RESEARCH ARTICLE   OPEN ACCESS

Classification of Melanoma and Nevus in Digital Images for Diagnosis of Skin Cancer

Nimma Manasa, Sunke Akshitha, Vallepu Dhaneshwari
Department of Electronics and Communications Engineering
GuruNanak Institute Of Technology (GNIT)
Ibrahimpatnam, Rangareddy -501506 India

ABSTRACT
Melanoma is considered a fatal type of skin cancer. However, it is sometimes hard to distinguish it from Nevus due to their identical visual appearance and symptoms. The mortality rate because of this disease is higher than all other skin related consolidated malignancies. The number of cases is growing amongst young people but if it is diagnosed at its earlier stage then the survival rates become very high. The cost and time required for the doctors to diagnose all patients for Melanoma are very high. In this research work, we propose an intelligent system to detect and distinguish Melanoma from Nevus by using state of the art image processing techniques. At first, Gaussian Filter is used for removing noise from the skin lesion of the acquired images followed by the use of improved K-mean clustering to segment out the lesion. A distinctive hybrid super feature vector is formed by the extraction of textural and color features from the lesion. Support Vector Machine (SVM) is utilized for the classification of skin cancer into melanoma and nevus. Our aim is to test the effectiveness of the proposed segmentation technique, extract the most suitable features and compare the classification results with the other techniques present in the literature. The proposed methodology is tested on DERMIS dataset having a total number of 397 skin cancer images where 146 are melanoma and 251 are nevus skinlesions. Our proposed methodology archives encouraging resultshaving 96% accuracy.
Keywords: Melanoma, Nevus, Feature, K-means clustering, and centroid selection.

I. INTRODUCTION

Skin cancer is measured as a major contributor to the causesof deaths around the world. There are various types of cancers that are discovered and battled with. However, skin cancer is amongst fast-growing cancer nowadays. According to modern research, patients with a skin cancer diagnosis is significantly increasing more than any other cancer form every year. Melanoma is the most common form of skin disease that affects the skin surface cells known as melanocytes. It consists of cells that cause the skin to turn to black colo. Melanoma can be found in dark or darker color yet at some point it might likewise be in the skin, pink, red, purple, blue or white color. This form of cancer is very disturbing due to its tendency to cause metastasis, i.e. ability to spread. Melanoma can be found anywhere on the human body, however, it is mostly developed on the back of human legs.

Figure 1. Types of skin cancer

Detecting skin cancer at the initial stage can help in reducing the risk factor in patients. Different skin cancer types can be found in figure 1. According to the research, the mortality rate may be reduced up to 90 %, if the skin cancer is diagnosed at an initial stage,Hence the diagnosis and classification of the skin cancer in its early stage are vitally importan. Among the conventional approaches followed by researchers to detect melanoma and nevus is ABCD rule. A total dermoscopic score is obtained for each of the ABCD features where Arepresents Asymmetry, B is for Border irregularity, C represents color variations and D is for the Diameter. Each respective feature is assigned an individual weight based on their significance in the feature space.

Based on the calculated score the lesion is identified as cancerous or benign. A 7-point Checklist is another technique that is used to identify skin cancer in dermoscopic images. The list target symptoms of atypical pigment network, grey-blue areas and atypical vascular pattern, streaks, blotches, irregular dots and globules, and regression patterns. At times when these symptoms are identified, a medical professional is consulted for the treatment. Later on, the checklist reduced to a lesser number of features of the different network, Asymmetry and blue-white structure]. Considering the complex nature of melanoma, it becomes hard for the researchers to detect skin cancer only on the basis of these geometrical features. Another problem is that the size of the image database is increasing dramatically. So the practicality of such information is dependent on how well it can be accessed, searched and how well the relevant knowledge can be extracted from it. With the advent of computer-aided diagnostic systems,researcher mainly emphasizes on the automatic detection and classification of skin cancer. Medical images in the form of textural features, geometric features, color features and in a combination have been used to identify and classify skin cancer diseases. However, it is still a challenging task to identify most discriminative features for identifying melanoma at its initial stage. Our research work aims to achieve high accuracy results in identifying and classifying skin cancer, contributing to the present literature.

• To develop a complete automated computer-aided system to detect melanoma cancer accurately.

• Design of an improved K-Mean approach for computationally efficient segmentation.

• Utilization of hybrid features incorporating both texture and color of the lesion. The remaining of our research paper is organized as a detail literature review of existing techniques of features extracting and classification is discussed in Section 2. The components of the proposed work are discussed in detail in section 3. Section 4 describes the experimental setup and evaluation metrics. Section 5 portrays the results and discussion on the given datasets. The conclusion is given in the last section.

II. RELATED LITERATURE

Extensive research has been done in the respective domain by the discovery and development of new approaches to accurately diagnose skin cancer. Related work can be divided into three types based on features extraction technique i.e. geometrical, textural, and color. These techniques are individually discussed in

detail below.

A. Geometrical features

The ABCD-E system, 7-point checklist 3-point checklist offers geometrical features techniques that are used for classification of the melanoma. .According to Johr, the features extracted by the ABCD rule is computationally less expensive with respect to the 7-point checklist. It is also observed that the consistency in the clinical diagnosis is at a higher rate for the ABCD rule. So,most of the computer-aided systems for melanoma detection used ABCD rule for features extraction. Though, ABCD technique is more prone to over classification of nevi as melanomas. In research work Reda et. al., extractedthe characteristics of ABCD attributes and combines with color asymmetry and dermoscopic structures. The proposed system achieved 91.25% sensitivity. In another study the author tried to investigate the possibility to automatically detect the dermoscopic patterns based on ABCD rule using deep convolutional neural networks. Experimental results demonstrated 88% accurate classification of skin cancer.In another research study, Rebecca et. al used ABD rule excluding the color from the traditional ABCD, because color requires an additional large amount of computer resources. They utilized KNN as classifier achieving 89% accuracy.It is also observed that ABCD rubric is sometimes very qualitative and subject manner that results in large interobserver and intra-observer bias. For this purpose,High-quality intuitive features (HLIF) approach is used that represents the asymmetry characteristic of ABCD rubric from skin cancer images. The system achieved 86% accuracy from the extracts set of features. Later Emerald et. al. extended his previous work by proposing six new HLIF features for different color channels. Experimental results show a better classification accuracy of 94%.

B. Color Features

The color features can be extracted on the basis of the statistical value calculated from color channels. They are Mean color, variance color and standard deviation value of the RGB or HSI color model. Some of the different color features extraction techniques consist of color asymmetry, centroid distance and LUV histogram distance. In another study, the author classified the melanoma on the basis of global and local descriptors. They combined the textural features with the color obtained classification scores of SE and SP are 93%, 95% respectively. Other researchersused the same approach using color features with a set of textural and shape-based features. Ganster et al.

used color and shape based features from skin lesion with KNN classifier. The number of images taken for the experimentation purpose is more than 5300 and they achieved 87% and 92% sensitivity and specificity respectively. Rubegni et al. also utilized textural and geometrical features and achieved a sensitivity of 96% and specificity of 96%. Celebi et al. used a large features vector consisted of color, shape and texture features. The SVM achieved a sensitivity of 93% and specificity of 92%.Almansour et. al in used color moments with textural features with SVM as classifier achieved 90 % accuracy with 227 melanoma and nonmelanoma skin cancer images.

C. Textural Features
Image texture represents the spatial distribution of pixel intensity levels in an image. Textural features represent the underlying pattern and layout of intensity levels acting as one of the most distinctive features for object or region of interest identification. When it comes to skin cancer, texturalfeatures are frequently used for image analysis as it helps in classifying between nevus and melanoma by calculating the irregularity of their structure [33]. It is observed that interest is increasing in the computerized examination of digital images taken by the Dermoscopic process. To improve early identification melanoma classification strategy for the computerized analysis of images obtained from ELM (Epiluminescence microscopy) has been established [6]. Region of interest was extracted using the segmentation algorithm and a combined features strategy was used belonging to shape and radiometric features. KNN classifier was used to achieve 87% and 92 % sensitivity and specificity respectively. The automatic data analysis for the melanoma early detection system (ADAM) [34] obtained 80% for both specificity and sensitivity using asymmetry and boundary descriptors with (SVM) as a classifier. Iyatomi et. al. [35] used a similar approach and achieved 100 % specificity and 95.9 % sensitivity. Recent systems in the literature [36] achieved 91 % specificity and 88.2 % sensitivity over database of 120 skin cancer images. For a proper treatment of the skin cancer computer-aided diagnosis (CAD) techniques helps the physicians to obtain a second opinion about the lesion. Extracting accurate region of interest from skin lesion alongside adjacent portion is required for proper diagnosis and analysis. The authors [37] used active contour and watershed mask for segmentation. They extracted features related to shape, textural and color. The proposed system was tested on 50 images of DERMIS dataset achieving 80% accuracy. Another novel approach to the

Computer Aided Diagnosis (CAD) of initial melanoma has appeared in the process of the web and smartphone applications. In these systems, images were captured with high-resolution cameras instead of using an image database[38]. The system used digital camera images with context knowledge such as type and nature of the skin, sex, age and the body part that is affected. The proposed system also triedto extract the features compatible with the Dermoscopy ABCD rule. The features were further classified in various steps, a processing step for correlation-based feature selection. The system produced specificity of 68% and a sensitivity of 94%, on images of 45 Nevus and 107 Melanoma skin cancer images. In another work [39], the author used some non-dermoscopic images of skin cancer where they extracted the ROI or lesion areas using k-mean clustering technique and then color and textural features were extracted. Further, a set of visual features wereidentified by inspecting the dermatologist.
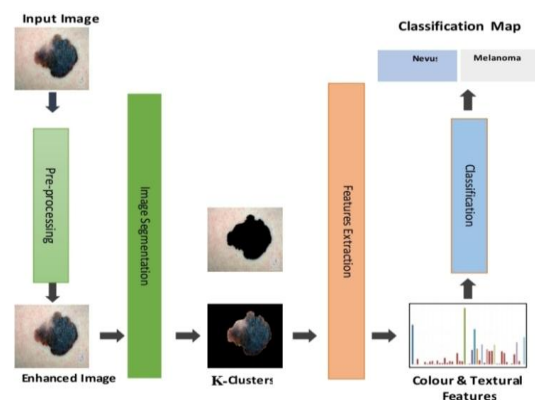

Figure 2. Proposed Methodology for skin Cancer Detection

For automatic prediction the physician attributes and extraction features were used. A majority vote of all the given predictions was used to achieve final classification. The suggested method resulted in higher diagnostic of 81% and achieved comparable results to the latest methods that are using skin cancer images. In [40], the melanoma lesion is depicted through a component vector containing texture data, shape, color and also global and local parameters. Some other automated Melanoma detection and classification systems discussed in [41] [42] for the purpose to improve the recognition and classification of the skin lesion. In [43], an approach was presented for feature selection criterion which was built on the arrangement of differential evolution and SVM. Currently, Convolutional Neural Network (CNN) and deep learning-based approaches were used for cancer detection. The computation cost for these approaches became barriers in clinical applications

[2], [44], [45]. In the latest research work ulzii et. al. [46] classified different skin cancer diseases using deep convolution network. For extracting features a Pre-trained AlexNet convolutional neural network model is used. An ECOC SVM classifier is utilized in the classification of skin cancer. The proposed approach obtained the results of 90% accuracy for melanoma detection. An enhanced computer automated system was proposed with pixel based technique is used for segmentation [47]. They extracted features using CNN and classify the images through SVM classifier achieving 93% accuracy on DermIS dataset.The afore-mentioned techniques provide promising results however, they are tested over an insufficient number of images for detailed analysis. From the literature, it is observed that different features extraction techniques have been used to classify the skin lesion. However, the combination of all these features is still not properly tested. As the skin cancer images can be differentiated based on color variations, therefore, color features should be combined with other features.

III. PROPOSED METHODOLOGY

We have discussed the components of our proposed methodology in this section. Input images acquired from the datasets undergo quality enhancement through preprocessing techniques. Later the ROI from the skin lesion is extracted which are further processed for significant featureextraction and lastly classifying them into melanoma or nevus. Figure 2 presents the proposed methodology where each stage is briefly discussed in the section below.

A. PRE-PROCESSING

Medical images are often susceptible to noise mainly due to bad illumination, hair and air bubbles [48]. This inclusion of noise in images results in the formation of artifacts. Due to such artifacts, the segmentation results may get affectedcausing inaccurate detection results. Therefore, noise removal is a significant step before applying any segmentation or feature extraction technique for an accurate diagnosis. To smoothen the image, Gaussian filter is highly recommended as it removes the speckle noise added during the process of acquisition. Gaussian kernel coefficients are sampled from the 2D Gaussian function.

B. SEGMENTATION

K-mean clustering is a common machine learning technique that is extensively used in many applications such as data mining, image processing, and pattern recognition. K-meanis considered as one of the basic methodologies for groupingand clustering of data-points into K number of clusters [18].It works by splitting the image into non-overlapping groups of pixels based on their intensity levels. The process initiates by selecting the centroids from the data-points either randomly or through a certain criterion. The pixels or datapoints are clustered based on their minimum distance from the selected centroids. After each iteration, the mean values of the formed clusters are found and are set as the centroids for the next iteration.
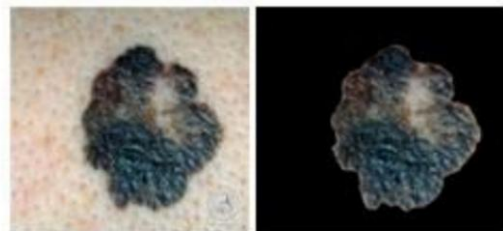

Figure 3. Input image on the left with its equivalent segmented image on the right

The iterative process repeats itself until there is no variation in the successive cluster centroids. The proposed K-Mean initializes its centroids by centroid selection technique. The centroid selection technique works by ensuring significant difference among the values of initialized centroids making it more efficient and robust by converging to the final position in a lesser number of iterations.In our proposed system, input images consist of affected lesion surrounded by the background skin. The value of K is taken as 2, such that the foreground lesion is extracted from the background skin as shown in Figure 3.

C. FEATURES EXTRACTION

Once the lesion is segmented out of the background skin, it is then classified as malignant or benign. For better classification results, it is required to use the best feature descriptors for machine learning modeling. The Increase in the number of features increases the computational cost, inspiring the description of precise decision boundaries. Thus, it is ensured that a distinctive feature set is used. A lesion is characterized by its texture and its color. In this research work three different features using Local BinaryPattern (LBP), Grey Level Co-occurrence Matrix (GLCM) and RGB color channel features, are extracted from the ROI of skin lesion. The techniques are utilized to extract the textural and color-based features from the input skin lesion.

i. Grey Level Co-occurrence Matrix

Grey level co-occurrence matrix (GLCM) is a global textural feature extraction technique computing the statistical distribution of intensities in combination at specific positions in the image. Based on the number of contributing intensity levels in the combination, the order of the statistics is determined. GLCM extracts the second order statistical texture features by considering the spatial relationship of two intensity levels. A GLCM is a square matrix with dimensions of the number of intensity levels G. The matrix element P(i,j) specified by the row and column position with i and j, gives the frequency with which the pixel values i and j have occurred overall directions specified by angle θ as shown in figure 4.
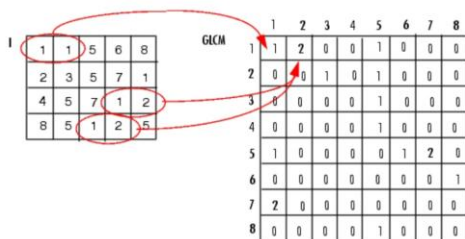

Figure 4. Formation of a GLCM matrix from the input image (shown as a matrix of intensity levels)

After the formation of the matrix, various statistical features defined by Haralick [49] are extracted specifying the underlying textural information of the image. Entropy,Angular second moment, variance, etc. are few of the listed features among the Haralick features.

ii. Local Binary Pattern

GLCM extracts the global texture of the image giving an overall spatial distribution of image texture however, local texture also needs to be part of the feature vector. Ojala presented a local texture feature extraction technique for aninvariant two-dimensional Grey scale analysis; Local Binary Pattern [50]. It is used by several applications and is an efficient way and simple operator to denote local patterns. The pixels are tagged by LBP to identify the eight neighborhood pixels with respect to the center value of the window image. The pixels are assigned a binary number based on the threshold value. By comparing the central pixel of the image window with respect to its neighbors LBP code .
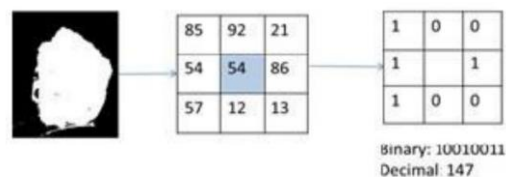

Figure 5. LBP represented by binary output matrix calculated from the input image.

The LBP algorithm comprises of its roots in 2D texture analysis. Summarization of the local structure of the image is done by comparing each pixel with its neighbor's pixels. First of all,take a pixel as center and threshold its neighbors against that pixel. Mark 1 if the center pixel intensity is greater or equal its neighbor and mark 0 if not. We will have a binary number for every pixel, much the same as 10010011. With 8 encompassing pixels, we will finish up with 256 possible combinations, which are named as Local Binary Patterns or LBP codes. The method to compute an LBP code is shown in Figure 5

The label image histogram is used as a texture feature following the labeling of pixels with their corresponding LBP codes.

iii. Color Features

When an object is exposed to a certain wavelength of light it reflects a corresponding color, forming its appearance. By examining the three main colors (red, green, blue), this phenomenon of color space can be better understood. In the segmented lesion, the color feature is used to identify itsvisible color, this is achieved by utilizing four statistical values such as mean, variation, standard deviation, and skewness through color spaces of RGB. These values are also referred to as color moments or features. Consider C as the color channel with image i, N represents a total number of pixels in a color segmented image, k is the kth pixel of that color channel P of an image i with N pixels in a color space. Hence the color features can be defined as below. Here mean shows the average value of each color in RGB color space.Standard deviation (SD) is the average of the squared differences from the Mean. Skewness calculates that how much the Asymmetry of the probability distribution of some are given real-valued random variable about the mean. Variance is the variation of the color distributionFor an accurate diagnosis of skin cancer, each discriminative feature is extracted individually. GLCM assesses the global texture of the image whereas, LBP provides the texture analysis of patches within the image.

Nimma Manasa, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 10, Issue 5, (Series-III) May 2020, pp. 42-49
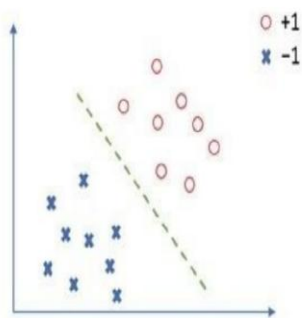
Figure 6 SVM Classifier Model

Their combination forms a distinctive feature descriptor providing an optimal measure of the texture. Alongside lesion is also characterized by its color. Thus, the textural along with the color features listed with the description in Table I, are combined to form a hybrid super feature vector. The feature-length as a result of the concatenation of the individual features are summed up to a total feature length of 294.

D. CLASSIFICATION

After segmentation and feature extraction, the hybrid featurevector is then provided to classifiers to identify the melanoma and nevus. Different classification algorithms are trained and tested at their default parameter settings to achieve high accuracy as discussed below.

i. Support Vector Machines

Support Vector Machines (SVM) consists of a set of supervised learning methods originally discussed by Vapnik in 1963. This technique is used to reduce the classification error and also maximizes the geometric boundary that separates the class values. They are also known as maximum margin classifier. In N-dimensional feature space, each data point is plotted at its respective coordinates. The classifier works by finding the right hyper-plane to separate the data points into required classes. Once the hyperplane isdetermined, the testing samples are predicted to be on either side of the plane. The hyperplane can be characterized as

$$w.x + b = 0$$

Here x denotes the N-dimensional input vector, w is the vector weight defined as $w = w1, w2, w3.....wn$, while b is the model bias function. As we have two classes melanoma and nevus, several decision boundaries may occur. SVM identifies the decision boundary, the hyper-plane having the most extreme separation from the two classes. The proposed SVM utilizes linear kernel with margin

constant C as The hyperplane dotted line and the binary classes +1 and -1 is shown in figure 6. Classification using SVM is a supervised machine learning approach that needs appropriate training on larger datasets for binary classes.

TABLE 1 FEATURE DESCRIPTION TABLE

S. No	Feature	Size	Description
1	GLCM	1x23	Generate texture features as second order method is provided by co-occurrence probabilities
2	LBP	1x257	Local Binary Pattern
3	Color Features	1x12	Mean, Standard Variation, skewness, and variance of RGB color channel

SVM selects the values of w and b from the training samples. SVM identifies a hyperplane for maximum separation between true and false training examples. From the below equations the hyperplane H over the training data samples are represented as

$$Xi * w + b \leq 1, yi = -1 \quad (1)$$
$$Xi * w + b \geq 1, yi = +1 \quad (2)$$

From equation 1 and 2, we get

$$yi(xi.w + b) - 1 \geq 0, \forall i \quad (3)$$

From the above equations, it is observed that all the trained samples may occur on both sides of the hyperplane. As the model is trained, the solution of the decision problem is found by evaluating the sign of the yi with the coefficient vector w. Classifiers take these textural and color features as input for classification. Detection and classification of melanoma is a binary classification problem with discrete values of the data-points calculated over a dataset of an adequate number of images making it suitable for SVM. The accuracy of the classifier is calculated mapping the classified results with the ground truth data.

ii. K-Nearest Neighbor

K-nearest neighbors (KNN) is a machine learning algorithm that keeps all training values and classifies new ones based on a similarity measure.

IV. RESULT

Nimma Manasa, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 10, Issue 5, (Series-III) May 2020, pp. 42-49

Input Image: The input image is given from the datasets of Melanoma and Nevus.
This input we have taken is the effected skin tissue.

Gaussian filtered Image: The input image is subjected to the gaussian filter. This filter is used to remove the noise and also to remove the hair on the input image which we have taken
The above shown figures shows the original and segmented images.
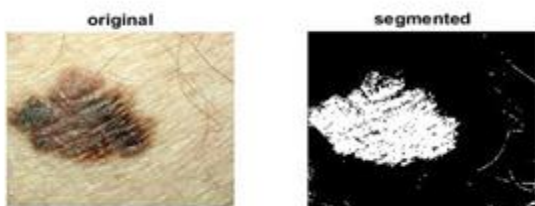
The above shown figures shows the original and segmented images.
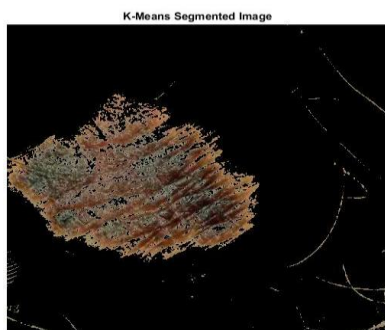
The above figure shows the k-means segmented output.

V. CONCLUSION

In this research paper, we presented an intelligent system for classification of skin cancer into melanoma and nevus. It is observed that major problem that causes the misclassification is lesion detection and segmentation. The K-mean clustering technique using centroid selection is used to extract the ROI from the cancer image more accurately and efficiently. Textural and color features extraction techniques are used to obtain best-suited features for classification. For texture features, GLCM and LBP features are combined with the color features to achieve a high classification accuracy of 96%. In this way, our proposed technique has been able to classify skin cancer images into melanoma and nevus more accurately and efficiently. The effectiveness and performance of the p

ACKNOWLEDGEMENT

REFERENCES

[1]. M. Aleem, N. Hameed, and A. Anjum, "m-Skin Doctor : A Mobile Enabled System for Early Melanoma Skin Cancer Detection Using Support Vector Machine," vol. 2, pp. 468–475, 2017.
[2]. B. Kong, S. Sun, X. Wang, Q. Song, and S. Zhang, "Invasive cancer detection utilizing compressed convolutional neural network and transfer learning," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018.
[3]. Q. Li, L. Chang, H. Liu, M. Zhou, Y. Wang, and F. Guo, "Skin cells segmentation algorithm based on spectral angle and distance score," Opt. Laser Technol., vol. 74, pp. 79–86, 2015.
[4]. R. Kasmi and K. Mokrani, "Classification of malignant melanoma and benign skin lesions : implementation of automatic ABCD rule," vol. 10, no. 1, pp. 448–455, 2016.
[5]. R. J. Friedman, D. S. Rigel, and A. W. Kopf, "Early Detection of Malignant Melanoma : The Roleof PhysicianExamination and of the Skin," 1985.
[6]. H. Ganster, A. Pinz, R. Röhrer, E. Wildling, M. Binder, and H. Kittler, "Automated

melanoma recognition," IEEE Trans. Med. Imaging, vol. 20, no. 3, pp. 233–239, 2001.

[7]. G. Betta, G. Di Leo, G. Fabbrocini, A. Paolillo, and M. Scalvenzi, "Automated application of the '7-point checklist' diagnosis method for skin lesions: Estimation of chromatic and shape parameters," in Conference Record - IEEE Instrumentation and Measurement Technology Conference, 2005, vol. 3, pp. 1818–1822

[8]. G. Argenziano et al., "Seven-point checklist of dermoscopy revisited," Br. J. Dermatol., vol. 164, no. 4, pp. 785–790, 2011.

[9]. I. Zalaudek et al., "Three-point checklist of dermoscopy: An open internet study," Br. J. Dermatol., vol. 154, no. 3, pp. 431–437, 2006.

[10]. P. Mohanaiah, P. Sathyanarayana, and L. Gurukumar, "Image Texture Feature Extraction Using GLCM Approach," Int. J. Sci. Res. Publ., vol. 3, no. 5, pp. 1–5, 2013.