A COMPREHENSIVE SURVEY ON SKIN CANCER DETECTION

Jerry Johnson[1],Dr. R. Satheeskumar[2], Dr. Rajeswari[3],J.Krishnadas[4]
[1]PG Scholar, Sahrdaya College of Engineerig and Technology, Kerala ,India
[234]Associate professor of CSE, Sahrdaya College of Engineering and Technology, Kerala,India

Abstract— Human Cancer is one of the most dangerous disease which is mainly caused by genetic instability of multiple molecular alterations. Among many forms of human cancer, skin cancer is the most common one. To identify skin cancer at an early stage we will study and analyze them through various techniques named as segmentation and feature extraction. Here, we focus malignant melanoma skin cancer, (due to the high concentration of Melanoma- Hier we offer our skin, in the dermis layer of the skin) detection. In this, We used our ABCD rule dermoscopy technology for malignant melanoma skin cancer detection. In this system different step for melanoma skin lesion characterization i.e, first the Image Acquisition Technique, pre-processing, segmentation, define feature for skin Feature Selection determines lesion characterization, classification methods. In the Feature extraction by digital image processing method includes, symmetry detection, Border Detection, color, and diameter detection and also we used LBP for extract the texture based features. Here we proposed the Back Propagation Neural Network to classify the benign or malignant stage.

Keywords—Skin cancer detection; dermoscopy; computer aided diagnosis; classification; computer vision; melanoma detection;

I. INTRODUCTION

Melanoma, also referred to as malignant melanoma, is a type of skin cancer caused by abnormal multiplication of pigment producing cells that give color to the skin. Among three types of skin cancer, viz., Squamous Cell Carcinoma (SCC), Melanoma and Basal Cell Carcinoma (BCC), Melanoma is most dangerous in which survival rate is very low. In USA, in every hour one person dies in melanoma. From an study, it is estimated that around 87,110 new cases of melanoma will be diagnosed in 2018. Among them, 9,730 will die because of melanoma. Melanoma consists of only 1% of all skin cancer cases but the majority of skin cancer death. The vast majority of melanomas are caused by the sun. From a survey done by a UK University, it is found that 86% of melanomas are exposed by ultraviolet (UV) radiation.

There are three main layers of the human skin i.e. Epidermis, dermis and hypodermis as shown in Fig 1.

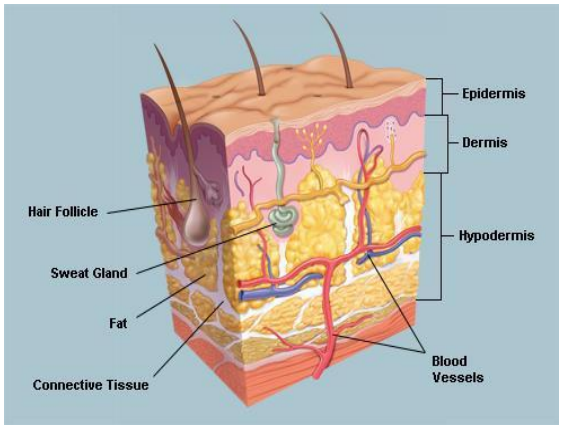

Fig. 1. Human skin anatomy [1]

In the United States, approximately 83,510 new malignant skin cancer cases were estimated in 2016. Out of these

There are mainly two types of skin cancer lesions i.e. benign and malignant lesions. In benign lesions (common nevi) melanin deposits are normally found in the epidermis layer. In malignant lesions, melanin is reproduce at a high abnormal stage. Malignant lesions are not life threatening till the melanocytes and their associated melanin remains in the epidermis layer but when they penetrate into the dermis and leave deposits then the nature of the skin colour change. According to the British Skin Foundation, about 100,000 new cases of skin cancer are diagnosed each year and around 2500 people died due to this fatal disease .

According to the Cancer Research UK, 14,509 melanoma cases were found in the year 2013 alone and it is predicted that this rate will increase in coming years as shown in Fig. 3.

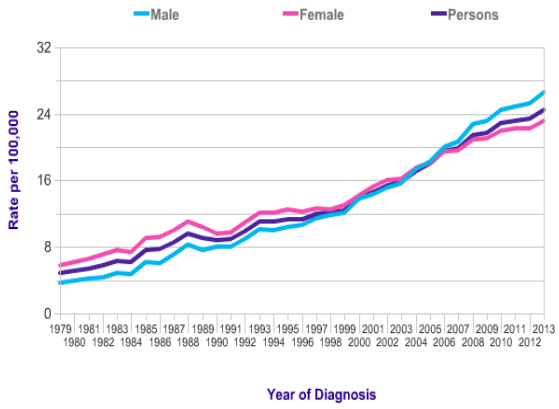

Fig. 2. UK skin cancer statistics from year 1979-2013 [5]

13,650 death cases were found [1]. Malignant melanoma is still incurable if detected in last stage therefore it should be detected in early stages to reduce the health risks, the mortality rates and the treatment costs. To differentiate the early melanoma from other pigmented skin lesion is not an easy task even for experience dermatologists; even in some cases, melanoma is underestimated in early stages by primary care physicians [2] hence attracted interest of researchers. For new researchers working in this area, survey paper of recent literature is highly demanded. For this reason author's addresses the most recent and state of the art techniques used for computer aided skin diagnosis system. Additionally an evaluation of the state of the art work is performed according to different performance metrics.

Almost same steps are involved in every skin disease diagnost ic system and follow the same sequence as shown in Fig 3.

Fig. 3. skin diagnosis systems

This work reviews the noise removal techniques as well as different segmentation, feature extraction and classification techniques. The following sections sequentially discuss about the different pre-processing, segmentation, feature extraction and classification techniques as well as mobile enabled skin disease diagnosis systems found in the literature. Dataset, computational time, accuracy were considered while evaluating the work done by researchers.

2. PRE-PROCESSING

This step includes Converting the RGB acquired skin image to gray image, Contrast enhancement, Histogram modification and, Noise Filtering. Contrast enhancement and histogram modification are proposed since some of the acquired images are not homogenous due to incorrect illumination during the image acquisition. While the histogram modification techniques such histogram equalization is used to enhance the contrast of the image and, therefore, making the segmentation more accurate. Noise filtering using median filter is implemented to reduce the impact of hair cover on the skin in the final image used for classification. The main of pigmented skin lesion is to remove the noise. In computer aided skin lesion diagnosis research it has been observed that the dermoscopy images contain some artifacts such as black frame, dermoscopic gel, air bubbles and some cutaneous features that can affect the lesion classification such as hairs, skin lines and blood vessels. Therefore, some pre -processing steps are required to remove these artefacts to facilitate the segmentation and classification process. Everything that might create hurdle in border detection must be detected and then replaced. In research many approaches were used by different researchers like image resizing [8], contrast adjustment, filtering [4][5][7], colour quantization [5], cropping and hair removal [4]

Usually RGB colours have thousands of colours which are very difficult to handle. Celebi et al. proved that the colour quantization should reduce the number of colours in skin lesion to 20 for precise quantization [4]. Another factor that affects border detection in dermoscopy images is insufficient image contrast. Contrast of skin lesion images are usually enhanced to ensure that lesion edges are prominent [3]. Hiam Alquran proposed that Converting the RGB acquired skin image to gray image, Contrast enhancement, Histogram modification and, Noise Filtering. Contrast enhancement and histogram modification are proposed since some of the acquired images are not homogenous due to incorrect illumination during the image acquisition. While the histogram modification techniques such histogram equalization is used to enhance the contrast of the image and, therefore, making the segmentation more accurate. Noise filtering using median filter is implemented to reduce the impact of hair cover on the skin in the final image used for classification.

Black frames are often part of dermoscopic images usually introduced during the digitization process. For black frame removal different researchers proposed and implemented different techniques. Celebi et al. proposed an iterative algorithm for removal of black frames in which they utilized the lightness component of HSL (Hue-Saturation-Lightness) colour space and the pixel was considered to be black if the value of L was less than 20. Using this criterion the image was scanned row by row and if a particular row contains greater than 60 % black pixels it was considered as part of black frame [8]. Same color space was used by Korjakowska and Tadeusiewicz where they used the L compnent value and considered it black if its value was less than 15. The input image given to the system can be obtained in any lighting condition or by using any camera such as mobile camera. Hence it needs to be pre-processed.

3. SEGMENTATION

Segmentation is process of removing region of interest from given image. Region of interest containing each pixel similar attributes. Partitioning of an image into disjoint regions that are homogenous w.r.t. a chosen property like shape, colour and texture is known as segmentation. The second stage after preprocessing is detecting and segmenting the region of interest (ROI) which represents the lesion region. The segmentation stage includes steps: Image thresholding, image filling, image opening, converting

extracted region to gray level, and then performing histogram equalization to the extracted gray level image. Uzma Bano Ansari using maximum entropy thresholding for segmentation [9]. First of all we have to take gray level of original image then calculate histogram of gray scale image then by using maximum entropy separate foreground from background. After maximum entropy we obtained binary image that is black and white image.

Shivangi Jain the image segmentation is performed by using our proposed automatic thresholding and masking operation in R,G and B planes. First, automatic thresholding proposed by Otsu12 is applied in each plane. Binary masks for each plane are obtained and then combined to produce a final lesion mask. We use 3-plane masking procedure to increase segmentation accuracy[10]. Then edge detection is applied to further segmentation. The main prerequisite for extracting the features is that the lesion must be separated from the surrounding normal skin. But the segmented image may contain other smaller blobs which are not the skin lesion. To overcome this, we find the biggest blob in the segmented image. The segmented image obtained contains only the skin lesion

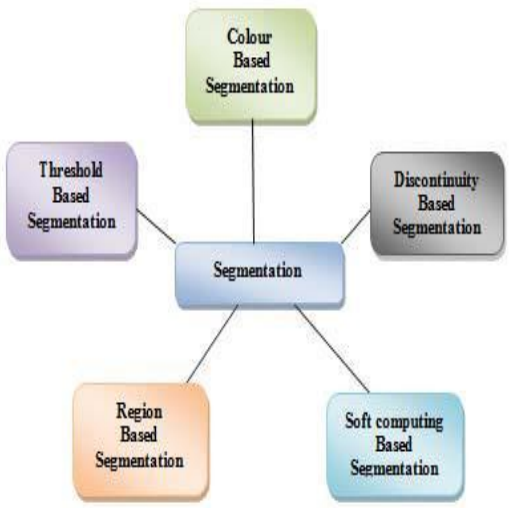

Fig.4. Segmentation types in Skin Cancer Detection Systems

4. FEATURE EXTRACTION

Feature extraction plays an important role in extracting information present in given image After extracting the lesion (ROI) in the segmentation stage, the predefined features will be extracted from the ROI for classification. The selected features are shape, color and various texture features. Since these images have some statistical texture features, we use one of the common algorithm to extract such features which is Gray Level CoOccurrence Matrix

The output image of enhancement is taken as the input for segmentation. Asha Gnana Priya.H A Global Threshold is computed by Otsu's method to be used to convert an intensity image to a binary image. Morphological operations are used to remove the lower pixels to suppress light structures connected to image border and to fill image region and holes.Otsu Threshold - Select an initial estimate as T which is the function of (x, y). Compute the means of the two regions determined by T. Set the new T as the average of the two means. Repeat this until the difference in T in successive iterations is smaller than a predefined parameter. Morphological Operation-The enhanced image is converted into binary image. Remove all the connected components which is fewer than 30 pixels from the binary image and fill the holes in the image. To preserve the circular nature of the object create a disk shaped structuring element. Closing operation is done to join the circles in an image by filling the gaps between them and smoothing their outer edges. To create image emphasizing edges and binary mask sobel filter is used. Morphological dilation and erosion operations are done and the final mask is generated.

(GLCM) [11]. where some of the colour and texture features are calculated based on the definition of three significant image regions – lesion, inner and outer periphery. Features play an important role in skin lesion classification. There are many features associated with skin and can be classified into different categories like color features, ABCD rule features,

Colour is one of the main characteristics used for determining the skin disease. In dermoscope images epidermis usually appears as white. Melanin colour is most important as it helps in determining the different structural and chromatic patterns. The pigmented skin lesion may have different distribution of pigmentation depending on the location of melanin in different skin layers Melanin usually appears as black in upper epidermis. Different color spaces are used for skin cancer detection. Among them prominent color spaces are RGB, CIE L*a*b*, CIE L*u*v*, HSV (Hue Saturation Value and HSL (Hue Saturation Luminance). ABCDE feature includes asymmetry, border irregularity, color and diameter. The feature extraction process includes 4 phases as follows:

- Phase 1: In this phase, we deal with the original image in RGB format, which contains three channels of colors, Red, Green, and Blue. Color feature extracted from image by calculating the density of specific colors in the lesion image.
- Phase 2: In this phase, we deal with the binary, where the features of Asymmetry, border irregularity, and circulation are obtained from the binary image TDS features are calculated with parameters as Asymmetry, Border irregularity, color and diameter

- Phase 3: in this phase, we deal with a lesion image in grayscale image. Energy, correlation, homogeneity and contrast features are obtained by applying graylevel co-occurrence matrix (GLCM) on the gray level image of the lesion
- Phase 4: in this phase, we deal with the histogram equalized image, where the features of entropy, skewness, kurtosis and mean are obtained

The main features of the Melanoma skin Lesion are its Geometric Feature. Shivangi Jain proposes to extract the Geometric Features of segmented skin lesion. Here, he used some standard geometry features are Area, Perimeter, Greatest Diameter, Circularity Index, Irregularity Index. From the Segmented image containing only skin lesion, the image blob of the skin lesion is analyzed to extract the its geometrical features. The Different Features extracted are as follows: Area (A): Number of pixels of the lesion. Perimeter (P): Number of edge pixels.

Uzma Bano Ansari uses GLCM for texture image analysis. GLCM is used to capture spatial dependency between image pixels. GLCM works on gray level image matrix to capture most common feature such as contrast, mean, energy, homogeneity. GLCM texture picks up the relation between two pixels at a time, called the reference and the neighbour pixel. GLCM expounds the distance and angular spatial relationship over an image sub- region of specific size. GLCM is prepared from gray scale values. It is taken into account how often a pixel with gray level(gray scale intensity or gray tone) values come either horizontally, vertically and diagonally to levelled the pixels.

Feature extraction is the second component of image classification module. Since this component produces a significant impact on the results of classification, it plays an important role in the performance of any image classification. Feature extraction is a type of dimensionality reduction that efficiently represents interesting parts of an image as a compact feature vector. This approach is useful when image sizes are large and a reduced feature representation is required to quickly complete tasks such as image matching and retrieval

V. CLASSIFICATION

Classification phase of the diagnostic systems is the one in charge of making the inferences about the extracted information. Classifier is used to classify cancerous image from other skin diseases. In literature review there are several classification methods some of them are clinical and others are computer aided. Prominent classification methods are ABCD rule of Dermatoscopy Researchers have used different machine learning techniques for automated skin lesion classification. Frequently used techniques in literature are support vector machine, artificial neural networks, k-nearest neighbours and decision trees. Classification techniques in different classifiers are shown in Fig.5. The training and test paradigm is very common in statistical studies. According to this paradigm the data is separated into two distinct sets one is training set and other is testing set. These two sets should be independent statistically. One set is used for developing or training the algorithm and the other set is used for testing the developed algorithm. Sensitivity and specificity are the common norms to find the efficiency of the system [12] but in some cases accuracy alone is also used.

Abdulaziz Namozov proposes We have created a simple CNN for melanoma classification using some of the ideas, which was applied to recognize hand-written characters. Considering the fact that we need a deeper network for this task we add more layers and the final version of our network consists of 9 layers. The network has four convolutional and two pooling layers for feature extraction, and three fully connected layers, in the end, for classification

A classifier assigns each object to a class. This assignment is generally not perfect and objects may be assigned to the wrong class. To evaluate a classifier, the actual class of the objects must be known. To evaluate the classification quality, the class assigned by the classifier is compared with the actual class. This allows the objects to be divided into the following four subsets:

1. True positive (TP): the classifier correctly predicts the positive class.
2. True negative (TN): the classifier correctly predicts the negative class.
3. False positive (FP): the classifier incorrectly predicts the positive class.
4. False negative (FN): the classifier incorrectly predicts the negative class.

Based on the cardinality of these subsets, statistical quantities for the classifier can now be calculated. A common and widely used quantity is accuracy, which is only a reasonable measure if the different classes in the dataset are approximately equally distributed. Accuracy is calculated by (TP + TN)/(TP + TN + FP + FN). It specifies the percentage of objects that have been correctly classified. Two other important metrics are sensitivity and specificity, which can be applied even if the different classes are not equally distributed. Sensitivity indicates the ratio of objects correctly classified as positive out of the total number of positive objects contained in the dataset and is calculated by TP/(TP + FN). Specificity indicates the ratio of negative

objects correctly classified as negative out of the total number of negative objects contained in the available dataset and is calculated by TN/(TN + FP).

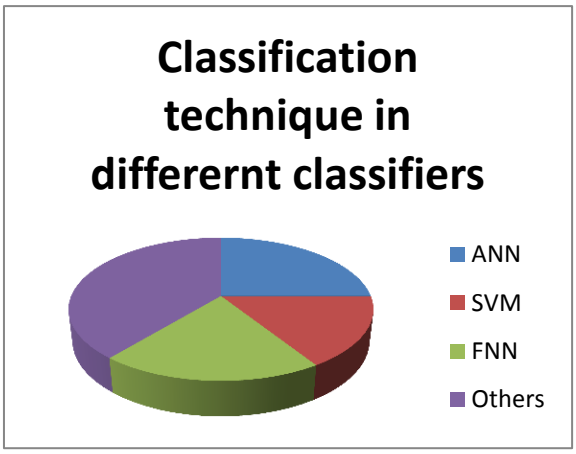

Fig.5. machine learning techniques used by existing systems

Uzma Bano Ansari uses For simplicity Support Vector machine classifier is used here. Svm takes set of images and predicts for each input image belongs to which of the two categories of cancerous and non-cancerous classes. The purpose of SVM is create hyper plane that separates two classes with maximum gap between them [9]. In our proposed system output of GLCM is given as input to SVM classifier which takes training data, testing data and grouping information which classifies whether given input image is cancerous or non-cancerous.The full set of features (11 features) and the selected 5 features using PCA are fed into the SVM model which is used to classify the image into binary classes benign and malignant. Hiam Alquran proposed simplest SVM plane is linear whenever the data can be linearly separated. But the data here is not linearly separable. therefore, kernel SVM with radial basis function [13] was used to classify the data into benign or malignant. In the training stage, the classifier model was built using cross validation procedure to find the optimize parameters of the hyperplane to avoid biasing with overfitting.

Shivangi Jain Using the ABCD rules for the melanoma skin cancer, we use some pre-defined thresholds in classification stage. The Feature Values Extracted in the Feature Extraction stage are compared and the skin lesion is classified as Melanoma Skin Cancer or normal skin or Mole. This classification method proves to be efficient for most of the skin images[10]. Fuzzy logic image analysis techniques were used to analyse three shade of blue in dermoscopic images for melanoma detection. For constructing the fuzzy set for different blue colour shades, training set of 22 additional melanoma images were examined. Fuzzy set colour representation was determined. 150 different features were derived and categorized into three different categories i.e. contour, colour and texture features. Ten - fold cross-validation was used training and testing set. SVM was utilized for skin lesion classification and 81.4% accuracy was achieved.

Ramlakhan et al. introduced an automated skin lesion classification system for smartphone devices. In the proposed system the skin image was first converted to greyscale image and then to monochrome image for finding the contour. A total of 33 colour and shape features were derived from the defined lesion area. The proposed system classified lesions as either malignant or benign using K-NN classifier. Experiments were performed on 83 images and the results showed that system was not highly efficient, achieving and average accuracy of 66.7% [15] Comparison of different classifiers results are shown in Table 1

Ref	Tech	Images	Results	Time
[9]	SVM	1358	Sensitivity: 80.76% Specificity: 85.57%	Seg 3.4938 sec Cla:10.3500 sec Full analysis: 15 sec
[14]	SVM	400	Sensitivity: 88.27% Specificity: 75.31%	Seg: 2.4010 sec Cla:7.4030 sec Full analysis: 15 sec
[15]	K-NN	135	Accuracy: 70.7% Sensitivity: 70.7% Specificity:88.5%	NA

[16]	ANN	250	NA	NA

Table. 1. Comparison of skin cancer diagnosis system

6. PUBLICALLY AVAILABLE DATASET

Some datasets are publically available. The details of the dataset are as follows.

1) PH^2: PH^2 is publically available dermoscopic image database acquired Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal. The PH^2 dataset has been developed to facilitate comparative studies on melanoma segmentation and classification (Mendonca, Ferreira, Marques, Marcal, & Rozeira, 2013). PH^2 dataset contains 200 images and available on http://www.fc.up.pt/addi/ph2database.html (PH^2 Database, 2013).

2) DermIS and DermQuest:Melanoma image dataset is also available on DermIS and DermQuest [17] Website. Dataset is freely available to use for educational purpose.

3) ISIC challenge dataset :The International Skin Imaging Collaboration (ISIC) is an international effort to improve melanoma diagnosis. This challenge provides dataset which is publically available. This dataset contains 900 images for the training purpose and 350 images are for the testing purpose [18].

VII. CONCLUSION

According to literature early detection of melanoma can reduce the mortality rates. In this paper, we review the state of the art in computer aided diagnosis system and examine recent practices in different steps of computer aided diagnosis systems. These systems employ various methods for pre-processing, segmentation, feature extraction and lesion classification by using the extracted features. Certain conclusions are drawn after the analysis of literature. Among machine learning techniques used for skin cancer diagnosis these days, SVM is the prominent and the diagnostic accuracy of these systems lies in between 60%-97%. In mobile based skin cancer diagnosis system computational time is as equal as accuracy. On average the computational time of mobile based systems is 13 – 15 sec. Work found in literature is trained and validated on different datasets which makes the proper comparison a challenging task as the dataset size and image acquisition techniques vary. In conclusion, there should be standard procedures and publically available datasets for the new researchers so that together we can fight against this deadliest disease. Noise-removal, segmentation and real-time classification in mobile based skin cancer diagnosis system are the hot research topics.

REFERENCES

[1]. American Cancer Society, “Cancer Facts & Figures 2016”, pp.1–9, 2016

[2]. R.J. Pariser and D. M. Pariser, “Primary care physicians errors in handling cutaneous disorders”, Journal of the American Academy of Dermatology, vol. 17, pp. 239-245, August 1987.

[3]. S. Jain, V. Jagtap and N. Pise, " Computer aided melanoma skin cancer detection using image processing" Computer Science, pp. 736–741, 2015'

[4]. Q. Abbas, M. E. Celebi and I. Fondón, "Hair removal methods : A comparative study for dermoscopy images", Biomedical Signal Processing and Control, vol. 6 issue 6, pp. 395–404,

[5]. Q. Abbas, I. F. Garcia, M. Emre Celebi, and W. Ahmad, "A Feature-Preserving Hair Removal Algorithm for Dermoscopy Images", Skin Research and Technology, vol. 19 issue 1,pp 27–36. 2013

[6]. M. E. Celebi., Y. A. Aslandogan and P. R. Bergstresser,. "Unsupervised border detection of skin lesion images", International Conference on Information Technology: Coding and Computing (ITCC'05) , vol. II, pp. 1–6, 2005

[7]. I. Maglogiannis, S. M. and K. Delibasis , "Hair Removal on Dermoscopy Images", 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2960–2963 ,2015

[8]. H. Zhou, M. Chen., R. Gass, J. M. Rehg, L. Ferris, J. Ho and L. Drogowski, "Feature-preserving artifact removal from dermoscopy images",. Proceedings of the SPIE Medical Imaging Conference, pp. 9–14, 2008

[9]. Uzma Bano Ansari, Tanuja Sarode," Skin Cancer Detection Using Image Processing",. International Research Journal of Engineering and Technology (IRJET), Apr -2017

[10]. Shivangi Jaina , Vandana jagtapb, Nitin Pise," Computer aided Melanoma skin cancer detection using Image
Processing", International Conference on Intelligent Computing, Communication & Convergence (ICCC-2014)
Conference Organized by Interscience Institute of Management and Technology, Bhubaneswar, Odisha, India

[11]. Maglogiannis, Ilias, and Charalampos N. Doukas. "Overview of advanced computer vision systems for skin lesionscharacterization." IEEE transactions oninformation technology in biomedicine 13.5 (2009): 721-733.

[12]. T. Wadhawan, N. Situ, H. Rui, X. Yuan, and G. Zouridakis, "SkinScan c : A PORTABLE LIBRARY FOR MELANOMA DETECTION ON HANDHELD DEVICES". 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 133–136, 2011

[13]. Hiam Alquran," The Melanoma Skin Cancer Detection and Classification using Support Vector Machine", 2017IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)

[14]. M. Lingala, R. Joe Stanley, R. K. Rader, J. Hagerty, , H. S. Rabinovitz, , M. Oliviero, and Stoecker, "Fuzzy logic color detection: Blue areas in melanoma dermoscopy images", Computerized Medical Imaging and Graphics, vol. 38 issue 5, pp. 403–410, 2014

[15]. K. Ramlakhan and Y. Shang, "A mobile automated skin lesion classification system", 23rd IEEE International Conference
on Tools with Artificial Intelligence, pp. 138–141, 2011

[16]. A. Karargyris and A. Pantelopoulos, "DERMA / care : An advanced image-processing mobile application for monitoring
skin cancer" , 2012 IEEE 24th International Conference on Tools with Artificial Intelligence, pp. 1–7, 2012

[17]. DermQuest Image Library (2016). DermQuest Image Library [Online] Avaiable https://www.dermquest.com/image library/
Retrieved June 29, 2016

[18]. ISBI 2016: Skin Lesion Analysis Towards Melanoma Detection [Online] Available at http://biomedicalimaging.org
November 14, 2016