

Principles of Biosignals and Biomedical Imaging - Bioengineering Department

MATLAB Project - Image Processing of Dermoscopic Images [EN]

Authors:

Armando Gonçalves (100290)
Patrícia Marques (100357)

armandogoncalves@tecnico.ulisboa.pt
patriciasmarques@tecnico.ulisboa.pt

Abstract:

This project uses MATLAB to classify 10 skin lesion images as benign or malignant based on their standard deviation and circularity. In order to achieve accurate results, there's a combination image processing and cluster analysis with knowledge on skin cancer and diagnostic techniques, specifically identifying melanomas and distinguishing them from keratosis-like lesions. The project highlights the importance of computer-aided analysis in dermatology and shows the potential for using MATLAB as a tool for image analysis and classification.

2022/2023 – 2º Semester, P3

1 Introduction

Skin cancer is one of the most common type of cancer in the world. It occurs when cells in the skin mutate and grow uncontrollably, forming a tumor [3]. These mutations are often caused by exposure to ultraviolet (UV) radiation from the sun or other sources, such as tanning beds. Other risk factors include having fair skin, a history of sunburns, a weakened immune system, and exposure to certain chemicals or radiation. There are several different types of skin cancer, including basal cell carcinoma, squamous cell carcinoma, and melanoma [8].



(a) Basal cell carcinoma [2]



(b) Squamous cell carcinoma [1]



(c) Melanoma [7]

Figure 1: Most common types of skin cancer.

Melanoma is the deadliest form of skin cancer, and it is responsible for the majority of skin cancer-related deaths. Unlike other types of skin cancer, which typically develop in sun-exposed areas of the body, melanoma can occur anywhere on the skin [5]. It can also develop in areas that are not exposed to the sun, such as the soles of the feet and the palms of the hands.

One of the challenges of identifying melanoma is distinguishing it from keratosis-like lesions, which are small and rough benign scaly growths that can look very similar to melanomas [6] (see figure 2). The **ABCDE rule** is a commonly used method for identifying melanomas.



(a) Atypical Seborrheic Keratosis [11]



(b) Melanoma ([depositphotos](#))

Figure 2: Keratosis like lesion vs. Melanoma

1.1 ABCDE rule

ABCDE stands for **A**symmetry - **B**order irregularity - **C**olor variation - **D**iameter greater than 6 millimeters - **E**volution (changes in size, shape, or color over time) [9]. By using this rule, individuals can be more confident in identifying potentially dangerous skin growths and seek medical attention if necessary.

We are given a dataset of 10 dermoscopic images. Each one is either a keratosis-like lesion (benign) or a melanoma (malignant), there being five of each category. For this work, the main goal is to **relate the ABCDE rule with skin lesion features** and **classify** them as "**Benign**" or "**Malignant**". For this purpose, we will aim at two objectives; on one hand, we will try to sort the data into two groups or clusters, on the other, we will try to correctly classify them.

2 State-of-the-Art

Dermoscopy is a non-invasive diagnostic technique that involves the use of a handheld device called a dermatoscope to examine skin lesions for signs of melanoma. It works by illuminating the skin with polarized light and magnifying the area of interest. This allows doctors to see the structure of skin lesions and identify features associated with melanoma, such as irregular borders, uneven color distribution, and abnormal blood vessels.

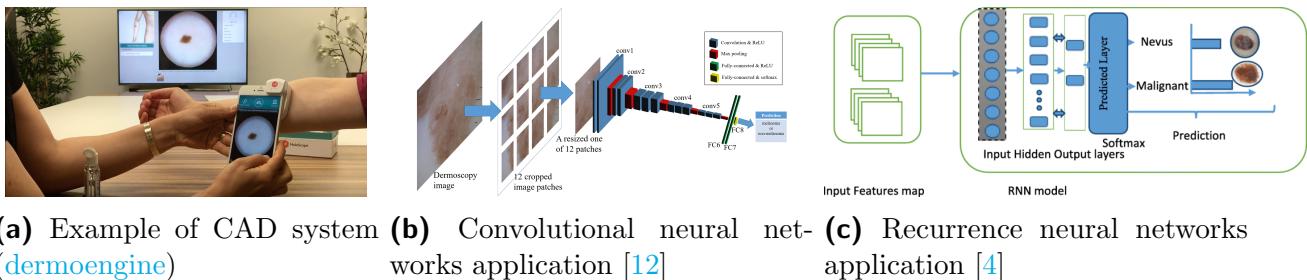
To enhance the accuracy of dermoscopy, various computer-aided diagnosis (CAD) systems have been developed in recent years. These systems use machine learning algorithms to analyze dermoscopic images and provide diagnostic assistance to dermatologists.

CAD systems typically rely on deep learning models that are trained on large datasets of dermoscopic images with annotations provided by expert dermatologists. The models learn to

identify features associated with melanoma and can make predictions about the likelihood of malignancy for a given lesion [10].

There are various techniques used in deep learning models for melanoma detection, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models that combine both. CNNs are designed to recognize patterns and features in an image by breaking it down into smaller pieces and analyzing each piece individually. The network applies a series of filters or "kernels" to the input image, which helps to identify features such as edges, corners, and textures. These features are then combined and passed through multiple layers of the network, where they are progressively refined and used to classify the image [12]. The final output of the network is a prediction of what the image contains, such as a certain object or class. RNNs are a type of neural network commonly used in natural language processing and speech recognition, but they can also be applied to medical imaging analysis. RNNs have the ability to analyze sequential data, meaning they can take into account the context of each data point based on the previous ones. In the context of diagnosis of skin lesions, RNNs can be used to analyze sequential images of skin lesions taken over time. For example, if a patient has a lesion that is being monitored for changes in size or color over a period of time, an RNN can be trained to detect these changes and alert the physician if they are potentially indicative of a malignant lesion [4].

These models have shown promising results in clinical studies and are being integrated into commercial dermatoscopes to assist in the early detection of melanoma.



3 Methods

In this study, we implement a classic image processing pipeline to classify dermoscopy images (figure 6.1) into two classes: benign (keratosis-like lesions) and malignant (melanomas).

It was used the programming language MATLAB. The first step is to import all the .jpg images from the working directory using `imagedatastore` and `readall` functions. Then, the images are converted to grayscale and inverted, using `rgb2gray` and `imcomplement`. Next, the images are denoised by applying a Gaussian filter (`imgaussfilt`) to remove any noise. The denoised images are then thresholded using methods such as Otsu to obtain binary masks that separate the foreground (lesion, white) from the background (skin, black), using for that `graythresh` and `imbinarize` functions. The pre-processed masks are cleaned using morphological operations such as binary dilation (`imdilate`), border clearing (`imclearborder`), small object removal (`bwareaopen`), and hole-filling (`imfill`) within the lesion masks. To keep only

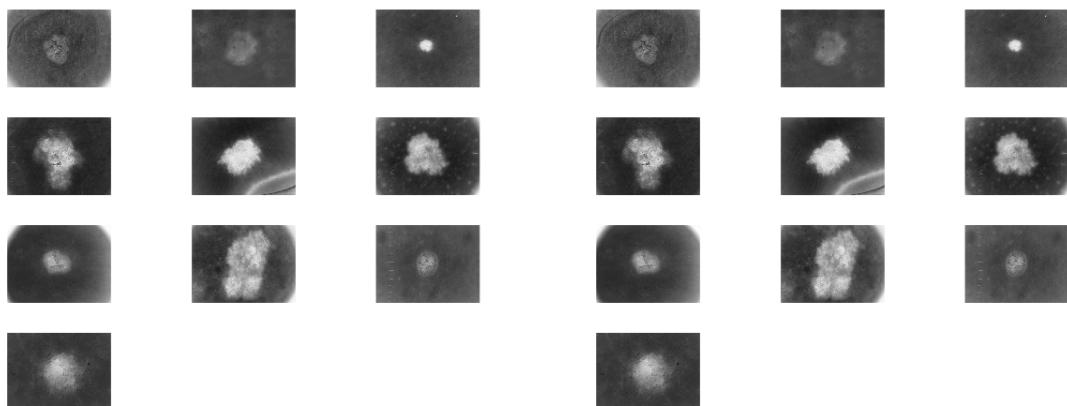
the largest connected component it was used `bwareafilt`. Features are extracted by multiplying the grayscaled and inverted images with the final binary masks. The standard deviation of the intensity distribution of each lesion is computed, and `regionprops` is used to obtain the circularity of each mask.

Finally, k-means clustering (`kmeans`) is performed to automatically separate the two classes (benign / malignant lesions). The features are plotted in the 2D feature plane using `scatter`, along with the corresponding estimated class label for each image. This allows for easy visualization and analysis of the classification results.

The implementation of the code can be seen in section [6.2](#).

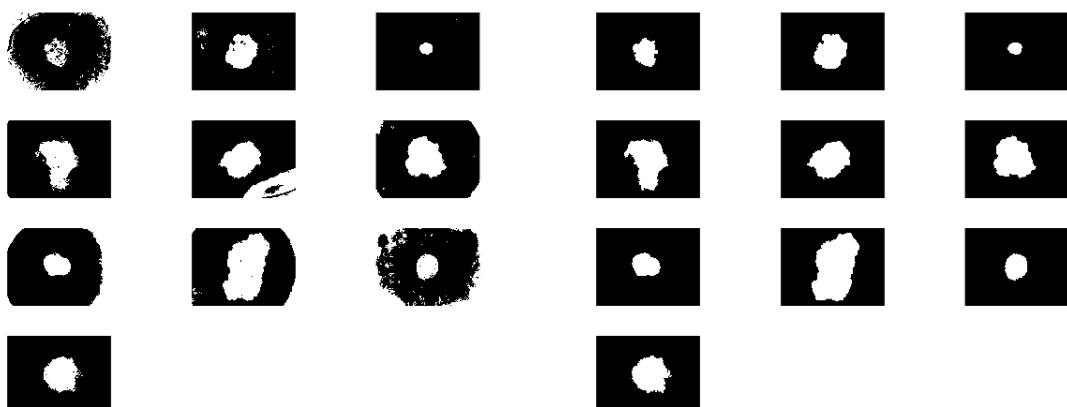
4 Results and Discussion

After applying the methods, we get the following results:



(a) Inverted grayscaled images.

(b) Filtered images.



(a) Thresholded images.

(b) Masked images.

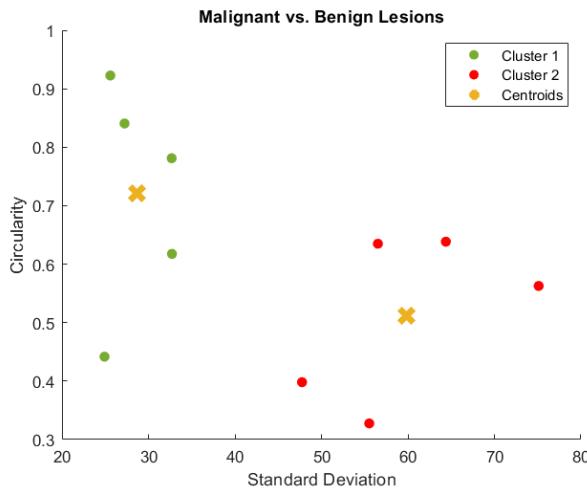


Figure 6: Final plot output.

Table 1: Features extracted from images

Image	Std dev	Circularity
”image0.jpg”	24.9175	0.4417
”image1.jpg”	32.7166	0.6175
”image2.jpg”	25.5942	0.9227
”image3.jpg”	55.5259	0.3276
”image4.jpg”	64.3821	0.6384
”image5.jpg”	56.5298	0.6349
”image6.jpg”	32.6795	0.7812
”image7.jpg”	75.1156	0.5627
”image8.jpg”	27.2339	0.8406
”image9.jpg”	47.7586	0.3981

As can be seen in the graph 6, two clusters were obtained in the end, as expected. To arrive at the final result, certain values for the parameters asked throughout the application of the methods were chosen. Through experimentation, it was concluded that the values that yielded the most satisfactory results (i.e., a clear division in the clusters) were the following:

- Denoising - Gaussian filter sigma: 1
- Denoising - Gaussian filter filterSize: 5
- Morphological Operations - disk-shaped element radius: 3
- Morphological Operations - maximum number of pixels for object removal: 100

We need now to classify each cluster as benign or malignant. To do that, we take into account the ABCDE rule 1.1 and the meaning of the studied features. The ABCDE rule suggests that malignant lesions tend to have asymmetry, irregular borders, uneven color, larger diameter, and evolving features. The features extracted (circularity and standard deviation) correspond to exactly two of them. Circularity is equivalent to asymmetry; the more the lesion looks circular, the less it is asymmetrical. Standard deviation of the intensity has a correspondence with uneven color, because the higher standard deviation value, the less homogeneity exists in the color of the lesion. We can further say

- Circular lesions, which tend to have higher circularity values, are more likely to be benign (keratosis-like lesions).
- Lesions with a larger spread of color values, which tend to have higher standard deviation values, are more likely to be malignant (melanomas).

Therefore, we can say **cluster 2**, which has lower circularity and higher standard deviation values, is more likely to correspond to **malignant** lesions based on these criteria. Following the same line of reasoning, **cluster 1** must correspond to **benign** lesions, as has higher values of circularity and lower values of standard deviation. Therefore, graph 6 ends up like:

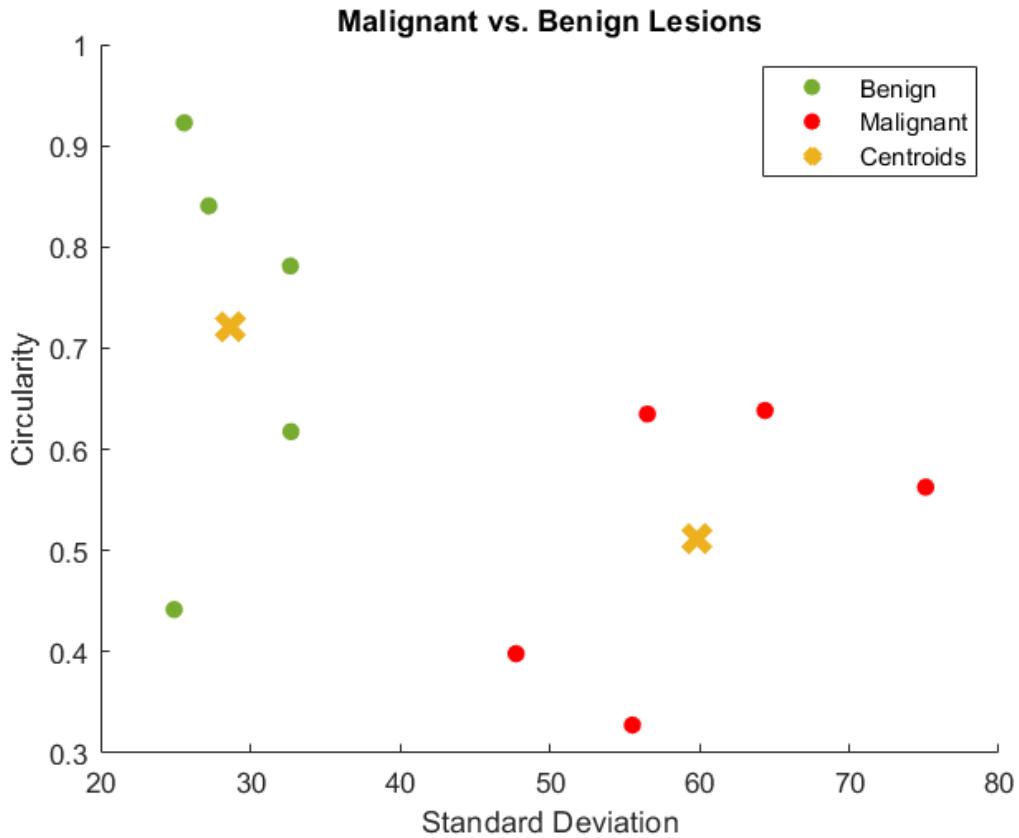


Figure 7: Final plot with the clusters identified.

Together with the table 1, we are now in position to classify every image in the dataset as benign or malign.

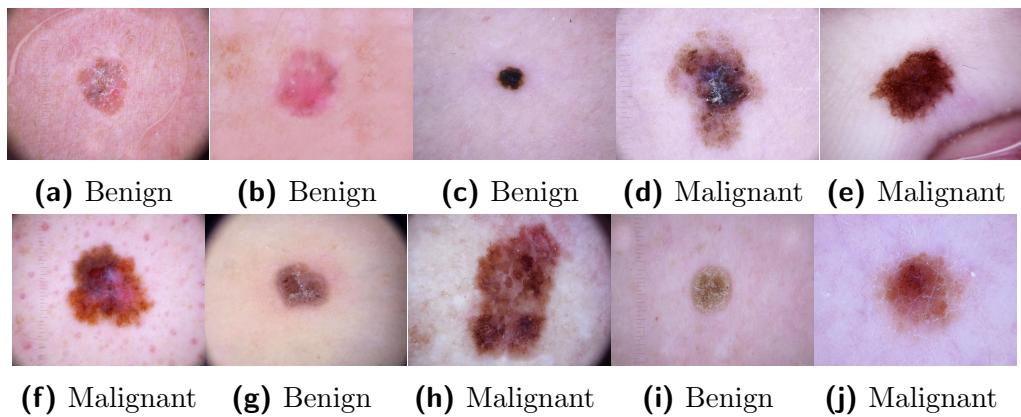


Figure 8: Classification of images in "benign" or "malignant".

To seek for validation, two images known to correspond to a benign and a malignant lesion were selected and included in the dataset along with the original images. The resulting plot is shown in the following figure:

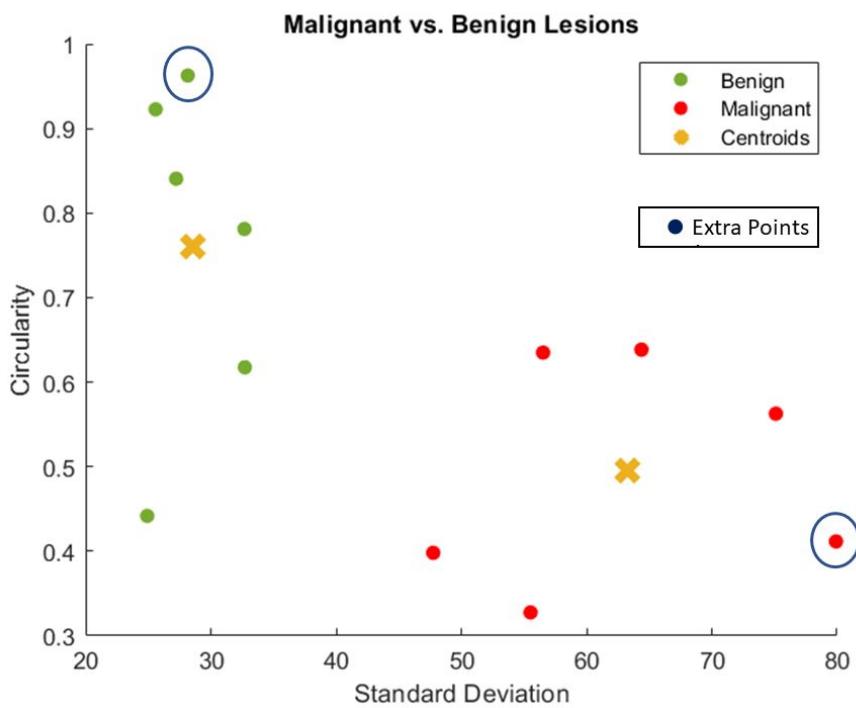


Figure 9: Plot using given dataset plus 2 selected images

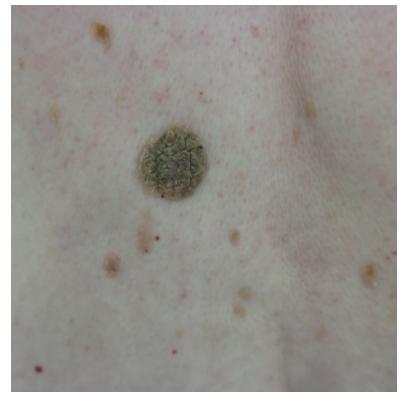


Figure 10: Benign-known

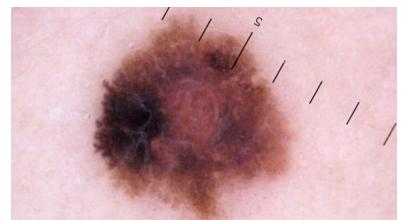


Figure 11: Malignant-known

What was observed is that the image of the benign lesion joined the cluster identified as benign and the image of the malignant lesion joined the cluster identified as malignant. This is further evidence of the correct identification of the clusters, which contributes to the level of confidence in their classification.

It is important to note that although our results are coherent, they are not necessarily exact. It is impossible to have absolutely sure the image lesion analysed by the code corresponds or not to a malignant lesion because the method that was implemented is solely based on the ABCDE rule. The ABCDE rule is a simple and useful tool for identifying potential melanomas, but it is not enough to definitively classify all melanomas. As we know, the rule stands for asymmetry, border irregularity, color variability, diameter larger than 6mm, and evolution over time [1.1](#), and while these are all important characteristics of melanomas, there are other features that can be present in melanomas or other skin lesions that are not captured by the ABCDE rule.

For example, a melanoma may not have all of the ABCDE features, or it may have additional features such as a shiny or scaly surface and a non-uniform distribution of pigment. In addition, some benign skin lesions may have some of the features of melanoma, such as irregular borders or multiple colors, making it difficult to rely solely on the ABCDE rule for classification. The depth of invasion, the mitotic rate, and the presence of ulceration are also factors to be into account. A definitive diagnosis of melanoma requires a biopsy and histopathologic examination by a dermatopathologist.

Therefore, only with additional tests such as a biopsy or dermoscopy one could determine whether the lesion is cancerous or not, alough the methods developed in this work can tell

with a high degree of confidence the lesions identified as malignant are indeed malignant.

5 Conclusions

Considering that our goals for this project were

1. to obtain two clusters that fit the data based on the features of standard deviation and circularity, and
2. correctly classify these clusters,
3. classify every image as "benign" or "malignant"

we can conclude that our objectives were successfully achieved.

To sum up, we have used MATLAB to analyze a set of skin lesion images to classify them as benign or malignant based on their standard deviation and circularity. We first plotted the data points on a scatter plot and identified two distinct clusters, which we further analyzed using the ABCDE rule. Our analysis indicates that benign lesions tend to have a lower standard deviation and higher circularity, while malignant ones tend to have opposite features.

This project demonstrates the potential of using MATLAB for image analysis and classification tasks. By leveraging the power of MATLAB's built-in functions, we were able to quickly and accurately analyze a large dataset of skin lesion images and make meaningful classifications.

It is important to note that our analysis was based solely on two features and further research could be done to incorporate other features that may increase the accuracy of the classification. Future work may also include applying more advanced machine learning techniques such as the ones referred in section 2 to improve the accuracy of classification.

References

- [1] [Accessed: 13-Apr-2023]. 2020. URL: ballaratskincancer.com.au/squamous-cell-carcinoma-scc/.
- [2] [Accessed: 11-Apr-2023]. 2022. URL: skinsight.com/skin-conditions/basal-cell-carcinoma-bcc/.
- [3] [Accessed: 11-Apr-2023]. 2023. URL: <https://www.skincancer.org/skin-cancer-information/>.
- [4] Qaisar Abbas and M. Emre Celebi. "DermoDeep-A classification of melanoma-nevus skin lesions using multi-feature fusion of visual features and deep neural network". In: *Multimedia Tools and Applications* 78.16 (2019), 23559–23580. DOI: <https://doi.org/10.1007/s11042-019-7652-y>.
- [5] Liga Portuguesa Contra o Cancro. *Melanoma*. [Accessed: 13-Apr-2023]. 2023. URL: <https://www.ligacontracancro.pt/melanoma/>.
- [6] American Academy of Dermatology Association. *Actinic Keratosis*. www.aad.org/public-diseases/scaly-skin/actinic-keratosis. [Accessed: 13-Apr-2023]. n.d.

- [7] C. dos. *Melanoma*. [Accessed: 13-Apr-2023]. 2006. URL: wikipedia.org/wiki/Melanoma.
- [8] Skin Cancer Foundation. *Melanoma*. www.skincancer.org/skin-cancer-information/melanoma/. [Accessed: 12-Apr-2023]. 2022.
- [9] Skin Cancer Foundation. *The ABCDEs of Melanoma*. www.skincancer.org/skin-cancer-information/melanoma/the-abcdes/. [Accessed: 13-Apr-2023]. n.d.
- [10] R;Van Ham. “Computer-assisted analysis of epiluminescence microscopy images of pigmented skin lesions”. In: *Cytometry* 37.4 (2018). URL: <https://pubmed.ncbi.nlm.nih.gov/10547610/>.
- [11] Curt Samlaska. *Seborrheic Keratosis: Atypical Forms... - Academic Dermatology of Nevada*. [Accessed: 12-Apr-2023]. 2019. URL: <https://acadderm.com/seborrheic-keratosis-atypical-forms/>.
- [12] Chanki Yu et al. “Acral melanoma detection using a convolutional neural network for dermoscopy images”. In: *PLOS ONE* 13.3 (2018). Ed. by Nikolas K. Haass, e0193321. DOI: <https://doi.org/10.1371/journal.pone.0193321>. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0193321>.

6 Appendix

6.1 Images used

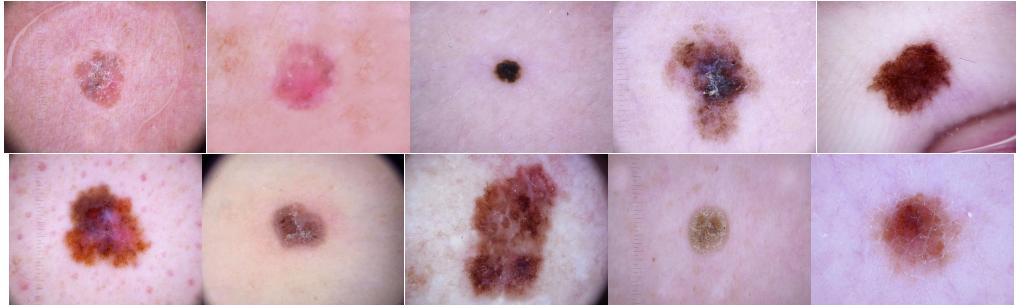


Figure 12: 10 dermoscopic images present in the *Data* folder

6.2 Matlab code

6.2.1 Dataset

```

1 % Define the folder path where the images are stored
2 folderPath = 'Data';
3
4 % Create an imageDatastore object to read all the .jpg images from the folder
5 imds = imageDatastore(folderPath, 'FileExtensions', '.jpg');
6
7 % Read all the images and convert them to grayscale and invert them
8 imds = readall(imds);
9 imds = cellfun(@(x) imcomplement(rgb2gray(x)), imds, 'UniformOutput', false);
10
11 % 'UniformOutput' is set to false to tell cellfun every output of the cell
12 % may not be uniform in size or shape (can have different data types)
13
14 % Display all inverted images in grayscale
15 figure;
16 for i = 1:numel(imds)
17     subplot(4, 3, i);
18     imshow(imds{i});
19 end
  
```

6.2.2 Denoising

```

20 % Define the standard deviation and filter size of the Gaussian filter
21 sigma = 1;
22 filterSize = 5;
23
24 % Apply a Gaussian filter to each image in the cell array
25 imds_filtered = cellfun(@(x) imgaussfilt(x, sigma, 'FilterSize', filterSize), imds, ,
26 % 'UniformOutput', false);
27
28 % Display all the denoised images in a loop
29 figure;
30 for i = 1:numel(imds_filtered)
31     subplot(4, 3, i);
32     imshow(imds{i});
  
```

6.2.3 Thresholding

```

33 % Apply Otsu's method to each image to obtain threshold values
34 thresholds = cellfun(@(x) graythresh(x), imds_filtered);
35
36 % Apply the obtained threshold values to each image to obtain binary masks
37 imds_binary = cellfun(@(x, t) imbinarize(x, t), imds_filtered, num2cell(thresholds), ,
38     'UniformOutput', false);
39
40 %imshow(imds_binary{1})
41 % Display all the denoised images
42 figure;
43 for i = 1:numel(imds_binary)
44     subplot(4, 3, i);
45     imshow(imds_binary{i});
46 end

```

6.2.4 Morphological Operations

```

46 % Apply morphological operations to each binary mask
47 % disk-shaped structuring element with radius 3
48 se = strel('disk', 3);
49 for i = 1:numel(imds_binary)
50     % Perform binary dilation
51     imds_binary{i} = imdilate(imds_binary{i}, se);
52
53     % Perform border clearing
54     imds_binary{i} = imclearborder(imds_binary{i});
55
56     % Perform small object removal (<100 pixels)
57     imds_binary{i} = bwareaopen(imds_binary{i}, 100);
58
59     % Perform hole-filling
60     imds_binary{i} = imfill(imds_binary{i}, 'holes');
61
62     % Keep only largest connected component
63     imds_binary{i} = bwareafilt(imds_binary{i}, 1);
64 end
65
66 % Show the cleaned and refined binary masks
67 figure;
68 for i = 1:numel(imds_binary)
69     subplot(4, 3, i);
70     imshow(imds_binary{i});
71 end

```

6.2.5 Feature Extraction

```

72 % Loop through each image
73 for i = 1:numel(imds)
74
75     % Multiply with binary mask
76     imds_masked = double(cell2mat(imds(i))) .* double(cell2mat(imds_binary{i}));
77
78     % Compute standard deviation of intensity distribution
79     std_intensity(i) = std(double(imds_masked(:)));
80
81     % Use regionprops to obtain circularity of mask
82     stats = regionprops('table', imds_binary{i}, 'Circularity');
83     circularity(i) = stats.Circularity;
84 end
85
86 % Concatenate the features into a matrix and display it
87 features = [std_intensity', circularity']

```

6.2.6 Data Visualization and Analysis

```
88 % Set the number of clusters (2 for benign/malignant)
89 k = 2;
90
91 % Perform k-means clustering on the features
92 [idx, centroids] = kmeans(features, k);
93
94 % Plot the results
95 figure;
96 scatter(features(idx==1,1), features(idx==1,2), 40, 'filled', 'DisplayName','Benign',...
97     'MarkerFaceColor','#77AC30');
98 hold on;
99 scatter(features(idx==2,1), features(idx==2,2), 40, 'filled', 'DisplayName','Malignant',...
100     'MarkerFaceColor','r');
101 scatter(centroids(:,1), centroids(:,2), 150, 'x', 'Linewidth', 4, 'DisplayName','Centroids');
102 legend();
103 title('Malignant vs. Benign Lesions')
104 xlabel('Standard Deviation');
105 ylabel('Circularity');
```