Skin Lesion Classification using Machine Learning Algorithms

Ilker Ali OZKAN *[1], Murat KOKLU [1]

Abstract: Melanoma is a deadly skin cancer that breaks out in the skin's pigment cells on the skin surface. Melanoma causes 75% of the skin cancer-related deaths. This disease can be diagnosed by a dermatology specialist through the interpretation of the dermoscopy images in accordance with ABCD rule. Even if dermatology experts use dermatological images for diagnosis, the rate of the correct diagnosis of experts is estimated to be 75-84%. The purpose of this study is to pre-classify the skin lesions in three groups as normal, abnormal and melanoma by machine learning methods and to develop a decision support system that should make the decision easier for a doctor. The objective of this study is skin lesions based on dermoscopic images PH^2 datasets using 4 different machine learning methods namely; ANN, SVM, KNN and Decision Tree. Correctly classified instances were found as 92.50%, 89.50%, 82.00% and 90.00% for ANN, SVM, KNN and DT respectively. The findings show that the system developed in this study has the feature of a medical decision support system which can help dermatologists in diagnosing of the skin lesions.

Keywords: *Skin lesion, classification, Medical Decision Support System, Machine Learning Methods*

1. Introduction

Cancer is a disease that occurs through multiplying of the body cells in an uncontrolled manner and occupying the peripheral tissues. Although the skin cancer disease occurs less frequent than many other cancer types, it is highly important because of its high mortality. The skin cancer has different of types, such as Malignant Melanoma, Squamous cell carcinoma, and Basal cell carcinoma [1].

Melanoma incidence is being reported to increase more rapidly than the other forms of cancer. Melanoma is responsible for 4% only of all skin cancers, whereas it is responsible for 75% total of skin cancer deaths [2].

Melanoma, which is thought to be stimulated by ultraviolet rays, is more commonly occurring in areas where exposure to sunlight is relative higher. In Europe are detecting 62.000 new cases each year [3]. According to the American Cancer Society's report 2016 was for the year 2016 foreseen that 76.380 cases will be diagnosed with melanoma in the United States and 10.130 people will die from it [4].

The natural development of melanoma takes place in two stages except for the nodular type. The horizontal or radial development stage that progresses along the epidermal surface, is defined as "single cancer melanoma" which has a critical importance for the early diagnosis. The melanoma progresses to the vertical development stage, if it couldn't be diagnosed in this period. It obtains at this stage a potential for metastatic spread [5]. The melanoma is a disease with a survival rate of approximately 90% on condition that it must be diagnosed early enough, whereas no effective treatment is available for the delayed cases [2, 6].

Dermoscopy is a non-invasive diagnostic method that allows us to examine in more detail the morphological structure of the pigmented skin lesions. The melanoma diagnosis is performed by interpreting the images obtained with the dermoscopy device. The dermoscopy device allows through the harsh light the detailed visualization of the morphological structures and patterns.

Dermatologists usually perform the diagnosis of melanoma through these images by ABCD (Asymmetrical Shape, Border, Color and Diameter) rule. ABCD is a highly subjective assessment that is dependent on the experience and knowledge of the related doctors [4].

Despite the use of dermoscopy to examine the subcutaneous tissue without surgery, successful results are dependent on intensive dermatology training and experience. A reliable diagnosis of the melanoma with this method is unfortunately often not possible, especially in early stages. For that reason, an automatic diagnostic tool becomes an inevitable need [7].

The melanoma diagnosis can be improved with the ABCD rule based and computer assisted systems. These systems usually consist of the separate units for the image segmentation, feature extraction and classification respectively [8-12]. Studies conducted in this field are as follows:

Baldrick et al. compared in their study the expert opinion and artificial neural networks when they classify the lesions. They obtained from the computer program a sensitivity of 95% and a specificity of 88%, while they measured the expert dermatological sensitivity and specificity as 95% and 90% respectively [13].

Moataz et al. practised upon a genetic algorithm with an artificial neural network technique for early detection of the skin cancers and obtained a sensitivity of 91.67% and a specificity of 91.43%. [14].

Kamasak et al. classified dermoscopic images by extracting the Fourier identifiers of the lesion edges after dividing the dermoscopic images. They obtained an accuracy of 83.33% in diagnosing of the melanoma [15].

Fidan et al. succeeded in an exact classification of 93.33% according to data extracted from the PH^2 data set by using an artificial neural network that was formed for the abnormal and melanoma skin cancers [17]. Baştürk et al. used a new method of detecting melanoma skin cancers called Deep Neural Network (DNN) in their study and reached an accuracy of 91.85% in disease diagnosis [18].

[1] Computer Eng., Faculty of Technology, Selcuk Uni., Konya –TURKEY
** Corresponding Author: Email: ilkerozkan@selcuk.edu.tr*

In the second part of this study were information given about data set, machine learning methods and performance measurement methods. In the third part, the classification studies and the results obtained from these studies were given by comparing with the studies in the literature.

2. Material and Methods

2.1. PH² Dataset

A diagnostic study were performed with the machine learning algorithms formed for melanoma diagnosis in the PH² data set. This data set was established by a group of researchers from the Technical Universities of Porto and Lisbon in the dermatology service to Pedro Hispano Hospital. The PH² dataset contains 200 dermoscopy images at 768x560 resolution. Each image has 8-bit RGB channels [16].
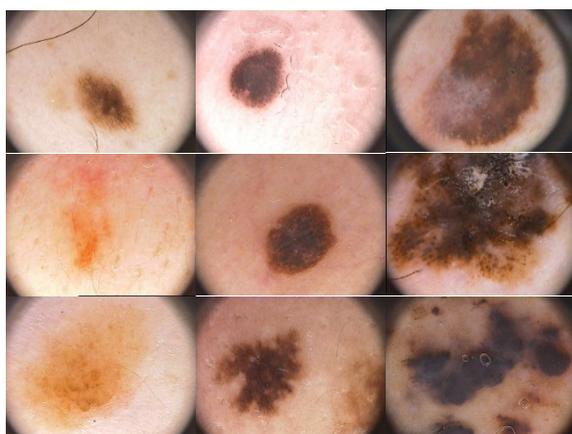
(a) Normal   (b) Abnormal   c) Melanom

Fig. 1. Views of the sample lesions in the PH² data set

In the PH² data set are available 80 images for the normal type, 80 images for the abnormal and 40 images for the melanoma respectively. Some examples of these are shown in Figure 1. Although the PH² data set was established by extracting the features according to the ABCD rule criteria, the criterion B was ignored hereby. For that reason, the features found in the dataset and used in the study are given in Table 1.

Table 1. Features of the PH² data set

ABCD Criterion	Criterion Name	Criterion Value
A	Asymmetry	0-Even-symmetric 1- Axially symmetric 2-Asymmetric
B	Edge	Not available in the PH² data set
C	Colours (white, red, light brown, dark brown, blue-gray, black)	1-Applicable 0-Not applicable
D	The vascular structure, associated regions, the blue-whitish area,	1- Applicable 0- Not applicable
	Pigment Network	0-Normal 1-Abnormal
	Dotted-beaded structure	0-Normal 1-Abnormal 2-No observed
ABCD	Diagnosis	0-Normal Nevus 1-Abnormal Nevus 2-Melanom

2.2. Machine Learning Methods

With this study, four different classifying techniques based on dermoscopic images of the data sets were applied on the skin lesions. Short information about each of the classifying techniques, i.e. ANN, SVM, KNN and DT are given in the following paragraphs.

Artificial Neural Network (ANN): ANNs are mathematical systems consisting of many process units (neurons) connected with each other in a weighted manner. The process unit receives signals from other neurons; combines, transforms them and generates a numerical result. In general, the process units are corresponding roughly to the real neurons and are interconnected in a network, so that this structure constitutes the artificial neural networks [19].

SVM: SVMs are nonparametric classifiers. Regarding their distribution is no preliminary information as a presupposition available. Inputs and outputs are paired in the training sets. Through the pairs, decision functions are obtained which classify the input variables in the test set and new data set. The task is here to be able to find out the line with the highest margin from the infinite number of lines that can classify the data, when a linearly separation were possible. It uses a non-linear mapping for transforming the original work data into a higher dimension, when a linearly separation were impossible. In the new transformed dimension is being investigated then the (optimal) separator plane with the maximum margin [20].

K-Nearest Neighbor (KNN): The KNN (K-Nearest Neighbor) algorithm is one of the most basic sample-based learning algorithms. In example based learning algorithms, the learning process is performed based on the data held in the training set. A new faced example is categorizing according to similarities with the examples in the available training set [21].

Decision tree (DT): The decision tree is a classifier algorithm in the structure form of a "tree". Decision Trees are simple, but very commonly used methods by moving the inductive logic into a programming environment. It works with discrete valued parameters. The basic intuition about the inductive philosophy on which the decision tree algorithms are based is that a "good" decision tree to be constructed with learning characteristics should be small as possible [22].

2.3. The Commonly-Accepted Performance Evaluation Measures

This is the case we focus on in this study. Classification performance without focusing on a class is the most general way of comparing algorithms. It does not favor any particular application. The introduction of a new learning problem inevitably concentrates on its domain, but omits a detailed analysis. Thus, the most used empirical measure and accuracy does not distinguish between the numbers of correct labels of the different classes. [23]:

TP = true positives: number of examples predicted positive that are actually positive

FP = false positives: number of examples predicted positive that are actually negative

TN = true negatives: number of examples predicted negative that are actually negative

FN = false negatives: number of examples predicted negative that are actually positive

Accuracy: It refers to the total number of records that are correctly classified by the classifier. Accuracy of a classifier is defined as the percentage of test set tuples that are correctly classified by the model [24].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} x100\% \qquad (1)$$

Sensitivity: Refers to the true positive rate that means the proportion of positive tuples that were correctly identified [24].

$$Sensivity = \frac{TP}{TP+FN} x100\% \tag{2}$$

Specificity: Indicates the rate at which a test or diagnostic method sets a correct (ie negative) diagnosis for a patient who is not ill.

$$Specifity = \frac{TN}{TN+FP} x100\% \tag{3}$$

Balanced accuracy (BACC) : The balanced accuracy, which can be defined as the average accuracy obtained on either class [25].

$$BACC = \frac{\frac{TP}{P}+\frac{TN}{N}}{2} x100\% \tag{4}$$

Precision: The fraction of retrieved instances that are relevant [24].

$$Precision = \frac{TP}{TP+FP} x100\% \tag{5}$$

F- measure: The F- measure also refers to F measures that combined both the measures Precision and Recall as the harmonic mean [24].

$$F-measure = \frac{2*precision*sensivity}{precision+sensivity} x100\% \tag{6}$$

ROC Curve: Receiver Operating Characteristics curved showed both sensitivity and specificity of the test. The comparison of TPR (True Positive Rate) and FPR (False Positive Rate) is defined as ROC curve. The TPR is the proportion of positive tuples that are correctly labeled by the model whereas FPR is of negative tuples misclassified as positive [24].
i.e. TPR = TP (TP+FN) and FPR = FP (FP+TN)

3. Results and Conclusions

In this study for a melanoma diagnosis; Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Decision Tree (DT) classifiers were compared with each other within a PH^2 data set. Categorical values are coded by "one-of-N coding" for entries performed in this study.

The experimental studies have shown that the optimum value for "k" amounted to 5 and 10 in the k-fold cross-validation method. [19, 26]. As shown in Fig. 2, in this study is the data set divided into 10 parts by using a 10-layer cross-validation method. The system is trained and tested with "k" different training and test clusters whereas for each case the "k" performance measures could be obtained. Thus, the arithmetic mean of the obtained "k" performance measures is calculated to determine the success of the cross validation [27].

This study has been performed using the functions of MATLAB Statistics and Machine Learning Toolbox and MATLAB Neural Network Toolbox [28, 29]. The ANN structure formed consists of three layers. All the input parameters (12 units total) in the data set are establishing the input vector at Layer 1. Layer 3 is the layer that indicates the output of the classification and the number of neurons in this layer depended on the number of classes in the output. Scaled conjugate gradient backpropagation algorithm is used as the learning algorithm [30]. Back propagation training parameters are given in Table 2.
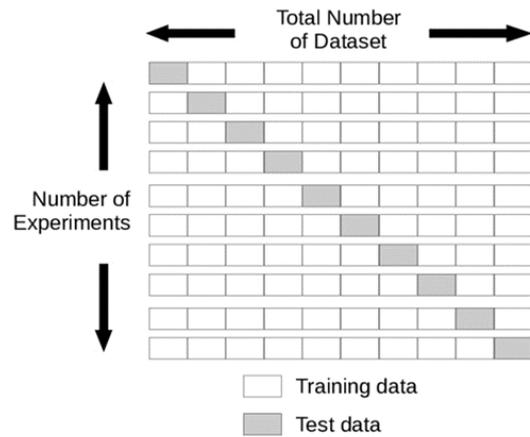

Fig. 2. 10-Fold cross validation method

Table 2. Backpropagation training parameters

Parameters	Value
Learning Rate	0.01
Momentum Constant	0.9
Epochs	1000-10000
Performance Function	Crossentropy
Minimum Performance Gradient	$1/(e^{-10})$

Network structures containing from 2 to 50 neurons were trained on the given network structure to find out the number of the hidden layer neurons with the best result. The established network structures in the study were examined and the ANN architecture with 18 neurons was used in the hidden layer with the best accuracy (92.50%).

Table 3 shows the performance results for classifying PH^2 data set obtained with Artificial Neural Network (ANN), Support Vector Machine (SVM), K-nearest neighbors (KNN) and decision tree (DT) classifiers by using the 10-fold cross-validation.

Table 3. Performance rates of the application

Algorithm	Accuracy (%)	Balance Accuracy (%)	Sensivity (%)	Specifity (%)	Precision (%)	F1-Score (%)
ANN	92.50	93.49	90.86	96.11	92.38	90.45
SVM	89.50	90,35	86.25	94.44	89.09	87.31
KNN	82.00	85,04	79.58	90.49	81.45	80.33
DT	90.00	90,97	87.08	94.86	88.58	87.70

Table 3 shows that ANN has an accuracy of 92.50%, SVM of 89.50%, KNN of 82.00% and DT of 90.00%. This suggests that the proposed ANN has a clearly better classification performance for the PH^2 data set. The accuracy values of the ANN, SVM, KN and DT algorithms for each classifier output are given in Fig. 3 for the classification of skin lesions according to the data from the PH^2 data set. The ANN classifier appears to be more successful in classifying each skin lesion than other algorithms.

Table 4. Sensitivity, specificity, accuracy and balanced accuracy of the proposed method and other comparison methods

| Methods (%)	Sensivity	Specifity	Accuracy	Balance Accuracy	References
ANN	90.86	**96.11**	**92.50**	**93.49**	**This study**
Joint Reverse Classification (JRC)	87.50	93.13	92.00	90.31	Bi et al [33]
Deep Learning	-	-	91.85	-	Basturk et al [18]
SVM	**94.10**	77.40	79.10	85.75	Marques et al [34]
SVM	84.00	94.00		89.00	Riaz et al [35]
BoW model with Harris Laplace features	85.00	87.00	-	86	Barata et al [36]
BoW model with Shades of Gray transform	92.50	76.30	-	84.4	Barata et al [37]
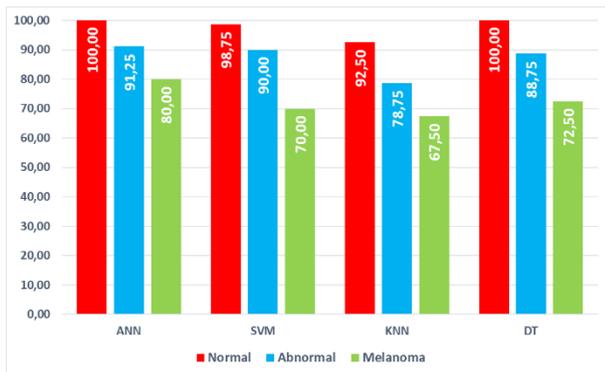
Fig. 3. Accuracy values for each classifier output according to algorithms

In Fig. 4 is given the ROC curve for ANN with the highest classification performance according to the classification results obtained by the algorithms.
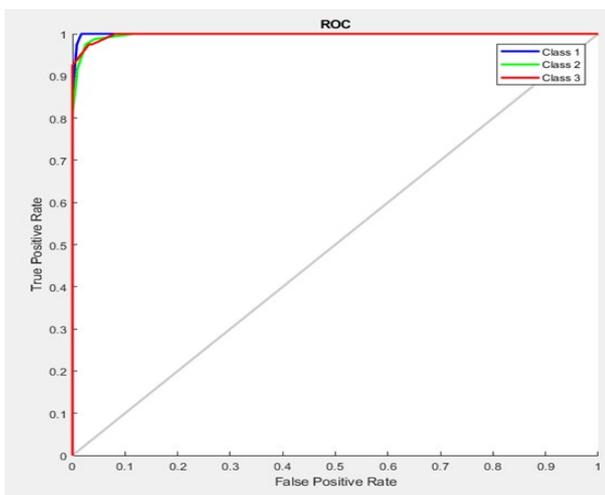

Fig. 4. Receiver Operating Characteristic (ROC) curves for ANN

According to a study conducted by Jain et al., even though expert dermatologists use dermatology images for diagnosis, the rate of correct diagnosis of experts is estimated at 75-84% [31]. In this study, which was performed with the different classification algorithms to classify the skin lesions, the normal skin lesions were by ANN and DT classifiers 100% correctly classified according to the data obtained from the PH^2 data set. All of the classifier algorithms used are revealed in terms of the classification outputs to be "better" than others in the normal type classification and to be "worse" than the others in the melanoma type classification. When the obtained data should be evaluated in terms of the output accuracy ratios and accuracy level of each class, will be observed that ANN has more successful classified the PH^2 data set than SVM, KNN and DT. An accuracy of 92.50% achieved with the ANN classifier reveals that this classifier is a medical decision support system which could help dermatologists to diagnose the skin lesions.

Previous studies including upper mentioned on the related data set are summarized with the accuracy ratios in Table 4.

Compared with the literature studies given in Table 4, the ANN structure, obtained in this study for classification of the skin lesions, has only in the sensivity of the skin lesions has a lower value than that of Barata et al. and Marques et al., whereas our study has a higher specifity value. Additionally this study has distinguished with its higher accuracy, specifity and balanced accuracy ratio compared to all other studies.

This study maybe further progressed by using the different preliminary data processing techniques and hybrid classification algorithms. In addition, this study can be combined with the related image processing techniques also to be able to make autonomous decisions in several medical issues.

References

[1] A. M. Y. Palomo, M. d. J. D. Pérez, O. R. P. Pérez, V. d. J. A. Yabor, and A. M. Fontaine, "Melanoma maligno cutáneo en pacientes de la provincia de Las Tunas," *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*, vol. 40, 2015.

[2] R. o. T. M. o. Health, "Türkiye Melanom Yol Haritası," 2012.

[3] T. I. C. A. Board, "The melanoma white paper: Reshaping EU healthcare for melanoma patients.," 2012.

[4] A. C. Society, "Cancer Facts & Figures," 2016.

[5] M. A. Sedat Özçelik, "Epidemiology of Melanoma," *TURKDERM*, vol. 41, pp. 1-5, 2007.

[6] B. Ozturk, E. Yaman, K. Ali Osman, R. Yildiz, U. Demirci, U. Coşkun, "Kutanöz malign melanomda adjuvan medikal tedavi yaklaşımları," *Türk Onkoloji Dergisi*, vol. 25, pp. 170-80, 2010.

[7] H. Yalcin, "Çeşitli Özniteliklerle Kötü Huylu Melanom Karakterizasyonu Characterization of Melanomas Using a Variety of Features," *Vogue*, vol. 15, 2015.

[8] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Archives of dermatology*, vol. 134, pp. 1563-1570, 1998.

[9] R. H. Johr, "Dermoscopy: alternative melanocytic algorithms—the ABCD rule of dermatoscopy, menzies scoring method, and 7-point checklist," *Clinics in dermatology*, vol. 20, pp. 240-247, 2002.

[10] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images," in *International Workshop on Machine Learning in Medical Imaging*, 2015, pp. 118-126.

[11] Z. Ma and J. M. R. Tavares, "A novel approach to segment skin lesions in dermoscopic images based on a deformable model," *IEEE journal of biomedical and health informatics*, vol. 20, pp. 615-623, 2016.

[12] J. Premaladha and K. Ravichandran, "Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms," *Journal of medical systems*, vol. 40, p. 96, 2016.

[13] J. C. Boldrick, C. J. Layton, J. Nguyen, and S. M. Swetter, "Evaluation of digital dermoscopy in a pigmented lesion clinic: Clinician versus computer assessment of malignancy risk," *Journal of the American Academy of Dermatology*, vol. 56, pp. 417-421, 2007/03/01/ 2007.

[14] M. Aboras, H. Amasha, and I. Ibraheem, "Early detection of melanoma using multispectral imaging and artificial intelligence techniques," *American Journal of Biomedical and Life Sciences*, vol. 3, pp. 29-33, 2015.

[15] E. Albay and M. Kamaşak, "Skin lesion classification using fourier descriptors of lesion borders," *in Medical Technologies National Conference (TIPTEKNO)*, 2015, 2015, pp. 1-4.

[16] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "PH 2-A dermoscopic image database for research and benchmarking," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013, pp. 5437-5440.

[17] U. Fidan, İ. Sarı and R. K. Kumrular, "Classification of skin lesions using ANN," in *Medical Technologies National Congress (TIPTEKNO) 2016*, 2016, pp. 1-4.

[18] A. Baştürk, M. E. Yüksel, H. Badem and A. Çalışkan, "Deep neural network based diagnosis system for melanoma skin cancer," in *Signal Processing and Communications Applications Conference (SIU)*, 2017 25th, 2017, pp. 1-4.

[19] R. J. Schalkoff, *Artificial neural networks* vol. 1: McGraw-Hill New York, 1997.

[20] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," 2003.

[21] T. Mitchell, "Machine Learning, McGraw-Hill Higher Education," *New York*, 1997.

[22] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, pp. 660-674, 1991.

[23] W. Dai and W. Ji, "A mapreduce implementation of C4. 5 decision tree algorithm," *International journal of database theory and application*, vol. 7, pp. 49-60, 2014.

[24] V. N. Vapnik and V. Vapnik, *Statistical learning theory* vol. 1: Wiley New York, 1998.

[25] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *Pattern recognition (ICPR), 2010 20th international conference on*, 2010, pp. 3121-3124.

[26] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and Regression Trees. Wadsworth & Brooks/Cole Advanced Books & Software," *Pacific California*, 1984.

[27] D. Thukaram, H. Khincha, and H. Vijaynarasimha, "Artificial neural network and support vector machine approach for locating faults in radial distribution systems," *IEEE Transactions on Power Delivery*, vol. 20, pp. 710-721, 2005.

[28] I. The MathWorks, "MATLAB and Neural Network Toolbox Release ", ed: Natick, Massachusetts, United States., 2016.

[29] I. The MathWorks, "MATLAB Statistics and Machine Learning Toolbox," ed: Natick, Massachusetts, United States., 2016.

[30] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural networks*, vol. 6, pp. 525-533, 1993.

[31] S. Jain and N. Pise, "Computer aided melanoma skin cancer detection using image processing," *Procedia Computer Science*, vol. 48, pp. 735-740, 2015.

[32] S. Turkeli, M. S. Oguz, S. B. Abay, T. Kumbasar, H. T. Atay, and K. K. Kurt, "A smart dermoscope design using artificial neural network," in *Artificial Intelligence and Data Processing Symposium (IDAP)*, 2017 International, 2017, pp. 1-6.

[33] L. Bi, J. Kim, E. Ahn, D. Feng, and M. Fulham, "Automatic melanoma detection via multi-scale lesion-biased representation and joint reverse classification," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium* on, 2016, pp. 1055-1058.

[34] J. S. Marques, C. Barata, and T. Mendonça, "On the role of texture and color in the classification of dermoscopy images," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 4402-4405.

[35] F. Riaz, A. Hassan, M. Y. Javed, and M. T. Coimbra, "Detecting melanoma in dermoscopy images using scale adaptive local binary patterns," in Engineering in Medicine and Biology Society (EMBC), *2014 36th Annual International Conference of the IEEE*, 2014, pp. 6758-6761.

[36] C. Barata, J. S. Marques, and J. Rozeira, "Evaluation of color based keypoints and features for the classification of melanomas using the bag-of-features model," in *International Symposium on Visual Computing*, 2013, pp. 40-49.

[37] C. Barata, M. E. Celebi, and J. S. Marques, "Improving dermoscopy image classification using color constancy," *IEEE journal of biomedical and health informatics*, vol. 19, pp. 1146-1152, 2015.