Artificial Neural Networks (ANN) for automatic detection of dendritic-shaped cancer cells of cutaneous melanoma in Reflectance Confocal Microscopy (RCM) images

Giorgio De Nunzio[1,2,*], Luana Conte[1,2], Massimo Federico[3], Francesca Farnetani[3], Michele Maffia[2,4], Benedetta Tafuri[1,2], Giovanni Pellacani[3]

[1]*Laboratory of Biomedical Physics and Environment, Department of Mathematics and Physics "E. De Giorgi", University of Salento, Lecce, Italy*
[2] *Interdisciplinary Laboratory of Applied Research in Medicine (DReAM), University of Salento and ASL Lecce, 'V. Fazzi' Hospital, Lecce, Italy*
[3]*University of Modena and Reggio Emilia, Department CHIMOMO, Modena, Italy*
4Laboratory of Physiology, Department of Biological and Environmental Sciences and Technologies, University of Salento, Lecce, Italy

* Corresponding author: giorgio.denunzio@unisalento.it

Abstract
Melanoma (MM) is one of the tumors with the highest incidence. In Italy, MM affected about 13,700 patients out of 373,000 new cases of cancer in 2018, with prognosis dependent on the degree of tumor invasion and presence of metastasis at diagnosis: only an early detection can lead to a better prognosis. Recent evidence suggests that MM is a family of different tumors with varying abilities to grow and metastasize: dendritic-shaped tumor cells were typically found in thin MM in situ. Reflectance Confocal Microscopy (RCM) is a non-invasive imaging tool that enables in vivo observation of the skin at a quasi-histological resolution, providing transverse-section grayscale images related to refractive index of different tissues. In this work, a dataset of RCM images, from 13 healthy subjects and 22 patients affected by MM in situ, were used to train a Multi-Layer Perceptron (MLP) artificial neural network. Each image was subdivided into sub-blocks, labeled as positive if containing significant clusters of dendritic-shaped tumour cells. In each block, various standard features were calculated, e.g. Haralick's and features from the run-length matrices. The MLP was trained to recognize the presence of clusters of dendritic-shaped cancer cells. The preliminary results are encouraging, giving AUC=0.81 with about 73% accuracy. Tests are currently underway to improve quality.

Keywords: Artificial Neural Networks, Reflectance Confocal Microscopy, Computer Aided Detection, Radiomics, Melanoma

1. INTRODUCTION

Cutaneous melanoma (MM) arises from the malignant transformation of melanocytes derived from normal skin or a pre-existing melanocytic nevus [1]. Since melanoma in advanced stages is still incurable, early diagnosis is indispensable to reduce mortality.

In the era of precision medicine [2], the identification and stratification of atypical nevi in the early stage of cancer development is an essential starting point for increasing the probability of therapeutic success. To date, however, identification of MM lesions is affected by an unsatisfactory rate of false-positive results due to the way in which diagnosis is performed. In fact, conventional clinical diagnosis of MM is based on some observable criteria such as Asymmetry, Borders, irregularity, non-uniform Colour, and Diameter >6mm (ABCD acronym), that only provide about 64% accuracy [3]. Therefore, diagnosis is often complemented with surgical excision for performing biopsy, an invasive and painful method that obviously leaves a scar. It is also time-consuming because many days are needed before the clinician gets the report back to complete the diagnosis [3].

In this scenario, clinical dermatology needs to undergo a technological revolution, through the development of novel imaging techniques that provide detailed information about the skin in a non-invasive manner [4]. Among these techniques, Reflectance Confocal Microscopy (RCM) enables *in vivo* visualization of the skin with a resolution at the cellular level, providing an alternative to histopathology [3]. This performant technique allows in vivo examination by producing an optical biopsy using a non-invasive procedure, which helps clinicians in real-time diagnosis of MM at an earlier and curable stage, avoiding unnecessary scars from surgical biopsies of benign lesions, reducing costs and time consuming medical procedures [3]. RCM works with incident light technology. The skin is illuminated from above with a focused laser in the near infrared band, and the reflected light is directed through a pinhole onto a detector. The contrast in the images relies on the differences in the reflectivity of the tissue that depends on the molecular and chemical structures. In RCM, structures with a higher refractive index appear bright and different refractive indices lead to particular reflection patterns in different shades of gray [5].

Clinically, melanomas appear as dark, flat or slightly raised mark on the skin, with different colours. Borders can be irregular, with indentations or notches [3]. On the contrary, on RCM, melanocytes appear bright due to their higher refractive index and can be depicted with great definition due to the clear endogenous contrast of melanin [5], leading to distinguish two different microscopic subtypes of MM: pagetoid MM and solar MM [3]. Unlike pagetoid MM, that normally occurs in adults with intermittent solar exposure and showing high number of nevi on the skin, solar MM is frequent in patients with low nevi, in areas with high solar exposition [3] and there is no evidence of occurring in pre-existing nevus [3]. In addition, on RCM, atypical melanocytes in solar MM appear as clusters of cells with dendritic branches [3].

The introduction of novel artificial intelligence (AI) technologies applied to the diagnosis and possibly to the prognosis of MM, could largely revolutionize the status of the management of MM patients. In this work, the development of a Computer-Aided Detection (CAD) software tool based on Artificial Neural Networks (ANN) for the automatic detection of dendritic-shaped cancer cells of solar cutaneous MM on a series of RCM images is described.

The results expected from this innovative automated system of assisted diagnosis concern the possibility to support the clinician in reading the images, typically difficult to interpret,

during the diagnostic process, identifying clusters of malignant cells that can be used as biomarkers of disease. In particular, the presence and the automatic detection of dendritic-shaped tumour cells can be immediately presented to the attention of the clinician in any doubt case, when it is difficult - even for the naked eye of the best dermatologist - to discriminate between atypical nevus vs. melanoma, leading to the correct therapeutic choice. A CAD tool can also be used as a first reader in the screening of large courts of patients, to focus the attention of the clinician primarily on images of suspicion or obvious malignancy. This software would also support the prognosis, since it would be able to quantitatively measure the reduction of these atypical cell clusters during follow-up or even assess whether the scar zone shows undamaged margins after surgery.

2. MATERIALS AND METHODS

A. RCM data

In this work, a dataset of RCM images, from 13 healthy subjects and 22 patients affected by solar melanoma in situ, were used to train a Multi-Layer Perceptron (MLP) ANN. The images were acquired in 'Città di Lecce Hospital' – Gruppo Villa Maria (GVM), site in Lecce, Italy.

Imaging was performed using Vivascope 1500 RCM which uses a laser diode with a near-infrared wavelength of 830 nm. Laser power is low so that the light causes no damage to tissue, but imaging depth is limited to 200–300 µm which corresponds more or less to the papillary dermis. VivaScope 1500 provides basic images with a 500×500 µm horizontal field of view (x, y) at a preselected imaging depth (z). RCM allows the scanning of the entire area of the lesion up to 8×8 mm. An automated stepper can then generate a grid of contiguous horizontal images (1000x1000-pixel each) forming a two-dimensional composite mosaic called VivaBlock.

After setting the referencing mark to zero, three mosaics were acquired from all the subjects at the epidermal level (z = 30 µm), DEJ (z = 60-90 µm), and upper dermis (z = 90–120 µm), respectively [3]. Figure 1 shows an example of RCM images.

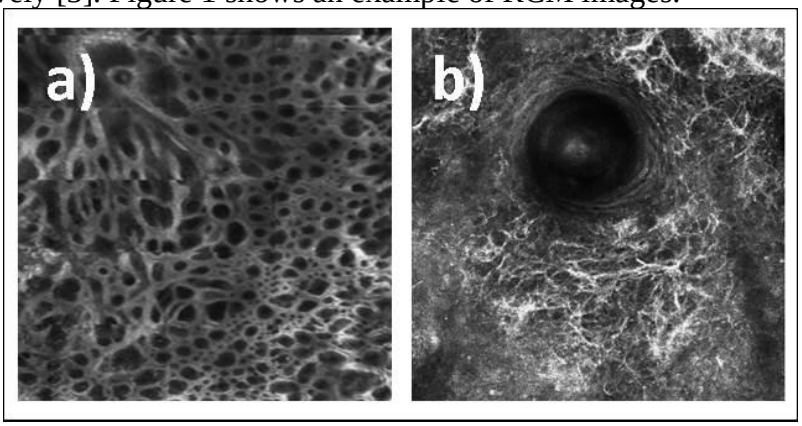

Figure 1. Example of dermis level of benign nevus, with regular and conserved architecture of dermal papillae (a) and dermis level of solar melanoma, with irregular architecture and clusters of dendritic-shaped cancer cells around a hair follicle (b).

B. Image Processing

A Graphical User Interface (GUI) was developed in MATLAB (The MathWorks, Inc. Natick, Massachusetts), in order to allow the dermatologists to easily label image parts

according to presence or absence of pathological tissues (the latter, characterized by dendritic-shaped tumour cells).
Vivablocks often show some blurring, variable illumination between mosaic blocks, and noise. To avoid calculations to be affected by the discontinuities existing between mosaic blocks, every image was split and each block was then considered as a single ROI, where features were calculated.
Each step of loading, labelling and splitting the images was achieved using the GUI (Figure 2).

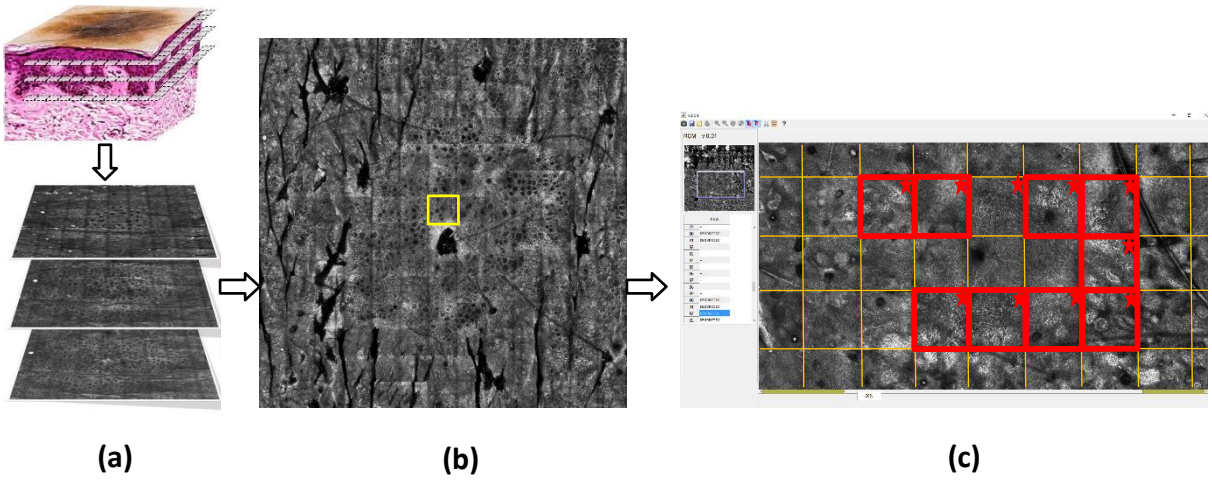

Figure 2. Scheme of RCM image processing. Figure (a) shows the acquisition of three (or more) RCM layers of the skin (epidermis/DEJ/derma). These images are mosaic-structure Vivablocks. Figure (b) shows a whole mosaic-structure Vivablock composed by blocks of 1000x1000 pixels. The yellow square shows a single block. Figure (c) shows the GUI interface: blocks are labelled by the dermatologists according to the presence of pathological blocks (red squares) and then split and placed in form of a 3D object (depth allowing to index single blocks) ready for easy access by the feature-calculation and MLP software. Yellow line is a convenient grid.

C. Feature Extraction

To compute the features that describe the lesion textures, the Radiomics MATLAB toolbox developed by Vallières et al. [6] was used. The computed features, used as input for the neural network, were extracted from the intensity histogram (first-order statistics), the gray-level co-occurrence matrix (GLCM, Haralick's second-order features), the gray-level run-length matrix (GRLM), the gray-level size zone matrix (GLSZM) and the neighborhood gray-tone difference matrix (NGTDM) [7]. Information of each method and the corresponding features can be found in [6].
Prior to the computation of texture features, the images were quantized to a lower number of gray levels (32) to improve the signal-to-noise ratio [8].
Finally, before being used in the training/validation loop, all the features were normalized (between 0 and 1) to avoid model computation being affected by differences in the feature scales [9].

D. Model Evaluation

The flow chart in Figure 3, illustrates the main software steps, starting with block labeling by the user interface (left), followed by feature calculation (centre) and classification (right).

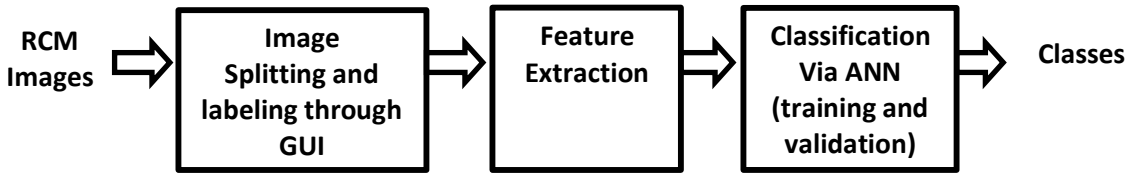

Figure 3. The main steps of the classification algorithm.

After various tests with several network architectures, a feed-forward neural network (MLP) composed of 1 hidden layer with 5 neurons, trained by supervised learning with backpropagation, was used to classify the RCM images into two classes (Negative vs presence of Dendritic-shaped tumour-cell clusters), .

In order to have random patient partitioning during hold-out, avoiding the arbitrary selection of training and validation sets by an operator, the patients were beforehand randomly permuted and then automatically partitioned through a subtle procedure that takes into account that the number of ROIs per patient (the 1000x1000-pixel blocks) may vary substantially and also the number of positive and negative blocks per patient may vary. Accordingly, the partitioning algorithm performs a kind of 'bin packing', a term that reminds the packing of arbitrary parcels in the most effective way [10]. In this way, positive and negative blocks were divided into training and validation groups in ways that approximate the request of particular proportions between dataset populations, paying attention to work with balanced datasets. The training set was then used to build the model and then this model was evaluated using the validation set. Model performance was measured using the Receiver Operating Characteristics (ROC) and the Area Under the ROC Curve (AUC).

3. RESULTS

Figure 4 shows the ROC curve obtained from the trained network applied to the validation set. In order to assign a value for the uncertainty on the ROC AUC, training and performance calculation were repeated 20 times, and the mean value of AUCs together with its standard deviation were calculated, giving AUC = 0.81±0.02. The accuracy, sensitivity and specificity values were calculated at the ROC point closest to (0,1) giving values around 73%.

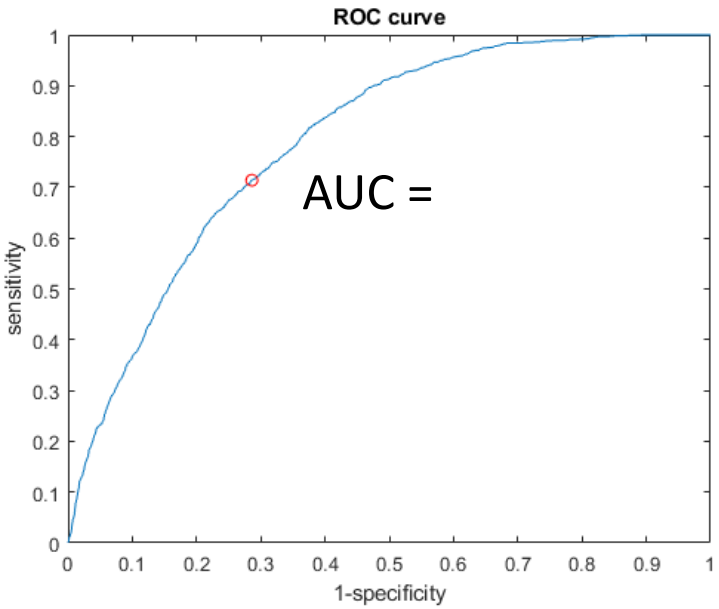

Figure 4. ROC curve and the related AUC

With the same particular choice of the ANN output cutoff, the results obtained by the CAD system on the patients in the validation dataset were mapped onto their images and compared with the truth. An example for one patient is shown in Figure 5. The figure shows a particular frame of a RCM image loaded in the GUI, which was previously labelled by an expert dermatologist according to the presence or absence of clusters of tumour cells ("DENDRITIC" vs "-" , respectively). Contextually, on the same image the output of the trained classifier for the same image is overlapped: red squares and green X symbols highlight the positive blocks found by the software, whereas the green minus identify the negative blocks.

It is easy to assess that there is a good match between positive blocks determined by the expert operator and positive blocks found by the software, for 14 out of 15 positive blocks. Similarly, 4 negative blocks out of 5 were correctly found as negative by the software. Only 1 block (n° 51) was considered positive in spite of being labelled as negative by the expert dermatologist, probably because there are still some cells on the bottom, that the software has considered as a cluster of tumour cells. Another single block was seen as negative (block n° 52), although it was tagged as positive by the operator.

These results correspond to an accuracy of 90%, which is better than the average accuracy of the software, previously shown. Of course, other regions of this image and of other images show inferior performance.

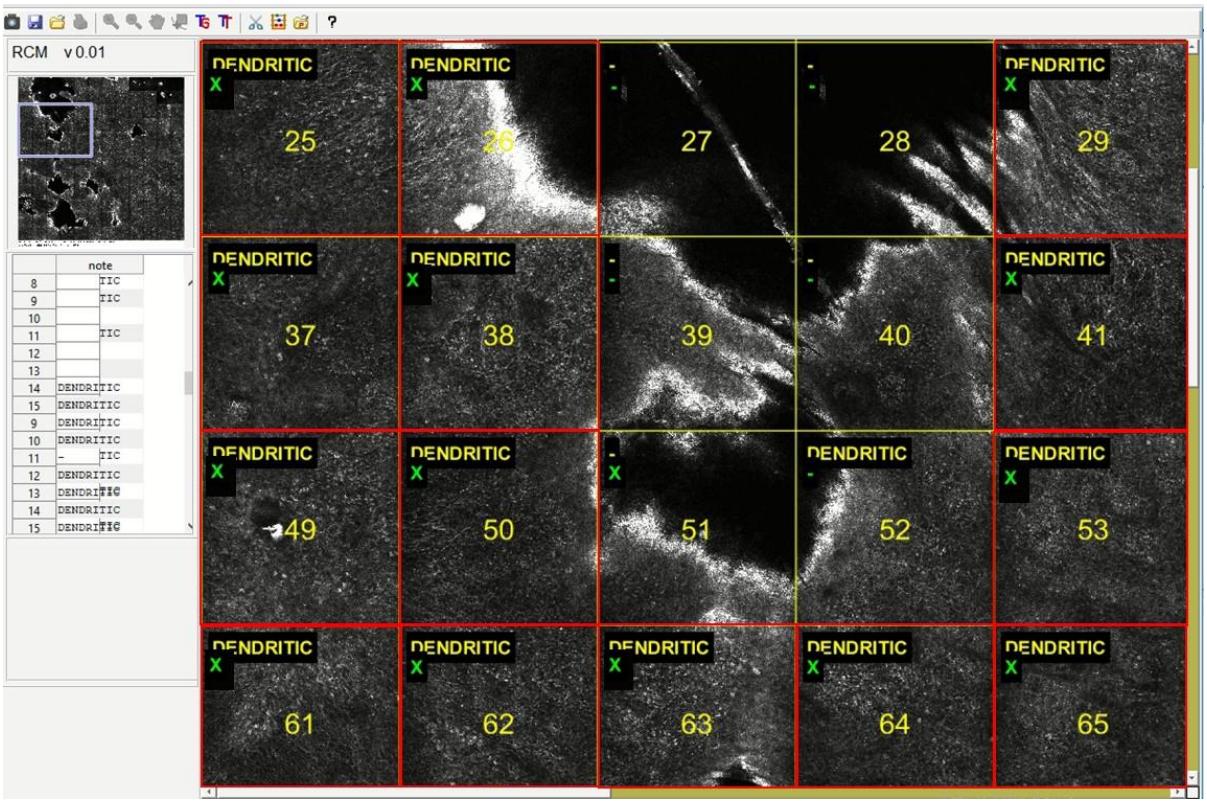

Figure 5. Output of the CAD system for dendritic-shaped tumour cells detection, for a patient of the validation set.

4. CONCLUSIONS

This work reports the preliminary results obtained from a CAD system developed to identify areas of dendritic-type tumour cells in RCM images. Dendritic-type cancer cells can be associated with the presence of solar melanoma, so in perspective the described system can be used as an aid to the dermatologist in the detection of this tumour.

Despite the immaturity of the software, the first results obtained are comforting. The quality of the system can be summarized by the area under the ROC curve, which stands at about 81%, and in the accuracy, sensitivity and specificity values (all around 73%) obtained with a particular choice of the work point in the ROC space.

Of course, the database used for its development is still limited, and has to be significantly expanded to cover a larger number of situations and learn to cope with the variability of the images of different patients.

A thorough investigation of the results obtainable on the dataset with the current architecture and choice of features is now in order. Further cases will be tested in the next future to implement the performance of the CAD system, and different network architectures will be investigated.

In addition, interesting developments go in the direction of a Deep Learning approach. Deep Learning systems allow pattern recognition without the need to choose the best

features, which are automatically detected. On the other hand, the size of the training dataset needs to be significantly increased.

In conclusion, this is an embryonic work, but with great potential and promising preliminary results.

References

[1] C. Bertolotto, Scientifica 2013, 635203, (2013).

[2] R.L.N. Godone, G.M. Leitão, N.B. Araújo, C.H.M. Castelletti, J.L. Lima-Filho, D.B.G. Martins, Biomedicine & Pharmacotherapy **106**, 14–34 (2018).

[3] R. Hofmann-Wellenhof, G. Pellacani, J. Malvehy, H. Soyer, Springer Science & Business Media (2012).

[4] S. González, Aula Médica 111 (2012).

[5] J. Welzel, M. Ulrich, S. Lange-Asschenfeldt, U. Hohenleutner, Current Rev.(2011).

[6] M. Vallières, C.R. Freeman, S.R. Skamene, I. El Naqa, Physics in Medicine and Biology **60**, 5471–96 (2015).

[7] R. Ortiz-Ramon, A. Larroza, E. Arana, D. Moratal, 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 493–6 (2017).

[8] P. Gibbs, L.W. Turnbull, Magnetic Resonance in Medicine **50**, 92–8 (2003).

[9] M. Kuhn, K. Johnson, A Short Tour of the Predictive Modeling Process, Applied Predictive Modeling: Springer New York, 19–26 (2013).

[10] B. Korte, J. Vygen, Combinatorial Optimization **2**, 471–488 (2006).