



*Research article*

## **Skin disorder diagnosis with ambiguity reduction assisted by lesion color adaptation**

**Nikos Petrellis\***

Department of Computer Science and Engineering, University of Thessaly (ex TEI of Thessaly), Geopolis, Larissa, 41110, Greece

\* **Correspondence:** Email: [npetrellis@teilar.gr](mailto:npetrellis@teilar.gr); Tel: +30-2410-684542; Fax: +30-2410-684573.

**Abstract:** A smart phone application based on a low complexity image processing technique and a novel fuzzy-like classification method are presented for skin disorder diagnosis. The proposed classification method takes into consideration the size and color features of skin lesions rather than their shape and texture. The classification rules are determined after processing statistically a small number of representative training photographs. Consequently, they can be defined by an end user that is not necessarily skilled in computer science. The application presented in this paper can serve as a complementary tool for a dermatologist to continuously monitor remotely his patients. The accuracy of the diagnosis that is based only on the image processing outcomes, ranges between 85.3% and 97.7% using 5 only representative photographs as a “training set” (corresponding from 9% to 24% of the test set per disease). The achieved accuracy can be improved (up to 17%), if the photographs are processed using a specific color adaptation technique. The small fraction of training photographs can be scaled up if the size of the test set is increased but it is expected that a limited number of training photographs will be sufficient in order to achieve an acceptable accuracy for a test set of any size. This accuracy can be further improved if other factors are taken into consideration (progression of the symptoms, information provided by the user, etc).

**Keywords:** skin infections; skin disorders; image processing; color adaptation; mobile apps; lesions; histograms

---

## 1. Introduction

The diagnosis of skin disorders can be assisted by several smart phone applications that have been recently presented in the market or described in the literature. Most of these approaches offer the opportunity to the users to access databases, textbooks and journals related to dermatology [1]. The rest of these applications can be based on information given by the user or can help in carrying out other procedures (e.g., biopsies). Such applications can be integrated in teledermatology platforms and they are capable of accessing databases of images, guidelines that are useful for diagnosis, etc [2]. In the following, we focus on image processing approaches.

Several approaches diagnose a single skin disorder like Psoriasis [3]. The skin color is processed using the moment of the R, G, B color planes in order to estimate the mean color, the standard deviation and the skewness of the color. Co-occurrence matrices are used [3] to extract texture features. The diagnosis of Acne is described in [4] and [5]. Color spaces different than RGB, such as Hue, Saturation, Value (HSV) can be employed [5], for the efficient segmentation of the input image. When a binary decision has to be taken, Support Vector Machines (SVM) is often adopted [5]. Smart phone implementations [5] have the advantage of portability and low cost but the limited processing speed, memory and power, have to be taken into consideration. The most important skin disease is of course the melanoma. A review of segmentation and classification techniques for melanoma diagnosis is presented in [6]. MATLAB framework is used in [7] for image enhancement by removing artifacts (e.g., hair, noise, etc). In [8], a sophisticated iPhone application called SkinScanc is presented for melanoma diagnosis and SVM is used for classification. A similar smart phone implementation is presented in [9]. In [10], the color particularity of the melanoma lesion is used by special color feature detection techniques that assess the variation of hues. In the same paper, a multi-threshold technique has been used for segmentation. Images of the back of a human body are used in [11] to monitor potentially malignant Pigmented Skin Lesions. A review about Light-Induced Fluorescence Spectroscopy (LIFS) and Optical Coherence Tomography (OCT) methods for melanoma diagnosis is presented in [12].

The most appropriate supervised techniques for applications where photographs have to be classified in different clusters are: Naïve Bayes, k-Means, k-Nearest Neighbors (kNN), Decision Trees like J48, Random Forests, Fuzzy Clustering, Neural Networks (NNs). One of the major issues concerning the supervised classification methods is their training, since hundreds of images have to be analyzed during the training phase. In many approaches (e.g., in [8]), about  $\frac{3}{4}$  of the input samples are used for training and  $\frac{1}{4}$  of them for testing.

Neural networks (e.g., MultiLayer Perceptron-MLP) are used in [7], [13–15] for dermatological disease diagnosis. Other approaches such as the ones presented in [16] and [17] are closer to the method described here since they classify an image as one of the supported skin diseases. In [16] the supported diseases include Eczema, Acne, Leprosy, Psoriasis, Vitiligo, etc. The features used for classification are the average color of the infected area, its size, shape, etc. Additional information is given by the user about gender, age, liquid type and color, etc. In [17], the presented image processing application discriminates between 3 skin disorders. The range of each one of the features used is defined for the supported diseases similarly to our approach. The most important features that should be used for the recognition of each disease are listed in [10]. Ten skin diseases are discriminated by a neural network in [18]. Similarly, six skin disorders are supported in [19].

The proposed skin disorder diagnosis was initially presented in [20] and [21]. The proposed classification method was inspired for plant disease diagnosis [22]. This previous work is extended here by employing a color adaptation method based on the average gray level of the normal skin and

the lesion. Moreover, extended experiments have been conducted using different thresholds than [20] in order to have comparable results with the ones obtained using the proposed color adaptation method. The proposed method has been evaluated in this paper by several metrics: sensitivity, specificity and accuracy different to the ones used in [20–22]. A comparison with many referenced approaches has been included. Moreover, several different classifiers (Naïve Bayes, MLP, J48 Decision Tree, Random Forest, available in the framework of the Weka tool [23]) have been tested using the same photographs and features, in order to compare the achieved sensitivity.

In the developed Skin Disease application, the following skin disorders are diagnosed: acne, melanoma, mycosis, papillomas, psoriasis, vitiligo and warts. A photograph displaying human skin is analyzed and a number of features are extracted. These features are different than the ones described in the referenced approaches. A deterministic fuzzy-like method is adopted for the classification of an image as one of the supported skin diseases. This classification method is based on a set of features and the expected range of these features forms the Color Signature of a specific disease. The Color Signatures can be easily defined or modified by the end user that does not need to have access to the source code of the application. These signatures can be defined based on simple statistical observation of a few representative photographs. This statistical processing can be performed manually or using a spreadsheet (it will be performed automatically in future versions of the developed tool). This process is similar to the “training” used in the referenced approaches. The number of these training photographs is much smaller compared to the ones required e.g., by a neural network. Of course, the number of the required training photographs gets larger if the test set is extended, in order to cover as many variations in the appearance of a skin disorder as possible. However, if the employed training photographs cover all of the possible color and size variations of the lesions, their number does not have to be further increased, regardless of the test set size. The low complexity of the proposed method makes it ideal for implementation on mobile phones. The experimental results presented here, indicate that the proposed classification method has good accuracy even if it is merely based on image processing. A skin disorder recognition tool like the one described in this paper does not intend to substitute the safe diagnosis performed by a dermatologist. However, the extensibility of the proposed application can be exploited in order to define or customize the supported diseases and their recognition rules. The progression of a skin disorder symptoms can be monitored remotely and the patient may be notified that he should visit his doctor for further medical tests if necessary.

The materials and methods (image processing, classification, color normalization) used in our approach are described in Section 2. Experimental results are presented in Section 3. In Section 4, the results are discussed and comparison is presented with the referenced approaches.

## **2. Materials and method**

### *2.1. Image sources*

Many of the photographs used in this study were retrieved from the Dermoscopic blog (<http://dermoscopic.blogspot.com/>). These images were cropped to exclude e.g., watermarks, stretched or simply resized to  $1024 \times 576$  pixels. In this way, multiple patches can be extracted by a single photograph. The specific size was selected because no critical information is lost in this resolution while the processing of the photograph is completed in a reasonable time using this size (less than 2 seconds). The proposed method processes these images in RGB format. The number of photographs tested for each skin disorder ranges between 21 and 54. The “training” of the

application i.e., the rules for the diagnosis of each disease were determined using 5 representative photographs of the specific disorder i.e., only the 9%–24% of the tested photographs were used for training. Larger test sets would require more training photographs but still fewer than the ones required by other classification methods.

Human body parts with skin disorders (lesions) are displayed in the photographs used. The user in the current implementation determines the displayed part of the body. It is assumed that the camera has focused on the lesion that has different color than the normal skin. In many skin disorders, a halo with distinct color exists around the lesion spots. If the photograph is captured within a short distance from the body, no background appears. However, if there is a background, it is assumed for simplicity that it is white or generally much brighter than the skin color. Complicated and time-consuming background separation algorithms are avoided in this way.

## 2.2. Image processing method and feature extraction

An image is divided into 4 distinct regions: Normal skin, Lesions, Halo and Background. The invariant features used for the classification of the skin diseases are the following: the number of spots and their gray level, the area of the lesion as well as several color features extracted from color histograms. The histograms used in this application indicate the number of pixels that have a specific red, green or blue color level. For example, if the histogram value for a specific region and color level  $c$  is  $H(c)$ , then there are  $H(c)$  pixels in this region with color level  $c$ . The employed classification method checks if the aforementioned features, fall within predefined strict or loose limits. Each disease is ranked based on how many features are found within these narrow or broad ranges. Since the shape and the texture of the lesion is not taken into consideration, the resolution of a photograph as well as the existence of noise or blurring does not affect significantly the performance of the employed classification method.

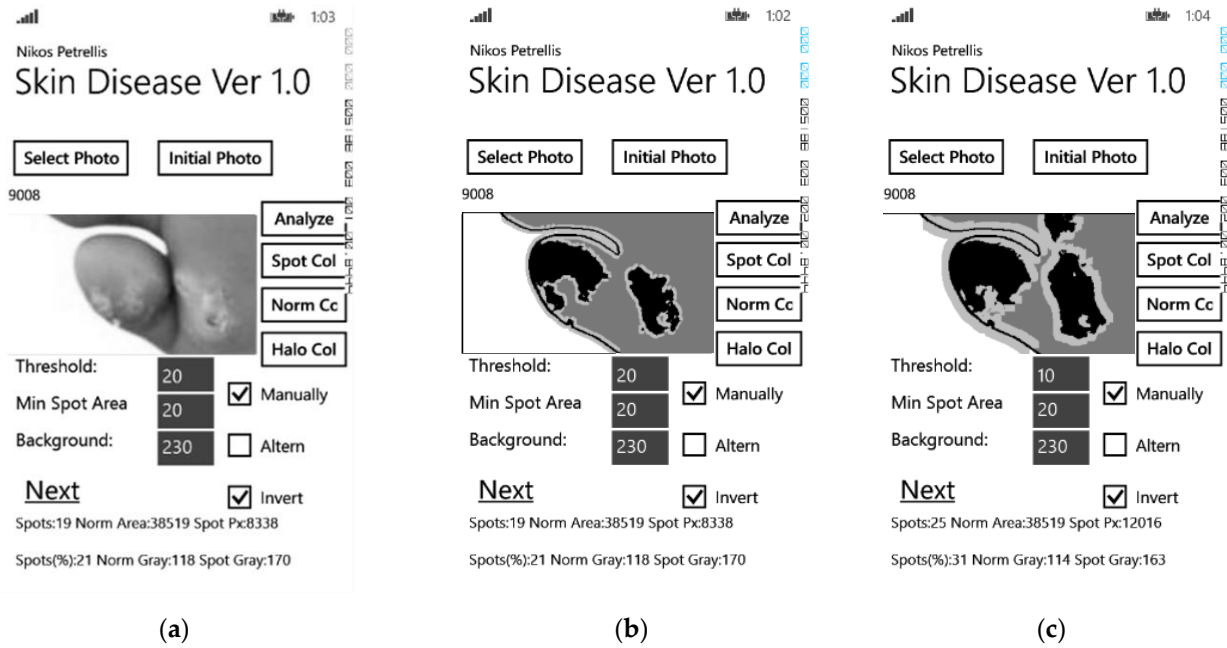
The segmentation is based on multiple thresholds and is applied to the gray version of the image. Although more advanced and precise segmentation techniques can be employed, there is actually no need for a more complicated segmentation method, since the classification method used does not take into consideration features such as the shape and the perimeter of the lesion. The pixels of the gray version of the photograph are swept top-down and from left to right. The threshold  $B_g$  is used to separate the Background from the skin. Pixels with gray level  $g > B_g$  are assumed to belong to the background. A second threshold ( $T_h$ ) is used to distinguish the Normal skin region from the Lesion. The gray level of all the pixels that do not belong in the Background is averaged ( $G_{av}$ ). If the Lesion region is brighter than the Normal skin, a lesion pixel  $i$  has a gray level  $g_i$  higher than the sum of the average  $G_{av}$  plus an offset  $T_h$ :

$$B_g > g_i > G_{av} + T_h, \quad (1)$$

If the lesion pixel is darker than the normal skin, then:

$$g_i < G_{av} - T_h, \quad (2)$$

The thresholds  $B_g$  and  $T_h$  can be interactively modified by the user in real time in order to achieve a higher precision during the segmentation process. The Halo can be defined as a zone (of  $H_p$  pixels width) around the lesion spots. If no Halo exists,  $H_p$  can be set to 0.



**Figure 1.** The main page (in inverted text/background color) of the Skin Disease application with the initial photo (a), the display of the R matrix with  $T_h = 20$ ,  $B_g = 230$ ,  $H_p = 5$  (b) and with  $T_h = 10$ ,  $B_g = 230$ ,  $H_p = 10$  (c).

The image pre-processing stage identifies the regions described earlier (Normal skin, Lesion, Halo, Background). The matrix R is defined where special values can be assigned to each pixel indicating the region it belongs to. In Figure 1a, the initial color photograph is selected. In Figure 1b, the R matrix is represented as a gray image. The background is displayed in white, while the normal skin in gray and the lesion spots in black. The halo is shown in brighter gray. If higher threshold value  $T_h$  is used than the optimal one then, some lesion spots are not recognized. On the contrary, when lower  $T_h$  values are used, some normal skin areas are misinterpreted as lesion. As was shown in [22] a variation by  $\pm 10$  in the optimal value of  $T_h$  can result in a degradation of the accuracy by up to 20%. This is also an indication about how the light exposure can affect the classification results. However, it is unlikely that the user will select such an inappropriate  $T_h$  value due to the interactive comparison of the segmented with the original photograph. In Figure 1b, an appropriate  $T_h$  value (20) is selected. However, if  $T_h = 10$  (Figure 1c), more area is recognized as lesion. The user can determine  $T_h$  and  $B_g$  values in order to recognize precisely the background and the lesion regions of each photograph.

**Table 1.** The R matrix after the first scan. The cell in the 8th row, 4th column can be initially assigned either to 3 or 1. Halo has not been defined yet.

N	N	N	N	B	B	B	B
N	1	1	N	N	N	N	B
N	N	1	N	N	N	N	N
N	N	1	1	N	2	2	N
N	N	1	1	N	2	2	N
N	1	N	N	N	N	N	N
N	1	N	3	N	N	N	N
N	1	1	?	N	N	N	N

**Table 2.** The R matrix after merging spots 1 and 3. 1 pixel Halo zone has been defined.

H	H	H	H	B	B	B	B
H	1	1	H	N	N	N	N
H	H	1	H	H	H	H	H
N	H	1	1	H	2	2	H
H	H	1	1	H	2	2	H
H	1	H	H	H	H	2	H
H	1	H	1	H	H	H	H
H	1	1	1	H	N	N	N

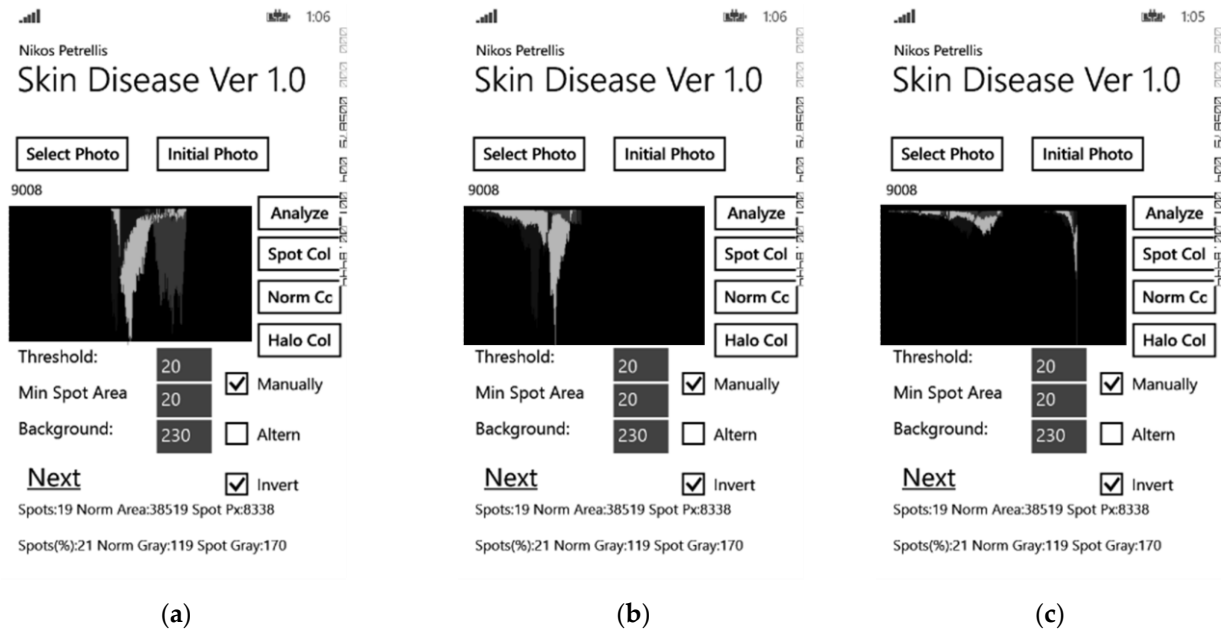
Tables 1 and 2 show how the R matrix is constructed. Initially, the gray image is scanned top-down and from left to the right. If the gray level of the current pixel in the gray version of the original image is higher than  $B_g$  then, it is marked as Background (“B”) otherwise, it is marked as Normal Skin (“N”). The pixels mapped as “N” are averaged to estimate  $G_{av}$ . Then, they are examined again to check if their gray level fulfils one of the conditions defined in eq. (1) or eq. (2). In this case, the corresponding pixels are marked with a spot identity ( $S_i$ ). Initially,  $S_i$  is set to 1. If the current pixel is adjacent to another spot as indicated by the pixels already visited in the current scan (left, top right, top, top left) their spot identity is also used for the current pixel. Otherwise, it is assumed that a new spot is found that has to be marked with a new identity  $S'_i$ :

$$S'_i = \max(S_i) + 1, \quad (3)$$

Table 1 shows the described mapping for an arbitrary image of only  $8 \times 8$  pixels. However, there are adjacent cells that have been assigned with different identity. In the next iteration, the spots No. 1 and 3 will be merged under the identity 1 as shown in Table 2. Moreover, a Halo zone with  $H_p = 1$  is also defined in Table 2 and the pixels belonging to this zone are marked with “H”. The area of the lesion and the halo can be directly estimated by the number of pixels of the matrix R. The parameters  $N_L$ ,  $N_N$  and  $N_H$  denote the number of pixels belonging to the lesion, normal skin and halo, respectively. The average gray level of each region ( $G_{LA}$ ,  $G_{HA}$ ,  $G_{NA}$ : lesion, halo and normal skin, respectively) can be estimated from the gray level of the corresponding pixels of the original image using the mapping of the R matrix. The values of these features are displayed at the bottom of the main application page shown in Figure 1. The relative area  $A_L$ , of the lesion is estimated as:

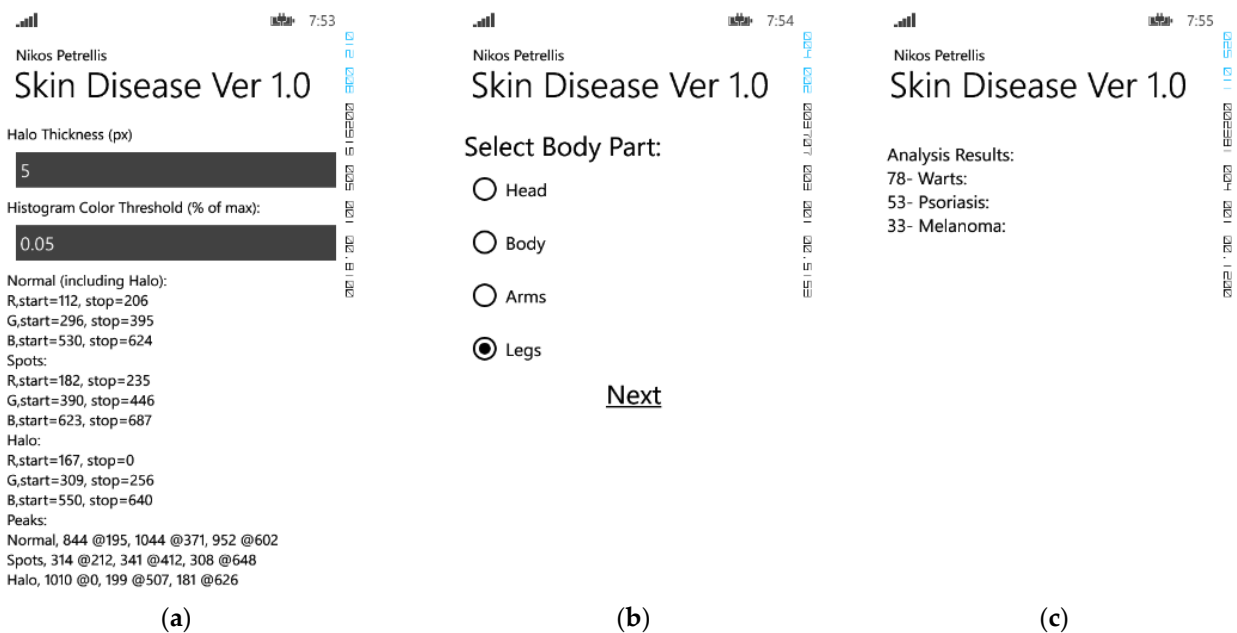
$$A_L = N_L / (N_L + N_N + N_H), \quad (4)$$

The relative area of the normal ( $A_N$ ) and the halo ( $A_H$ ) regions can also be estimated in a similar way. The highest  $S_i$  value denotes the number of spots recognized in an image. In the specific example of Table 2, the lesion consists of 2 spots. The rest of the features used in the classification process, are extracted by the color histograms of each region. These color histograms are displayed in the Skin Disease application as shown in Figure 2a (Lesion), Figure 2b (Normal skin) and Figure 2c (Halo). Overall, there are 9 histograms (one for each color plane: red, green, blue and each region: Lesion, Normal skin, Halo). The horizontal axis in each histogram indicates the color level (0.255). As can be seen from Figure 2a and Figure 2b, each histogram is a single lobe because most of the pixels belonging to the same region have similar color.



**Figure 2.** The color histograms of the Lesion (a), the Normal skin (b) and the Halo (c) for the photograph of Figure 1.

The peak, the starting and the ending point of the histogram lobes are used as invariant features in the classification process. The Begin and the End of the lobe are defined as the color levels where a histogram lobe crosses a predefined threshold. The number of these features is  $9 \times 3 = 27$  since there are 3 color spaces and 3 regions of interest. The symbol of each one of these features has the following format: “crf” where c can be R(ed), G(reen) or B(lue), r can be S(pot), N(ormal skin) or H(alo) and f can be B(egin), E(nd) or P(eak). All of these features are estimated by the Skin Disease application and the user is notified about them through the administrative page shown in Figure 3a.



**Figure 3.** Features extracted from the color histograms (a), selection of body part, (b) and the results of the diagnosis (c) for the photograph of Figure 1.

### 2.3. Employed classification method

The features shown by the Skin Disease application in Figure 1 and Figure 3a can be used in order to extend the supported set of skin disorders or customize the Color Signatures of the already supported diseases. For example, the analysis of 5 photographs that display papillomas, can show that the feature GSE is between 70 and 130. These values can be considered as the strict limits of the GSE feature and this feature is also expected to be found within these limits in other photographs that display papillomas. Looser limits are heuristically defined e.g., 60 and 140. If a new photograph is analyzed and its GSE value is found within those loose limits, it gets a grade  $R_{L\_GSE}$  while if it is found to be within the strict limits an additional  $R_{S\_GSE}$  grade is added. If the GSE of the new photograph is not found in the loose limits then, no grade is given for this feature. Let us formally define the loose limits of a feature “crf” as  $K_D(\text{crfL}) = (\text{crfLn}, \text{crfLx})$  and its strict limits as  $K_D(\text{crfS}) = (\text{crfSn}, \text{crfSx})$ . Its rank  $R_D$  of a disease  $D$  can be estimated then, as:

$$R_D = \sum_{c,r,f} R_{L\_crf} x_{L\_crf} + \sum_{c,r,f} R_{S\_crf} x_{S\_crf} \quad (5)$$

$$x_{L\_crf} = \begin{cases} 1, & \text{crfLn} < \text{crf} < \text{crfLx}, \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

$$x_{S\_crf} = \begin{cases} 1, & \text{crfSn} < \text{crf} < \text{crfSx}, \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The exact features used in the Skin Disease application are listed in Table 3. As can be seen by this table, the normal skin region features are not used because the color of the skin does not provide an indication about a disease. However, the brightness of the normal skin region can be useful for adjusting the brightness of the lesion and the halo regions according to the environmental conditions. The body part displayed in the photograph can be selected by radio boxes (see Figure 3b). A critical feature is the body part displayed since e.g., warts appear more often at the feet and the hands.

The Color Signature of a disease includes the strict and loose limits of the features listed in Table 3. The skin disorders that are recognized in the present version of the application are the following: Acne, Melanoma, Mycosis, Papillomas, Psoriasis, Vitiligo, Warts. Consequently, seven Color Signatures have to be defined. As shown in Figure 3c, the three disorders with the highest rank according to equation (5) are listed.

If  $K_D$  is the range of a specific feature in the Color Signature of the skin disorder  $D$  then, it is desirable to distribute the acceptable range  $L$  of this feature into  $K_D$  ranges without overlaps and gaps. For example, Figure 4a shows 4 diseases where all the  $K_D$  ranges are equal. In this case, any feature value corresponds to a single disease and there is no ambiguity. The ranges do not necessarily have to be equal as long as they do not overlap and they do not leave gaps. If gaps exist as in Figure 4b, then the feature values in the gaps do not allow a safe disease diagnosis. This may occur if the ranges have been defined based on a very small number of training samples. If too many training samples are used then, the extracted feature values can span to a wide range. The overlapping ranges that can appear in this case have feature values that belong to multiple diseases and do not allow a safe disease diagnosis too.

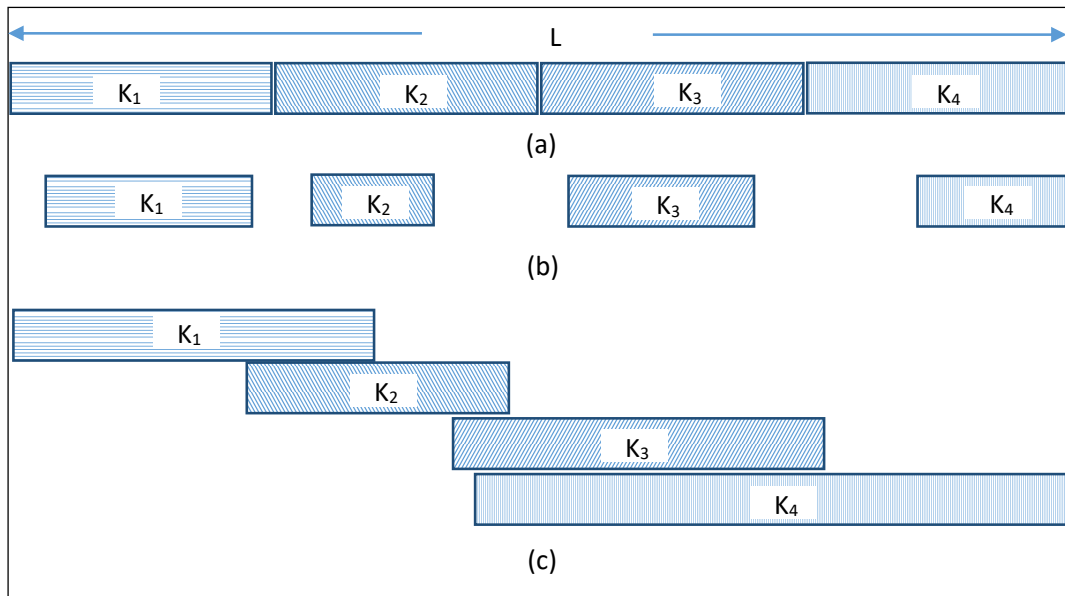


**Table 3.** List of features.

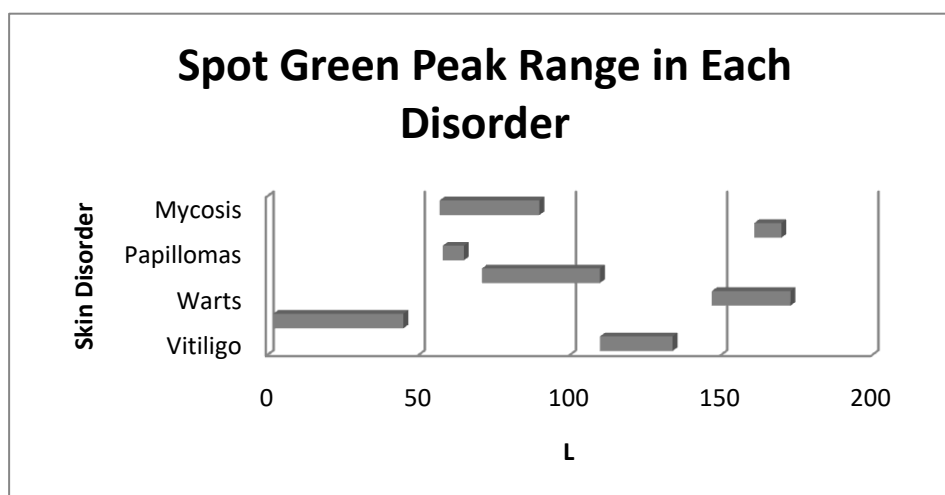
Feature	Notes
Number of Spots ( $N_S$ )	Too small spots are assumed to be noise and they are ignored. The minimum acceptable number of pixels is defined in the field “Min Spot Area” of Figure 1
Relative Lesion Area ( $A_L$ )	See eq. (4)
Relative Normal Skin Area ( $A_N$ )	Estimated similarly to eq. (4)
Relative Halo Area ( $A_H$ )	Estimated similarly to eq. (4)
Lesion Average Gray ( $G_{LA}$ )	The average gray level of the pixels assigned with a spot identity in matrix R
Halo Average Gray ( $G_{HA}$ )	The average gray level of the pixels mapped as “H” in matrix R
Normal Skin Average Gray ( $G_{NA}$ )	The average gray level of the pixels mapped as “N” in matrix R
Histogram features $K_D(crfL)$ and $K_D(crfS)$	$c = R, G, \text{ or } B$ $r = \text{Lesion or Halo, } f = B, E, \text{ or } P$ (18 features) The corresponding Normal Skin Histogram features are not used
Body part	Different grades may be assigned to each body part in a disease Signature

The following Ambiguity metric A, measures how the distribution of the feature ranges favors a “one-to-one” mapping between feature values and diseases. It is defined as:

$$A = L - \sum_{i=1}^N K_i \tag{8}$$



**Figure 4.**  $K_D$  ranges fitting without overlaps and gaps (a), with gaps (b) and with overlaps (c).



**Figure 5.** Example ranges of the histogram peak of the green color of the spots.

If  $K_i = LN$ , as in Figure 4a, then  $A = 0$  and there is no ambiguity in the disease diagnosis even if a single feature is used. In Figure 4b,  $L > \sum_{i=1}^N K_i$  and  $A > 0$  while in Figure 4c,  $L < \sum_{i=1}^N K_i$  and  $A < 0$ . In these cases, there is a “one-to-many” mapping of feature values to diseases. Since there are several features that are taken into consideration in the classification method described in the previous sub-section, the ambiguity is cleared by the ranking defined with equations (5)–(7). For example, if the value of a feature  $F1$  is extracted from a specific photograph it may correspond to 2 skin disorders  $\{S1, S2\}$  due to a potential overlapping in the ranges defined for this feature in the Color Signatures of these two disorders. A second feature  $F2$  extracted from the same photograph may correspond to a different set of disorders  $\{S1, S3, S4\}$ . It is clear that in this example,  $S1$  is the dominant disorder since it appears in both sets. However, if e.g., the ranges of these features did not overlap, then the system would be capable of diagnosing this disorder with a higher confidence even by a single feature. A real example of range overlapping for a specific feature used in skin disorder diagnosis is presented in Figure 5.

#### 2.4. Color normalization

Several color normalization techniques can be found in [24] that have been used in medical imaging and precision agriculture especially in gray scale. The brain MRI scans, mammograms and ultrasound images can be enhanced by Gaussian low pass filtering for noise removal and normalization of pixel intensity. Histogram stretching and shifting may be used to cover the full gray scale and increase the contrast. A different normalization is performed in RGB format where the color of each pixel is derived by its previous value divided by the sum of the rest of the colors. This kind of color normalization has been employed e.g., to count seedlings in a wheat field. Relative color can also be estimated in order to moderate the effect of different light exposure in the detection of skin disorders. For example, the average background skin color is subtracted from each lesion pixel in melanoma diagnosis. The use of normalized or relative color compensates the effect of variation in ambient light, the errors in digitizing images, the skin color variations and mimics the way the mammals perceive color. External tools like ColorChecker can be employed in order to implement a specific color normalization (<https://xritephoto.com/colorchecker-classic>). However, an

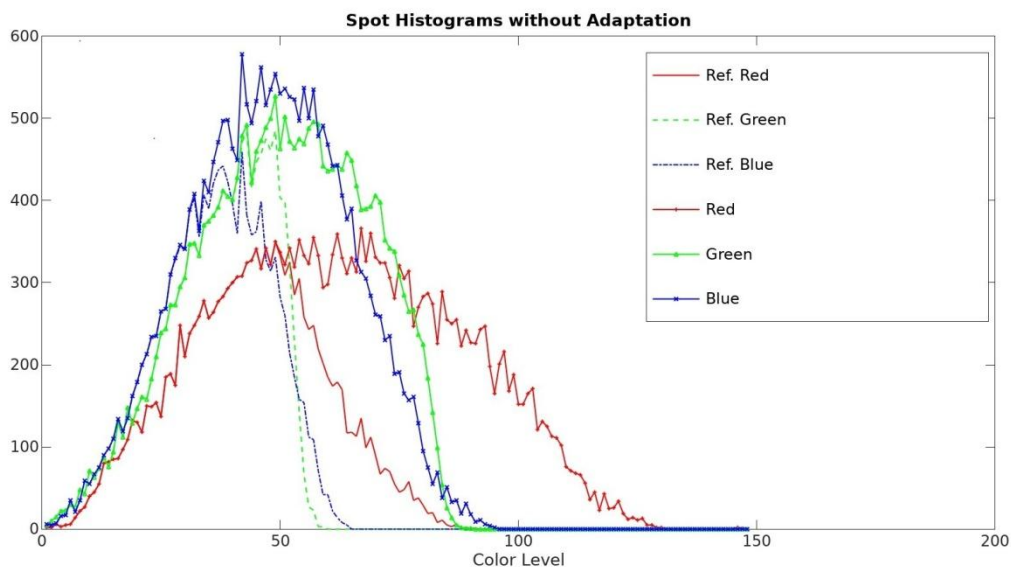
integrated solution was preferred to make this process transparent to the end user.

As explained in the previous subsection and Figure 4, both the gaps between the ranges of a feature and their overlapping are undesirable for safe disorder diagnosis. If the determined ranges are wide due to variations in the environmental conditions (e.g., light exposure) then, it is better to exploit some kind of normalization that would lead to the shrinkage of these ranges. If on the other hand, the shrinking of the feature ranges leaves gaps then, these ranges can be extended to cover the gaps by employing a larger number of training photographs. However, much larger training sets do not favor the advertised extendibility of the developed system since large training sets are more difficult to be statistically processed by the end user. The employed color adaptation process attempts to narrow the range of the color histogram features during the definition of the Color Signatures (training process).

The training photograph with the lower normal skin gray level for a specific skin disorder, serves as a reference (ref). If the lesion average gray level of each training sample  $i$ , is  $G_{LA}(i)$  and the lesion average gray level of the reference photograph is  $G_{LA}(\text{ref})$  then, their difference  $D_i$  is estimated:

$$D_i = G_{LA}(i) - G_{LA}(\text{ref}) \quad (9)$$

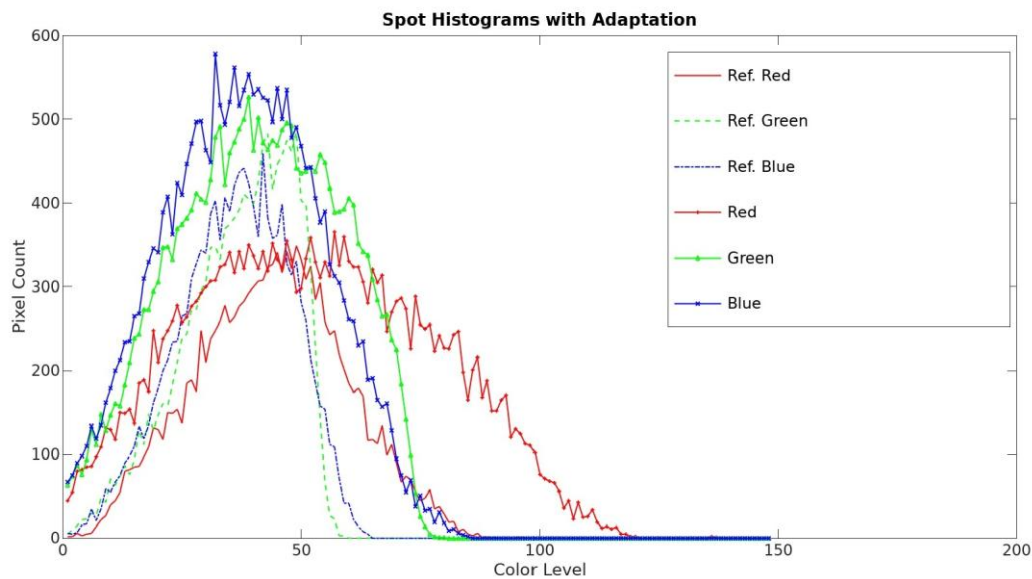
Each one the histogram features “crf” of the trainings samples  $i$ , are adjusted by subtracting  $D_i$ . The loose limits  $K_D(\text{crfL})$  of a feature “crf” and its strict limits  $K_D(\text{crfS})$  are defined in the Color Signature after this subtraction.



**Figure 6.** The red, green, blue color histograms of a reference lesion spot and a spot under test. No color adaptation is performed to the red, green, blue color histograms of the spot under test.

The histograms of Figure 6 are used to demonstrate the lesion spot color adaptation method. They display the histograms of a reference (with average gray level: 149) and a spot under test (average gray level: 159). Figure 7 shows the adapted color histograms after a left shift by  $D = 159 - 149 = 10$  positions. The strict and loose range of the feature  $f$  after the described color

adaptation is (crfSn-D, crfSx-D) and (crfLn-D, crfLx-D), respectively or (0, crfSx-D) and (0, crfLx-D) if for example,  $\text{crfSn-D} < 0$  and  $\text{crfLn-D} < 0$ . The differences  $D_i$  of the photographs in the training set are not identical. If lesion color adaptation is activated, the strict and loose limits defined in a Color Signature are determined as described in the previous paragraphs while the  $G_{LA}(\text{ref})$  value is also stored in this signature. In real time operation, when a new photograph is examined, its normal skin average gray value  $G_{NA}$  is estimated as well as the difference:  $D = G_{LA} - G_{LA}(\text{ref})$ . Then, all of the “crf” feature values extracted from the specific photograph under test, are updated as:  $\text{crf}' = \text{crf} - D$ . The resulting  $\text{crf}'$  values are compared with the adjusted strict and loose limits defined in each one of the disease Color Signatures.



**Figure 7.** The red, green, blue color histograms of a reference lesion spot and a spot under test. Color adaptation is performed to the red, green, blue color histograms of the spot under test.

### 3. Experimental results

Four metrics have been used to assess the proposed skin disorder diagnosis method and the color adjustment method explained in the previous subsection. The Sensitivity metric, measures how many of the photographs that display a specific disorder have been recognized correctly. The Specificity is an indication about the quality in the recognition of the photographs that do not display this disease. The Accuracy is a combination of the two metrics indicating the quality in the recognition of both cases. All these metrics can be expressed as a combination of the following parameters: TP, TN, FP, FN. The True Positive (TP) is the number of photographs that have been correctly recognized with a specific skin disorder. The False Negative (FN) is the number of photographs that display a specific disorder and failed to be recognized. The True Negative (TN) is the number of photographs that were correctly recognized as negative to a specific disease. Finally, the False Positives (FP), are the photographs that were falsely recognized as displaying the specific disease [10].

**Table 4.** Number of test and training photographs.

Skin Disorder	Test Photos	Training Photos	Fraction of Test Photos used for Training
Acne	21	5	24%
Melanoma	54	5	9%
Mycosis	33	5	15%
Papillomas	22	5	23%
Psoriasis	41	2+3(INV)	10%–14%
Vitiligo	21	5	24%
Warts	32	2+3(INV)	12%–18%

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \quad (10)$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \quad (11)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP}, \quad (12)$$

The number of photographs tested for each skin disorder in this paper (ranging between 21 and 54) is listed in Table 4. Five of the most representative photographs of each skin disorder are used for the definition of the corresponding Color Signature. All the tested photographs are somehow represented in this test set in terms of disorder progression, brightness of lesion, color skin, etc. Of course if the test set is extended with new photographs that display the same disease with very different features (such as lesion size, brightness), the training set should also be extended accordingly, to cover these cases too. However, the fraction of training photographs in the test set is always quite small as explained in the previous subsection, giving to the end user the opportunity to perform by himself the statistical processing needed for the Color Signature definition. In some photographs displaying Psoriasis and Warts there are cases where the lesion is darker and some others where it is brighter (marked as “INVerted” in Table 4) than the normal skin. The most appropriate solution for this case is to define two different Color Signatures for the same disease. Only 2 or 3 photographs were used here to define each one of the two signatures.

Three methods related to the color adaptation are experimentally tested. The 1st method does not use color adaptation as described in subsection 2.1. This case is similar to the experiments carried out in [20] and [21] where the success rate metric in [20] is actually the sensitivity as defined in eq. (10) [10]. The 2nd method uses the color adaptation technique described in subsection 2.4 which shrinks the feature ranges but can increase the gaps between them. The 3rd method combines the results of both techniques to counterbalance the effect of extended gaps between feature ranges without using a larger training set. More specifically, in the 3rd method the grades that each skin disorder received from the 1st and the 2nd are added and then the diseases are sorted. The sensitivity achieved in all these three methods is presented in Table 5. Focusing initially on the two methods described in Section 2, we can see that mixed results are obtained. Similarly, in the specificity results listed in Table 6, there are 3 cases where color adaptation achieves better results and 2 cases where it leads to worse specificity. The accuracy results (Table 7) are also balanced.

**Table 5.** Sensitivity comparison.

Skin Disorder	Sensitivity without Color Adaptation	Sensitivity with Color Adaptation	Sensitivity of Combined Results
Acne	85.7%	100%	100%
Melanoma	75.9%	68.5%	77.7%
Mycosis	81.8%	75.6%	90.1%
Papillomas	45.5%	54.5%	68.2%
Psoriasis	36.6%	27%	31.7%
Vitiligo	100%	95.2%	100%
Warts	28.1%	37.5%	43.7%

**Table 6.** Specificity comparison.

Skin Disorder	Specificity without Color Adaptation	Specificity with Color Adaptation	Specificity of Combined Results
Acne	92%	90.6%	90.6%
Melanoma	94.7%	94.7%	94.7%
Mycosis	97.4%	99%	98.9%
Papillomas	99.5%	100%	100%
Psoriasis	100%	100%	100%
Vitiligo	84.2%	83.7%	83.7%
Warts	97.9%	98.9%	98.9%

**Table 7.** Accuracy comparison.

Skin Disorder	Accuracy without Color Adaptation	Accuracy with Color Adaptation	Accuracy of Combined Results
Acne	91.4%	91.4%	91.4%
Melanoma	90.1%	88.4%	90.6%
Mycosis	95%	95.5%	97.7%
Papillomas	94.2%	95.6%	96.8%
Psoriasis	88.3%	86.5%	87.4%
Vitiligo	85.7%	84.8%	85.3%
Warts	87.9%	90.2%	91%

The features used in our approach were tested with the following classification algorithms that have been implemented in the Weka framework [23]: Naïve Bayes, MultiLayer Perceptron (MLP) neural network, J48 Decision Tree and Random Forest [24]. The Sensitivity metric is used to compare these methods with the corresponding results of the proposed method, in Tables 8 and 9. Two cases were examined for the training of these supervised classification methods. In the first case, 20% of the samples were used for training of each algorithm in Weka. The results of this case are presented in Table 8 and are comparable to the results listed in Table 5 that have been derived when 5 photographs are used for training from a test set of 21–54 photographs per disease. The selected classification methods were also trained with 75% of the samples in Weka and the results are presented in Table 9. Although this comparison is not fair for the proposed method that it is trained with a much smaller number of samples, it has been included since  $\frac{3}{4}$  of the test samples are often used for training in machine learning [24].

**Table 8.** Sensitivity comparison with other classifiers (training with 20% of the samples).

Skin Disorder	This method (Combined Results)	Naïve Bayes	MLP	J48	Random Forest
Acne	100%	22.2%	55.6%	0%	5.6%
Melanoma	77.7%	80.1%	81.8%	91%	95.5%
Mycosis	90.1%	51.5%	33.3%	36.4%	57.6%
Papillomas	68.2%	47.1%	70.6%	17.6%	17.6%
Psoriasis	31.7%	60%	55%	30%	60%
Vitiligo	100%	64.7%	76.5%	64.7%	76.5%
Warts	43.7%	50%	62.5%	62.5%	50%

**Table 9.** Sensitivity comparison with other classifiers (training with 75% of the samples).

Skin Disorder	This method (Combined Results)	Naïve Bayes	MLP	J48	Random Forest
Acne	100%	100%	83.3%	33.3%	50%
Melanoma	77.7%	93.8%	81.3%	75%	100%
Mycosis	90.1%	70%	80%	80%	90%
Papillomas	68.2%	40%	60%	60%	40%
Psoriasis	31.7%	100%	100%	100%	100%
Vitiligo	100%	80%	100%	60%	100%
Warts	43.7%	16.7%	100%	50%	100%

#### 4. Discussion

From the experimental results presented in the previous section, it is obvious that the employment of the color adaptation method is not justified because worse results are achieved in some cases, since it shrinks the feature ranges but may leave larger gaps between them. We have not compensated these gaps by increasing the size of the training set. On the contrary, the complementary achievements of both methods are combined as shown in the 3rd column of Tables 5, 6, and 7 and higher accuracy is achieved without having to extend the training set. As far as Table 5 is concerned, the combination of the two methods improves the sensitivity in all cases but psoriasis. The best specificity achieved by the first two methods is also achieved by their combination in most of the cases as shown in Table 6. Finally, the accuracy results of the combination of the two methods presented in Section 2 are improved in most cases

The developed application seems to fail to recognize efficiently the psoriasis and warts skin disorders due to the very low sensitivity results obtained in Table 5. However, this is explained by the very small number of photographs used for training (only 2 or 3). A much better sensitivity can be achieved if both the Color Signatures for bright or dark lesions of these diseases are defined using at least 5 photographs as in the rest of the cases. As was shown in [21] extending the training set from 3 or 5 to 8 photographs improves the accuracy (in most of the cases) by up to 53% (e.g., Warts: from 32% with 3 training photographs, to 85% with 8 training photographs). The absolute number of training photographs is expected to have an upper limit regardless of the test set size when most of the variations in the size and color of the lesions has been covered. Using more training photographs beyond this limit won't lead to higher accuracy.

As can be seen from Table 8, Naïve Bayes, MLP and Random Forest achieve better sensitivity from the proposed method in only 3 of the 7 diseases. A better sensitivity is achieved by the J48 decision tree in only 2 of the 7 skin disorders. The sensitivity results achieved by each classification algorithm are improved in most of the cases when 75% of the samples are used for training as shown in Table 9. However, even in this case, a higher sensitivity than the proposed classification method is achieved by MLP and Random Forest in 3 of the 7 diseases while the other two classifiers (Naïve Bayes and J47) achieve a higher sensitivity in only 2 of the 7 diseases. The cases where the referenced classification algorithms achieve a higher sensitivity than the proposed approach are highlighted with italics in Tables 8 and 9. The accuracy results achieved by the referenced approaches are listed in Table 10. The discriminated diseases are listed in the second column of Table 10 while the accuracy (and/or sensitivity, specificity) achieved is listed in the third column. The 4th column of Table 10 is used to describe mainly the classification method used.

Each referenced approach is tested on different photographs and databases, but the comparison of different classifiers that use the same set of photographs and features has already been performed in the Weka framework as described above. The experimental results presented in Table 10, provide a good indication about the level of accuracy that has been achieved in similar applications in the literature. Higher accuracy is achieved in the approaches where a single disease has to be verified as in [9] and [10] for melanoma. However, accuracy higher than 90% is also reported for applications where multiple diseases are supported ([13,15,19]). As can be seen from Table 10, an accuracy higher than 80% is acceptable in most of the approaches presented in the literature. The accuracy achieved by the proposed approach using the combination of the results of the two methods described in Section 2 is also higher than 80% and comparable with the referenced methods.

Concerning the type of training that has been used in each one of the referenced approaches, it is worth mentioning that in [3], 20497 iterations are needed to achieve a very small Sum Square Error, lower than  $10^{-6}$ . In [5], each classification method was applied ten times and 10-fold cross-validation was used for SVM while in [8], 100 experiments were performed to estimate the sensitivity and specificity listed in Table 10. The database used in [9] consisted of 200 image lesions: 80 normal moles, 80 atypical moles and 40 melanoma moles. The 75% of these images were used for training and the remaining 25% were used for testing. In [13], 704 images were used in the experiments performed. In [15], 2/3 of the images were used for training while the 1/3 for testing. The number of photographs displaying melanoma, psoriasis and dermo in [17] is 32, 30 and 25, respectively. The number of images used for each one of the 9 supported diseases in [19] is between 24 and 277. The test set size used in our approach is comparatively sufficient and the achieved accuracy can be considered representative of the quality of the employed classification method.

The main advantage though, of the proposed method is that it does not require a large number of photographs for the definition of Color Signatures (i.e., for the training). In most of the referenced approaches (e.g., in [8–9]),  $\frac{3}{4}$  of input samples are used for training and only  $\frac{1}{4}$  for test. Moreover, the proposed method gives access to the end user in order to customize the existing diagnosis rules or extend the supported set of skin disorders.



**Table 10.** Comparison of the achieved accuracy with the referenced approaches.

Reference	Skin Disorders	Accuracy	Notes
[3]	Psoriasis	Sum Square Error= $10^{-6}$ (20497 iterations)	Feed Forward Back Propagation NN
[4]	Eczema, Acne, Leprosy, Benign, Dandruff, Syringoma, Mastitis, Scabies, Vitiligo, Diapercandi	90%	NN for Classification
[5]	Acne	70%	k-means (Segmenation)
		66.6%	Acne/ Inflammatory discrimination(SVM Or Fuzzy C-means)
		80%	Acne/normal skin discrimination with Fuzzy C-means
		100%	
[8]	Melanoma	80.7% (Sensitivity) 85.6% (Specificity)	SVM classification
[9]	Melanoma	93.5% (accuracy) 95% (sensitivity) 92% (specificity)	SVM classification
[10]	Melanoma	96.8% (accuracy) 99.9% (sensitivity) 95.2% (specificity)	Segmentation performed with multilevel thresholding
[13]	Acne/Eczema	97%, 98% 89%, 88% 98%, 99%	Feed Forward, Back Propagation NN
[15]	Psoriasis/ Scabies/Vitiligo		
[15]	Psoriasis, Seborrheic Dermatitis, Lichen Planus, Pityriasis Rosea, Chronic Dermatitis, Pityriasis Rubra Pilaris	98%/97%/82%, 93%/92%/88%, 97%/95%/87%, 85%/89%/75%, 92%/91%/83%, 95%/97%/-	The experimental results for each disease were obtained using NN / Decision Trees / k-Nearest Neighbors
[17]	Melanoma, Psoriasis, Dermo	90%	AdaBoost classification framework
[18]	Eczema, Acne, Leprosy, Benign, Dandruff, Syringoma, Mastitis, Scabies, Vitiligo, Diapercandi	90%	Neural Network
[19]	psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis and pityriasis rubra pilaris	91%	Back Propagation NN
This work	Acne, Melanoma, Mycosis, Papillomas, Psoriasis, Vitiligo, Warts	100/91/91%, 78/95/91%, 90/99/98%, 68/100/97%, 32/100/87%, 100/84/85%, 44/99/91%	Sensitivity/Specificity/ Accuracy Combined methods for color adaptation

## 5. Conclusion

A skin disorder classification method based on Color Signatures that can be defined using a

small number of training photographs is presented in this paper. This method can be exploited e.g., by a dermatologist in order to customize/optimize the database of the supported skin disorders and offer this tool to his customers for the remote monitoring of the skin disorders' progression. The combination of multiple image processing techniques such as the described color adaptation, can improve the achieved accuracy without using a large training set. The proposed method is characterized by extensibility and low complexity. The accuracy is quite high compared with referenced approaches and other classifiers in the Weka framework.

Future work will focus on testing different classification methods such as multiple strict/loose feature limits or pure fuzzy rules. Image processing techniques like color normalization applied in color spaces different than RGB will also be examined. The information that can be given by the user about the progression, the feel of the lesion, etc, is also expected to improve the achieved accuracy.

### Acknowledgements

This work is protected by the provisional patents 1009346/13-8-2018 and 1008484/12-5-2015 (Greek Patent Office).

### Conflict of interest

The authors declare that there is no conflict of interest.

### References

- 1 Becevic M, Anderson B, Han JG, et al. (2013) TeleMDID: Mobile Technology Applications for Interactive Diagnoses in Teledermatology Clinics. *Proceedings of the IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 429–433.
- 2 Mannaro K, Baralla G, Ibba S, et al. (2017) Towards a smart region: The case study of a teledermatology platform in sardinian region (Italy). *Proceedings of the IEEE Wireless and Mobile Computing, Networking and Communications (WiMob)*, 370–377.
- 3 Al Abbadi NK, Dahir NS, Al-Dhalimi MA, et al. (2010) Psoriasis Detection Using Skin Color and Texture Features. *Journal of Computer Science* 6: 648–652.
- 4 Kabari LG, Bakpo FS (2009) Diagnosing skin diseases using an artificial neural network. *Proceedings of the IEEE International Conference of Adaptive Science & Technology*, 187–191.
- 5 Alamdari N, Tavakolian K, Alhashim M, et al. (2016) Detection and Classification of Acne Lesions in Acne Patients: A Mobile Application. *Proceedings of the 2016 IEEE International Conference on Electro Information Technology (EIT)*, 739–734.
- 6 Baig R, Bibi M, Hamid A, et al. (2019) Deep Learning Approaches Towards Skin Lesion Segmentation and Classification from Dermoscopic Images – A Review. *Curr Med Imaging Rev* 15: 1–20.
- 7 Bhardwaj A, Bhatia JS (2014) An Image Segmentation Method for Early Detection and Analysis of Melanoma. *IOSR Journal of Dental and Medical Sciences* 13: 18–22.
- 8 Wadhawan T, Situ N, Lancaster K, et al. (2011) SkinScan ©: A Portable Library for Melanoma Detection on Handheld Devices. *Proceedings of the IEEE International Symposium On Biomedical Imaging* 2011: 133–136.

- 9 Joseph S, Panicker JR (2016) Skin Lesion Analysis System for Melanoma Detection with an Effective Hair Segmentation Method. *Proceedings of the International Conference on Information Science (ICIS)*, 91–96.
- 10 Santy A, Joseph R (2015) Segmentation Methods for Computer Aided Melanoma Detection. *Proceedings of the Global Conference on Communication Technologies (GCCT)*, 490–493.
- 11 Mirzaalian H, Lee TK, Hamarneh G (2014) Spatial Normalization of Human Back Images for Dermatological Studies. *IEEE Journal of Biomedical and Health Informatics* 18: 1494–1501.
- 12 Ghervase L, Carstea EM, Borisova E, et al (2015) Bringing Light into the Diagnosis of Skin Disorders - Short Review on Laser Induced Fluorescence Spectroscopy and Optical Coherence Tomography in Dermatology. *Curr Med Imaging Rev* 10: 297–303.
- 13 Arifin MS, Kibria MG, Firoze A, et al. (2012) Dermatological Disease Diagnosis Using Colour-skin Images. *International Conference on Machine Learning and Cybernetics* 5: 1675–1680.
- 14 Islam MN, Gallardo-Alvarado J, Abu M, et al. (2017) Skin disease recognition using texture analysis. *Proceedings of the IEEE International Conference on Control and System Graduate Research Colloquium (ICSGRC)*, 144–148.
- 15 Kumar VB, Kumar SS, Saboo V (2016) Dermatological Disease Detection Using Image Processing and Machine Learning. *Proceedings of the International Conference on Artificial Intelligence and Pattern Recognition (AIPR)*, 1–6.
- 16 Yasir R, Rahman A, Ahmed N (2014) Dermatological Disease Detection using Image Processing and Artificial Neural Network. *Proceedings of the 8th International Conference on Electrical and Computer Engineering*; 687–690.
- 17 Ambad PS, Shirsat AS (2016) An Image analysis System to Detect Skin Diseases. *IOSR Journal of VLSI and Signal Processing* 6: 17–25.
- 18 Kabari LG, Bakpo FS (2009) Diagnosing skin diseases using an artificial neural network. *International Conference on Adaptive Science & Technology*, 187–191.
- 19 Abdul-Rahman S, Norhan AK, Yusoff M, et al. (2012) Dermatology Diagnosis with Feature Selection Methods and Artificial Neural Network. *IEEE EMBS Confence on Biomedical Engineering and Sciences*, 371–376.
- 20 Petrellis N (2018) Using Color Signatures for the Classification of Skin Disorders. *Proceedings of the IEEE International Conference on Modern Circuits and Systems Technology (MOCASST)*, 1–4.
- 21 Petrellis N (2018) The Effect of the Training Set Size in a Skin Disorder Classification Application. *2018 41st International Conference on Telecommunications and Signal Processing*, 1–5.
- 22 Petrellis N (2017) A Smart Phone Image Processing Application for Plant Disease Diagnosis. *International Conference on Modern Circuits and Systems Technology (MOCASST)*, 1–4.
- 23 Frank E, Hall M and Witten I (2016) The WEKA workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, Fourth Edition, 2016.
- 24 Petrellis N (2018) A Review of Image Processing Techniques Common in Human and Plant Disease Diagnosis. *Symmetry* 10: 270.

