Final report on the ICT for Health laboratories

Luca Gioacchini
s257076

A.A.2018/19

Contents

1 Skin Conditions: Nevi

Skin conditions are medical conditions which affect the integumentary system (skin and other organs which aim to protect the body). They comprehend several diseases and thousands of skin conditions have been described.

Moles, or *nevi*, are a very common example of skin conditions. They are lesions of skin and are related to an alteration in the melanocytes' functionality. Melanocytes are neural crest-derived cells of the bottom layer of the skin's epidermis, producing melanin, a dark pigment responsible of the skin color. These alterations are skin cells mutations, or tumors. Depending on the mutation type, the tumor can be benign or malignant (*melanomas*).

Melanomas can be diagnosed through the nevi observation. There are five parameters:

- **A**: Asymmetrical shape. Melanomas are asymetric nevi;
- **B**: Borders. Melanomas have irregular borders;
- **C**: Color.
- **D**: Diameter. Melanomas have larger ones;
- **E**: Evolving. Melanomas change over time.

To automatically detect skin cancers a teledermatological tool such as a nevi picture classifier can be useful to improve the diagnosis quality, since, for example, the waiting lists and times are even increasing and there are too few dermatologists.

Since the tool is a classifier, a procedure called *feature extraction* is needed. This procedure consists on assigning a value to each one of the five "ABCDE" features. A good example of feature extraction is the development of a tool which can assign values to different nevi pictures by observing the B feature, or the nevi border.

The dataset is made of 58 .jpeg photos of nevi. 13 are classified as "low risk", 17 as "medium risk" and 25 as "melanoma".

2 Main Procedure

In order to perform the feature extraction on the moles pictures, it is necessary to turn the dataset from a collection of .jpeg images (which can be seen as 3D matrices storing the pixels amount of Red, Blue and Green) into a set of 2D matrices representing the images converted into two-tone ones. To perform this conversion it is necessary to apply the *K-means* algorithm, which allows to define clusters of data by observing their attributes. In this case the attribute is the color and the data are the pixels of the moles pictures. The K-means algorithm detects three clusters, whom centroids represent the three darkest colors in the .jpeg image. The pixels are assigned to the clusters on the basis of their distance from the cluster centroids.

After that a user must manually select the rectangular area containing the three-colored mole and the image is converted into the final two-tones matrix containing pixels marked as 0 if they belong to the background, as 1 if they belong to the mole.

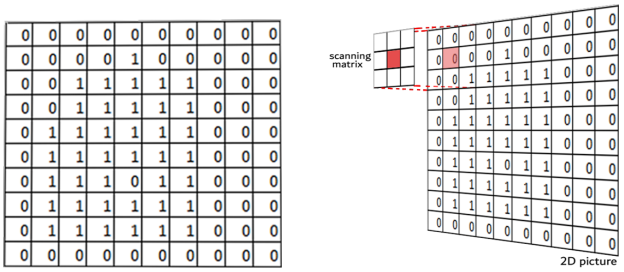
The value associated to the "border" feature is a number called *perimeter ratio*. It is defined as the ratio between the the mole perimeter and the one of a circle having the same area of the mole. By observing Figure 3(a) it is clear that the picture need to be cleaned to remove all the impurity due to the low resolution of the original image and to the shadows in order to determine the perimeter ratio. Section 3 described the cleaning algorithm.

3 Cleaning Algorithm

The idea at the basis of the cleaning algorithm is to scan the original 2D matrix made of pixels and modify some "irregular" ones analyzing the n nearest elements. This is because the probability that a 1 would be a 0 if the n surrounding ones are 1 too is very low. However, after having scanned the whole matrix, some clusters of irregular pixels might be still irregular. In this way, after having detected the contour of the main picture, another scan is performed, in order to correct these clusters.

The algorithm is divided into three phases: The subject definition and superficial cleaning one, the contour finding one and the internal cleaning one.

3.1 Subject Definition



(a) original 2D picture

(b) 1st scan. The red pixel is the center of the scanning matrix

Figure 1: First phase of the cleaning algorithm: defining the main picture by using a scanning matrix

By considering the matrix shown in Figure 1(a), the aim of the first phase is to clean the main irregular pixels (e.g. the element marked as 1 in position [1, 4] and the 0 in position [6, 4][1]). Together with the original matrix, a smaller and empty one is created. It will be called *scanning matrix* and it is centered in the pixel which is going to be processed. By considering the *radius* of the analysis process as the number of pixels to consider from the centered one, the scanning matrix dimensions are fixed by setting the radius (this means that a 4px ray leads to a matrix $\in \mathbb{R}^{9x9}$). A scanning matrix with a 1px ray is shown in Figure 1(b).

[1]Start counting from 0

The task of the $\mathbb{R}^{N\text{x}N}$ scanning matrix is to set the central pixel value as the most frequent one, determined by analysing the $N^2 - 1$ pixels around it (from an implementation point of view it consists of determining the mode of the values inside the scanning matrix). This procedure is repeated for all the 2D matrix elements.

3.2 Contour Finding

After having defined the main image and having cleaned it superficially, it is necessary to find the contour of the image. By recalling that the new picture is composed by a set of 0s (background) and 1s (the mole), the contour is defined as the first pixel marked as 1 centered in the set [0, 1, 1] or [1, 1, 0].

In this phase, an adaptive algorithm has been developed. The idea of the algorithm is to choose randomly the first pixel marked as 1 after a 0 when the image is parsed from left to right, mark it as 3, which means that it is a contour pixel, and move in one of the 9 possible directions[2] by sliding a scanning matrix $\mathbf{S} \in \mathbb{R}^{3x3}$ similar to the Section 3.1 one. When the sliding is performed, the pixel in position $\mathbf{S}_{2,2}$ is considered as a contour one and it is marked as a 3. This process is repeated until the first contour pixel found is not reached again.

The algorithm is based on a training phase and on an application phase.The task of the training phase is to manually classify all the possible $\mathbf{S}$ matrices by considering the direction as the class. These information are recorded in a file which is consulted during the application phase in order to obtain a new image with pixels marked as 0s (background), 1s (mole) and 3s (contour), as the one shown in Figure 2.

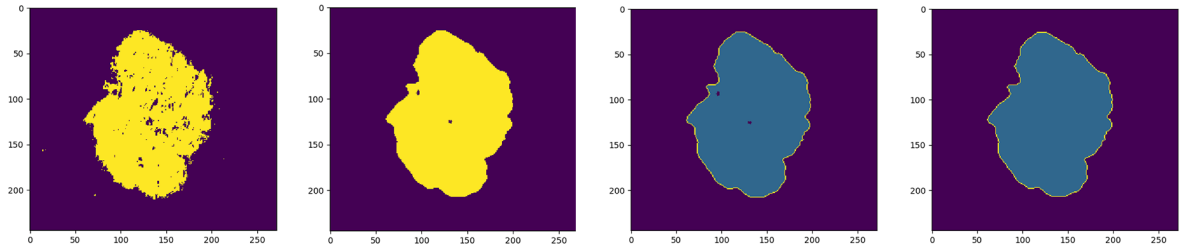
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	3	3	3	3	3	0	0	0
0	0	3	1	1	1	3	0	0	0
0	3	1	1	1	1	3	0	0	0
0	3	1	1	1	1	3	0	0	0
0	3	1	1	1	1	3	0	0	0
0	3	1	1	1	1	3	0	0	0
0	3	3	3	3	3	3	0	0	0
0	0	0	0	0	0	0	0	0	0

Figure 2: Second phase of the cleaning algorithm: contour finding

3.3 Internal Cleaning

In the third phase all the remaining irregularities, which are the clusters made of $x > N^2$ pixels, are replaced, by scanning again the matrix row-by-row. If the actual pixel is a 3, so it belongs to the contour, it is used as a trigger: until another 3 is detected, all the encountered 0s are replaced with a 1. Of course, there are some exceptions, e.g. if the image has a little indentation, probably there will be two 3s in succession, so they are skipped and the trigger is not considered. To improve the third phase performances and to fix some possible errors due to the moles shape, the matrix is scanned once from left to right and once form right to left.

[2]Possible directions: $\uparrow, \nearrow, \rightarrow, \searrow, \downarrow, \swarrow, \leftarrow, \nwarrow$



(a) Original 2D picture (b) 1st phase: cleaning (c) 2nd phase: contour finding (d) 3rd phase: internal cleaning

Figure 3: Step-by-step results of the cleaning algorithm application 4px radius

Figures 3(b) to 3(d) show the resulting matrices after each phases of the cleaning algorithm applied to the sample image shown in Figure 3(a). The radius of the scanning matrix defined in Section 3.1 is set equal to 4px.

4 Results and Conclusions

Tables 1 to 3 show the perimeter ratio results after having performed the algorithms described in Sections 2 and 3. Three samples (medium_risk_1, melanoma_17 and melanoma_27) have failed the algorithm application because of the pictures low quality.

According to what has been said in Section 1, the moles classification depends on five items, so the perimeter ratio is the result of the feature extraction procedure. By observing Figure 4 which shows the probability density functions (pdf) of the three classes (low risk, medium risk and melanoma) determined from the values of the perimeter ratio, it is clear that, even if it is possible to determine a threshold to distinguish between medium risk and melanoma, the low risk pdf completely overlays into the medium risk one, so the threshold determination is not possible.

The obtained result are consistent with the feature extraction procedure: to perform a good moles classification more features are required; the perimeter ratio alone based on the mole contour is not sufficient.

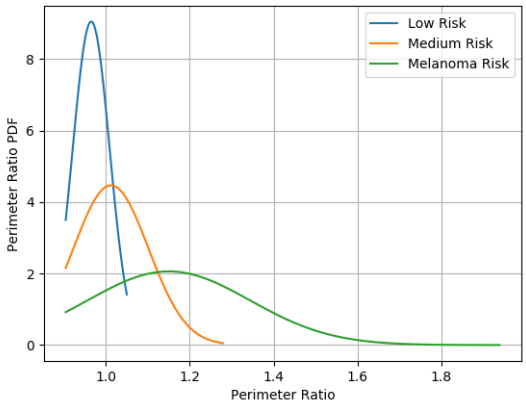


Figure 4: Probability Density Functions obtained from the perimeter ratio values reported in Tables 1 to 3

File Name	Ratio
medium_risk_2	0.94
medium_risk_3	0.93
medium_risk_3_h	0.93
medium_risk_3_s	0.91
medium_risk_4	0.98
medium_risk_5	1.27
medium_risk_6	1.10
medium_risk_7	1.04
medium_risk_8	1.02
medium_risk_9	1.05
medium_risk_10	1.09
medium_risk_11	1.02
medium_risk_12	0.94
medium_risk_13	0.97
medium_risk_14	0.95
medium_risk_15	1.02
medium_risk_16	0.98

Table 1: Perimeter Ratio obtained from the algorithm application. Medium risk moles.

File Name	Ratio
melanoma_1	1.00
melanoma_2	1.07
melanoma_3	1.32
melanoma_4	1.21
melanoma_5	1.18
melanoma_6	1.27
melanoma_7	1.06
melanoma_8	1.18
melanoma_9	1.32
melanoma_10	1.07
melanoma_11	1.11
melanoma_12	0.98
melanoma_13	0.98
melanoma_14	1.02
melanoma_15	1.25
melanoma_16	1.13
melanoma_18	0.97
melanoma_19	1.10
melanoma_20	1.13
melanoma_21	1.18
melanoma_22	1.01
melanoma_23	1.93
melanoma_24	1.21
melanoma_25	0.97
melanoma_26	1.12

Table 2: Perimeter Ratio obtained from the algorithm application. Melanoma.

File Name	Ratio
low_risk_1	1.01
low_risk_2	0.94
low_risk_3	1.05
low_risk_3_h	0.91
low_risk_3_s	0.94
low_risk_4	0.98
low_risk_5	0.98
low_risk_6	0.95
low_risk_7	0.98
low_risk_8	1.00
low_risk_9	0.90
low_risk_10	0.90
low_risk_11	0.94

Table 3: Perimeter Ratio obtained from the algorithm application. Low risk moles.