# A Survey of Feature Extraction in Dermoscopy Image Analysis of Skin Cancer

Catarina Barata, M. Emre Celebi, and Jorge S. Marques,

***Abstract*—Dermoscopy image analysis (DIA) is a growing field, with works being published every week. This makes it difficult not only to keep track of all the contributions, but also for new researchers to identify relevant information and new directions to be explored. Several surveys have been written in the past decade, but these tend to cover all of the steps of a CAD system, which can be overwhelming. Moreover, in these works, each of the steps is briefly discussed due to lack of space. Among the different blocks of the CAD system, the most relevant is the one devoted to feature extraction. This is also the block where existing works exhibit the most variability. Therefore, we believe that it is important to review the state-of-the-art on this matter. This work thoroughly explores the several types of features that have been used in DIA. A discussion on their relevance and limitations, as well as suggestions for future research are provided.**

***Index Terms*—Dermoscopy, skin cancer, melanoma, CAD systems, feature extraction.**

## I. Introduction

The World Health Organization estimates that skin cancer accounts for one third of all the diagnosed cancers worldwide[1]. Over 5 million non-melanoma and more than 87,000 melanoma cases are diagnosed every year in the US [154], while the UK and the Australian societies report nearly 13,000 melanoma diagnoses each [143]. Moreover, the incidence rates of skin cancer have been increasing for the past decades, as can be seen in countries such as UK, where the rate of melanoma has increased 119% since the 1990's, or USA (from 27,600 cases in 1990 to 91,270 in 2018) [177], [176]. The explanation of this trend lies not only in the reduction of the ozone layer, which has diminished the protection against the UV radiation, but also on the abusive exposure to the sun or the solarium and the use of tanning.

The medical community has invested a lot of time and money in prevention campaigns, raising the awareness of the population. However, changing irresponsible behavior may not guarantee safety, as the probability of getting skin cancer also depends on the number of sunburns that people got throughout their life. Therefore, it is also important to invest in the development of technologies that can be used for early diagnosis of skin cancer.

Among the different non-invasive techniques that are used by dermatologists [65], the two most popular ones allow the acquisition of color images of the skin lesions. The images can either be macroscopic or dermoscopic depending on the acquisition setup. Macroscopic (clinical) images are acquired using standard cameras or mobile phones, while dermoscopy ones are obtained using specific magnification devices and an oil/gel interface (immersion contact dermoscopy) or using cross-polarizing light filters (non-contact dermoscopy) [65]. This paper focuses on dermoscopy images not only because these images allow the visualization of additional color and pattern properties of the skin lesions, which increases diagnostic accuracy, but also because dermoscopy is extensively used worldwide [10], [141], [102]. Nonetheless, it is important to stress that some of the techniques discussed in this paper can also be applied to macroscopic images.

One model for the diagnosis of dermoscopy images is a hierarchical process, where a distinction is made first between melanocytic and non-melanocytic lesions and then between malignant and benign [10]. This diagnosis is based on scoring rules that have been proposed to reduce the subjectivity of analysis. Despite their differences, all of the medical scoring approaches share a common denominator: they rely on a set of dermoscopic criteria. Some of the methods focus on the identification of all the possible criteria and their density inside the lesion (pattern analysis [148]), while others aim at recognizing only criteria that are associated with melanoma (*e.g.*, 7-point checklist [9] and the Menzies method [131]). A third group of methods combines the identification of dermoscopic criteria with a broader analysis of the lesion, taking into account the degree of asymmetry, border sharpness, lesion architecture, and color distribution (ABCD [186] and CASH [76] rules). Among the aforementioned methods, only pattern analysis is suitable for both melanocytic and non-melanocytic lesions. Additional information based on clinical covariates, such as the age, gender, and familial history of the patients, is also taken into account [191].

Two of the major limitations of dermoscopy are its subjectivity and requirement of extensive training. A great effort has been made by the research community in the development of computer-aided diagnosis (CAD) tools that can be used by dermatologists to overcome the aforementioned issues [100], [145], [147]. These systems follow a pipeline: i) image pre-processing, ii) lesion segmentation, iii) feature extraction, iv) feature selection (optional), and v) classification.

*i) Image pre-processing* is a required step to deal with images that do not have sufficient quality to be analyzed. This lack of quality can be due to the presence of artifacts (*e.g.*, hair) that can negatively influence the performance of the subsequent steps. Another important issue is color normal-

This work was supported by the FCT project and plurianual funding: [PTDC/EEIPRO/0426/2014], [UID/EEA/50009/2013].

C. Barata and J. S. Marques are with Institute for Systems and Robotics, Instituto Superio Técnico, Lisboa, Portugal. e-mail: ana.c.fidalgo.barata@ist.utl.pt

M. E. Celebi is with the Department of Computer Science, University of Central Arkansas, Conway, AR, USA.

[1]Source http://www.who.int

ization. Dermoscopy images may be acquired using different devices and illumination conditions, rendering unreliable color information [15]. Therefore, it may be important to include a color correction step in the pre-processing phase.

*ii) Lesion segmentation* is a challenging task that has been thoroughly investigated in the literature. The great variety of lesion shapes, sizes, and colors as well as different skin types and textures make it difficult to develop a robust segmentation algorithm. An accurate segmentation is believed to be necessary to achieve a proper extraction of features and consequent lesion characterization. To prevent potential misclassification due to inaccurate segmentation, several CAD systems include semi-automatic segmentation.

*iii) Feature extraction* is a crucial step to obtain a discriminative representation of the skin lesions. Finding appropriate features is a difficult process and a lot of research has been performed in this field, making it possible to identify a large spectrum of features that characterize skin lesions. These features can be divided into four classes: i) hand-crafted features, which are the most popular ones and comprise global image descriptors of shape, symmetry, color, and texture; ii) dictionary-based features, where methods such as bag-of-features or sparse coding are used to obtain local descriptions of the skin lesions; iii) deep learning features that use convolutional neural networks to automatically learn good image representations; and iv) clinically inspired features that aim at attributing a medical meaning to the features used by the CAD system.

*iv) Feature selection* is used to reduce the dimensionality of the feature space in some of the CAD systems by eliminating irrelevant, redundant, or noisy features.

*v) Lesion classification* is the final step of the CAD system. Here, a classification algorithm is trained to predict a diagnosis. Most of the CAD systems focus on the distinction between melanoma and benign or atypical nevi, due to high degree of malignancy associated with the former type of cancer. Nonetheless, it is also possible to find CAD systems that aim at distinguishing between melanocytic and non-melanocytic lesions as well as identifying more than one type of skin cancer. Several classifiers have been used for the diagnosis task: instance-based, decision trees, Bayesian classifiers, artificial neural networks (ANNs), support vector machines (SVMs), and ensemble methods.

The number of works describing different methods to perform one or more of the aforementioned steps is large. Thus, it is important to summarize them in surveys that guide new researchers in the field [100], [36], [136], [145], [147]. To the best of our knowledge, only one of these surveys focus solely on one of the blocks of the CAD system [36], while the remaining try to extensively cover all of the blocks. This can be overwhelming when one is trying to find better approaches to this problem. This issue has motivated us to write an overview focused on only one of the blocks of the CAD system: feature extraction. The choice of this block is not coincidental, since finding appropriate descriptors is of major importance both in segmentation and diagnosis. Moreover, feature extraction is by far the most variable step across related works.

The remainder of this paper is organized as follows. Section II provides a summary of pre-processing methods that can be used to improve the feature extraction process. The four feature extraction methodologies (hand-crafted, dictionary based, deep learning, and clinically inspired) are respectively described in Sections III through VI. Section VII discusses these methodologies and gives suggestions for future directions. Finally, Section VIII concludes the paper.

## II. Pre-Processing

The goal of the pre-processing step is to improve the quality of dermoscopy images, ensuring a better performance in the lesion segmentation and feature extraction blocks. A summary of pre-processing methods can be found in [36]. These can be divided into two different groups: i) image enhancement and ii) artifact removal.

### A. Image enhancement

This group comprises the following operations: color normalization/calibration, contrast enhancement, and color space transformation. Dermoscopy images can be acquired using different types of digital cameras and illumination conditions. This introduces significant variability in the color properties of the images, making unreliable the use color information to diagnose the lesions. To deal with this issue, it is possible to perform a color normalization step that will greatly improve the contribution of color information, as was experimentally demonstrated in [14]. Color normalization techniques can be hardware or software-based [36]. The former require knowledge about specific properties of the acquisition device (*e.g.*, camera offset, color gain, and aperture), in order to estimate a transformation matrix that can be used to convert the images to a device independent color space [72], [67], [192], [151]. Sometimes it is not possible to access the aforementioned information, which limits the application of hardware-based methods. On the other hand, software-based methods do not require prior knowledge of the acquisition setup, and use image properties to perform the normalization [31], [78], [14], [112].

Contrast enhancement is a popular technique to improve the performance of lesion segmentation methods, since their primary goal is to increase the contrast between the border of the lesion and the surrounding skin [169], [4]. As we will see in the following sections, an accurate segmentation may be important for the extraction of informative features, even though the relationship between the accuracies of segmentation and feature extraction has not been established conclusively yet, see, for example, [63], [70].

The default color space of dermoscopy images is RGB (red-green-blue). However, there are several drawbacks associated with this color space, including: i) it is not perceptually uniform; ii) it is not device independent; iii) there is a high correlation between the channels. This has motivated the use of alternative color spaces, such as CIE L*a*b*, CIE L*u*v, HSV, and opponent, to extract relevant color information [147]. Recently Madooei and Drew [115] proposed the use of a bioinspired color space that models the skin coloration and allows the separation of the melanin and

hemoglobin components. Their color space is called MHG, which stands for melanin, hemoglobin, and geometric mean (intensity information). The experimental results showed that the proposed color space outperformed the RGB, HSV, and L*a*b* color spaces in the lesion diagnosis task. In addition to the multichannel processing, it is also common to convert the RGB dermoscopy image to a single-channel to either perform lesion segmentation or to extract texture features [36]. The following methods can be used: i) a predefined or adaptive luminance transformation; ii) performing Karhunen-Loéve transformation and then retaining the channel with the highest variance; iii) the selection of the blue channel or the one with the highest entropy.

### B. Artifact removal

The artifacts usually found in dermoscopy images can be of two different types: i) acquisition artifacts (air bubbles, reflections, ruler and ink markings, and black frames); and ii) cutaneous artifacts (skin lines, blood vessels, and hairs) [36]. These artifacts not only hamper the lesion segmentation process, which in turn leads to misleading shape and symmetry features (see Section III), but also add color and texture information that can be misconstrued to the lesion features.

The simplest strategy to deal with these artifacts is to apply a smoothing operator (*e.g.*, median or mean filter) to the dermoscopy image [59]. Such approaches work fairly well for some artifacts (skin lines, blood vessels, and thin hairs). However, smoothing should be exercised with care, since the use of large filter masks may result in the blurring of the image and consequent loss of texture properties and border definition. Artifact-specific algorithms have been proposed to deal with bubbles, reflections, black frames, and hairs. Hair removal is by far the most popular topic of research (*e.g.*, [59], [104], [2], [21]), and these algorithms comprise two main steps: hair identification and repair. The latter is usually performed using an inpainting algorithm to replace the space occupied by the hair with proper intensity or color values. Although hair removal has been shown to be very useful, it is necessary to pay special attention to the output of the algorithms, as they might introduce some blurring or texture deformation in areas covered with a lot of hairs. Moreover, the algorithms may treat dark and linear dermoscopic structures (*e.g.*, pigment network) as hairs, and remove them.

## III. Hand-Crafted Features

Hand-crafted features are extracted from the images to characterize desired properties. The selection of the most appropriate type of features and descriptors is problem specific and is usually based on the knowledge of the practitioners. Alternatively, one can use learned features, which are automatically derived from the datasets and do not require any prior knowledge of the problem. In the remainder of this section, we will address the different types of hand-crafted features used in DIA, while the learned features will be presented in Sections IV and V.

As stated above, the extraction of relevant hand-crafted features requires a certain level of knowledge about the classification problem. In the case of DIA, this knowledge comes from the methodologies designed by dermatologists. Among the different medical procedures, the one that has inspired the development of the early CAD systems and that still remains popular is the ABCD rule of dermoscopy. According to this method, skin lesions can be characterized based on four properties: asymmetry, border irregularity, color distribution, and number of dermoscopic structures. Therefore, CAD systems try to characterize this information using appropriate features, as we will see later. Most of these features are also suitable to reproduce the CASH rule, as exemplified in [1]. Although the ABCD method was proposed for melanocytic lesions, some of the highlighted color and texture features have also been used in the development of CAD systems that diagnose non-melanocytic lesions (*e.g.*, [175]).

### A. Asymmetry Features

The ABCD rule of dermoscopy gives the highest weight to the asymmetry criterion, making it a relevant cue for melanoma diagnosis. Experts consider that the asymmetry of a lesion should be evaluated with respect to its shape and especially colors and patterns. To achieve this goal, they start by identifying the major and minor axes of the lesion and then visually compare the opposite halves. It is possible for the lesion to be: i) fully symmetric (0 points); ii) asymmetric on one axis (1 point); or iii) asymmetric on two axes (2 points). The automatic assessment of asymmetry has been extensively addressed in the literature. For the sake of simplicity we will divide our analysis into shape and color/structures symmetry.

*1) Shape symmetry:* is computed with regard to the segmentation binary mask. Some works characterize the overall shape of the lesion, taking into account the assumption that benign lesions usually have small dimensions and are approximately circular. Thus, simple metrics that characterize these properties (area, perimeter, compactness index, rectangularity, bulkiness, major and minor axis length, convex hull, comparison with a circle, and eccentricity) have been used (*e.g.*, [41], [142], [34], [81], [156], [133], [87]). Other strategies include: Hu's moment invariants, wavelet invariant moments, Žunić compactness, symmetry maps, symmetry distance, and adaptive fuzzy symmetry distance [142], [156].

An alternative to the aforementioned methods is to mimic the clinical analysis and automatically find the two symmetry axes of the lesion [137]. Various strategies have been used to determine the axes, namely Fourier transform [44], longest and shortest diameter [137], principal component analysis [183], and symmetry maps [142]. After finding the two axes, the mirror-image of one of the halves is overlapped with its correspondent and the area of intersection between the two overlapping folds is computed. Finally, an asymmetry index calculated as twice the ratio between the areas of intersection and the lesion. This approach is then repeated for the second axis.

Recently, Satheesha et al. [167] proposed a methodology to extract depth information from dermoscopy images, in order to obtain a 3-D reconstruction of the lesion. The authors then used several moment invariants to characterize the 3-D shape, obtaining promising classification results.

Although shape features have been shown to provide discriminative information for melanoma diagnosis [167], [156], such features can only be computed when the lesion is contained within the image [34].

*2) Color and structures symmetry:* can be evaluated using an approach similar to the one described above, *i.e.*, symmetry with respect to the two axes [137]. Various strategies have been used to evaluate the color and pattern symmetry. The simplest one consists of comparing the color components or intensity values of symmetric pixels [203], [137]. More complex methods require the positioning of a regular grid on top of the lesion. This grid is oriented along one of the axes and allows the extraction of symmetric lesion blocks. The idea is to compare the color and texture features of the blocks (*e.g.*, by computing the Euclidean or Kullback-Leibler distance between the feature vectors of the symmetric blocks), in order to obtain a statistical-based characterization (*e.g.*, maximum distance, minimum distance, and standard deviation) of their symmetry. Simple color and texture descriptors have been used to characterize the blocks, such as the mean color vector in the L*a*b* or HSV color spaces [171], [156].

Other works propose the use of fractal dimension [122] or radial analysis [43] to assess the pigment distribution on the surface of the lesion and its irregularity. Liu et al. [108] proposed a methodology to divide the image into the melanin and hemoglobin color components, followed by the analysis of asymmetry based on the distribubtion of these two components.

Similarly to shape, color and structures symmetry can only be assessed when the lesion is fully contained in the image.

### B. Border Features

The clinical evaluation of the border is a semiquantitative process in which the dermatologist is asked to determine the sharpness of the transition between the skin and the contour of the lesion. To perform this analysis, the physician starts by dividing the lesion into 8 slices and for each of them determines if there is an abrupt or indistinct (blurred) cut-off of the peripheral pigment pattern. If the transition is abrupt, the slice is scored 1, otherwise it is scored 0. It is assumed that melanomas exhibit a higher border score than benign nevi, but this is also the ABCD criterion with the lowest weight (only 0.1) [10].

Automatic methods have tried to reproduce the clinical analysis using two distinct approaches. The first one consists of assessing the degree of sharpness, inverse of blurriness, of the border. Similarly to the clinical analysis, the goal is to compare the color of the border against that of the surrounding skin [57]. Some systems resort to the computation of the gradient along the border points [79], using the luminance or the blue channel. Then, they divide the lesion into octancts and for each compute the mean and standard deviation values of the gradient. The Day [49] method uses the luminance image to examine the transition of intensity between the border and the skin. First, the lesion is divided into 8 slices, followed by the identification of five equally spaced points along the contour of each slice. For each of these points an intensity profile is extracted, using a length of 30 pixels inside and outside the border, *i.e.*, 60 values are identified for each point. Finally, the least-squares method is applied to find the slope of each intensity profile, which is then used to quantify the blurriness of the border. A simple method was proposed by Manousaki [122], where they computed the standard deviation of the border using the luminance channel, and used that information to estimate a coefficient of variation. Korjakowska [82] proposed a methodology based on distance between border points and a bounding box containing the lesion segmentation, from which it is possible to compute the number of irregularities.

The second approach to evaluate the border of the lesion consists of assessing its degree of irregularity. Some works propose the compactness index (thinness ratio) $C = \frac{4A\pi}{P^2}$, where $A$ and $P$ are respectively the lesion's area and perimeter, as an indicator of irregularity. However, $C$ seems to be more appropriate to describe the shape symmetry of the lesion, for several reasons: i) lesions with different boundaries may have similar compactness indexes; ii) the index is too sensitive to segmentation noise; and iii) it does not describe well the overall shape of the border. More elaborate methods include the use of the fractal dimension [1], [98], wavelet transform [42], [62], Anova-based analysis [203], [137], and Fourier descriptors [156].

It is generally believed that border features may only be extracted from lesions that are: i) fully contained in the images; and ii) correctly segmented. However, recently Xie et al. [194] proposed a methodology to extract border features from incomplete lesions, showing that incomplete feature information can still improve the performance of the CAD system. One of the described features uses the convex hull to identify border concavities. Statistical measures are then computed for each of the concavities: mean and standard deviation or the diameter, span, and average thickness. The other proposed feature requires the identification of inner and outer borders and the estimation of the distance between the two.

### C. Color Features

Color has been extensively studied in DIA [117] and there are several works solely devoted to finding discriminative color features (*e.g.*, [182], [37]). The goal of these works is to characterize the color distribution inside the lesion, in order to mimic the color quantization performed by dermatalogists. Several approaches have been proposed, ranging from simple statistical measures to more complex color quantization based methods.

Popular statistical measures are the average, standard deviation, variance, skewness, maximum, minimum, and entropy, 1- or 3-D color histograms, and the autocorrelogram [122], [34], [23], [163]. All of these descriptors are computed over at least one of the color spaces presented in Section II. Besides the use of multiple color spaces, it is also common to divide the lesion into inner and border parts and compute separate color descriptors for each of them [203]. Whenever this information is available, it is possible to add features that express the

difference between the regions, *e.g.*, the ratio of the statistical measures from the border and inner regions [34], [30].

Quantization methods reduce the number of colors in the image to a small predefined value. The idea is to group pixels with similar color properties into regions, which can then be used to characterize the lesions. Several techniques have been used to group the pixels, namely k-means [35], [37], Gaussian mixture model (GMM) [137], and multi-thresholding [103]. Lee et al. [103] proposed a dark-middle-bright (DMB) color system. All of these techniques are unsupervised and thus do not try to establish an association between the identified regions and the six colors considered in the ABCD rule.

A strategy used in many of the works is the estimation of a relative color image, from which the aforementioned features can be extracted [182], [34]. This image is obtained by subtracting the average color of the surrounding skin from the lesion color. The goal of this step is to handle, among other things, different skin colors and pigmentations as well as the Tyndall effect [5].

### D. Visual Texture Features

The D in the ABCD rule stands for the identification of five dermoscopic structures. Each of these structures exhibits a specific visual pattern (*e.g.*, pigment network is a network like structure with dark lines over a lighter background, and dots/globules are round or oval structures of dark coloration and variable size). This has motivated the use of various descriptors that characterize the texture of a lesion, *i.e.*, the existence of repeated visual patterns.

The gray-level co-occurrence matrix (GCLM) has been extensively used to characterize the texture of the lesions (*e.g.*, [34], [81], [168], [83]). This matrix is computed over the grayscale image and allows the estimation of the joint probability of two pixels that are separated by a fixed distance. Many statistical measures, such as variance, correlation, homogeneity, contrast, and entropy can be estimated using this descriptor. Several of these descriptors may also be computed directly from the gray scale image. Higher order statistics can be computed using the gray level run-length matrix (GLRLM) [83]. The gradient information (amplitude and orientation histograms and the histogram of oriented gradient) was also used to characterize the texture of the lesions [23], [152], as well as variants of local binary patterns (LBP) that compare the intensities of neighbor pixels inside a cell [137], [5]. Other texture descriptors that have been used include the wavelet and Fourier transforms [5], [163], fractal dimension [62], multidimensional receptive fields histograms, Markov random fields [164], and Gabor filters [152]. Similarly to color features, texture ones can also be computed separately for the inner and border parts of the lesion (*e.g.*, [30], [23]).

## IV. Dictionary-Based Features

The various descriptors presented in Section III allow a global characterization of the lesion. Such methodology is appropriate to represent some of the criteria of the ABCD rule (*e.g.*, the border or the overall color distribution). However, the D criterion, which stands for dermoscopic structures, corresponds to localized color and/or texture patterns. By using a global description of the lesion, one might miss these relevant cues. This has motivated several authors to pursue a different direction, and use local features to characterize the color and texture of the lesions [23], [152], [45], [153], [200], [19], [6].

The basic idea of local features is that an image (skin lesion) can be represented as a collection of elements (atoms) of a dictionary of *visual words*. This dictionary is usually learned in an unsupervised manner, as we will discuss later. The simplest and most popular method for local feature extraction is called bag-of-features (BoF), which is inspired by the traditional bag-of-words for text analysis [179]. Assuming that a dictionary $D$ of $K$ atoms is known, a dermoscopy image is processed as follows: i) divide the image into $M$ patches and characterize each of them using a feature vector $x_m \in \mathbb{R}^p$; ii) match the features to the closest dictionary atom, typically using the Euclidean distance; and iii) count the number of times each atom was selected and store this information in a histogram of occurrences, which is then used as a lesion descriptor.

In most cases, $D$ is unknown and must be estimated using a training set of $N$ feature vectors $\{x_1, ..., x_N\} \in \mathbb{R}^p$, extracted from the patches of several images. This task is accomplished using a clustering algorithm, such as k-means [33]. Variations to the BoF model can also be found in the literature, such as the work of Fornaciali et al. [60] that combines the BossaNova algorithm with a spatial circular pooling to replace the traditional vector quantization of BoF.

The assumption that each patch corresponds to a single atom of $D$ may be too restrictive. Therefore, some groups have explored the sparse coding (SC) formulation [121]. As in BoF, patches are extracted from the images and characterized using various features. However, SC assumes that each of these vectors is approximated by a combination of a small number of atoms, instead of being associated with a single one. This increases the complexity of the problem, requiring alternative methods to: i) find the small set of atoms that best represents a patch; ii) estimate $D$; and iii) represent the images [121]. Points i) and iii) have been addressed similarly in most of the dermoscopy works, while some variability has been observed in the estimation of $D$. Most of the works use either the K-SVD [153] or ODL algorithms [45], [19], which learn a single dictionary to represent all types of skin lesions. However, Yao et al. [197] suggest that this representation is not discriminative enough to distinguish melanomas from benign lesions, since both share a number of common patterns. To tackle this issue, they investigated the learning of class-specific dictionaries (that contain only discriminative information) as well as a common dictionary (to aggregate the shared patterns), as described in [99].

Recently, Yu et al. [201] explored the use of the Fisher vectors [166], which have been shown to outperform BoF in classification tasks. The first difference between this strategy and the previous ones is the use of GMM to estimate a probabilistic dictionary to represent the data. Each image is then characterized by a Fisher vector, which contains information about the deviation of the image patches with respect to the parameters of the estimated GMM [166].

Patch extraction and representation have also been thoroughly investigated in dermoscopy works. The simplest strategy is to divide the lesion into squares using a regular grid, this is called dense sampling. Alternatively, one can inspect the lesion for the presence of salient points that are associated with specific color and texture patterns, and extract square patches centered on them (sparse sampling). Some popular keypoint detectors include the difference of Gaussians and Harris-Laplace [22]. Each of the extracted patches will comprise a set of neighbor pixels, forming a so-called superpixel. In addition to the methods described above, there is extensive literature on the extraction of superpixels, in particular in grouping together pixels that share color and/or texture properties. To the best of our knowledge such approaches has never been applied to dictionary-based features.

Multiple descriptors have been used to represent the patches, ranging from raw color and intensity pixel values [45] to many of the color and texture descriptors discussed in Section III (*e.g.*, color and gradient histograms, Gabor filters, and LBP [23], [24], [152], [6]). Specialized local descriptors such as SIFT, color-SIFT, and GHIST have also been applied to dermoscopy [152], [5]. More recently, deep learning features have been tested as well [157], [201].

## V. Deep Learning Features

Since Krizhevsky et al. [101] won the 2012 ImageNet challenge, it is undeniable that deep learning, and convolutional neural networks (CNN) in particular, have become the technique of choice in many computer vision problems [69]. The community of medical imaging has also embraced this technique, with an ever increasing number of applications that use this methodology to either diagnose or segment organs and structures in medical images [107]. To the best of our knowledge, the first work to apply deep learning to dermoscopy images was published by Codella et al. [45] in 2015, and since then several authors have explored deep learning to learn suitable image representations and achieve (near) human expert diagnosis performance (*e.g.*, [58]). An indication of the popularity of deep learning in the dermoscopy field is the 2017 ISIC challenge [46], where 22 out of 23 works used at least one type of CNN architecture.

The basis of deep learning methods is an ANN. This kind of learning algorithm is composed of units (called neurons), each with a specific activation function $f$ and parameters $\theta = \{w, b\}$, where $w$ is a set of weights and $b$ is a set of biases. The activation function consists of a linear combination of the neuron's input $x$ with the parameters, followed by the application of an element-wise nonlinearity $\eta$: $f = \eta(w^T x + b)$.

A common choice for $\eta$ is the rectified linear unit (ReLu) function that sets the negative values to zero. Deep learning methods comprise several layers of these transformations. During the training phase, it is necessary to determine the set of model parameters. The most popular strategy is the stochastic gradient descent using mini-batches of the data, in which the goal is to minimize a specific loss function (*e.g.*, cross-entropy loss for classification problems and reconstruction loss for unsupervised methods).

There are two types of deep learning methods: unsupervised (*e.g.*, stacked auto-encoders and deep belief networks) and supervised (*e.g.*, CNNs and recurrent NNs). Almost all of the works in dermoscopy use CNNs to diagnose skin lesions. Such methods exhibit a particular architecture, where it is possible to observe three main layers (convolutional, pooling, and fully connected), organized in a hierarchical framework. In the convolutional layers, the image as well as intermediate feature maps are convolved with kernels of various dimensions. This is followed by the application of a nonlinearity, usually ReLu. Each of these layers produces increasingly abstract feature maps. A pooling layer appears after a convolutional layer, and the idea is to reduce the dimensions of the feature maps and network parameters. Average and max pooling strategies are the most popular. Fully connected layers appear at the end of the network, after the last pooling layer, and perform like a traditional ANN. The goal of these layers is to convert the 2D feature maps into a 1D vector that can be used for classification. The output of the final layer is usually fed to a *softmax* function to obtain a distribution over classes, which can be used for classification [69]. Alternatively, one can use the outputs of the fully connected layers as inputs for more sophisticated classification methods, such as SVM [45]. Some architectures also include a batch normalization layer before applying the nonlinearities (*e.g.*, [187]), which regularizes the model and speeds up its training.

CNNs are trained end-to-end in a supervised fashion (where the only ground truth is the diagnosis), usually using stochastic gradient. Since ANNs are prone to overfitting, it is common to include a dropout layer during the training phase, in which some of the activations or neurons are randomly omitted at each epoch [181]. The training process requires a large amount of data, which unfortunately is not available in dermoscopy. To deal with this issue, several authors have adopted a transfer learning strategy. This method consists of using one or more of the available pre-trained CNN architectures to extract features, and retrain only the fully connected layers to diagnose dermoscopy images (*e.g.*, [45], [128], [150], [89], [64]. Similarly to other research fields, the most popular pre-trained architectures are those that were the winners or runner-ups of the ImageNet challenges (AlexNet [101], VGG [178], GoogLeNet - Inception [187], and ResNet [75]). AlexNet [101] won the challenge in 2012, setting the tone for the following years. It consists of five convolutional and three fully connected layers. GoogLeNet (twenty one convolutional and one fully connected layers) [187] and VGG (thirteen/fifteen convolutional and three fully connected layers) [178] ranked respectively the first and second in the 2014 challenge, both demonstrating the possibility of training deeper networks to achieve better results. Inception layers that consist of applying multiple kernels with various dimensions at the same time, were also introduced [187]. The winner of the 2015 challenge, ResNet [75], proposed the use of residual layers to train even deeper networks, setting the record of more than 100 layers. ResNet is currently the state-of-the-art network in many applications.

Pre-trained networks can also be fine-tuned, meaning that the known CNN parameters are used as an initialization for a

TABLE I
CNN ARCHITECTURES USED IN DERMOSCOPY CAD SYSTEMS

| CNN Architecture | From Scratch | Transfer Learning | Transfer Learning + Fine Tuning |
|---|---|---|---|
| AlexNet [101] | [27][a] | [150][b][89][a][120][ac][201][a] | [180][ac][74][ac] |
| VGG [178] | [110][a] | [110][a][120][ac] | [129][a][110][a][64][ac][74][ac][85][a] |
| GoogleLeNet - Inception [187] | [140] | [38][140][190][ab] | [58][134][a][52][140][196][a][74][ac][180][ac][130][a] |
| CaffeCNN [86] | - | [45] [46][c] | [202][a] |
| ResNet [75] | - | [46][c] | [64][ac][200][a][61][b][124][130][a][55][a][29][74][ac] |

[a] Geometry-based data augmentation. [b] Color-based data augmentation. [c] Ensemble of network architectures.

new model. This model is trained using a labeled dermoscopy dataset (*e.g.*, [58], [130], [110], [61]). Most of the works include a data augmentation step when using fine-tuning to: i) increase the amount of data; ii) deal with class imbalance; and iii) prevent overfitting. Some methods apply geometric transformations to the original images (rotation, scaling, flipping, shift, and cropping) [64], [110], while others have created artificial examples through color transformations [150], [61]. Vasconcelos et. al. [190] combine geometric and color augmentation strategies, while Yoshida et al. [198] compare the performance of various geometric transformations. However, at this point at this point, the most suitable strategy to deal with dermoscopy images remains unclear. Data augmentation also allows the training of a CNN architecture from scratch, as explored in [128], [110]. However, results were worse than that of transfer learning. Some authors obtained improved results by combining more than one CNN architecture (*e.g.*, [64], [120]). Table I summarizes the works that use each type of CNN architecture, as well as their training and data augmentation strategies.

Recently, other deep learning architectures have been used to extract features from dermoscopy images, namely auto-encoders [26] and fully-convolutional neural networks (FCN) [109]. The former is a type of unsupervised deep learning method that has been used to extract meaningful features from dermoscopy images [95] and in combination with the BoF model [157]. Both works have shown improvements over more traditional CNN frameworks. FCNs allow a pixel-based analysis of the skin lesions, as has been explored in [105]. U-net [155] is an FCN architecture that has been used in dermoscopy mostly to segment skin lesions, but Codella et al. [47] have also used this architecture to provide information about the shape of the lesion.

## VI. CLINICALLY INSPIRED FEATURES

In Sections III-V we presented a chronological survey of the different types of low-level features that have been used in dermoscopy. These features led to promising experimental results, some of which were stated to be on par with the performance of dermatologists (*e.g.*, [58]). However, the development of CAD systems based on a combination of the aforementioned features has always been met with caution by the medical community [56]. Hand-crafted features have been criticized for their lack of medical meaning, even if they are inspired by the ABCD rule, as it is not easy to translate descriptors such as color histograms or Gabor filters into dermoscopic criteria that can be understood by the physicians. Dictionary based features try to mimic the localized search and analysis of dermoscopic criteria, but they lead to abstract descriptors. As the works start to converge towards deep learning features, the original concept of using representations that somehow characterize medical properties is becoming less relevant. Moreover, it is not easy to interpret the outputs of the different layers of a CNN.

The aforementioned feature-related problems, as well as the desire of the dermatologists to work with CAD systems that can serve as decision support and learning tools [56], fostered an alternative trend in DIA: the development of clinically inspired CAD systems. The core of these systems is the design of strategies to detect dermoscopic criteria, followed by the use of those criteria to diagnose skin lesions. This is believed to mimic the analysis performed by an expert and at the same time provides them with grounds to understand the diagnosis. The number of works related to this topic has been slowly increasing over the past decade [100], [145], [147] (in the 2017 ISIC challenge [46] only three submissions attempted to detect a subset of dermoscopic criteria).

There are two classes of dermoscopic criteria that have been addressed in the literature: i) global patterns and ii) localized dermoscopic structures and colors. The works that cover each of these categories can be seen in Table II.

Two strategies have been used to detect global patterns. One of them consists of extracting several of the hand-crafted features described in Section III for the entire lesion, followed by the application of a classification method to identify the pattern [80], [3], [195]. The color features include quantization and relative color methods [80], [3], as well as color statistics [80], all computed in several color spaces. GLCM [80], steerable pyramids transform [3], and gradient features [195] have been used as texture descriptors, while shape and border have been respectively characterized using simple shape descriptors and the blurriness metric proposed in [79]. Alternative methods have been developed to work with small lesion patches. First, each patch is characterized by a feature vector, such as Laws' masks [160], GLCM [188], Markov random fields [172], [165], or CNN [51]. Then, the vectors can be used either directly in a multiclass classification strategy [188], [51], or grouped to create templates that represent each of the patterns [160], [172], [165].

According to Table II, the majority of the clinically inspired methods attempt to detect localized dermoscopic criteria. A popular topic of research is the detection of colors, namely the six colors identified by the ABCD rule [170], [149], [171], [123], [106], [20], [158], [17], [16]. The traditional starting point of these methods is the construction of a palette that contains a representative number of examples of each of the colors. This palette is built by asking one or more

TABLE II
WORKS ON THE DETECTION OF CLINICALLY INSPIRED FEATURES. SYSTEMS IDENTIFIED WITH AN "$^a$" USE THE DETECTED CRITERIA TO DIAGNOSE SKIN CANCER.

| Dermoscopic Criteria | Works |
|---|---|
| Global patterns | [188], [172], [127], [160], [51], [195],[80]$^a$ [77]$^a$[1]$^a$[165]$^a$ |
| Colors | [149], [171], [123], [173], [158], [88],[170]$^a$[77]$^a$[132]$^a$[106]$^a$[20]$^a$[17]$^a$[16]$^a$ |
| Blue-whitish veil | [53], [116], [118],[32]$^a$[54]$^a$[16]$^a$ |
| Regression structures | [50], [53], [185],[54]$^a$[48]$^a$[16]$^a$ |
| Hypopigmentation | [50],[48]$^a$ |
| Blotches | [139], [149], [184], [114] |
| Pigment network | [59], [8], [66], [162], [161], [193], [21], [11], [113], [73], [25], [90], [13],[28]$^a$[54]$^a$[7]$^a$[105]$^a$[16]$^a$[146] |
| Dots/globules | [59], [50], [199], [189], [84][119]$^a$[16]$^a$ |
| Streaks | [135], [159], [90],[54]$^a$[105]$^a$ |
| Vascular structures | [28]$^a$[54]$^a$[39]$^a$[94]$^a$[96]$^a$ |
| Negative network | [90][105]$^a$ |
| Non-melanocytic criteria | [92]$^a$[174]$^a$[40]$^a$[71]$^a$[93]$^a$[90][105]$^a$ |

dermatologists to segment small color patches from different dermoscopy images. Each of the patches is then characterized by a small feature vector, such as the mean and standard deviation values of one or more colors channels - the RGB [170], HSV [20], and L*a*b* [158] spaces have been used in different works. Some methods use the aforementioned patches to label pixels or regions in test images, according to the closest patch (*e.g.*, [158]). Other works perform an intermediate clustering operation, using k-means [170] or GMMs [20], to group patches of the same color into a smaller set of centroids. These centroids are then used to label pixels or patches of new images. Similar patch-based methodologies have been proposed to detect color structures, namely blue-whitish veil [32], [53], [54], [116], regression areas [53], [54], [48], [185], and regions of hypopigmentation [48], [12], [91]. Other investigated types of abnormal pigmentation are dark blotches [139], [149], [184], [114] and blue or gray areas [173], [106]. Various methodologies have been proposed to detect them, ranging from a simple thresholding algorithm applied to absolute and relative color features [184], [173], to fuzzy clustering [114].

The most thoroughly analyzed dermoscopic structure is the pigment network, as can be seen in Table II. Due to the particular geometry of pigment network (dark connected lines over a lighter background), several works applied filtering techniques to either highlight and extract the lines or the holes, followed by a classification strategy to decide if the network is present or not (*e.g.*, [161], [21], [11], [73]), and, in some cases, identify atypical networks (*e.g.*, [28], [162], [146]). Strategies that use manual segmentations of pigment network regions to extract features and train a classifier have emerged, namely using FCN [90], [105] and fuzzy clustering [13]. Streaks, dots, and globules also exhibit specific shapes and colorations. Thus, their detection has also been accomplished using filtering methods [199], [84], [135], [159] and, more recently, FCN [105], [90]. The detection of vascular structures has been attempted using color [28], texture [54], a combination of both types of features [94], and more recently auto-encoders [96].

The ISIC challenge [46] led to the proposal of the first methodologies to detect negative network and milia-like cysts [90], [105], both based on FCN. The latter criterion is associated with non-melanocytic lesions. A few works attempted to detect other criteria associated with either basal cell carcinoma (dirt trail [40], specific coloration [71], and ulcers [92]), or squamous cell carcinoma (scale crust [174]).

The major limitation of clinically inspired features is that most of the methods rely on the existence of manual segmentations of the criteria. These segmentations are used to train and validate methods. However, it is very hard to obtain them, since dermatologists find manual segmentation time consuming. The absence of segmentations makes it impossible to apply the algorithms. Recently, various research groups tried to deal with this problem, by developing methodologies that are based on the use of weakly annotated data, where the training data comprises the entire images and a set of text labels [118], [17], [16], [111].

## VII. DISCUSSION AND FUTURE DIRECTIONS

We have described four classes of features that have been used in DIA. This is an ever growing field, and a quick search of the literature will provide the reader with a multitude of CAD systems, each using a combination of descriptors belonging to one or more of the described classes. The experimental results are usually very promising. However, they should be approached with caution, mainly because the performance of a CAD system is strongly dependent on the dataset. To exemplify this issue we have selected three non-commercial and publicly available datasets ($PH^2$ [125], [126], ISBI 2016 [70], and ISBI 2017 [46]), with varying degrees of difficulty. For each dataset we have then selected a set recent works that use them (see Table III). In particular, for ISBI 2017 we have selected the three top and bottom ranked submissions of the associated challenge [46]. $PH^2$ is the smallest set, comprising only 200 images. ISBI 2016 and 2017 are larger and are divided into training, validation, and test sets. Additionally, ISBI 2017 contains lesions of three classes (melanoma, nevi, and seborrheic keratosis), while the other two datasets only contain melanoma and nevi. These levels of difficulty clearly influence the results: the scores for $PH^2$ are significantly higher (the exception is [194], which combines this dataset with some images from [10]).

Table III also shows two trends: i) there is a convergence towards the use of deep learning features; and ii) several of the works rely on the softmax, SVM, and random forests classifiers. Deep learning achieved impressive results in many medical problems [107], particularly in dermoscopy [58].

TABLE III
MELANOMA DIAGNOSIS SCORES USING THREE PUBLICLY-AVAILABLE DATASETS.

| Dataset (Lesions/Melanomas) | Ref. | Features | Classifier | SE | SP | Pre | ACC | AUC |
|---|---|---|---|---|---|---|---|---|
| PH$^2$ [125] (200/40) | [18] | Hand-crafted (C,D) + Dictionary (BoF) | Random Forests | 98.0% | 90.0% | - | - | - |
| | [153][a] | Dictionary (SC) | Random Forests | 100% | 93.0% | - | - | - |
| | [156][a,c] | Hand-crafted (A) | kNN | 96% | 83.0% | - | - | - |
| | [167] | Hand-crafted (B,C,D) | SVM | 96% | 97.0% | - | - | - |
| | [6] | Dictionary (BoF) | kNN | 99.4% | 98.2% | - | - | - |
| | [194][b] | Hand-crafted (B,C,D) | Softmax | 83.3% | 95.0% | - | - | - |
| | [16] | Clinically inspired | Random Forests + SVM | 100% | 88.2% | - | - | - |
| ISBI 2016 [70] | [200][a] | Deep Learning (ResNet) | Softmax+SVM | 54.7% | 93.1% | 62.4% | 85.5% | 78.3% |
| | [47] | Hand-crafted (C,T) + Dictionary (SC) + Deep Learning | SVM | 69.3% | 83.6% | 64.9% | 80.7% | 83.8% |
| | [129][a] | Deep Learning (VGG) | Softmax | 47.6% | 88.1% | 54.9% | 79.2% | 80.7% |
| Train. Set (900/273) | [144][c,d] | Hand-crafted (A,B,C,D) | OPF | 91.8% | 96.7% | - | - | - |
| | [190][a] | Deep Learning (GoogleLeNet) | Committees | 74.6% | 84.5% | 66.9% | 82.5% | - |
| Test Set (379/75) | [201][b] | Deep Learning (AlexNet) + Dictionary (FV) | SVM | - | - | 53.5% | 83.1% | 79.6% |
| ISBI 2017 [46] | [124][a,b] | Deep Learning (ResNet)+Metadata | Softmax | 73.5% | 85.1% | 71.0% | 82.8% | 86.8% |
| | [55][a] | Clinically Inspired + Deep Learning + Metadata | Softmax | 10.3% | 99.8% | 65.4% | 82.3% | 85.6% |
| | [29][a,b] | Deep Learning (ResNet) + Metada | Softmax | 42.7% | 96.3% | 69.4% | 85.8% | 87.0% |
| Train. Set (2000/374) | [202][a] | Deep Learning (CaffeCNN) | Softmax | 0% | 100% | 59.8% | 80.5% | 50.0% |
| Val. Set (150/30) | [134] | Deep Learning (GoogleLeNet) | Softmax | 41.9% | 82.8% | 45.2% | 74.8% | 62.3% |
| Test. Set (600/117) | [68] | Deep Learning (VGG) + Hand-crafted (C,T) | Softmax + Random Forests | 6.8% | 88.2% | 18.7% | 72.3% | 47.5% |

[a] Data augmentation. [b] Additional training images. [c] Exclusion of some images.[d] Combines training and test sets.

From the results of ISIC 2017, it is possible to postulate that deep learning features seem to be the best for the melanoma problem [46]. However, such claim may only hold in some scenarios, since the best and worse results were both achieved using these features. This makes it critical to understand the real strengths and limitations of deep learning in the context of dermoscopy. A comparison between deep learning and the other three classes of features is also needed. Based on Table III and on our literature survey, we have identified a set of open questions:

**i) Given the same conditions (amount of training data and classifier) which is the best class of features?** The comparison between at least two classes of features has been explored. However, some of the results are contradictory, *e.g.*, in [152] it was concluded that BoF performed worse than global hand-crafted features, while in [18] the opposite was reported. In this case, both works used random forests, but used different training and test datasets. A comparison of hand-crafted, SC, and deep learning features was conducted in [45], using the same classifier and dataset. Dictionary-based features performed poorly compared to the others. This may be due to the strategy used to built the dictionaries (raw patches), since [197], [19] experimentally demonstrated that the selected descriptors have a great influence on the performance of SC. Similar observations were made for hand-crafted features, where it was shown that selecting different descriptors to characterize the same type of feature can lead to completely different classification scores. Although we focused on the feature extraction block in this survey, we are fully aware of the importance of the classifier in the diagnosis, as was recently shown in [138], using just two classifiers.

**ii) How can we establish synergies between the different classes of features?** Some works have already explored this issue at the classification level and reported increased performance (*e.g.*, [77], [203], [18], [45], [46], [111], [55]), which suggests that all classes convey relevant information. Moreover, feature can be combined at other levels, as will be discussed next.

**iii) What are the most robust classes of features?** The goal of the CAD systems is to help the dermatologists in their practice. This means that these systems may be required to analyze images that were acquired in significantly different conditions than those of the images they were trained with. Thus, identifying if these four classes of features are equally robust or if they are strongly dependent on the training set is critical and is still unexplored in the literature.

**iv) Are all of the features equally suitable to deal with any type of skin lesion (melanocytic and non-melanocytic)?** Melanoma is the less common type of skin cancer, but it is undeniably the most aggressive [10]. This and the number of scoring rules specifically designed to diagnose melanomas, fostered the development of CAD systems solely devoted to distinguishing melanomas from nevi. Non-melanocytic lesions have been disregarded, but the more recent datasets (*e.g.*, ISBI 2017), include lesions of this type. This makes one wonder if all of the four classes of features are equally discriminative in this multi-class scenario. A clear example are the clinically inspired ones, since both melanocytic and non-melanocytic lesions are associated with different dermoscopic structures. Another example are the hand-crafted border features. Melanomas are assumed to have irregular borders, when compared with nevi, but there are non-melanocytic lesions whose borders are even more irregular, such as the solar lentigos [97].

In addition to the aforementioned issues, which we believe are the most relevant, the following feature-specific points require proper study:

**i) Hand-crafted features:** Asymmetry, border, color, and texture descriptors tend to be grouped together into a single feature vector that is used for classification purposes. Such an approach is called early fusion, and has been shown to lead to poorer performance than training a classifier for each descriptor and then performing a score fusion (late fusion) [18].

**ii) Dictionary-based features:** have become the least explored class of features. Nonetheless, we believe that their value is yet to be determined. Recent works demonstrated the good performance of these features (either alone [6] or combined with deep learning [157], [200]). There are some aspects that should be properly studied: i) learning class-

specific dictionaries (as was briefly explored in [197]), a suitable approach to deal with imbalanced data, such as dermoscopy data; ii) learning discriminative dictionaries and the parameters of a classifier simultaneously [121], this is similar to a deep learning framework where the system learns discriminative local information for a type of lesion and the best classifier to diagnose it simultaneously; and iii) patch/superpixel extraction. Regarding the latter, all of the works have either used a regular grid or keypoint detectors to divide the lesion into small parts, which may not guarantee that all of the pixels in the extracted region share the same properties. More elaborate methods that enforce color/texture similarities between the pixels should be investigated, since this will allow the sampling of the lesion into regions that may have some medical relevance, without requiring expert annotation.

**iii) Deep learning features:** The most popular architectures have all been applied to dermoscopy images, but with a bias towards ResNet [75] and GoogLeNet [187] (see Table III and [46].) Although this could hint at a better performance of these architectures against others, the truth is that it is still not clear which is the best one among all of them. Another interesting issue is the importance of training data. From Table III, one observes that a difference between the top and bottom ranked submissions of the 2017 challenges is the use of additional training data, both from data augmentation and external sources It is well known that a proper training of a CNN requires large amount of data, but the questions in our field should be: i) How much data is enough to train the network?; ii) Should we use all of the augmentation strategies or some of them are more appropriate than others (*e.g.*, geometric vs. color manipulations); and iii) Is all of the data equally relevant or are we augmenting (repeating) examples that are not informative? This last question is valid for any type of feature-classifier configuration. Moreover, the full potential of deep learning has not been explored yet, such as using it to improve the detection of clinically inspired features or to characterize the patches/superpixels used to compute dictionary-based features.

**iv) Clinically inspired features:** have been addressed in several works. However, few of these works attempted to use these features to diagnose skin lesions. This leads to the following question: are these features useful enough to be included in a CAD system? Some works experimentally demonstrated that clinically inspired features are discriminative (*e.g.*, [149], [32], [54], [94], [16], [111], [146]), and in [16] it was shown that clinically inspired features not only achieved similar performance to that of hand-crafted and dictionary-based ones, but also that those results were verified on various datasets. Thus, we advocate the need to include a lesion-diagnosis based evaluation in works that detect clinically inspired features, since this will provide the readers with critical information. Moreover it is important to augment the number of detected criteria, since there is a set of them that is clearly associated with malignancy [97]. However, most of the works focus on a single criteria, and not many efforts have been made to integrate several detectors into a CAD system (*e.g.*, [54], [48], [16], [111], [105]). There are also criteria that have not been addressed, while other criteria were studied in multiple works, as can be seen in Table II. Other relevant criteria considered by experts are clinical covariates, such as the age, gender, and familial history of the patients [191]. Such information has been scarcely used in CAD systems, but the ISBI 2017 dataset contained this information for some of the images [46]. The results in Table III suggest that this information could have also played a role in the performance of the best systems, but further research is needed to clarify this.

## VIII. Conclusions

This paper presented a thorough review of the four classes of features used in DIA: i) hand-crafted; ii) dictionary-based; iii) deep learning; and iv) clinically inspired. We briefly explained relevant pre-processing techniques, reviewed all of the features, and provided a comprehensive explanation of their use and importance in dermoscopy. Finally, we presented a critical discussion of the various features and provided guidelines for future research.

We believe that feature extraction is one of the most important parts of DIA, and that there is still room for improvements and relevant contributions. We hope that this work will be a valuable guide for researchers to make advances in the field.

## References


[1] Q. Abbas, M. E. Celebi, and I. Fondón, "Computer-aided pattern classification system for dermoscopy images," *Skin Research and Technology*, vol. 18, pp. 278–289, 2012.

[2] Q. Abbas, M. E. Celebi, and I. F. García, "Hair removal methods: a comparative study for dermoscopy images," *Biomedical Signal Processing and Control*, vol. 6, pp. 395–404, 2011.

[3] Q. Abbas, M. E. Celebi, C. Serrano, and et al., "Pattern classification of dermoscopy images: A perceptually uniform model," *Pattern Recognition*, vol. 46, pp. 86–97, 2013.

[4] Q. Abbas, I. Fondón, and M. Rashid, "Unsupervised skin lesions border detection via two-dimensional image analysis," *Computer methods and programs in biomedicine*, vol. 104, pp. e1–e15, 2011.

[5] M. Abedini, Q. Chen, N. C. F. Codella, and et al., "Accurate and scalable system for automatic detection of malignant melanoma," *Dermoscopy Image Analysis*, pp. 293–343, 2015.

[6] N. Alfed and F. Khelifi, "Bagged textural and color features for melanoma skin cancer detection in dermoscopic and standard images," *Expert Systems with Applications*, vol. 90, pp. 101–110, 2017.

[7] N. Alfed, F. Khelifi, A. Bouridane, and et al., "Pigment network-based skin cancer detection," in *IEEE EMBC 2015*, 2015, pp. 7214–7217.

[8] M. Anantha, R. Moss, and W. Stoecker, "Detection of pigment network in dermatoscopy images using texture analysis," *Computerized Medical Imaging and Graphics*, vol. 28, pp. 225–234, 2004.

[9] G. Argenziano, G. Fabbrocini, P. Carli, and et al., "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Archives of Dermatology*, vol. 134, pp. 1563–1570, 1998.

[10] G. Argenziano, H. P. Soyer, V. De Giorgi, and et al., *Interactive Atlas of Dermoscopy*. EDRA Medical Publishing & New Media, 2000.

[11] J. L. G. Arroyo and B. G. Zapirain, "Detection of pigment network in dermoscopy images using supervised machine learning and structural analysis," *Computers in biology and medicine*, vol. 44, pp. 144–157, 2014.

[12] ——, "Hypopigmentation pattern recognition in dermoscopy images for melanoma detection," *Journal of Medical Imaging and Health Informatics*, vol. 5, pp. 1875–1879, 2015.

[13] ——, "Recognition of pigment network pattern in dermoscopy images based on fuzzy classification of pixels," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 61–69, 2018.

[14] C. Barata, M. E. Celebi, and J. S. Marques, "Improving dermoscopy image classification using color constancy," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, pp. 1146–1152, 2015.

[15] ——, "Towards a robust analysis of dermoscopy images acquired under different conditions," *in Dermoscopy Image Analysis*, pp. 1–22, 2015.

[16] ——, "Development of a clinically oriented system for melanoma diagnosis," *Pattern Recognition*, vol. 69, pp. 270–285, 2017.

[17] C. Barata, M. E. Celebi, J. S. Marques, and et al., "Clinically inspired analysis of dermoscopy images using a generative model," *Computer Vision and Image Understanding*, vol. 151, pp. 124–137, 2016.

[18] C. Barata, M. E. Celebi, and J. Marques, "Melanoma detection algorithm based on feature fusion," in *IEEE EMBC 2015*, 2015, pp. 2653–2656.

[19] C. Barata, M. A. T. Figueiredo, M. E. Celebi, and et al., "Local features applied to dermoscopy images: Bag-of-features versus sparse coding," in *IbPRIA 2017*, 2017, pp. 528–536.

[20] C. Barata, M. A. T. Figueiredo, M. E. Celebi, and J. S. Marques, "Color identification in dermoscopy images using gaussian mixture models," in *IEEE ICASSP 2014*, 2014, pp. 3611–3615.

[21] C. Barata, J. S. Marques, and J. Rozeira, "A system for the detection of pigment network in dermoscopy images using directional filters," *IEEE Transactions on Biomedical Engineering*, vol. 59, pp. 2744–2754, 2012.

[22] ——, "Evaluation of color based keypoints and features for the classification of melanomas using the bag-of-features model," in *ISVC 2013*, 2013, pp. 40–49.

[23] C. Barata, M. Ruela, M. Francisco, and et al., "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE Systems Journal*, vol. 8, pp. 965–979, 2014.

[24] C. Barata, M. Ruela, T. Mendonça, and et al., "A bag-of-features approach for the classification of melanomas in dermoscopy images: The role of color and texture descriptors," in *Computer vision techniques for the diagnosis of skin cancer*. Springer, 2014, pp. 49–69.

[25] A. Benam, M. S. Drew, and M. S. Atkins, "A CBIR system for locating and retrieving pigment network in dermoscopy images using dermoscopy interest point detection," in *IEEE ISBI 2017*, 2017, pp. 122–125.

[26] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, pp. 1798–1828, 2013.

[27] M. Berseth, "ISIC 2017-skin lesion analysis towards melanoma detection," *arXiv preprint arXiv:1703.00523*, 2017.

[28] G. Betta, G. Di Leo, G. Fabbrocini, and et al., "Dermoscopic image-analysis system: estimation of atypical pigment network and atypical vascular pattern," in *IEEE MeMea 2006*, 2006, pp. 63–67.

[29] L. Bi, J. Kim, E. Ahn, and et al., "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks," *arXiv preprint arXiv:1703.04197*, 2017.

[30] G. Capdehourat, A. Corez, A. Bazzano, and et al., "Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions," *Pattern Recognition Letters*, vol. 32, pp. 2187–2196, 2011.

[31] M. E. Celebi, H. Iyatomi, and G. Schaefer, "Contrast enhancement in dermoscopy images by maximizing a histogram bimodality measure," in *IEEE ICIP 2009*, 2009, pp. 2601–2604.

[32] M. E. Celebi, H. Iyatomi, W. V. Stoecker, and et al., "Automatic detection of blue-white veil and related structures in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 32, pp. 670–677, 2008.

[33] M. E. Celebi, H. A. Kingravi, and P. A. Vela, "A comparative study of efficient initialization methods for the k-means clustering algorithm," *Expert systems with applications*, vol. 40, pp. 200–210, 2013.

[34] M. E. Celebi, H. Kingravi, B. Uddin, and et al., "A methodological approach to the classification of dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 31, pp. 362–373, 2007.

[35] M. E. Celebi, Q. Wen, S. Hwang, and G. Schaefer, "Color quantization of dermoscopy images using the k-means clustering algorithm," in *Color Medical Image Analysis*. Springer, 2013, pp. 87–107.

[36] M. E. Celebi, Q. Wen, H. Iyatomi, and et al., "A state-of-the-art survey on lesion border detection in dermoscopy images," *in Dermoscopy Image Analysis*, pp. 97–129, 2015.

[37] M. E. Celebi and A. Zornberg, "Automated quantification of clinically significant colors in dermoscopy images and its application to skin lesion classification," *IEEE Systems Journal*, vol. 8, pp. 980–984, 2014.

[38] H. Chang, "Skin cancer reorganization and classification with deep neural network," *arXiv preprint arXiv:1703.00534*, 2017.

[39] B. Cheng, D. Erdos, R. J. Stanley, and et al., "Automatic detection of basal cell carcinoma using telangiectasia analysis in dermoscopy skin lesion images," *Skin Research and Technology*, vol. 17, pp. 278–287, 2011.

[40] B. Cheng, R. J. Stanley, W. V. Stoecker, and et al., "Automatic dirt trail analysis in dermoscopy images," *Skin Research and Technology*, vol. 19, pp. e20–e26, 2013.

[41] E. Claridge, P. N. Hall, M. Keefe, and et al., "Shape analysis for classification of malignant melanoma," *Journal of biomedical engineering*, vol. 14, pp. 229–234, 1992.

[42] K. M. Clawson, P. Morrow, B. Scotney, and et al., "Analysis of pigmented skin lesion border irregularity using the harmonic wavelet transform," in *IMVIP 2009*, 2009, pp. 18–23.

[43] K. M. Clawson, P. J. Morrow, B. W. Scotney, and et al., "Computerised skin lesion surface analysis for pigment asymmetry quantification," in *IMVIP 2007*, 2007, pp. 75–82.

[44] ——, "Determination of optimal axes for skin lesion asymmetry quantification," in *IEEE ICIP 2007*, vol. 2, 2007, pp. 453–456.

[45] N. C. F. Codella, J. Cai, M. Abedini, and et al., "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," in *MLMI 2015*, 2015, pp. 118–126.

[46] N. C. F. Codella, D. Gutman, M. E. Celebi, and et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1710.05006*, 2017.

[47] N. C. F. Codella, Q. B. Nguyen, S. Pankanti, and et al., "Deep learning ensembles for melanoma recognition in dermoscopy images," *IBM Journal of Research and Development*, vol. 61, pp. 5:1–5:15, 2017.

[48] A. Dalal, R. Moss, R. Stanley, and et al., "Concentric decile segmentation of white and hypopigmented areas in dermoscopy images of skin lesions allows discrimination of malignant melanoma," *Computerized Medical Imaging and Graphics*, vol. 35, pp. 148–154, 2011.

[49] G. R. Day and R. H. Barbour, "Automated melanoma diagnosis: where are we at?" *Skin Research and Technology*, vol. 6, pp. 1–5, 2000.

[50] O. Debeir, C. Decaestecker, J. Pasteels, and et al., "Computer-assisted analysis of epiluminescence microscopy images of pigmented skin lesions," *Cytometry*, vol. 37, pp. 255–266, 1999.

[51] S. Demyanov, R. Chakravorty, M. Abedini, and et al., "Classification of dermoscopy patterns using deep convolutional neural networks," in *IEEE ISBI 2016*, 2016, pp. 364–368.

[52] T. DeVries and D. Ramachandram, "Skin lesion classification using deep multi-scale convolutional neural networks," *arXiv preprint arXiv:1703.01402*, 2017.

[53] G. Di Leo, G. Fabbrocini, A. Paolillo, and et al., "Towards an automatic diagnosis system for skin lesions: estimation of blue-whitish veil and regression structures," in *IEEE SSD 2009*, 2009, pp. 1–6.

[54] G. Di Leo, A. Paolillo, P. Sommella, and et al., "Automatic diagnosis of melanoma: a software system based on the 7-point check-list," in *IEEE HICSS 2010*, 2010, pp. 1–10.

[55] I. G. Díaz, "Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions," *arXiv preprint arXiv:1703.01976*, 2017.

[56] S. Dreiseitl and M. Binder, "Do physicians value decision support? a look at the effect of decision support systems on physician opinion," *Artificial intelligence in medicine*, vol. 33, pp. 25–30, 2005.

[57] R. Erol, M. Bayraktar, S. Kockara, and et al., "Texture based skin lesion abruptness quantification to detect malignancy," *BMC bioinformatics*, vol. 18, p. 484, 2017.

[58] A. Esteva, B. Kuprel, R. A. Novoa, and et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.

[59] M. Fleming, C. Steger, J. Zhang, and et al., "Techniques for a structural analysis of dermatoscopic imagery," *Computerized Medical Imaging and Graphics*, vol. 22, pp. 375–389, 1998.

[60] M. Fornaciali, S. Avila, M. Carvalho, and et al., "Statistical learning approach for robust melanoma screening," in *IEEE SIBGRAPI 2014*, 2014, pp. 319–326.

[61] A. Galdran, A. Alvarez-Gila, M. I. Meyer, and et al., "Data-driven color augmentation techniques for deep skin image analysis," *arXiv preprint arXiv:1703.03702*, 2017.

[62] R. Garnavi, M. Aldeen, and J. Bailey, "Computer-aided diagnosis of melanoma using border-and wavelet-based texture analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, pp. 1239–1252, 2012.

[63] R. Garnavi, M. Aldeen, M. E. Celebi, and et al., "Border detection in dermoscopy images using hybrid thresholding on optimized color channels," *Computerized Medical Imaging and Graphics*, vol. 35, no. 2, pp. 105–115, 2011.

[64] Z. Ge, S. Demyanov, B. Bozorgtabar, and et al., "Exploiting local and generic features for accurate skin lesions classification using clinical and dermoscopy imaging," in *IEEE ISBI 2017*, 2017, pp. 986–990.

[65] A. G. Goodson and D. Grossman, "Strategies for early melanoma detection: Approaches to the patient with nevi," *Journal of the American Academy of Dermatology*, vol. 60, pp. 719–735, 2009.

[66] C. Grana, R. Cucchiara, G. Pellacani, and et al., "Line detection and texture characterization of network patterns." in *In ICPR'06: Proceedings of the 18th International Conference on pattern Recognition*, 2006.

[67] C. Grana, G. Pellacani, and S. Seidanari, "Pratical color calibration for dermoscopy applied to a digital epiluminescence microscope," *Skin Research and Technology*, vol. 11, pp. 242–247, 2005.

[68] S. Guo, Y. Luo, and Y. Song, "Random forests and vgg-net: An algorithm for the isic 2017 skin lesion classification challenge," *arXiv preprint arXiv:1703.05148*, 2017.

[69] Y. Guo, Y. Liu, A. Oerlemans, and et al., "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[70] D. Gutman, N. C. F. Codella, M. E. Celebi, and et al., "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1605.01397*, 2016.

[71] P. Guvenc, R. W. LeAnder, S. Kefel, and et al., "Sector expansion and elliptical modeling of blue-gray ovoids for basal cell carcinoma discrimination in dermoscopy images," *Skin Research and Technology*, vol. 19, pp. e532–e536, 2013.

[72] Y. V. Haeghen, J. M. A. D. Naeyaert, and I. Lemahieu, "An imaging system with calibrated color image acquisition for use in dermatology," *IEEE Transactions on Medical Imaging*, vol. 19, pp. 722–730, 2000.

[73] A. Hajdu, B. Harangi, R. Besenczi, and et al., "Measuring regularity of network patterns by grid approximations using the LLL algorithm," in *IEEE ICPR 2016*, 2016, pp. 1524–1529.

[74] B. Harangi, "Skin lesion detection based on an ensemble of deep convolutional neural network," *arXiv preprint arXiv:1705.03360*, 2017.

[75] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE CVPR 2016*, 2016, pp. 770–778.

[76] J. S. Henning, S. W. Dusza, S. Q. Wang, and et al., "The cash (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy," *Journal of the American Academy of Dermatology*, vol. 56, pp. 45–52, 2007.

[77] A. G. Isasi, B. G. Zapirain, and A. M. Zorrilla, "Melanomas non-invasive diagnosis application based on the abcd rule and pattern recognition image processing algorithms," *Computers in Biology and Medicine*, vol. 41, pp. 742–755, 2011.

[78] H. Iyatomi, M. E. Celebi, G. Schaefer, and et al., "Automated color calibration method for dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 35, pp. 89–98, 2011.

[79] H. Iyatomi, H. Oka, M. E. Celebi, and et al., "Parameterization of dermoscopic findings for the internet-based melanoma screening system," in *IEEE CIISP 2007*, 2007, pp. 189–193.

[80] ——, "Computer-based classification of dermoscopy images of melanocytic lesions on acral volar skin," *Journal of Investigative Dermatology*, vol. 128, pp. 2049–2054, 2008.

[81] ——, "An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm," *Computerized Medical Imaging and Graphics*, vol. 32, pp. 566–579, 2008.

[82] J. Jaworek-Korjakowska, "Novel method for border irregularity assessment in dermoscopic color images," *Computational and mathematical methods in medicine*, vol. 2015, 2015.

[83] J. Jaworek-Korjakowska and P. Kłeczek, "Automatic classification of specific melanocytic lesions using artificial intelligence," *BioMed research international*, 2016.

[84] J. Jaworek-Korjakowska and R. Tadeusiewicz, "Assessment of dots and globules in dermoscopic color images as one of the 7-point check list criteria," in *IEEE ICIP 2013*, 2013, pp. 1456–1460.

[85] X. Jia and L. Shen, "Skin lesion classification using class activation map," *arXiv preprint arXiv:1703.01053*, 2017.

[86] Y. Jia, E. Shelhamer, J. Donahue, and et al., "Caffe: Convolutional architecture for fast feature embedding," in *ACM MM 2014*, 2014, pp. 675–678.

[87] A. B. Katapadi, M. E. Celebi, S. C. Trotter, and et al., "Evolving strategies for the development and evaluation of a computerised melanoma image analysis system," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, pp. 465–472, 2018.

[88] R. Kaur, P. P. Albano, J. G. Cole, and et al., "Real-time supervised detection of pink areas in dermoscopic images of melanoma: importance of color shades, texture and location," *Skin Research and Technology*, vol. 21, pp. 466–473, 2015.

[89] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in *IEEE ISBI 2016*, 2016, pp. 1397–1400.

[90] J. Kawahara and G. Hamarneh, "Fully convolutional networks to detect clinical dermoscopic features," *arXiv preprint arXiv:1703.04559*, 2017.

[91] S. Kaya, M. Bayraktar, S. Kockara, and et al., "Abrupt skin lesion border cutoff measurement for malignancy detection in dermoscopy images," *BMC bioinformatics*, vol. 17, p. 367, 2016.

[92] S. Kefel, P. Guvenc, R. LeAnder, and et al., "Discrimination of basal cell carcinoma from benign lesions based on extraction of ulcer features in polarized-light dermoscopy images," *Skin Research and Technology*, vol. 18, pp. 471–475, 2012.

[93] S. Kefel, S. P. Kefel, R. W. LeAnder, and et al., "Adaptable texture-based segmentation by variance and intensity for automatic detection of semitranslucent and pink blush areas in basal cell carcinoma," *Skin Research and Technology*, vol. 22, pp. 412–422, 2016.

[94] P. Kharazmi, M. I. AlJasser, H. Lui, and et al., "Automated detection and segmentation of vascular structures of skin lesions seen in dermoscopy, with an application to basal cell carcinoma classification," *IEEE journal of biomedical and health informatics*, vol. 21, pp. 1675–1684, 2017.

[95] P. Kharazmi, S. Kalia, H. Lui, and et al., "A feature fusion system for basal cell carcinoma detection through data-driven feature learning and patient profile," *Skin Research and Technology*, vol. 24, pp. 256–264, 2018.

[96] P. Kharazmi, J. Zheng, H. Lui, and et al., "A computer-aided decision support system for detection and localization of cutaneous vasculature in dermoscopy images via deep feature learning," *Journal of medical systems*, vol. 42, p. 33, 2018.

[97] H. Kittler, , A. A. Marghoob, G. Argenziano, and et al., "Standardization of terminology in dermoscopy/dermatoscopy: Results of the third consensus conference of the international society of dermoscopy," *Journal of the American Academy of Dermatology*, vol. 74, pp. 1093–1106, 2016.

[98] S. Kockara, M. Mete, T. Halic, and et al., "Fractals for malignancy detection in dermoscopy images," in *IEEE ICHI*, 2015, pp. 115–121.

[99] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," *ECCV 2012*, pp. 186–199, 2012.

[100] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: a review," *Artificial intelligence in medicine*, vol. 56, pp. 69–90, 2012.

[101] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS 2012*, 2012, pp. 1097–1105.

[102] Y. W. Kuo, Y. J. Chang, S. H. Wang, and et al., "Survey of dermoscopy use by taiwanese dermatologists," *Dermatologica Sinica*, vol. 33, pp. 215–219, 2015.

[103] G. Lee, O. Lee, S. Park, and et al., "Quantitative color assessment of dermoscopy images using perceptible color regions," *Skin Research and Technology*, vol. 18, pp. 462–470, 2012.

[104] T. Lee, V. Ng, R. Gallagher, and et al., "Dullrazor®: A software approach to hair removal from images," *Computers in biology and medicine*, vol. 27, pp. 533–543, 1997.

[105] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," *arXiv preprint arXiv:1703.00577*, 2017.

[106] M. Lingala, R. Stanley, R. Rader, and et al., "Fuzzy logic color detection: Blue areas in melanoma dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 38, pp. 403–410, 2014.

[107] G. Litjens, T. Kooi, B. E. Bejnordi, and et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[108] Z. Liu, J. Sun, L. Smith, and et al., "Distribution quantification on dermoscopy images for computer-assisted diagnosis of cutaneous melanomas," *Medical & biological engineering & computing*, vol. 50, pp. 503–513, 2012.

[109] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE CVPR 2015*, 2015, pp. 3431–3440.

[110] A. R. Lopez, X. Giro-i Nieto, J. Burdick, and et al., "Skin lesion classification from dermoscopic images using deep learning techniques," in *BioMed 2017*, 2017, pp. 49–54.

[111] J. López-Labraca, M. Á. Fernández-Torres, I. González-Díaz, and et al., "Enriched dermoscopic-structure-based cad system for melanoma diagnosis," *Multimedia Tools and Applications*, pp. 1–32, 2017.

[112] Y. Lu, F. Xie, Y. Wu, and et al., "No reference uneven illumination assessment for dermoscopy images," *IEEE Signal Processing Letters*, vol. 22, pp. 534–538, 2015.

[113] M. Machado, J. Pereira, and R. F. Pinto, "Classification of reticular pattern and streaks in dermoscopic images based on texture analysis," *Journal of Medical Imaging*, vol. 2, pp. 044 503–044 503, 2015.

[114] V. K. Madasu and B. Lovell, "Blotch detection in pigmented skin lesions using fuzzy co-clustering and texture segmentation," in *IEEE DICTA 2009*, 2009, pp. 25–31.

[115] A. Madooei and M. Drew, "A bioinspired color representation for dermoscopy image analysis," *in Dermoscopy Image Analysis*, pp. 23–66, 2015.

[116] A. Madooei, M. S. Drew, M. Sadeghi, and et al., "Automatic detection of blue-white veil by discrete colour matching in dermoscopy images," in *MICCAI 2013*, 2013, pp. 453–460.

[117] A. Madooei and M. Drew, "Incorporating colour information for computer-aided diagnosis of melanoma from dermoscopy images: A retrospective survey and critical analysis," *International Journal of Biomedical Imaging*, vol. 2016, 2016.

[118] A. Madooei, M. Drew, and H. Hajimirsadeghi, "Learning to detect blue-white structures in dermoscopy images with weak supervision," *accepted for publication in IEEE Journal of Biomedical and Health Informatics*, 2018.

[119] I. Maglogiannis and K. K. Delibasis, "Enhancing classification accuracy utilizing globules and dots features in digital dermoscopy," *Computer methods and programs in biomedicine*, vol. 118, pp. 124–133, 2015.

[120] A. Mahbod, R. Ecker, and I. Ellinger, "Skin lesion classification using hybrid deep neural networks," *arXiv preprint arXiv:1702.08434*, 2017.

[121] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Foundations and Trends in Computer Graphics and Vision*, vol. 8, pp. 85–283, 2014.

[122] A. G. Manousaki, A. G. Manios, E. I. Tsompanaki, and et al., "A simple digital image processing system to aid in melanoma diagnosis in an everyday melanocytic skin lesion unit. a preliminary report," *International journal of dermatology*, vol. 45, pp. 402–410, 2006.

[123] A. R. S. Marcal, T. Mendonca, C. S. P. Silva, and et al., "Evaluation of the menzies method potential for automatic dermoscopic image analysis," in *CompIMAGE 2012*, 2012, pp. 103–108.

[124] K. Matsunaga, A. Hamada, A. Minagawa, and et al., "Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble," *arXiv preprint arXiv:1703.03108*, 2017.

[125] T. Mendonça, P. M. Ferreira, J. S. Marques, and et al., "PH2: A dermoscopic image database for research and benchmarking," in *IEEE EMBC 2013*, 2013, pp. 5437–5440.

[126] T. Mendonça, P. M. Ferreira, A. R. S. Marçal, and et al., "PH2 A public database for the analysis of dermoscopy images," *Dermoscopy Image Analysis*, pp. 419–439, 2015.

[127] C. S. Mendoza, C. Serrano, and B. Acha, "Scale invariant descriptors in pattern analysis of melanocytic lesions," in *IEEE ICIP 2009*, 2009, pp. 4193–4196.

[128] A. Menegola, M. Fornaciali, R. Pires, and et al., "Towards automated melanoma screening: Exploring transfer learning schemes," *arXiv preprint arXiv:1609.01228*, 2016.

[129] ——, "Knowledge transfer for melanoma screening with deep learning," *arXiv preprint arXiv:1703.07479*, 2017.

[130] A. Menegola, J. Tavares, M. Fornaciali, and et al., "RECOD titans at isic challenge 2017," *arXiv preprint arXiv:1703.04819*, 2017.

[131] S. W. Menzies, K. A. Crotty, C. Ingvar, and et al., *Dermoscopy: an atlas of surface microscopy of pigmented skin lesions*. McGraw-Hill Sydney, Australia, 2009.

[132] M. Mete and N. M. Sirakov, "Dermoscopic diagnosis of melanoma in a 4d space constructed by active contour extracted features," *Computerized Medical Imaging and Graphics*, vol. 36, pp. 572–579, 2012.

[133] A. Mikołajczyk, A. Kwasigroch, and M. Grochowski, "Intelligent system supporting diagnosis of malignant melanoma," in *Polish Control Conference*, 2017, pp. 828–837.

[134] P. Mirunalini, A. Chandrabose, V. Gokul, and et al., "Deep learning for skin lesion classification," *arXiv preprint arXiv:1703.04364*, 2017.

[135] H. Mirzaalian, T. Lee, and G. Hamarneh, "Learning features for streak detection in dermoscopic color images using localized radial flux of principal intensity curvature," in *IEEE MMBIA 2012*, 2012, pp. 97–101.

[136] N. K. Mishra and M. E. Celebi, "An overview of melanoma detection in dermoscopy images using image processing and machine learning," *arXiv preprint arXiv:1601.07843*, 2016.

[137] K. Møllersen, M. Zortea, K. Hindberg, and et al., "Improved skin lesion diagnostics for general practice by computer aided diagnostics," *in Dermoscopy Image Analysis*, pp. 247–292, 2015.

[138] K. Møllersen, M. Zortea, T. R. Schopf, and et al., "Comparison of computer systems and ranking criteria for automatic melanoma detection in dermoscopic images," *PloS one*, vol. 12, p. e0190112, 2017.

[139] A. Murali, W. Stoecker, and R. Moss, "Detection of solid pigment in dermatoscopy images using texture analysis," *Skin Research and Technology*, vol. 6, pp. 193–198, 2000.

[140] D. H. Murphree and C. Ngufor, "Transfer learning for melanoma detection: Participation in isic 2017 skin lesion classification challenge," *arXiv preprint arXiv:1703.05235*, 2017.

[141] E. C. Murzaku, S. Hayan, and B. K. Rao, "Methods and rates of dermoscopy usage: a cross-sectional survey of US dermatologists stratified by years in practice," *Journal of the American Academy of Dermatology*, vol. 71, pp. 393–395, 2014.

[142] V. T. Y. Ng, B. Y. M. Fung, and T. K. Lee, "Determining the asymmetry of skin lesion with fuzzy borders," *Computers in biology and medicine*, vol. 35, pp. 103–120, 2005.

[143] A. I. of Health & Welfare, "Cancer in australia 2017," vol. Cancer series no. 101, 2017.

[144] R. B. Oliveira, A. S. Pereira, and J. M. R. S. Tavares, "Computational diagnosis of skin lesions from dermoscopic images using combined features," *Neural Computing and Applications*, pp. 1–21.

[145] R. Oliveira, J. Papa, A. Pereira, and et al., "Computational methods for pigmented skin lesion classification in images: review and future trends," *Neural Computing and Applications*, vol. 29, pp. 613–636, 2018.

[146] S. Pathan, K. G. Prabhu, and P. C. Siddalingaswamy, "A methodological approach to classify typical and atypical pigment network patterns for melanoma diagnosis," *Biomedical Signal Processing and Control*, vol. 44, pp. 25–37, 2018.

[147] ——, "Techniques and algorithms for computer aided diagnosis of pigmented skin lesions - a review," *Biomedical Signal Processing and Control*, vol. 39, pp. 237–262, 2018.

[148] H. Pehamberger, A. Steiner, and K. Wolff, "In vivo epiluminescence microscopy of pigmented skin lesions. i. pattern analysis of pigmented skin lesions," *Journal of the American Academy of Dermatology*, vol. 17, pp. 571–583, 1987.

[149] G. Pellacani, C. Grana, and S. Seidenari, "Automated description of colours in polarized-light surface microscopy images of melanocytic lesions," *Melanoma Research*, vol. 14, pp. 125–130, 2004.

[150] V. Pomponiu, H. Nejati, and N. M. Cheung, "Deepmole: Deep neural networks for skin mole lesion classification," in *IEEE ICIP 2016*, 2016, pp. 2623–2627.

[151] J. Quintana, R. Garcia, and L. Neumann, "A novel method for color correction in epiluminescence microscopy," *Computerized Medical Imaging and Graphics*, vol. 35, pp. 646–652, 2011.

[152] M. Rastgoo, R. Garcia, O. Morel, and et al., "Automatic differentiation of melanoma from dysplastic nevi," *Computerized Medical Imaging and Graphics*, vol. 43, pp. 44–52, 2015.

[153] M. Rastgoo, G. Lemaitre, O. Morel, and et al., "Classification of melanoma lesions using sparse coded features and random forests," in *SPIE Medical Imaging 2016*, 2016, pp. 97 850C–97 850C.

[154] H. W. Rogers, M. A. Weinstock, S. R. Feldman, and et al., "Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the us population, 2012," *JAMA dermatology*, vol. 151, pp. 1081–1086, 2015.

[155] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI 2015*, 2015, pp. 234–241.

[156] M. Ruela, C. Barata, J. S. Marques, and et al., "A system for the detection of melanomas in dermoscopy images using shape and symmetry features," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 5, pp. 127–137, 2017.

[157] S. Sabbaghi, M. Aldeen, and R. Garnavi, "A deep bag-of-features model for the classification of melanomas in dermoscopy images," in *IEEE EMBC 2016*, 2016, pp. 1369–1372.

[158] S. Sabbaghi, M. Aldeen, R. Garnavi, and et al., "Automated colour identification in melanocytic lesions," in *IEEE EMBC 2015*, 2015, pp. 3021–3024.

[159] M. Sadeghi, T. Lee, H. Lui, D. McLean, and S. Atkins, "Detection and analysis of irregular streaks in dermoscopic images of skin lesions," *IEEE Transactions on Medical Imaging*, vol. 32, pp. 849–861, 2013.

[160] M. Sadeghi, T. K. Lee, D. McLean, and et al., "Global pattern analysis and classification of dermoscopic images using textons," in *SPIE Medical Imaging 2012*, 2012, pp. 83 144X–83 144X.

[161] M. Sadeghi, M. Razmara, T. K. Lee, and et al., "A novel method for detection of pigment network in dermoscopic images using graphs," *Computerized Medical Imaging and Graphics*, vol. 35, pp. 137–143, 2011.

[162] M. Sadeghi, M. Razmara, P. Wighton, and et al., "Modeling the dermoscopic structure pigment network using a clinically inspired feature set," in *International Workshop on Medical Imaging and Virtual Reality*. Springer, 2010, pp. 467–474.

[163] A. R. Sadri, S. Azarianpour, M. Zekri, and et al., "Wn-based approach to melanoma diagnosis from dermoscopy images," *IET Image Processing*, vol. 11, pp. 475–482, 2017.

[164] A. Sáez, J. Sánchez-Monedero, P. A. Gutiérrez, and et al., "Machine learning methods for binary and multiclass classification of melanoma thickness from dermoscopic images," *IEEE transactions on medical imaging*, vol. 35, pp. 1036–1045, 2016.

[165] A. Sáez, C. Serrano, and B. Acha, "Model-based classification methods of global patterns in dermoscopic images," *IEEE Transactions on Medical Imaging*, vol. 33, pp. 1137–1147, 2014.

[166] J. Sánchez, F. Perronnin, T. Mensink, and et al., "Image classification with the fisher vector: Theory and practice," *International journal of computer vision*, vol. 105, pp. 222–245, 2013.

[167] T. Y. Satheesha, D. Satyanarayana, M. N. G. Prasad, and et al., "Melanoma is skin deep: A 3d reconstruction technique for computerized dermoscopic skin lesion classification," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 5, pp. 1–17, 2017.

[168] G. Schaefer, B. Krawczyk, M. E. Celebi, and et al., "An ensemble classification approach for melanoma diagnosis," *Memetic Computing*, vol. 6, pp. 233–240, 2014.

[169] G. Schaefer, M. I. Rajab, M. E. Celebi, and et al., "Colour and contrast enhancement for improved skin lesion segmentation," *Computerized Medical Imaging and Graphics*, vol. 35, pp. 99–104, 2011.

[170] S. Seidenari, G. Pellacani, and C. Grana, "Computer description of colours in dermoscopic melanocytic lesion images reproducing clinical assessment," *British Journal of Dermatology*, vol. 149, pp. 523–529, 2003.

[171] ——, "Colors in atypical nevi: a computer description reproducing clinical assessment," *Skin Research and Technology*, vol. 11, pp. 36–41, 2005.

[172] C. Serrano and B. Acha, "Pattern analysis of dermoscopic images based on markov random fields," *Pattern Recognition*, vol. 42, pp. 1052–1057, 2009.

[173] G. Sforza, G. Castellano, S. A. Arika, and et al., "Using adaptive thresholding and skewness correction to detect gray areas in melanoma in situ images," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, pp. 1839–1847, 2012.

[174] N. M. Shakya, R. W. LeAnder, K. A. Hinton, and et al., "Discrimination of squamous cell carcinoma in situ from seborrheic keratosis by color analysis techniques requires information from scale, scale-crust and surrounding areas in dermoscopy images," *Computers in biology and medicine*, vol. 42, pp. 1165–1169, 2012.

[175] K. Shimizu, H. Iyatomi, M. E. Celebi, and et al., "Four-class classification of skin lesions with task decomposition strategy," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 274–283, 2015.

[176] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA: a cancer journal for clinicians*, vol. 68, pp. 7–30, 2018.

[177] E. Silverberg, C. C. Boring, and T. S. Squires, "Cancer statistics, 1990," *CA: a cancer journal for clinicians*, vol. 40, pp. 9–26, 1990.

[178] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[179] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *IEEE ICCV 2003*, 2003, p. 1470.

[180] R. T. Sousa and L. V. de Moraes, "Araguaia medical vision lab at isic 2017 skin lesion classification challenge," *arXiv preprint arXiv:1703.00856*, 2017.

[181] N. Srivastava, G. E. Hinton, A. Krizhevsky, and et al., "Dropout: a simple way to prevent neural networks from overfitting." *Journal of machine learning research*, vol. 15, pp. 1929–1958, 2014.

[182] R. J. Stanley, W. V. Stoecker, and R. H. Moss, "A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images," *Skin Research and Technology*, vol. 13, pp. 62–72, 2007.

[183] W. V. Stoecker, W. W. Li, and R. H. Moss, "Automatic detection of asymmetry in skin tumors," *Computerized Medical Imaging and Graphics*, vol. 16, pp. 191–197, 1992.

[184] W. Stoecker, K. Gupta, R. Stanley, and et al., "Detection of asymmetric blotches (asymmetric structureless areas) in dermoscopy images of malignant melanoma using relative color," *Skin Research and Technology*, vol. 11, pp. 179–184, 2005.

[185] W. Stoecker, M. Wronkiewiecz, R. Chowdhury, and et al., "Detection of granularity in dermoscopy images of malignant melanoma using color and texture features," *Computerized Medical Imaging and Graphics*, vol. 35, pp. 144–147, 2011.

[186] W. Stolz, A. Riemann, and A. B. Cognetta, "ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma," *European Journal of Dermatology*, vol. 4, pp. 521–527, 1994.

[187] C. Szegedy, W. Liu, Y. Jia, and et al., "Going deeper with convolutions," in *IEEE CVPR 2015*, 2015, pp. 1–9.

[188] T. Tanaka, S. Torii, I. Kabuta, and et al., "Pattern classification of nevus with texture analysis," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 3, pp. 143–150, 2008.

[189] K. Thon, H. Rue, S. O. Skrøvseth, and et al., "Bayesian multiscale analysis of images modeled as gaussian markov random fields," *Computational Statistics & Data Analysis*, vol. 56, pp. 49–61, 2012.

[190] C. N. Vasconcelos and B. N. Vasconcelos, "Experiments using deep learning for dermoscopy image analysis," *to appearn in Pattern Recognition Letters*, 2018.

[191] C. G. Watts, C. Madronio, R. L. Morton, and et al., "Clinical features associated with individuals at higher risk of melanoma: a population-based study," *JAMA dermatology*, vol. 153, pp. 23–29, 2017.

[192] P. Wighton, T. K. Lee, H. Lui, and et al., "Chromatic aberration correction: an enhancement to the calibration of low-cost digital dermoscopes," *Skin Research and Technology*, vol. 17, pp. 339–347, 2011.

[193] ——, "Generalizing common tasks in automated skin lesion diagnosis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, pp. 622–629, 2011.

[194] F. Xie, H. Fan, Y. Li, , and et al., "Melanoma classification on dermoscopy images using a neural network ensemble model," *IEEE transactions on medical imaging*, vol. 36, pp. 849–858, 2017.

[195] S. Yang, B. Oh, S. Hahm, and et al., "Ridge and furrow pattern classification for acral lentiginous melanoma using dermoscopic images," *Biomedical Signal Processing and Control*, vol. 32, pp. 90–96, 2016.

[196] X. Yang, Z. Zeng, S. Y. Yeo, and et al., "A novel multi-task deep learning model for skin lesion segmentation and classification," *arXiv preprint arXiv:1703.01025*, 2017.

[197] T. Yao, Z. Wang, Z. Xie, and et al., "A multiview joint sparse representation with discriminative dictionary for melanoma detection," in *IEEE DICTA 2016*. IEEE, 2016, pp. 1–6.

[198] T. Yoshida, M. E. Celebi, G. Schaefer, and et al., "Simple and effective pre-processing for automated melanoma discrimination based on cytological findings," in *IEEE Big Data 2016*, 2016, pp. 3439–3442.

[199] S. Yoshino, T. Tanaka, M. Tanaka, and et al., "Application of morphology for detection of dots in tumor," in *IEEE SICE 2004*, vol. 1, 2004, pp. 591–594.

[200] L. Yu, H. Chen, Q. Dou, and et al., "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE transactions on medical imaging*, vol. 36, pp. 994–1004, 2017.

[201] Z. Yu, D. Ni, S. Chen, and et al., "Hybrid dermoscopy image classification framework based on deep convolutional neural network and fisher vector," in *IEEE ISBI 2017*, 2017, pp. 301–304.

[202] W. Zhang, L. Gao, and R. Liu, "Using deep learning method for classification: A proposed algorithm for the isic 2017 skin lesion classification challenge," *arXiv preprint arXiv:1703.02182*, 2017.

[203] M. Zortea, T. R. Schopf, K. Thon, and et al., "Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists," *Artificial intelligence in medicine*, vol. 60, pp. 13–26, 2014.