Transactions on Machine Intelligence 1(4) (2018) 220-227

ISSN Online: 2821-1693

Transactions on Machine Intelligence

Journal Homepage: https://tmachineintelligence.ir/index.php/TMI

Skin Cancer Detection from Dermoscopic Images with Emphasis on Shape, Color, and Texture Feature Extraction Using a Two-Stage Classification Algorithm

Gh. Shakourian [1,*], H. Montazery Kordy [2]

[1] MSc student, Department of Biomedical Engineering, Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran

[2] Assistant Professor, Department of Biomedical Engineering, Babol Noshirvani University of Technology, Babol, Iran

ARTICLE INFO

Article History:
Received 15 June 2018
Received in revised form 27 August 2018
Accepted 12 December 2018
Available online 20 December 2018

Keywords:
Skin Cancer, Melanoma, Feature Extraction, Shape, Color, Texture, Classification, Optimal Features, KNN

ABSTRACT

Malignant melanoma is one of the most aggressive and life-threatening forms of skin cancer, with a high potential for metastasis if not diagnosed and treated early. The definitive treatment for melanoma is possible when it is accurately detected by a trained specialist in a timely manner. Early detection can lead to a simple excision of the tumor, which can often result in complete cure. However, current diagnostic procedures typically involve a biopsy of the lesion, an invasive and often painful procedure that can cause discomfort to the patient. Given these challenges, this study aims to develop a more efficient, non-invasive method for the early detection of melanoma, leveraging advanced machine learning and image processing techniques. The proposed method utilizes a set of features based on the shape, color, and texture of dermoscopic images, which are extracted to capture critical characteristics of the lesion. These features are then analyzed through a two-stage classification process, designed to categorize the lesion into one of three categories: common nevus, atypical nevus, and melanoma. The method was tested on the PH2 dataset, a well-known dermatological dataset containing images of skin lesions. Results demonstrate that the two-stage classification model achieved an accuracy of approximately 90%, significantly outperforming traditional single-stage classification models for multi-class lesion classification. This innovative approach holds promise for enhancing the accuracy of melanoma detection and reducing the need for invasive procedures, ultimately improving patient outcomes.

1. INTRODUCTION

Skin cancer is one of the most common types of cancer worldwide. There are various types of skin cancer, among which melanoma is the most dangerous and life-threatening form. According to statistics, one person dies from melanoma every hour. In 2017, a total of 87,110 new cases of melanoma were identified in the United States,

* Corresponding Author: ghasemsh@stu.nit.ac.ir
MSc student, Department of Biomedical Engineering, Faculty of Electrical and Computer Engineering, Babol Noshirvani University of Technology, Babol, Iran

http://dx.doi.org/10.47176/TMI.2018.220

resulting in 9,730 deaths [1]. Accurate and early detection of melanoma significantly impacts its definitive treatment, as in such cases, a simple excision can effectively cure the disease.

Dermoscopy is one of the key tools for diagnosing melanoma and other pigmented skin lesions. It is a non-invasive imaging technique that reduces surface reflection and provides improved visualization of subsurface skin structures compared to conventional clinical images. This leads to a reduction in diagnostic errors and enhances the differentiation of various skin lesions [2]. Since melanoma diagnosis heavily relies on the experience and expertise of the physician, it has been established that the detection rate is lower when the examining physician has limited expertise [3]. Given these considerations, computer-aided approaches based on machine learning have gained attention to reduce diagnostic errors.

Most computer-aided dermoscopic image analysis approaches typically consist of four stages: preprocessing, segmentation, feature extraction, and classification [3]. Figure 1 illustrates an automated melanoma detection system. In this study, we specifically focus on the feature extraction and classification stages.

Pigmented skin lesions are typically assessed by dermatologists based on the "ABCD" rule, where each component represents asymmetry, border irregularity, color variation, and structural disparity within a lesion, respectively [3].

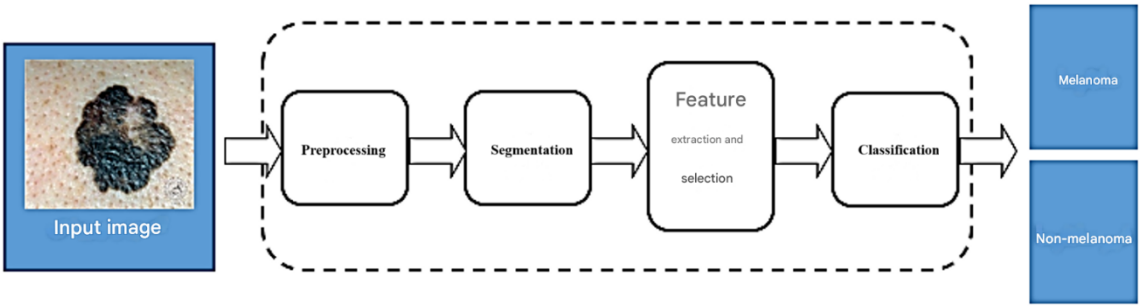

Fig. 1. Overview of an Automated Melanoma Detection System

Over the past decade, numerous studies have been conducted on melanoma detection using computer-aided image analysis. In this context, Mojdeh Rastgoo et al. introduced an automated framework that integrates and compares multiple texture features alongside shape and color features using multiple classifiers. Their proposed framework achieved a sensitivity of 98% and a specificity of 70% [4].

In [5], J. C. Kawisa et al. focused on extracting both global and local texture features, evaluating classification performance based on each type of feature. Their findings demonstrated that robust, high-speed features yielded better results, achieving an accuracy of 87.3%. Kawimasi, in his proposed system, compared various classification methods, including Support Vector Machines (SVM), ensemble classifiers, Probabilistic Neural Networks (PNN), and Adaptive Neuro-Fuzzy Inference Systems (ANFIS). He concluded that ANFIS exhibited superior performance [6].

Hanan Al-Asadi introduced an algorithm based on Artificial Neural Networks (ANN) that incorporated texture, color, and shape feature extraction. This approach achieved an accuracy of 98% for benign vs. malignant classification and 93% for distinguishing between different types of malignant melanoma [7].

2. DATABASE

In this study, we utilized images from the PH2 dataset to evaluate classification performance. This three-class dataset contains 200 lesion images, including 40 cases of melanoma, 40 cases of common nevi, and 40 cases of atypical nevi. Additionally, each image in the dataset is accompanied by a corresponding segmented image, which we used for lesion region separation and feature extraction.

3. PROPOSED METHOD

In this study, for data classification, we first utilized the segmented images from the database to isolate the lesion area from other regions. Subsequently, we extracted a set of shape, texture, and color features from the lesion images, which will be described in detail in the following sections. Finally, classification was performed using the K-Nearest Neighbors (KNN) classifier and a two-stage classification approach. The classification task was conducted for three classes: melanoma, common nevus, and atypical nevus.

4. FEATURE EXTRACTION

4.1. Shape Feature Extraction

The extracted shape features from the segmented image include lesion area, aspect ratio, eccentricity index, compactness coefficient, asymmetry, roughness coefficient, elongation coefficient, rectangularity index, equivalent diameter, convexity, and circularity index. Each of these features is described below.

The lesion area is defined as the number of pixels within the lesion region. The aspect ratio is calculated as the ratio of the major axis length to the minor axis length [8], where this index approaches zero for highly elongated objects.

The eccentricity index, also referred to as the ellipticity coefficient, is defined as the ratio of the minor axis length to the major axis length. A value closer to 1 indicates a shape that more closely resembles an ellipse [9].

$$Eccentricity = \frac{MinorAxisLength}{MajorAxisLength} \tag{1}$$

The compactness coefficient is computed as the ratio of the object's area to the area of a circle with the same perimeter. This index essentially compares the object to a circle, which has the highest compactness. It is calculated using the following formula:

$$Compactness = \frac{4\pi \times Area}{(Perimeter)^2} \tag{2}$$

The asymmetry index (AI) is calculated as follows:

$$AI = \frac{\Delta A}{A} \times 100 \tag{3}$$

where:

- A represents the total area of the image,
- ΔA is the difference between the total image area and the lesion area [10].

The irregularity coefficient, which indicates border irregularity, is obtained from the ratio of the lesion area to its convex hull area [8]. A lower value of this coefficient signifies a more irregular border.

$$Solidity = \frac{Area}{ConvexArea} \tag{4}$$

The elongation coefficient is defined as the ratio of the width to the length of the smallest bounding rectangle around the shape. The more elongated the object, the lower this coefficient. Notably, a square has an elongation coefficient of one.

The rectangularity index is calculated as the ratio of the lesion area to the area of its smallest bounding rectangle. This coefficient is typically lower for malignant lesions compared to benign ones [9].

$$Rectangularity = \frac{Area}{BoundingrectangleArea} \quad (5)$$

The **equivalent diameter** is defined as the diameter of a circle with an area equal to that of the lesion [8].

The **convexity coefficient** represents the extent to which an object deviates from convexity and is calculated as the ratio of the convex perimeter to the actual perimeter of the shape.

Finally, the **circularity index** quantifies how similar an object is to a perfect circle. It is obtained by dividing the shape's area by the area of its equivalent circle [9].

$$Roundness = \frac{4\pi \times Area}{(Convex\ perimeter)^2} \quad (6)$$

4.2. Colour Feature Extraction

The extracted color features from the lesion image include the mean, standard deviation, and variance for each channel of the O1O2O3, RGB, HSV, and Lab color spaces, as well as the variance of the I channel in the HSI color space.

4.3. Texture Feature Extraction

To extract texture features, it is essential to compute them within the lesion's texture region. Therefore, we first determine the largest inscribed rectangle within the lesion and then extract the desired texture features for its characterization. Figure 2 illustrates the largest inscribed rectangle within the lesion.

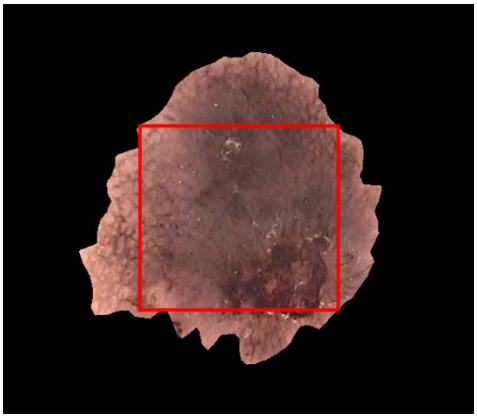

Fig. 2. The Largest Inscribed Rectangle within the Lesion Area

From the grayscale image, statistical features such as mean, standard deviation, smoothness, third-order moment, uniformity, and entropy were extracted [11]. The mathematical formulations for each of these descriptors are as follows:

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \quad (7)$$

$$\sigma = \sqrt{\mu_2(2)} = \sqrt{\sigma^2} \quad (8)$$

$$R = 1 - \frac{1}{(1+\sigma^2)} \quad (9)$$

$$\mu_3 = \sum_{i=0}^{L-1}(z_i - m)^2 p(z_i) \quad (10)$$

$$U = \sum_{i=0}^{L-1} p^2(z_i) \quad (11)$$

$$e = -\sum_{i=0}^{L-1} p(z_i) Log_2 p(z_i) \quad (12)$$

The variable z_i represents the random variable corresponding to the grayscale intensity level, *p(z)* is the histogram of the image in the given region, L is the number of possible intensity levels, and μ_n is the nth-order moment.

From the Gray-Level Co-occurrence Matrix (GLCM), features such as contrast, homogeneity, correlation, energy, mean, and variance were computed.

Next, a 2D wavelet transform with the "Haar" wavelet was applied to the lesion image, and statistical features were extracted from its approximation coefficient. Additionally, features were derived from the GLCM generated by the horizontal detail coefficient (H) of the wavelet transform.

The same process was repeated for the derivative and Fourier transform of the lesion image, resulting in the extraction of 60 texture features from each image. In total, 108 features, including color, shape, and texture, were extracted from each lesion image, which were subsequently used for the classification task.

5. CLASSIFICATION

Various methods are available for classifying data. Since our data is divided into three classes melanoma, common mole, and atypical mole we used a two-stage classification approach. In this method, two KNN classifiers with different parameters were used. In the first stage, Euclidean distance with 3 neighbors was employed, and in the second stage, the Chebyshev distance with 12 neighbors was utilized. In the first classifier, we considered the common and atypical moles as one class and melanoma as the other class. In the second stage, classification was performed for the two classes: common mole and atypical mole.

Using all the extracted features for classification led to low accuracy. Therefore, we first used the KNN classifier for each feature to determine the classification accuracy, retaining the features that performed well and eliminating the others. This process reduced the feature vector dimension to 23 optimized features, leading to an increase in classification accuracy. The classification system algorithm used is shown in Figure 3.

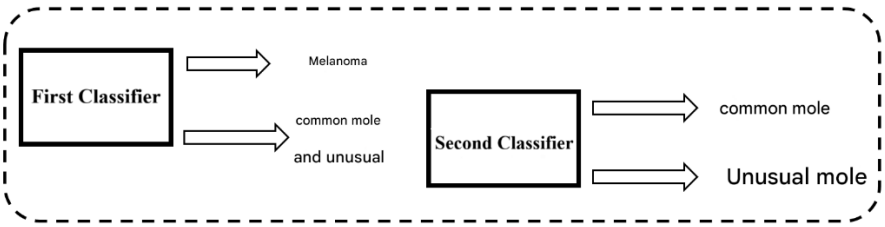

Fig. 3. Proposed framework for the classification of dermoscopic images.

6. RESULTS

To evaluate the performance of the proposed system, we selected 90 images from the PH2 database, with 30 images corresponding to each class. Features mentioned in Section 4 were extracted from the images, and classification was performed using the introduced method.

The first classifier, which distinguished between the classes melanoma and common/atypical mole, achieved an accuracy of 93.3%, sensitivity of 96.6%, and specificity of 86.6%. The second classifier, for categorizing common and atypical moles, achieved an accuracy of 93.33%, sensitivity of 100%, and specificity of 86.66%. Finally, after analyzing the results from both classifiers for the three-class problem, we achieved an overall accuracy of 88.8%.

Table 1 shows the results for each classifier, including sensitivity, specificity, and accuracy metrics. Table 2 presents the confusion matrix for the two-stage classification (first classifier), while Table 3 shows the confusion matrix for the second stage of classification. Table 4 provides the overall confusion matrix for the three classes.

Additionally, the classification was also performed using a single-stage classifier, which achieved an accuracy of 77.8%. The results from this single-stage classification are shown in Table 5.

Table 1. Results of Classification Based on the Presented Method

Classifier	Accuracy (%)	Sensitivity (%)	Certainty (Specificity) (%)
Melanoma – Common + Unusual Mole	93.33	96.67	86.67
Common Mole – Unusual Mole	93.33	100	86.6

Table 2. Scattering Matrix for the First Classifier

Desired Class	Predicted: Common & Uncommon Moles (%)	Predicted: Melanoma (%)
Common & Uncommon Moles	96.7	3.3
Melanoma	13.33	86.7

Table 3. Scattering Matrix for the Second Classifier

Desired Class	Predicted: Common Mole (%)	Predicted: Unusual Mole (%)
Common Mole	100	0
Unusual Mole	13.3	86.7

Table 4. Scattering Matrix for Three-Class Classification

Desired Class	Predicted: Common Mole (%)	Predicted: Unusual Mole (%)	Predicted: Melanoma (%)
Common Mole	100	0	0
Unusual Mole	13.33	80	66.6
Melanoma	6.66	6.66	86.7

Table 5. Scattering Matrix for One-Stage Classification

Desired Class	Predicted: Common Mole (%)	Predicted: Unusual Mole (%)	Predicted: Melanoma (%)
Common Mole	66.7	26.66	6.66
Unusual Mole	0	86.7	13.3
Melanoma	0	20	80

7. CONCLUSION

In this paper, we presented a method for melanoma detection based on the extraction of shape, color, and texture features, and a two-stage classification approach. First, we separated the lesion area from the rest of the image using the segmented image from the PH2 database. Then, we extracted a set of shape, color, and texture features from the

lesion image. In the next step, we employed a two-stage classification method using the KNN classifier to classify the data. Additionally, to reduce the dimensionality of the feature vector, we evaluated the classification accuracy for each feature and selected those with higher accuracy, while eliminating the rest. This process not only reduced the feature vector's dimensionality but also improved the classification accuracy.

We achieved a classification accuracy of 88.9% for the three-class data using the proposed method. In contrast, a single-stage method resulted in an accuracy of 77.8%. This indicates that the proposed method outperforms the one-stage classification approach.

Transparency Statement

The data supporting this study are available upon reasonable request to the corresponding author, subject to ethical and confidentiality considerations.

Acknowledgments

We would like to express our gratitude to all individuals who contributed to this project.

Declaration of Interest

The authors declare that they have no competing interests.

REFERENCES

[1] American Cancer Society. (2017). Cancer facts and figures 2017. Retrieved from http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc048738.pdf. Accessed January 10, 2017.

[2] Scheafer, G., Krawczyk, B., Celebi, M. E., & Iyatomi, H. (2014). An ensemble classification approach for melanoma diagnosis. Memetic Computing, 6(4), 233–240. https://doi.org/10.1007/s12293-014-0144-8

[3] Xie, F., et al. (2017). Melanoma classification on dermoscopy images using a neural network ensemble model. IEEE Transactions on Medical Imaging, 36(3), 849–858. https://doi.org/10.1109/TMI.2016.2633551

[4] Rastgoo, M., Garcia, R., Morel, O., & Marzani, F. (2015). Automatic differentiation of melanoma from dysplastic nevi. Computerized Medical Imaging and Graphics, 43, 44–52. https://doi.org/10.1016/j.compmedimag.2015.02.011

[5] Kavitha, J. C., Suruliandi, A., & Nagarajan, D. (2017). Melanoma detection in dermoscopic images using global and local feature extraction. International Journal of Multimedia and Ubiquitous Engineering, 12(5), 19–28. https://doi.org/10.14257/ijmue.2017.12.5.02

[6] Kavimathi, P. (2016). Comparative analyses of classifiers for diagnosis of skin cancer using dermoscopic images. Indian Journal of Science and Technology, 9(43). https://doi.org/10.17485/ijst/2016/v9i43/103824

[7] AlAsadi, A. H. H., & AL-Safy, B. M. R. (2016). Early detection and classification of melanoma skin cancer. LAP LAMBERT Academic Publishing.

[8] Celebi, M. E., Kingravi, H., Uddin, B., Iyatomi, H., Aslandogan, A., Stoecker, W. V., & Moss, R. H. (2007). A methodological approach to the classification of dermoscopy images. Computers in Medical Imaging and

Graphics, 31(6), 362–373. https://doi.org/10.1016/j.compmedimag.2007.01.003

[9] Fatih, M., Beigi, M. M., Fatih, M., & Mansouri, P. (2009). Melanoma detection using extraction of appropriate features from dermatoscopic images. Research Paper, Medical Lasers, 6(1).

[10] Bhuiyan, M. A. H., Azad, I., & Uddin, M. K. (2013). Image processing for skin cancer features extraction. International Journal of Scientific & Engineering Research, 4(2), 1–6.

[11] Gonzalez, R. C., & Woods, R. E. (2008). Digital image processing (2nd ed.). Prentice-Hall.