch size used to extract the feature is $10\,\text{px} \times 10\,\text{px}$. The three low-level features are sparsely encoded considering three sparsity levels $\lambda = \{2, 4, 8\}$ and different number of atoms $K = \{100, 200, \cdots, 1000\}$. The classification is performed in a 10-fold cross-validation model in which 80% of the data is used for training and 20% for testing.

4. RESULTS

The results of the experiment are shown in terms of Sensitivity (SE) and Specificity (SP) in Table 1. The highest classification rate in respect of each feature type and each sparsity level are highlighted in different shades of gray from dark to light. The results show that *C1* and SIFT sparse coded features perform better with sparsity levels of 2 and 4, respectively, while *C2* performs better with sparsity level of 8. In general, larger dictionary sizes lead to better classification performance, independently to the feature type and the sparsity level. More precisely, dictionaries with more than 600 atoms are preferable. Figure 2 illustrate the 12 best results. Although, SIFT and *C1* sparse coded features achieve the best classification performance in comparison with *C2*, it can be noted that *C2* features represent the dermoscopic images in their simplest form and create comparable results.

Table 1: The obtained results with different number of atoms and sparsity levels. The first, second, and third highest results for each sparsity level are highlighted in different shades of gray from dark to light color, respectively.

Features	Color1						Color2						SIFT					
Sparsity level	2		4		8		2		4		8		2		4		8	
Dictionary size	SE	SP	SE	SP	SE	SP	SE	SP	SE	SP	SE	SP	SE	SP	SE	SP	SE	SP
100	65.6	49.6	51.5	49.4	41.4	59.7	100	71.7	48.5	71.3	59.9	60	1.4	99.4	0	100	1.4	99.7
200	50.1	59.3	52.7	53.5	51.5	50	57.2	64	95.8	86.6	71.4	72	65.6	47.5	34.3	78	20	95
300	59.8	65.4	52.8	71	57.1	62.3	30	80	65.7	75.4	85.7	85.6	58.6	47.7	64.3	51.1	8.6	88.7
400	67	78.6	62.6	81.3	69.9	76	38.5	66	85.8	77.3	78.6	91.4	62.8	74.6	59.9	64.7	71.3	58.2
500	78.7	79	71.4	78.3	51.3	84	54.2	59.7	61.4	69.6	82.9	83.4	58.5	92	61.2	72.9	54	56.7
600	98.6	82.5	68.7	89.6	64	89.9	48.7	78.4	50	64.6	91.4	89.3	85.8	86.6	61.4	73.3	51.3	53.4
700	92.8	89.9	72.8	91.9	54.4	95.9	37.1	75.4	72.8	72	80	82.6	98.6	84.6	73	94.8	47	62.4
800	92.9	81.4	100	88.4	78.5	89.7	40	70.9	58.6	80.1	97.2	83.9	100	90.3	97.1	93	48.5	72.7
900	90	88	80	92	79.9	95.4	25.7	81.3	19.9	91.1	95.7	73.1	95.7	81.8	80	94.5	54.2	78.5
1000	100	86.8	80	89.6	94.3	91.7	34.3	70.7	42.7	76.5	100	73.8	90	83.5	71.4	89.8	51.3	90.3

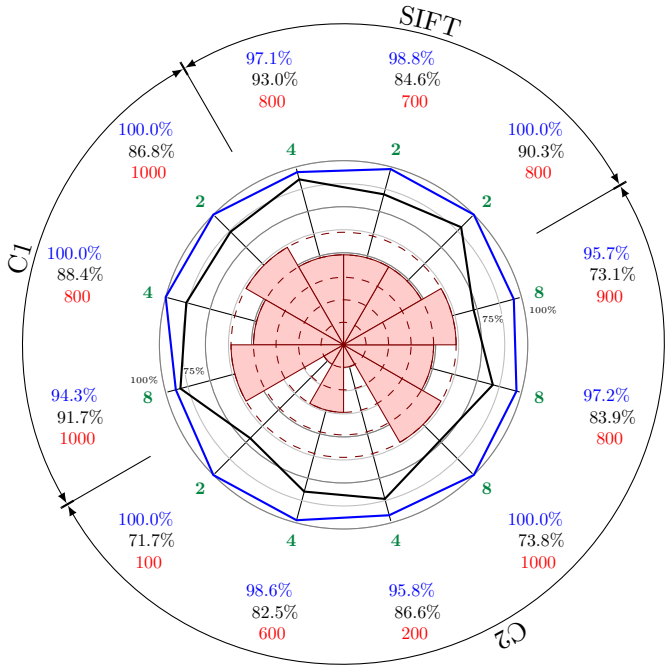


Figure 2: The 12 highest result achieved by RF classifier and sparse representation of SIFT, *color1*, and *color2* features with different **sparsity levels** are illustrated in blue and black as sensitivity and specificity, respectively, while their dictionary size is represented in red. A comparison of the dictionary sizes is also presented in middle of the graph, which contains five levels with maximum dictionary sizes of 1000, 800, 600, 400, 200 for each level respectively.

5. CONCLUSION AND FUTURE WORK

In this work, we proposed a novel classification framework of melanoma lesions, based on sparse representation of the low-level features. Our framework does not need the primary steps of pre-processing and segmentation of the lesions and provide more general algorithm to solve this problem. We proposed to use a well-known color descriptor based on Hue and opponent color space histograms, and SIFT as a texture descriptor. We also consider to represent the images in their simplest form and consider the second color descriptor as R,G and B intensity values from the image. An extensive comparison based on different dictionary sizes and several sparsity

levels were carried out on the PH^2 dataset. The results highlighted the advantage of the proposed method where an RF classifier and a sparse representation of SIFT features with a dictionary size of 800 and sparsity level of 2 achieved the highest performance in terms of SE and SP of 100% and 90.3%, respectively. As avenues for future research, a comparison of sparse learned dictionary with Bag-of-Word models can be performed.

Note. This work has not been submitted for publication or presentation elsewhere

REFERENCES

1. "World health organization, skin cancer."
2. A. C. Society, "Cancer facts & figures 2014," 2014.
3. A. Forsea, V. Del Marmol, E. de Vries, E. Bailey, and A. Geller, "Melanoma incidence and mortality in europe: new estimates, persistent disparities," *British Journal of Dermatology* **167**(5), pp. 1124–1130, 2012.
4. N. R. Abbasi, H. M. Shaw, *et al.*, "Early diagnosis of cutaneous melanoma: revisiting the abcd criteria," *Jama* **292**(22), pp. 2771–2776, 2004.
5. M. Rastgoo, R. Garcia, O. Morel, and F. Marzani, "Automatic differentiation of melanoma from dysplastic nevi," *Computerized Medical Imaging and Graphics* **43**, pp. 44–52, 2015.
6. L. Breiman, "Random forests," *Machine learning* **45**(1), pp. 5–32, 2001.
7. D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, **2**, pp. 1150–1157, Ieee, 1999.
8. A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," in *Proceedings of the international conference on Multimedia*, pp. 1469–1472, ACM, 2010.
9. J. Van De Weijer and C. Schmid, "Coloring local feature extraction," in *Computer Vision–ECCV 2006*, pp. 334–348, Springer, 2006.
10. J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **31**(2), pp. 210–227, 2009.
11. M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *Image Processing, IEEE Transactions on* **15**(12), pp. 3736–3745, 2006.
12. D. Sidibé, I. Sadek, and F. Mériaudeau, "Discrimination of retinal images containing bright lesions using sparse coded features and svm," *Computers in biology and medicine* **62**, pp. 175–184, 2015.
13. R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," *CS Technion* **40**(8), pp. 1–15, 2008.
14. M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer Publishing Company, Incorporated, 1st ed., 2010.
15. S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on* **41**(12), pp. 3397–3415, 1993.
16. Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pp. 40–44, IEEE, 1993.
17. G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive approximation* **13**(1), pp. 57–98, 1997.
18. C. Barata, M. Ruela, M. Francisco, T. Mendona, and J. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE Systems Journal,* **8**, pp. 965–979, Sept 2014.