Automatic Extraction of Dermatological Parameters from Nevi Using an Inexpensive Smartphone Microscope: A Proof of Concept

Automatic Extraction of Dermatological Parameters from Nevi Using an Inexpensive Smartphone Microscope: A Proof of Concept

Kristen M. Meiburger, Federica Veronese, Vanessa Tarantino, Massimo Salvi, Matteo Fadda, Silvia Seoni, Elisa Zavattaro, Bruno De Santi, Nicola Michielli, Paola Savoia, Filippo Molinari

***Abstract*— The evolution of smartphone technology has made their use more common in dermatological applications. Here we studied the feasibility of using an inexpensive smartphone microscope for the extraction of dermatological parameters and compared the results obtained with a portable dermoscope, commonly used in clinical practice. Forty-two skin lesions were imaged with both devices and visually analyzed by an expert dermatologist. The presence of a reticular pattern was observed in 22 dermoscopic images, but only in 10 smartphone images. The proposed paradigm segments the image and extracts texture features which are used to train and validate a neural network to classify the presence of a reticular pattern. Using 5-fold cross-validation, an accuracy of 100% and 95% was obtained with the dermoscopic and smartphone images, respectively. This approach can be useful for general practitioners and as a triage tool for skin lesion analysis.**

I. Introduction

Skin cancer incidence is increasing among fair-skinned individuals worldwide [1]. Teledermatology is the practice of sending digital images of skin lesions with relevant clinical information to a dermatologist for their opinion, and promising results in terms of diagnostic accuracy have been shown [2]. The technological development in smartphones has also led to an increase of their use in medical applications [2], [3].

Numerous studies in literature have faced the issue of the digital analysis of dermatological images [4]. Recently, a dermatologist-level classification of skin cancer was obtained using deep neural networks [5], showing the high level of accuracy that can be achieved with the same networks. Numerous smartphone apps exist for melanoma detection, but they rely only on the smartphone camera and studies have shown how these applications are inferior to in-person consultations and can have a high rate of incorrect melanoma classification [6]. Studies have also been done using dermoscopes coupled with smartphones [7], but the device is costly (~$900). Recently, an inexpensive derma-specific microscope was proposed by the South Korean company Nurugo (Nurugo Derma, Nurugo^{TM}, cost: ~$50). The images acquired with the Nurugo Derma can provide high resolution images and can possibly allow the extraction of important dermatological features.

Dermatologists typically rely on global or peculiar features for the differentiation between benign melanocytic lesions and melanoma. Among important global features is the presence of a reticular pattern with a symmetric distribution, which can be defined as a pigment network covering most parts of the nevus [3]. If just a reticular pattern with a symmetric distribution is observed, the lesion is more likely diagnosed as a melanocytic nevus, whereas if the reticular pattern is observed without symmetry and in combination with two or more other patterns or vessels or veil, the lesion is more likely diagnosed as melanoma [3].

The analysis of texture features has been used in many applications [8], and specifically in dermatology [9]. The main objective of this study is to compare the extraction of dermatological parameters (i.e., reticular pattern presence), from melanocytic lesions using a portable dermoscope and the Nurugo Derma, as a proof of concept. In particular, we show it is possible to determine the presence of a reticular pattern that is not observed with the naked eye, using smartphone-acquired images, texture features, and neural networks.

II. Materials and Methods

A. Image database

Forty-two cutaneous nevi were imaged with both a portable dermoscope (HEINE DELTA®20 T, Heine Optotechnik, Herrsching, Germany) and the Nurugo Derma (Nurugo^{TM}), making a total of 84 acquired images. The study was approved by the local ethics committee.

The dermoscopic images had a calibration factor equal to 0.012 mm/pixel (image size: 1704 x 2272 pixel^2). The Nurugo Derma images were acquired using an iPhone 6s, and the calibration factor was equal to 0.0047 mm/pixel (image size: 4032 x 3024 pixel^2).

An expert dermatologist (F.V.) with over 10 years of experience manually segmented the nevi in all images and annotated if the presence of a reticular pattern was visually observable or not. Fig. 1 shows examples of the images used. A reticular pattern was visually observed in 22 dermoscopic

K.M. Meiburger, M. Salvi, M. Fadda, S. Seoni, B. De Santi, N. Michielli and F. Molinari are with the PoliTo^{BIO}Med Lab, Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy. (phone: +39 011-090-4207; e-mail: kristen.meiburger@ polito.it).

F. Veronese, V. Tarantino and P. Savoia are with the Dermatology Unit, Dept. of Health Science, University of Eastern Piedmont, Novara, Italy.

E. Zavattaro is with the Dermatology Unit, Dept. of Translational Medicine, University of Eastern Piedmont, Novara, Italy.

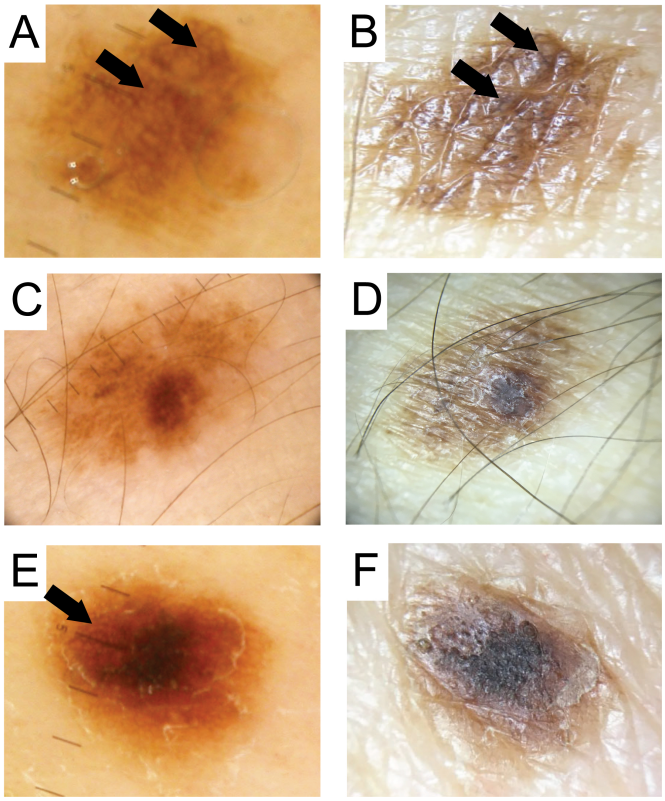

Figure 1. Example of images used. First column: dermoscopic images; second column: Nurugo Derma images. First row (A, B): example of reticular pattern visible in both images. Second row (C, D): example of reticular pattern not visible in either image. Last row (E, F): example reticular pattern visible in dermoscopic image (E) but not in Nurugo image (F).

images (52%), whereas in the Nurugo Derma images, it was visually observed in only 10 images (24%). A reticular pattern was never observed in the smartphone-acquired images when not observed in the dermoscopic images.

B. *Automatic algorithm architecture*

Image preprocessing and segmentation:

The first step of our automatic algorithm paradigm is the segmentation of the nevi within the image. In order to do so, some preprocessing steps were first done on both the dermoscopy and smartphone images (Fig. 2A and 2E).

First of all, since the smartphone-acquired images present artifacts due to illumination, they were first pre-processed to remove any glare that may be present on the image due to the smartphone flash. To do so, an adaptive thresholding technique was done on the image in order to obtain a mask that defines the glare pixels. The image was then interpolated with neighborhood pixel values to remove the glare. Secondly, a hair removal technique using Gabor filters was employed for both the dermoscopy and Nurugo Derma images so as to remove these features which can modify the texture feature calculation (Fig. 2B and 2F).

Subsequently, both dermoscopic and Nurugo Derma images were automatically segmented using the same technique. First the image was converted to the LAB color space and to gray scale, I_{gray}, by intersecting the color planes: $I_{gray} = (a + b) - L$. Then, the grayscale image was low-pass filtered (average filter, 0.18mm x 0.18mm) and the contrast was enhanced (Fig. 2C and 2G). Finally, a k-means algorithm with two classes was employed to obtain the final segmentation (Fig. 2D and 2H).

Texture feature extraction:

After automatic segmentation, our paradigm extracts texture features from within the segmented region-of-interest (ROI) of the preprocessed image on the three color layers (i.e., red, green and blue layers) and on the grayscale image. The parameters were computed by custom developed software in MATLAB (The MathWorks, Natick, MA, USA).

First-order statistical descriptors

Five features that are based on first-order statistics were extracted from the ROI: pixel intensity mean, variance, skewness, kurtosis and entropy. These features depend on the single gray level of the pixel (Table 1).

Haralick and Galloway descriptors

Haralick's features are based on the calculation of the gray level co-occurrence matrix (GLCM) [10], which measures the number of times a specific intensity pattern between adjacent pixels is repeated. Since adjacency can be measured in four principal directions, the GLCM is computed using four angles: 0°, 45°, 90° and 135°. The Haralick features mathematically describe the GLCM through the calculation of the features reported in Table 1.

The Galloway features are based on the run length matrix (RLM) *R,* where each element *R(i,j)* corresponds to the number of pixels with run length *j* and intensity *i* in a given direction [11]. The Galloway features are mathematical descriptors of the runs of the RLM (Table 1). These features are also directional, so they are computed using the same 4 angles (0°, 45°, 90°, and 135°).

Since both the Haralick and Galloway features are computed for four directions, a total of 40 of these descriptors are extracted for each ROI.

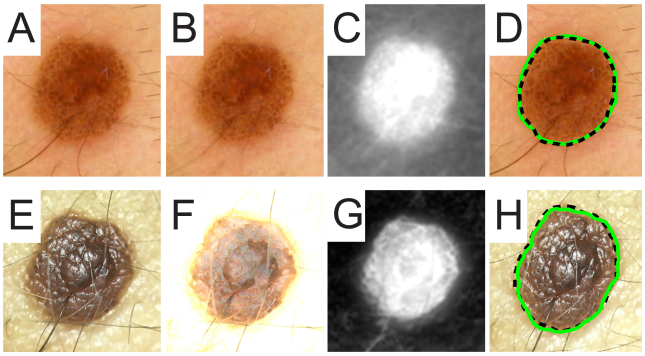

Figure 2. Summary of segmentation method. First row: dermoscopic image, second row: Nurugo Derma image. (A,E): Original images. (B,F): Preprocessed images. (C,G): LAB grayscale filtered and contrast-enhanced images. (D,H): Final segmentation. The manual segmentation is shown in the dotted black line, whereas the automatic segmentation is shown with the solid green line.

Table 1. Mathematical description of texture features

Feature name	Mathematical description
Mean (m)	$m = \sum_{x=1}^{M} \sum_{y=1}^{N} \frac{I(x,y)}{M \times N}$
Variance (σ^2)	$\sigma^2 = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} \{I(x,y) - m\}^2}{M \times N}$
Skewness (S_k)	$S_k = \frac{1}{M \times N} \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} \{I(x,y) - m\}^3}{\sigma^3}$
Kurtosis (K_t)	$K_t = \frac{1}{M \times N} \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} \{I(x,y) - m\}^4}{\sigma^4}$
Entropy$_1$ (E_1)	$E_1 = -\sum_{i=1}^{Nl-1} hist(i) * \log_2(hist(i))$
Haralick Symmetry (I_{sym})	$I_{sym} = 1 - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \lvert i-j \rvert P(i,j)$
Haralick Contrast (I_{con})	$I_{con} = \sum_{n=0}^{N-1} n^2 \left\{ \sum_{i=0}^{N} \sum_{j=0}^{N-1} P(i,j) \right\}$
Haralick Homogeneity (I_{hmg})	$I_{hmg} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{1}{1+(i-j)^2} P(i,j)$
Haralick Entropy (I_{Entr})	$I_{Entr} = -\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j) \log P(i,j)$
Haralick Correlation (I_{cor})	$I_{cor} = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i,j) P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Galloway short run emphasis (SRE)	$SRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{R(i,j)}{j^2}}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} R(i,j)}$
Galloway long run emphasis (LRE)	$LRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} j^2 R(i,j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} R(i,j)}$
Galloway gray-level non-uniformity (GLNU)	$GLNU = \frac{\sum_{i=1}^{N_g} \left(\sum_{j=1}^{N_r} R(i,j) \right)^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} R(i,j)}$
Galloway run length non-uniformity (RLNU)	$RLNU = \frac{\sum_{i=1}^{N_r} \left(\sum_{j=1}^{N_g} R(i,j) \right)^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} R(i,j)}$
Galloway run percentage (RP)	$RP = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} R(i,j)}{N_g \cdot M_r}$

I(x,y) denotes the input ROI. *Nl* is the number of gray levels, *hist* is the normalized histogram counts. $\sigma_x, \sigma_y, \mu_y, \mu_y$ are the standard deviations and means of P_x, P_y, the marginal probability density functions. $p_x(i) = i^{th}$ entry in the marginal-probability matrix obtained by summing the rows of $P(i,j)$. N_g represents the number of gray values in the image (i.e., the numberof rows of *R*). N_r represents the number of runs (i.e., the number of columns of *R*).

Reticular pattern classification:

There were a total of 180 features for each image: 5 first-order and 40 Haralick and Galloway descriptors, which were calculated on 4 images (i.e., red, green, blue layers and grayscale image). Due to the limited data in our current database and the large number of features that were calculated, feature selection was first employed to help reduce the problem of overfitting of the model. Therefore, collinear variables were removed by the Belsley collinearity diagnostics technique [12].

The remaining features were then used to train and test a neural network (NN) using 5-fold cross-validation. Since we are interested in determining if the reticular pattern can be extracted from the images acquired with an inexpensive smartphone microscope even when visually not observed, the visual reticular pattern presence or absence of the dermoscopic images was taken to be the output class for both NNs. Then, two different NNs were used, one for the images acquired with the clinical dermoscope and one for the Nurugo Derma-acquired images. The same test and train cross-validation folds were used for the two NNs. Several network configurations were tested, varying the network parameters (#units, #layers, activation functions and learning rate) and, for each configuration, the train and test set errors for each fold were evaluated. The optimal neural network was chosen in two steps: first, the networks that had a classification error standard deviation over the 5 folds below 0.10 for both train and test sets were selected. Among these, the network that gave forth the highest classification accuracy was selected as the final NN configuration.

Performance validation:

To validate the segmentation, the automatic segmentation mask was compared to the manually segmented mask and the precision, recall, and Dice coefficient were calculated.

On the other hand, the quality of the reticular pattern presence classification was evaluated by calculating the average confusion matrix of the final classification obtained over the 5-fold cross-validation.

III. Results

A. Segmentation results

All 84 images were correctly processed. The dermoscopic image segmentation showed a final precision, recall, and Dice coefficient equal to 91%, 89%, and 90%, respectively. On the other hand, the images acquired with a smartphone and the Nurugo Derma had final performance values equal to 92%, 85%, and 88%, respectively.

B. Texture and classification results

After removing the collinear variables, there was a total of 38 final features employed for the dermoscopic images (m^R, E_1^R, S_k^R, K_t^R, I_{con0}^R, I_{cor0}^R, I_{Entr0}^R, I_{hmb0}^R, I_{con45}^R, I_{cor45}^R, I_{Entr45}^R, I_{hmb45}^R, I_{sym90}^R, I_{con90}^R, I_{cor90}^R, I_{Entr90}^R, I_{con135}^R, I_{hmb135}^R, SRE_0^R, LRE_0^R, $GLNU_0^R$, $RLNU_0^R$, RP_0^R, SRE_{45}^R, LRE_{45}^R, $RLNU_{45}^R$, SRE_{90}^R, LRE_{90}^R, $GLNU_{90}^R$, $RLNU_{90}^R$, RP_{90}^R, SRE_{135}^R, LRE_{135}^R, E_1^G, σ^{2B}, E_1^B, S_k^B, E_1^{gray}).

For the Nurugo Derma images, on the other hand, a final number of 32 features were used (σ^{2R}, S_k^R, K_t^R, I_{sym0}^R, I_{cor0}^R, I_{Entr0}^R, I_{hmb0}^R , I_{sym45}^R, I_{cor45}^R, I_{hmb45}^R, I_{con90}^R, I_{cor90}^R, I_{sym135}^R, I_{con135}^R, $I_{Entr135}^R$, SRE_0^R, LRE_0^R, $RLNU_0^R$, SRE_{45}^R, $GLNU_{45}^R$, $RLNU_{45}^R$, SRE_{90}^R, LRE_{90}^R, $GLNU_{90}^R$, $RLNU_{90}^R$, RP_{90}^R, SRE_{135}^R, LRE_{135}^R, E_1^B, S_k^B, K_t^B, I_{con45}^B).

The architectures for the final neural networks employed for the reticular pattern presence classification are shown in Fig. 3A and 3B. The trained NN obtained a 100% accuracy rate in all 5 cross-validations folds for both the train and the test set when considering the dermoscopic images. For the Nurugo Derma images, the final average accuracy for the training set was equal to 95 ± 2%., whereas for the test set the average accuracy was equal to 95 ± 6%.

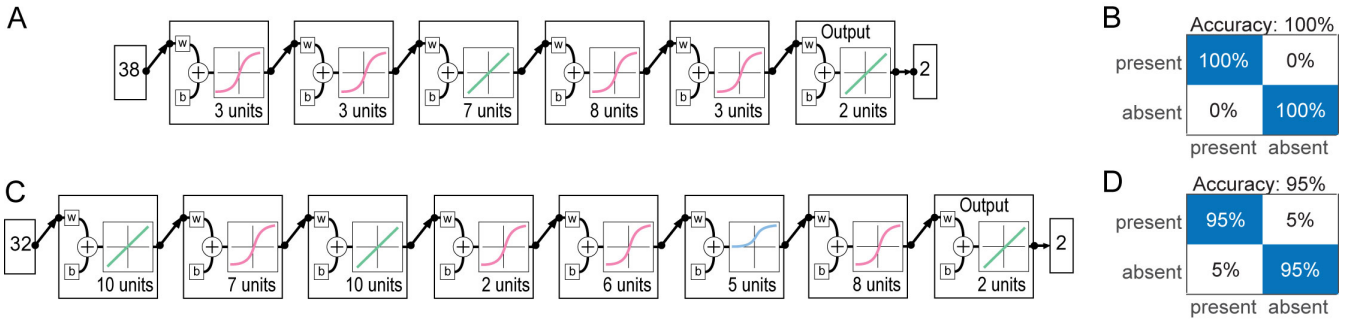

Figure 3. Neural network (NN) configurations and classification results. (A) NN architecture for dermoscopic images. (B) Average confusion matrix for dermoscopic images. (C) NN architecture for Nurugo images. (D) Average confusion matrix for Nurugo images.

IV. Discussion and Conclusions

This work has shown that the proposed approach can effectively segment and extract texture parameters from dermoscopic and Nurugo Derma smartphone-acquired images and subsequently classify the presence or absence of a reticular pattern. Specifically, it can be appreciated how the deep learning technique is able to correctly classify 100% of the dermoscopic images in each cross-validation fold. This result can be expected since the reticular pattern was able to be visually appreciated in the images, confirming how the NN was able to correctly learn which features reflect the reticular pattern presence. More importantly, however, the deep learning approach was able to reach a very high accuracy (95%) also with the Nurugo Derma images, in which a reticular pattern was visually appreciated only in 45% (10/22) of the images that actually presented a reticular pattern. This shows how the extracted texture features coupled with an effectively trained NN can extract information from the image that is unable to be appreciated by the naked eye. In fact, the extracted texture features, which present a majority of red layer features, are able to effectively describe neighboring pixel patterns relevant to the reticular pattern.

The study proposed here is to be considered as a proof-of-concept, demonstrating how it is possible to extract important dermatological features from Nurugo Derma smartphone-acquired images. The study is limited by its small database size and the lack of a classification between benign and malignant lesions. However, the results presented here are an important demonstration of the potential capabilities of this approach. We are currently working to amplify the database to include a much larger number of images and to include also melanocytic malignant lesions and epithelial lesions (like Non Melanoma Skin Cancer), and plan to further validate the proposed technique and compare it with other classification methods.

In conclusion, the proposed method is very inexpensive, it is easily accessible and usable, and could find a good use by general practitioners to help them address expert dermatologists only in doubtful or urgent cases, and it could also be useful as a triage tool.

References

[1] R. J. T. van der Leest *et al.*, "The Euromelanoma skin cancer prevention campaign in Europe: characteristics and results of 2009 and 2010," *J. Eur. Acad. Dermatology Venereol.*, vol. 25, no. 12, pp. 1455–1465, Dec. 2011.

[2] S. Kroemer *et al.*, "Mobile teledermatology for skin tumour screening: diagnostic accuracy of clinical and dermoscopic image tele-evaluation using cellular phones," *Br. J. Dermatol.*, vol. 164, no. 5, pp. 973–979, May 2011.

[3] G. Argenziano *et al.*, "Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet," *J. Am. Acad. Dermatol.*, vol. 48, no. 5, pp. 679–693, May 2003.

[4] S. J. Coates, J. Kvedar, and R. D. Granstein, "Teledermatology: From historical perspective to emerging techniques of the modern era: Part II: Emerging technologies in teledermatology, limitations and future directions," *J. Am. Acad. Dermatol.*, vol. 72, no. 4, pp. 577–586, Apr. 2015.

[5] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[6] J. V. Wang and L. W. Chapman, "Challenges to smartphone applications for melanoma detection," *Dermatol. Online J.*, vol. 23, no. 2, p. no pagination, 2016.

[7] A. Börve, K. Terstappen, C. Sandberg, and J. Paoli, "Mobile teledermoscopy-there's an app for that!," *Dermatol. Pract. Concept.*, vol. 3, no. 2, pp. 41–8, Apr. 2013.

[8] K. M. Meiburger *et al.*, "Quantitative analysis of patellar tendon abnormality in asymptomatic professional 'Pallapugno' players: A texture-based ultrasound approach," *Appl. Sci.*, vol. 8, no. 5, 2018.

[9] M. A. Wahba, A. S. Ashour, Y. Guo, S. A. Napoleon, and M. M. A. Elnaby, "A novel cumulative level difference mean based GLDM and modified ABCD features ranked using eigenvector centrality approach for four skin lesion types classification," *Comput. Methods Programs Biomed.*, vol. 165, pp. 163–174, Oct. 2018.

[10] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[11] M. Galloway, "Texture analysis using grey level run lengths," *Comp Graph. Image Process*, vol. 4, pp. 172–179, 1975.

[12] D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression diagnostics : identifying influential data and sources of collinearity*. Wiley, 2004.