Detection And Recognition Method Of Benign Skin Cancer In Dermoscopic Images Using Hybrid Feature Fusion Model

Dr. Manjunatha Hiremath[1*], Dr. Seema Prashant Nehete[2], Dr. Shila Sumol Jawale[3], Dr. Mostaque Md. Morshedur Hassan[4], Divyang Pankajbhai Raval[5], Hitarth Dharmendrakumar Patel[6], Jayashri Deb Sinha[7], Santanu Koley[8]

[1*]Department of Computer Science, CHRIST (Deemed to be University), Bangalore560076, Karnataka, India manju.gmtl@gmail.com
[2]Department of Computer Engineering, Don Bosco Institute of Technology, Kurla, India. seema@dbit.in
[3]Department of Artificial Intelligence, Datta Meghe College of Engineering, Airoli, India. shilaph@gmail.com
[4]Department of Computational Sciences, Brainware University Barasat, India. mostaq786@gmail.com
[5]Department of Computer Engineering, Shree Swaminarayan Polytechnic Gandhinagar, India. ravaldivyang3201@gmail.com
[6]Department of Computer Engineering, Shree Swaminarayan Polytechnic Gandhinagar, India. patelhitarth1711@gmail.com
[7]Department of Computer Science(Data Science), Haldia Institute of Technology Haldia, India. jdebsinha@gmail.com
[8]Department of Computer Science & Engineering, Haldia Institute of Technology Haldia, India santanukoley@gmail.com

Abstract:
Image Processing and computer-aided diagnostics have become increasingly important in the modern world of medical diagnosis. This diagnostic procedure has been widely accepted by the medical community for its biological imaging applications. According to a recent survey conducted by the American Academy of Dermatology Association (AAD), it is evaluated that 192,310 new instances of melanoma and 95,830 favorable cells(noninvasive) cases will be analyzed in the U.S in 2019. In order for this diagnosis to take place and in order to identify the reasons behind this, immense technological support along with a thorough analysis by medical experts is required. The objective of this paper focuses on developing a recognition system to differentiate and identify benign cells from the affected part of the skin. So that the skin lesion can be identified as affected or not. This paper proposes a prediction system using the SVM classifier. The dataset being used for classification is collected from the ISIC archive. The proposed system is implemented with the help of hybrid segmentation techniques named Snakes Algorithm and Marker controlled watershed segmentation with geometrical feature extraction. The model is being implemented on the matlab tool and provides the 86% accuracy rate, it helps to give an adequate and easy identification of cancerous cells. This model can be improved and enhanced in the future with the help of further research.

Keywords: Image processing, Hybrid Segmentation, Benign Cells, recognition.

1. INTRODUCTION

Skin cancer is one of the most fatal syndromes to affect human beings. Additionally, being a form of cancer, it is not an easily curable disease. Failure in recognizing these affected skin cells can prove to be fatal in nature. To aid the process of diagnosis and detection, computer-aided designs are available to help in the recognition of these hazardous cells. Dermoscopy is one method used to diagnose skin lesions and the various associated conditions related to human skin[1]. It is discovered that dermoscopy has a lower symptomatic exactness in the hands of unpracticed dermatologists. These errors have been rectified by computer-aided analysis. Skin malignant growth is seen as of different kinds, for example, melanoma, basal and squamous cells. Among these, melanoma is considered the most unpredictable in its prognosis. If detected in its early stages, there are higher chances that melanoma cancers can be cured. Computer visualizations play a major role in identifying these hazardous cells by capturing images of these skin cells[2]. On an increased analysis of these images (such as texture analysis, detection, shape analysis, etc.), one can gain a better understanding of the cancerous cells.

A dermatologist may not always be readily available for patients who have skin cancer. Hence, computer-aided systems can fill in for these periods of absences and help detect any issues that may arise[3]. Proper health care treatments can be given to affected individuals if it is diagnosed early as well. Hence, there should exist a system, which can help in diagnosing the disease early on. This is where computer-aided diagnostics come into play. Computer-aided diagnostics can help save several lives if the proceedings applied to the affected regions correctly[4].

With the help of health records from the medical professionals and using the above proposed model - early identification of disease is possible to carry on further medications. Many algorithms and techniques that detect these cancerous cells are already available. However, these methods have a high rate of error as well.

The main focus of this model is to identify the affected regions more efficiently and classify them accordingly than the other available models.

2. RELATED WORK

Prashant Bhati and Manish Singhal [5] have proposed a system that helps to classify skin lesions as malignant or benign. The system is based on the Total Dermoscopic Score. The preparing and OTSU's division technique is utilized for the grouping of sores as nonmalignant or dangerous.This along with the ABCD algorithm can lead to the result being accurate enough in classifying the skin lesion. Additionally, Farzam Kharaji, Nezhadian, and Saeid Rashidi proposed a new algorithm to categorize the dermoscopic images into malignant and benign. The snake’s Algorithm method was used for segmentation purposes. From the segmented image, the color features and textural features were extracted. Texture-based features were analyzed first and it wasseen that this method was highly efficient. By the use of the support vector classifier, they managed to get an accuracy of 97% (as was mentioned in the dataset of international skin imaging collaboration)[6].

M.Chaitanya Krishna, S.Ranganayakulu and Dr. P. Venkatesan proposed a methodology through the clustering technique and feature extraction of skin cancer. This Lesion Image analysis tools checked for the various Melanoma factors Like Asymmetry, Border, Color, Diameter (ABCD), etc. This was done by a surface, estimate and shape examination for image segmentation and highlight. The mined feature parameters are used to classify the image as Normal skin or Melanoma cancer lesion[7].

Pratik Dubal, Sankirtan Bhatt, Chaitanya Joglekar and Dr. Sonali Patil proposed a new procedure for Skin cancer detection and classification. The images are segmented, and features are extracted using ABCD rule. A neural network is used for the classification of the lesion to attain a higher degree of accuracy. This trained Neural Network achieved an overall classification accuracy of 76.9% on a dataset of 463 images (divided into six distinct classes)[8].

Abdol-al-Jalil Et Al identified skin cancer using the neural network. He used artificial intelligence and image processing techniques for the identification of skin cancer. The images were taken through the preprocessing methods by various image enhancement techniques in order to remove any undesirable objects and noises. This led to improved image quality. 2-D wavelet transformation was used to extract the features from the segmented image. Researchers used the extracted features as the input to the neural network and backpropagation neural network was used to classify the images into those that have the presence of melanoma and those which don’t[9].

Hamd and Asa predicted skin cancer melanoma by treating the pigment. A digital method based on symmetric color pigments of the skin lesion was used. For additional processing, the edge of the lesion was identified and segmented. The symmetry level of all the images was then calculated to identify the category of the tumor. The images of the skin lesions were classified into three groupings that were Melanoma, Basal cell carcinoma, and Squamous cell carcinoma. The accuracy of the results after classification using the first method is 80% and using the second method is 92.5%. The results can be increased if the segmentation of the skin lesion is increased[10].

3. PROPOSED METHODOLOGY

Block Diagram:

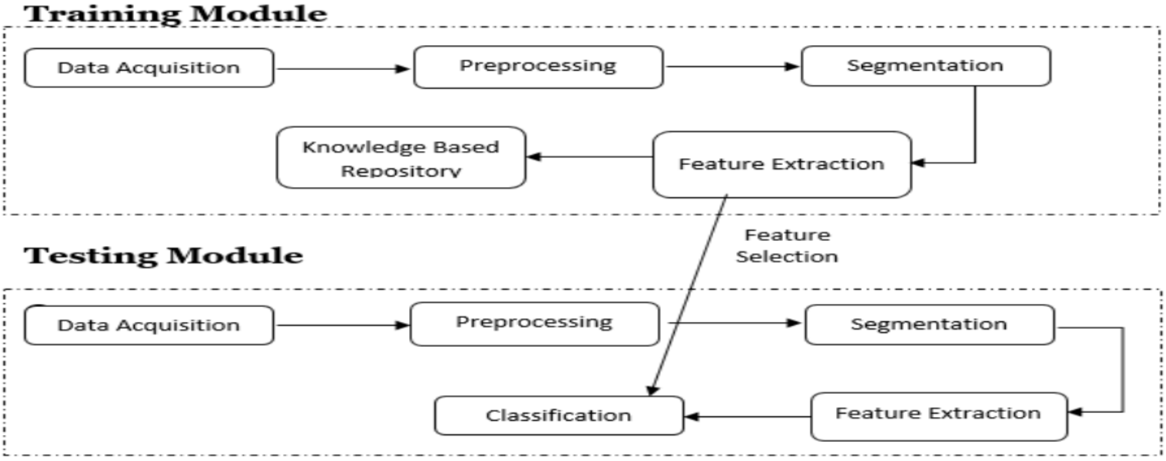

Fig. 1. Proposed research model – block diagram.

Stage 1 - Data Collection

This stage includes collecting the image data from ISIC archives and other archives. Also there is a need to check whether the collected images are a combination of benign and melanoma cells. Digital images obtained using photo dermatoscope have adequate resolution that allows for a precise analysis in terms of differential structures appearance[11]. A dermatologist can make exact documentation of the accumulated images, opening a way for PC investigation, where images are regulated so as to separate data that can later use to order those images.. The dataset flooded in the ISIC archive has been used by many researchers for developing and proposing various models to identify the cancerous cells.

Stage 2 - Pre-Processing

Pre-processing ought to be performed on every one of the images procured. This movement is applied so as to ensure that every one of the images are dependable in the ideal trademark. The collected dermoscopic images have to be pruned for noise removal. The generic procedures that are highly applied through the image pre-processing technique, that are of conversion of the image from the color format to a grayscale image, which makes the extraction of needful features from the converted images, a helping hand for further segmentation proceedings. Pre-processing techniques such as resizing, equalization and morphological processing can be applied[12]. So implying more than one pre-processing technique over the collected image generally helps to classify the lesion either as melanoma or benign.

Stage 3 - Segmentation

Segmentation helps to segment the region of interest from the pre-processed image. There is continuous regression of segmentation algorithms that have been imposed on the images, to acquire a proposed model. The algorithms such as the Snakes algorithm, Marker controlled watershed algorithm and a combination of both algorithms. Representing the individuality among the algorithms gives better refined segmented regions, but the proposed model implies to attain a more refined region from the hybridization of the algorithms. The generic ideology behind using the segmentation technique is to achieve the border of the skin lesions[13].

Stage 4 - Feature Extraction

Feature extraction is considered as the crucial step that helps towards the classification. Based on the extracted features the classification of the images is made out to rectify whether the lesions are cancerous or non-cancerous[14]. The features can be extracted using geometrical parameters. The geometrical parametric features include:

- Area
- Major Axis
- Minor Axis
- Roundness.
- Perimeter.
- Totality
- Eccentricity
- Length.
- Compactness.

Stage 5 - Classification

Classification plays a major role in identifying the lesions by categorizing the extracted regions[15]. The various classification algorithms being used are SVM, K-NN and decision tree algorithms. The three algorithms have been used to obtain a well precise accuracy from the model. The models are being constructed in such a way, that each model uses three classification models in the testing phase. The more accurate value is obtained from the classifier which helps for easy identification of the lesions from the affected regions of the skin.

Stage 6 - Knowledge Repository

A knowledge-based repository is constructed from the features being extracted. These features are being stored as knowledge for further data interpretations and enhancements. These values are being cross verified with testing parameters chosen from the single images. The Knowledge base helps to improve the efficiency of the features being extracted. Repository helps for easy identification of the image.

4. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the proposed technique is explained in detail. The model briefly works on the two segmentation techniques and the hybrid method of those algorithms, which marked as a third proposed algorithm. The segmentation techniques are followed by feature extraction and classification algorithms. The classification algorithms considered are of three algorithms, named K-NN, SVM, and Decision tree algorithms. Each classification algorithm has been used on the extracted features to obtain a knowledge base with higher accuracy level, the more accurate classifier is chosen for the identification of benign cells. There are similar techniques being used to find out the cancerous cells but hybridized model are a handful.

Method 1:
Snakes Algorithm and Decision tree algorithm.
Snakes algorithm segments an image into the foreground and background. Snakes algorithm is considered as the best method for shape recognition. The snake's algorithm can accurately detect the boundary of an object during the image analysis[16]. The mask used in the segmentation model will return us the initial contour location and based on the location, the boundary of the lesion is identified and segmented. The decision tree has been widely used for problems like classification and regression in data mining. As the name suggests the decision tree will make a hierarchical model according to some predicted outcomes. From the input data, it makes certain conditions to move on. The conditions will be on a yes/no basis "if this occurs then this occurs". While training a decision tree the output generated rules [17]. The algorithm forms a tree that can be linearized into different rules. In a decision tree, each node represents an attribute from this attribute rules are generated and each outcome can be considered as a leaf. The features are split into different branches, it includes the root node and child node. Splitting of branches is dependent on certain rules formed by the algorithms. The dataset which we have consists of 9 attributes excluding the target attribute.

Table 1. **This** is the features extracted from the segmented image using Snakes algorithm.

Area	Mj.Axis	Mn.Axis	Eccentricity	Perimeter	Roundness	Totality	Length	Compactness
402.13	2.5471	78498	0.7738	1.5057	0.4350	0.2670	1.5787	2.2984
3.8953	2.7056	79998	0.7194	1.5732	0.4061	0.2475	1.4397	2.4621
2.9403	2.6553	60162	0.4294	1.2683	0.4699	0.2318	1.1073	2.1278
2.3241	1.9975	35512	0.5112	8.6682	0.5939	0.2681	1.1635	1.6837
2.9413	2.4931	56469	0.5305	1.4848	0.3218	0.1980	1.1797	3.1068

The accuracy obtained from the classification is 80%.

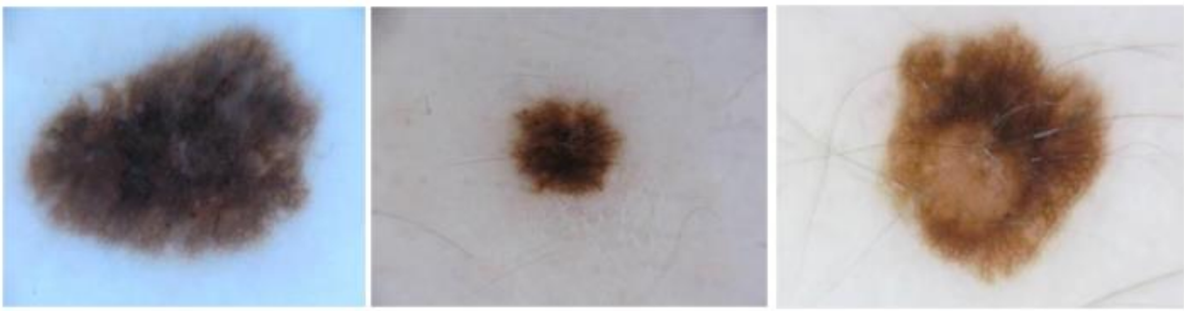

Fig. 2. Sample Dataset of Benign cells used for Snakes Algorithm.

Fig. 2.1. Segmented Image obtained from Sample Dataset of Benign cells used for Snakes Algorithm.

Method 2:
Marker controlled Watershed algorithm and K-NN Classifier
The marker-controlled watershed division has been demonstrated to be a hearty and adaptable technique for the division of articles with shut forms, where the limits are communicated as edges[18]. The marker pictured utilized for the watershed division is a paired picture comprising of either single marker focuses or bigger marker areas, where each associated marker is set inside an object of intrigue. Each underlying marker has

a coordinated relationship to a particular watershed area, in this way, the number of markers will be equivalent to the last number of watershed locales.

After division, the limits of the watershed districts are masterminded on the ideal edges, in this way isolating each article from its neighbors. The markers can be physically or naturally chosen, yet high throughput analyzes frequently utilize consequently produced markers to spare human time and assets. We have used the k-nearest neighbor approach as a classifier in this work. A dissent is gathered by a bigger part vote of its neighbors, with the inquiry being doled out to the class which is most ordinary among its k nearest neighbors. The motivation for this classifier is that models that are close to one another in the component space is most likely going to have a spot with a comparative model class.

The neighbors are taken from a course of action of tests for which the correct portrayal is known. It is customary to use the Euclidean partition, anyway other division measures, for instance, the City square, Cosine detachments could be used. In this work we have used three different partition measures viz., Euclidean, City peace and Cosine expel measure to focus the effect on the course of action exactness.The dataset which we have consists of 9 attributes excluding the target attribute.

Table 2. This is the features extracted from the segmented image using Marker controlled watershed algorithm.

Area	Mj.Axis	Mn.Axis	Eccentricity	Perimeter	Roundness	Totality	Length	Compactness
4.7476	4.5181	166888	0.3071	3.4914	0.1720	0.1359	1.0507	5.812
4.6728	4.2344	150410	0.4228	3.3234	0.1711	0.1406	1.1035	5.843
4.8879	4.1414	155310	0.5311	3.4415	0.1647	0.1420	1.1802	6.068
3.898	3.0985	92620	0.6068	1.9408	0.3089	0.2008	1.2581	3.236
3.3772	2.4803	65403	0.6786	1.2560	0.5209	0.2688	1.3616	1.9195

Fig. 3. Sample Dataset of Benign cells used for Marker Controlled Watershed algorithm.

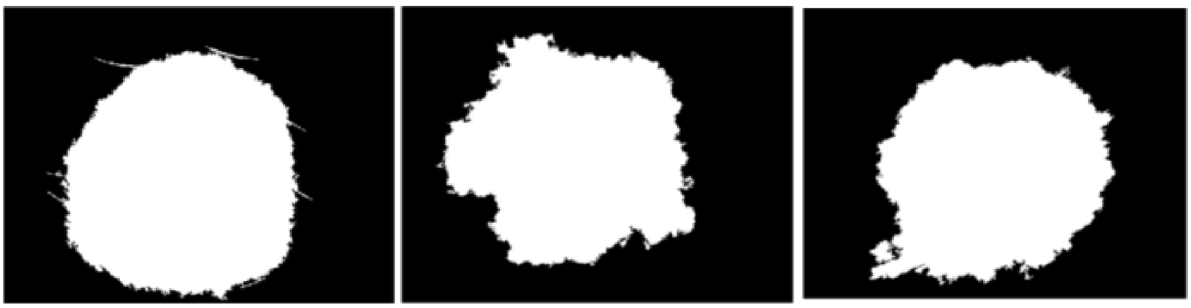

Fig. 3.1. Segmented Image obtained from sample dataset of benign cells used for Marker Controlled Watershed algorithm.

The accuracy obtained from the classification is 90%.

Method 3:
Hybridization algorithm and K-NN Classifier

The above method is considered as the proposed method. The method combines two (2) segmentation algorithms named, Snakes algorithm and Marker controlled watershed algorithm. The hybridized result helps to identify the cells from the images. The method follows a procedure of creating a knowledge base, and the single image data are being tested with the knowledge base. If the tested image values fall under or over the boundary with the values of the knowledge base, then the classifier helps to give a categorization of whether the values are of cancerous cells or non-cancerous data. The Segmentation method follows the looping structure initially, it goes through the snake's algorithm followed by marker controlled watershed algorithm. When the looping is being carried out the features are being extracted. These features are processed with the help of python programming to classify the data. Once classified the accuracy of the extracted features is being taken into consideration. These accurate values determine the efficiency of the model.

In the proposed model K-NN marks the more accurate value among the other two (2) classification algorithms. The difficulty being associated with the hybridization was the continuous monitoring of the segmentation technique. If images are being excluded from the expected regions, those images are pruned out. The data set comprises 110 images which are of a combination of melanoma as well as benign cells. Comparing other models the hybridized model stands out by giving an accuracy level of 86% from 110 images. Consider that three (3) images are taken into consideration from the dataset and the features extracted from the image are of six (6) which belongs as the feature of two (2) segmentation algorithms.

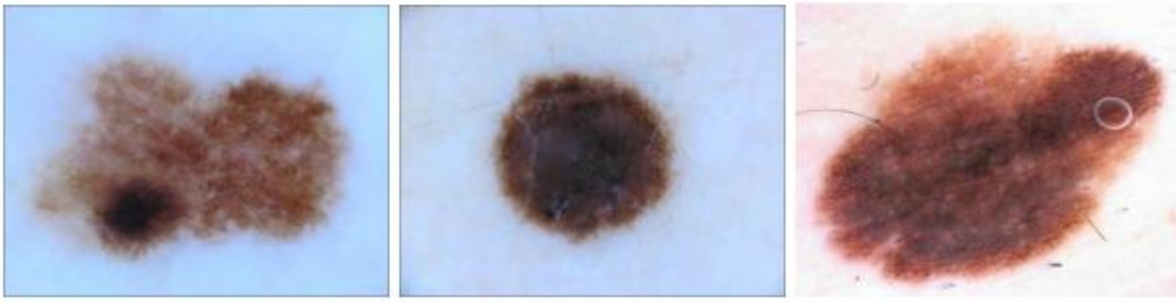

Fig. 4. Sample Dataset of Benign cells used for Hybrid algorithm

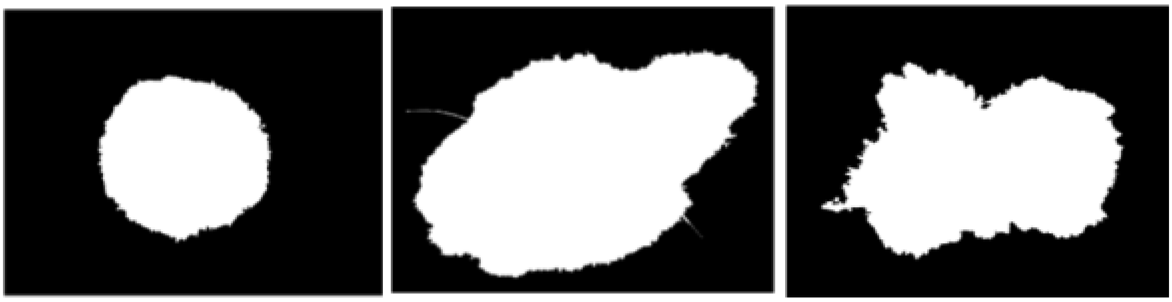

Fig. 4.1. Segmented Image obtained from sample dataset of benign cells used for Hybrid algorithm.

The features extracted, generates a knowledge base and the knowledge base made to test with other respective images to check the accuracy of the model. If the values fall under or over the knowledge base the model attains the correctness in identifying the benign cells. The dataset which we have consist of 10 attributes including the target attribute.

Table 3. This is the features extracted from the segmented image using Hybrid algorithm.

Area	Mj.Axis	Mn.Axis	Eccentricity	Perimeter	Roundness	Totality	Length	Compactness	Target
665.90	595.61	307833	0.447	2909.2	0.457	0.228	1.117	2.188	0
52.89	21.507	380	0.9135	134.07	0.265	0.3945	2.459	3.7641	0
206.71	141.70	22846	0.7280	793.667	0.455	0.2604	1.458	2.1941	1
106.84	80.40	6318	0.658	476.24	0.350	0.224	1.328	2.85	1
440.31	302.54	103208	0.726	1738.5	0.429	0.253	1.455	2.330	1
613.48	577.12	274753	0.3391	2604.68	0.508	0.2355	1.063	1.9649	0

The above values form a knowledge base, these values are being tested with a single image feature. The values match or in a boundary with the knowledge base it categorizes the respective data as either benign or melanoma.

5. CONCLUSION

In this paper, we have proposed a hybrid segmentation algorithm as a conclusion to cancer instigating cells, by means of the prevailing data of different reported patients. The model followed an approach of creating a knowledge base repository by visiting various stages, such as pre-processing, segmentation, feature extraction, and classification. This knowledge base has been tested with single images for easy identification of cancerous cells.

The proposed method follows a variant in pre-processing and opts for a hybrid segmentation technique. To have better clarity the images are being tested with melanoma images and through the finest classification, only benign cells have been separated out from the considered data set of images. Proposal for implementation is carried out with three classification algorithms and a hybrid segmentation algorithm. The best accuracy obtained through a process of hybrid segmentation is achieved with the help of a K-NN classifier. It provides 86% of accuracy in predicting the cells found are benign cells or melanoma cells. The error may depend upon the importance and contribution of each feature extracted from the images. In the future, improvement in dermatoscopic capturers may help for the increased accuracy.

References:

1. Farzam Kharaji Nezhadian and Saeid Rashidi, "Melanoma skin cancer detection using color and new texture features", artificial intelligence and signal processing.
2. M.H.Hamd,K.A.EssaandA.Mustansirya,"SkinCancerPrognosisBasedPigment Processing", International Journal of Image Processing, vol. 3, no. 7, pp.227,2013.
3. M.Chaithanya Krishna 1 , S.Ranganayakulu 2 , DR.P.Venkatesan3, "Skin Cancer Detection and Feature Extraction through Clustering Technique Vol. 4, Issue 3, March 2016"
4. Prashant Bhati, Manish Singhal, "Early Stage Detection and Classification of Melanoma"
5. Pratik Dubal, Sankirtan Bhatt, Chaitanya Joglekar, Dr. Sonali Patil, "Skin Cancer Detection and Classification"
6. J.A. Jaleel, S. Salim and R.B. Aswin, "Artificial neural-network-based detection of skin cancer", International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering, vol. 1, no. 1, pp. 200-205, 2012.
7. An Introduction to Digital Image Processing with Matlab Notes for SCM2511 Image Processing 1 Alasdair McAndrew School of Computer Science and Mathematics Victoria University of Technology.
8. Sagar Adatrao, Mayank Mittal, "An Analysis of Different Image Preprocessing Techniques for Determining the Centroids of Circular Marks Using Hough Transform".
9. R.M. Nishikawa, M.L. Giger, K. Doi, C.J. Vyborny and R.A. Schmidt. 1995. "Computer aided detection of clustered microcalcifications in digital mammograms," Med. Biol. Eng. Comp. Vol. 33, pp. 174-178.
10. Chandrahasa M, Varun Vadigeri and Dixit Salecha,"Detection Of Skin Cancer Using Image Processing Techniques " International Journal of Modern trends in engineering and research
11. DermLite, 2017a. https://dermlite.com/. (Accessed 20 July 2017).
12. Erdem Okur a, Mehmet Turkan, "A survey on automated melanoma detection" Engineering Applications of Artificial Intelligence 73 (2018) 50–67.