Identification of dermatological disease by using Fuzzy logic and Evolutionary Algorithm

T.D.Srividya
Ph.D Research Scholar
Tiruppur Kumaran College for women
Post box no.18S.R.Nagar, Mangalam Road,
Tiruppur - 641687

Dr. V. Arulmozhi
Associate professor
Tiruppur Kumaran College for women
Post box no.18S.R.Nagar, Mangalam Road,
Tiruppur - 641687

Abstract: Skin cancer detection is a very crucial task. A process of selecting a subset of minimum, significant features is Feature subset selection, which is said to be a pre-processing task in many applications. A Compact set of features help to make the patterns easier to understand. An attempt to find the feature subset for optimal clusters is employed in this paper. The Fuzzy C-Means clustering algorithm finds the centroid of the cluster groups. The histogram equalization calculates the intensity values of the grey level images and decomposition of images are done by PCA. By applying Fuzzy C-Means clustering and histogram, the detection of skin cancer from lesion area is carried out in this work. The PCA is used to reduce the dimensionality of the wavelet coefficient. For this two different genetic search techniques, CFC and CF/RSC are applied. The projected Automatic Feature Subset Selection (AFSGA) with Genetic Algorithm decreases the computational cost to estimate a good cluster and automatically recognize the required features. The computational cost is estimated by the experimental results which show improvement in the efficiency of the algorithm with optimal clusters.

Keywords: Fuzzy C-Means, Clustering, Histogram Equalization, Segmentation.

I. INTRODUCTION:

The skin being the outermost layer protects us from heat, sunlight, injury, and infection. Skin stores water and fat also helps control body temperature[1]. The most common type of cancer nowadays is skin cancer due to overexposure to sunlight[2]. The fair complexioned inhabitants have been liable for melanoma skin cancer. If diagnosed early, it is medicable else leads to death. The prominent source of skin cancer is exposure to UV radiation[2].
Image processing techniques are used nowadays to detect skin cancer in recent years. Image processing plays a fundamental role in producing digital image with a good contrast images which are a strong requirement in medical field[3]. Clustering is an unsupervised process of grouping objects into classes of similar objects[4].

Feature subset selection is the process of selecting a subset of features d, from the larger set D features, to maximize the classification performance over all possible subsets[5].
The objects need to be differentiated based on similar characteristics. Searching the accurate subset of features is a difficult search problem.

This paper focuses on the problem of feature subset selection using two genetic-based search algorithms. An advanced evolutionary algorithm, CHC for solving the 3D image registration problem is employed in this work[8][15].

A cluster is a group of objects with high similarity and dissimilar to the objects belonging to other clusters.The new scheme will be affected by noise. using different shapes Results also suggest that CHC is better when used for problems with noise added to the features[17].

The paper is organized as follows. First, a literature review of the material involved in the research is presented. Second we present a description of the proposed methodology. Third an overview of the existing work followed by a conclusion.

II. LITERATURE REVIEW:

Munya A Arasi et al proposed a survey paper in which the classification methods such as K-NN, ANN & SVM are analysed .Proximal Support Vector Machine(PSVM) was adopted in this work. The PNN & Neuro-Fuzzy system was proposed here.

M.Chaithanya Krishna et al proposed a paper in which preprocesssing is done using image illumination equalization, color range normalization & image resolution normalization. The segmentation techniques of a threshold based , clustering techniques & edge-detection based methods are used. For signal discontinuities, the step detection & change discontinuities are used.The TDS calculation formula is given in this work.

A S Deshpande et al put-forth a methodology in which the image is pre-processed using a median filter for noise removal. For segmentation, Fuzzy C-Means is used. GLCM is used and classification is done by SVM. He concludes that SVM always correct.

Ruchika Sharma et al proposed segmentation methods such as edge detection, thresholding, region based, based on clustering. The unsupervised learning algorithms such as K-Means clustering & Fuzzy C-means are used. ANN segmentation also used. This work shows the comparison between all available segmentation techniques.

Christopher Baldassano et al paper uses Network clustering here. Two reverse inference meta-analyses were identified. Based on data-driven connectivity analyses and analysis a unifying framework for understanding the neural systems involved in processing both visual and non- visual properties of natural scenes.

Nishima Sachdeva et al analyses the ABCD feature extraction and used Otsu segmentation and PCA method. The proposed scheme uses wavelet transformation for image improvement, denoising and histogram analysis. In this paper, GA are used to simultaneously select significant features as input to ANN and automatically determine the optimal number of the hidden node. Here classification is done by PCA method.

M. Vimaladevi et al imposes the feature selection using 2D Wavelet transformation. The existing system uses parallel hierarchical and median filter and K-NN classifier.

Douglas Zonger et al proposes a methodology to evaluate the quality of feature subsets generated by various algorithms and compare their computational requirements. The SFFS algorithm is used. Various algorithms are compared. For pattern classification, a multilayer feedforward network with a back propagation learning algorithm is used. Various algorithms for feature selection are compared in terms of classification error and run time on a 20-D, 2-Class Gaussian data. A total of 13 feature selection algorithms are listed and applied to SD-3 hand printed character set.

Mineichi Kudo et al developed a feature selection based on reducing the cost of extracting features and improvement of the classification accuracy of a practical classifier. The algorithms used are SFS, SBS, GSFS(g), GSBS(g).
The GA methods of mutation and crossover used. PARA- A parallel algorithm devised to compare with GA. He studied that the criterion curve of the algorithm is possible only when the algorithms are sequential. It is concluded that SFFS and SBFS are effective for small, medium problems. GA is well suited for finding an optimal solution, BAB+ is very efficient in monotonic problems and by using RARB there is a high possibility to find the best solution in appropriate monotonic problems.

John P.Eakins et al proposed an automatic scene analysis. A model based on a statistical approach to object classification and adaptive learning from user feedback. A Content-based image retrieval(CBIR) is done using PCA.
He concluded that the image retrieval a semantic level can be achieved only by reference to some KB of prior experience.

Ebtihal Aalmansour et al discusses the death rate of melanoma is three times than other cancers. The proposed method uses two types of texture feature and compared it with the state of the art method. Four color feature formulas are given. GLCM and SVM classifiers are used. The dataset contains 69 dermoscopic images, 43 melanoma images and 26 non-melanoma images collected from Dermatology Information System(Dermis). He concludes that color is very important to distinguish melanoma and non-melanoma.

Sonali et al combined thresholding segmentation techniques to establish boundaries in segmentation technique with Fuzzy C-Means segmentation.

Hina Sood et al devised a method for merging of segmentation using a Genetic algorithm. The lesion segmentation is compared with other algorithms. Various segmentation methods employed here are adaptive thresholding, Fuzzy based split and merge, Gradient Vector Flow(GVF) and Expectation Maximization Level (EM-LS) .He also concluded another approach to detect skin lesion is to find out the best thresholds with the help of mutlilevel adaptive thresholding. The formulas for sensitivity, specificity, accuracy are given and he concluded by providing a unique algorithm.

O.Cordon et al proposes the concept of CHC, to solve the 3D image registration problem. In image, registration mapping is carried out between two images both spatially and with respect to intensity. The most known feature- based algorithm for IR is Iterative Closest Point(ICP). The results are obtained by two CHC variant .

Sri Krishna et al introduce the feature subset for optimal clusters with Automatic Feature Subset Selection(AFSGA) using GA. Clustering is classified as hierarchical and partitional. This algorithm deals with selection of optimal initial seeds and deals with the process of feature subset selection while clustering by selecting CS measure as the fitness function. A new GA based wrapper feature selection method for classification of hyperspectral image data using SVM is proposed. The AFSGA methods and procedures are given. In the result, the AFSGA is compared with the classical clustering algorithm. The results are demonstrated to improve the efficiency of AFSGA.

Cesar Guerra-Salcedo et al deals with object classification. The previous study was based on decision tables as accurate as C4.5 for classification purposes. Two different genetic search techniques such as CHC and CF/RSC are used. In this paper, he introduced the feature subset selection and case-based classification. The bit climber procedure is given, which is successfully applied for better results. The CF/RSC provides good feature subsets less than 50% of time completed by CHC.

S. Gopinathan et al proposes an Otsu segmentation methodology that segments the lesion from the entire image. For further segmentation, the Boundary tracing algorithm is used. For classification, Stolz algorithm is used and results are presented in the form of tables 7 graphs. The filters used here are Gaussian noise with standard deviation of 0.5.

Sivaramakrishnan et al develop the existing method based on K-NN. In this work, the detection of brain tumor region is accompanied by Fuzzy C-Means clustering and histogram. The histogram equalization calculates the intensity values of gray level images and decomposition is extracted using PCA to reduce the dimensionality of the wavelet coefficient. The Fuzzy C-Means clustering algorithm finds centroids of cluster groups. The histogram equalization is mainly to suspect the tumor. In the discrete wavelet transform two quadrature mirror filters are used. The Fuzzy C-Means is used for simplicity and for faster clustering.

Ramjeet Singh Yadav et al gives a comparative analysis of K-Means, Fuzzy C-Means, Subtractive clustering, hybrid subtractive clustering. The K-Means clustering algorithm is given. For all of the above soft computing techniques equations are provided.

III. PROPOSED METHODOLOGY:

A. Feature Subset Selection:

Feature subset selection is the process of identifying a subset of features by removing redundant features. It is the process of selecting subset of features sufficient to describe the target[3]. The process of feature selection also reduces the Dimensionality of data sets. A good feature set contain a minimum subset of features able to model the target that contains a highly relevant feature to improve the efficiency of the classification algorithm. Feature selection plays a crucial role in several fields to create a model.[6]

The purpose of feature subset selection is to discover a minimum set of features such that the resulting chance of allocating the data classes is as close as possible to the original distribution obtained using all features [16]. Removal on a reduced set of features has an added advantage. It helps to minimize the number of features to make the patterns easier to understand [17].

1. Segmentation based on clustering:

Segmentation is the process of dividing an image into its constituent images. It is considered a vital task in image analysis in which each and every pixel contain different attributes[4][5]. Clustering methods are generally unsupervised, used to classify the data into a group, based on resemblances of individual data items. The major goal of clustering is compressing data, gaining insight into data which includes identifying essential features, detecting anomalies, classifying data and clustering by Fuzzy C-Means[3][4].

The clustering algorithms considered previously are not based on assumptions common to conventional statistics methods. The two well- known methods are Hierarchical and Partitioning[16]. The first one builds the clusters by repeatedly partitioning the objects and the latter one divide a dataset with or without overlap.

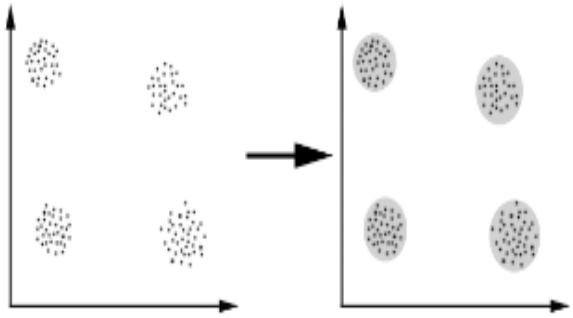

Fig. 1 Clustering

2. Fuzzy C-Means Clustering:

FCM clustering allows data to belong to more than one cluster[3]. Commonly used in soft computing and image processing[4]. Fuzzy C-Means is a clustering algorithm for membership point in order to determine data at each location that belongs to a certain cluster. FCM separate the N number of vectors by initializing i=1 up to n to make it into a C fuzzy group.

And compute the cluster center of each group making the value of the non-similarity index to achieve the maximum[14]. Fuzzy C-Means(FCM) is a technique which allows data to belong to more than one cluster, where m is any real number greater than 1,

Uij – degree of membership of xi in cluster j

Xi – ith of d-dimensional data which is measured data

Cj – the d-dimension center of cluster and

||*|| - any rule expressing the similarity between measured data and the centre.

Depending on the results of fuzzy the FCM algorithm obtain segmentation.

Various segmentation algorithms based on fuzzy are included in the literature.

3. Histogram Equalization:

A new innovative method of increasing the dynamic range of an image histogram is achieved by histogram equalization[16]. For an input image, it assigns the intensity

values of pixels so that the output image contains uniform intensity[19].
If the image histogram contains many peaks and valleys, it will be shifted after equalization.

4. Genetic algorithm:

Among the various categories of feature selection algorithm the genetic algorithm, is a recent innovation. Genetic Algorithms (GAs) are stochastic search mechanisms based on natural selection concepts [14]. The genetic algorithm approach for feature subset selection appears first in 1998. The GA is biologically stimulated evolutionary algorithm[16].Since the search problem is a subset of features, GA is applied to the feature selection[13]. Many works of literature were published illustrating the advantages of GA for feature selection

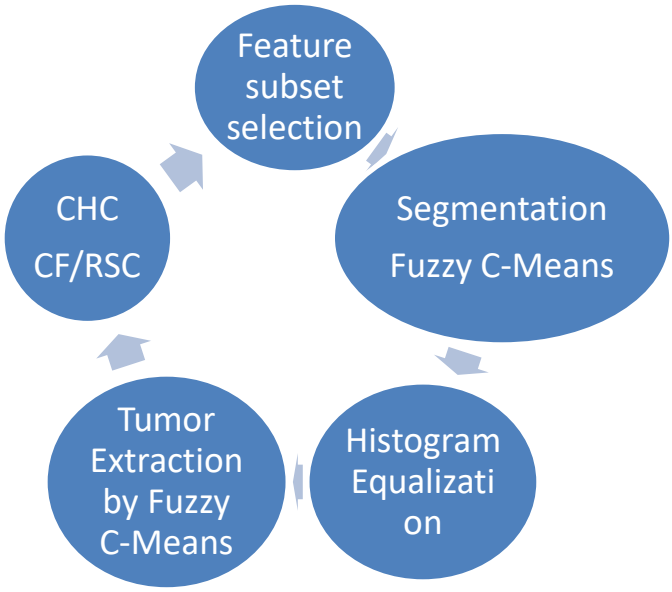

Fig. 2 Overview of the proposed system:

B. Numerical Illustration:

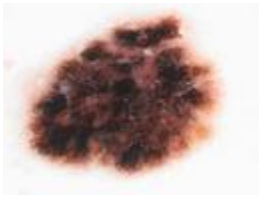

Fig. 3(a) Benign Image

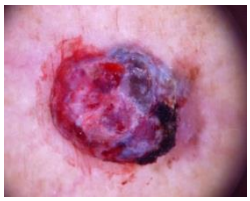

Fig. 3(b) Malignant Image

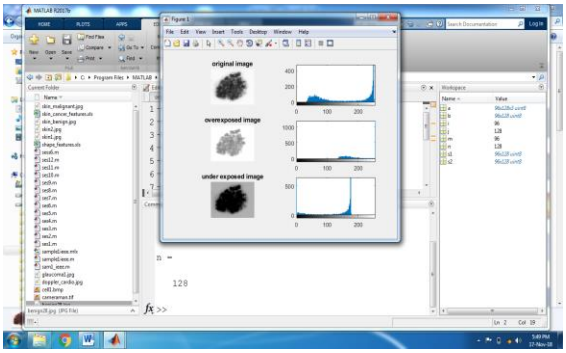

Fig. 3(c) Input Image, Overexposed, Underexposed with corresponding Histograms

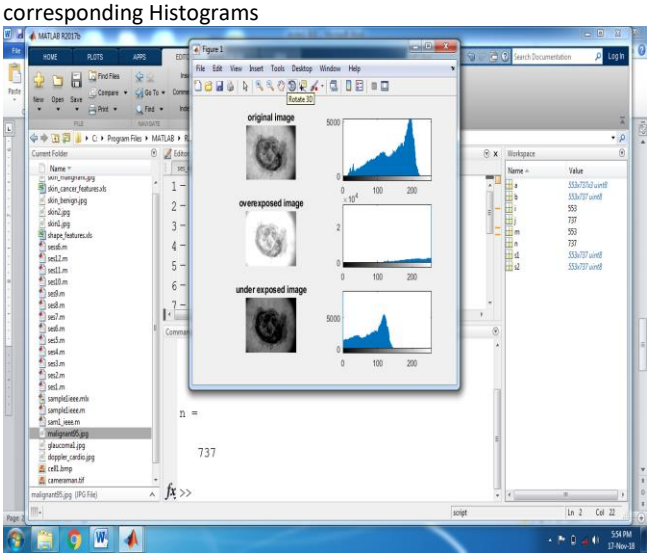

Fig. 3(d) Malignant Input Image, Overexposed, Underexposed with corresponding Histograms

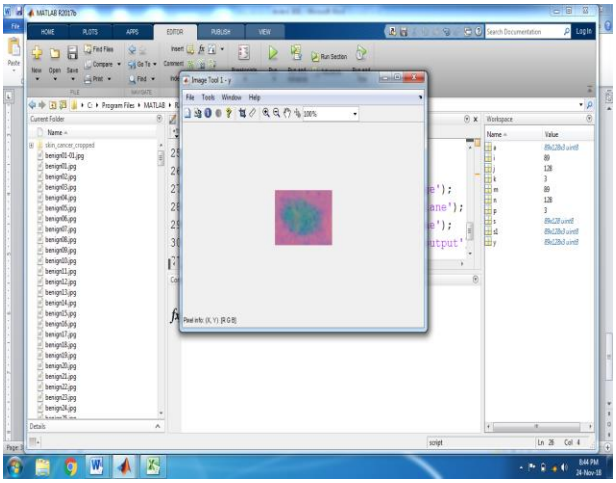

Fig. 4 Original image

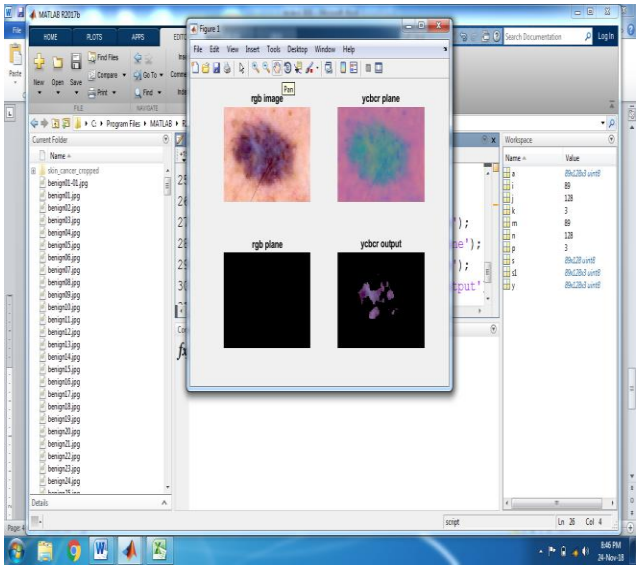

Fig. 5 Image in YCBCR Plane

A sample benign and malignant image is considered and the corresponding histogram is shown. A total of 20 samples are considered in which 10 are benign and 10 are malignant images. A total of 10 features are calculated for all the images.

Table 1 Skin cancer features

r1_start	r3_start	y_start	mean2(g)	std2(g)	g_start	Skew	kurt	g_high	en
0.77	0.43	0.98	0.41	0.59	0.51	0.4	0.24	0.59	1.7
0.78	0.42	0.82	0.34	0.55	0.49	0.37	0.22	0.15	1.45
1	0.82	1	0.57	0.52	0.45	0.67	0.19	0.37	1.63
0.73	0.55	0.67	0.41	0.5	0.47	0.47	0.17	-0.55	1.81
0.88	0.65	0.9	0.49	0.52	0.47	0.55	0.24	-0.28	1.51
1	0.97	1	0.93	0.51	0.5	0.94	0.15	-2.85	12.6
1	0.89	0.8	0.61	0.52	0.47	0.67	0.15	-1.59	4.87
1	0.87	0.74	0.43	0.5	0.43	0.58	0.14	-1.19	4.62
1	0.86	0.83	0.61	0.56	0.47	0.67	0.16	-1.52	4.8
0.72	0.65	0.46	0.35	0.45	0.42	0.52	0.05	-1.33	3.89
1	0.63	1	0.55	0.58	0.51	0.53	0.27	-0.72	2.6
0.68	0.55	1	0.64	0.65	0.54	0.56	0.15	-0.43	2.03
0.99	0.58	1	0.58	0.59	0.54	0.51	0.07	0.01	3.07
0.82	0.6	0.75	0.47	0.55	0.5	0.48	0.16	-0.46	2.06
0.96	0.48	0.73	0.3	0.52	0.47	0.36	0.16	0.67	2.47
0.87	0.44	0.6	0.21	0.51	0.45	0.3	0.12	1.53	4.62
1	0.61	0.84	0.39	0.52	0.45	0.49	0.2	0.34	1.84
0.83	0.44	0.58	0.22	0.51	0.46	0.31	0.1	0.27	2.58
0.83	0.44	0.58	0.22	0.51	0.46	0.31	0.1	0.27	2.58
0.95	0.5	0.73	0.3	0.51	0.46	0.38	0.16	0.79	2.37

The features calculated for the skin cancer images are r1_stat, r3_stat, y_stat, mean, standard deviation, g_stat, skewness, kurtosis, g_high, energy, entropy. The overall features include color and texture features.

C. Genetic Algorithm Review:

1. CHC:

CHC is a genetic search algorithm which uses truncation selection. The CHC algorithm indiscriminately pairs parents whose string pairs differ by some number of bits. This is said to be a mating threshold allowed to replicate. The threshold initially is set as l = 4, where l is the length of the string. The threshold is decreased by 1 when no child is inserted into the new population during truncation selection. In the succeeding generation when no offspring are inserted into the population and mating threshold reaches value 0, CHC injects new diversity into the population through a restart form known as cataclysmic mutation. Uniform crossover is performed here that haphazardly exchange half of the bits that differ between two parents.

Cataclysmic mutation uses the best individual in the population as a model to re-initialize the population [14]. The new population includes one replica of the template string; the remainder of the population is generated by mutating some percentage of bits (e.g 35%) in the template string.

2. Common Features/Random Sample Climbing (CF/RSC):

Another GA variant is CF/RSC which is a two-stage algorithm. It is usually employed for a population of small individuals. When recombining two individuals in the population CF is used. In case of a local improvement is needed for an operator Random Sample Climbing (RSC) is employed to heuristically improve the individual in the population. In generation 0, all those in the population are initialized to strings of all 0's. RSC is applied to all individual, where n bits are mutated randomly. If n is 10, for k times, thus producing k new solutions for each member in

the set. In the successive generations, recombinations are applied. When recombining two strings, the common selected features in two parents are passed to offspring. All other features are set to z.

GA maintains a set of solutions called population. Each solution vector is a chromosome. The process of picking individuals that survive for next generation is Biological evolution. The results of crossover and mutation genetic operations are considered as survival individuals of the fittest chromosomes.

Having found the candidate solutions (parents) the crossover takes place, where the parent's genetic information involved in generating new offspring (children individual) [14].

Agreeing on a small probability some new individuals will suffer mutation, and mutation operation is functional to the offspring. A new fitness value is intended to the individuals that underwent mutation. The process of calculating new offspring is continued until a stopping condition. The principle is a value of the fittest chromosome in the population or a maximum generation or lapsed processing time.

3. Automatic Feature Subset Selection Using Genetic Algorithm for Clustering:

Automatic Feature Subset Selection using Genetic Algorithm (AFSGA) recommends a Genetic Algorithm centered feature selection method for clustering data. This system is carried out in two steps. The optimal initial centroids using CS measure is carried out in the first step. The second step is to discover optimal clusters in selecting the features based on GA. For n element dataset, the CS measure randomly selects n/10 sets of centroids in the first step. The second step constructs a GA based algorithm to find the minimal set of required features for clustering[15].

IV. Conclusion:

Feature subset selection is the technique of selecting features depending on some criterion. In this paper, we propose a technique for evaluating the accuracy of a subset of features based on the probability of randomly generated subset. FCM approach is quite effective for color-based image segmentation. The GA technique of CHC and CF/RSC is proposed for image processing using Fuzzy C-Means clustering. The AFSGA used in this work is for minimum data and it can be extended for more datasets in the future.

REFERENCES:

1. Munya A. Arasi, 2El-Sayed A. El-Dahshan, 3El- Sayed M. El-Horbaty, 4Abdel-Badeeh M. Salem " Malignant Melanoma Detection Based on Machine Learning Techniques: A Survey" Egyptian Computer Science Journal (ISSN: 1110 – 2586) Volume 40 – Issue 03, September 2016
2. M.Chaitanya Krishna, S.Ranganayakulu, Dr.P.Venkatesan "Skin Cancer Detection and Feature Extraction through Clustering Technique" IJIRCCE vol 4, Issue 3, March 2016 (ISSN: 2320-9801)
3. A.S.Deshpande, Gajbar Aamruta M "Automated Detection of skin cancer and skin Allergy", IJARCSMS , vol 4, Issue 1, January 2016(ISSN:2321-7782)
4. Ruchika Sharma*, Dr. Pankaj Mohindru, Dr. Pooja Review of Segmentation Techniques for Melanoma Detection Volume 6, Issue 7, July 2016 ISSN: 2277-128X International Journal of Advanced Research in Computer Science and Software Engineering
5. Christopher Baldassano, Andre esteva, Li Fei-Fei, Diane M.Beck " Two Distinct Scene-Processing Networks Connecting Vision And Memory" doi:e0178-16.2016 1-14
6. Nishima Sachdeva(Mtech scholar), Professor Rohan Gupta "Hybrid Approach To Investigate The Probability Of Skin Cancer By ABCD And PCA Method" *ISSN: 2278 – 909XInternational Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 6, Issue 4, April 2017*
7. M.VimalaDevi, B.Kiruba "K-Nn Classifier For Skin Cancer Classification" (IJCSE)
8. Douglas Zongker, Anil Jain "Algorithms for Feature Selection: An Evaluation"
9. Mineichi Kudo1 And Jack Sklansky "A Comparative Evaluation Of Medium and Large-Scale Feature Selectors For Pattern Classifiers" KYBERNETIKA — VOLUME *34* (1998), NUMBER 4, PAGES 429-434
10. John P.Eakins "Towards Intelligent Image Retrieval" Pattern Recognition 35 (2002)3 }14
11. Ebthihal Almansour and M.Arfan Jaffar "Classification Of Dermoscopic Skin Cancer Images Using Color And Hybrid Texture Features" IJCSNS International Journal of Computer Science and Network Security, VOL.16 No.4, April 2016
12. Sonali Raghunath Jadhav, D.K.Kamat, "segmentation based detection of skin cancer", irf international conference, 2014
13. Hina Sood#1, Manshi Shukla*2, "Segmentation of Skin Lesions from Digital Images using an optimized Approach: Genetic A lgorithm" IJCSIT Vol 5 (5),2014, 6831-6837 ISSN:0975-9646
14. Ruchika Sharma*, Dr. Pankaj Mohindru, Dr. Pooja Review of Segmentation Techniques for Melanoma

Detection Volume 6, Issue 7, July 2016 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering

15. O.Cordon, S.Damas, J.Santa Maria "Feature-based image registration by means of the CHC evolutionary algorithm" Image and Vision Computing 24 (2006) 525–533
16. A. Srikrishna1, B. Eswara Reddy2, V. Sesha Srinivas1 "Automatic Feature Subset Selection using Genetic Algorithm for Clustering" Int. J. on Recent Trends in Engineering and Technology, Vol. 9, No. 1, July 2013
17. Cesar Guerra Salcedo, Stephen Chen, Darrell Whitley, Stephen Smith "Fast and Accurate Feature Selection Using Hybrid Genetic Strategies".
18. Dr. S.Gopinathan1, S. Nancy Arokia Rani2 The Melanoma Skin Cancer Detection and Feature Extraction through Image Processing Techniques- (IJETTCS) Volume 5, Issue 4, July - August 2016 ISSN 2278-6856
19. A.Sivaramakrishnan1, Dr.M.Karnan2 "A Novel Based Approach for Extraction of Brain Tumor in MRI Images Using Soft Computing Techniques" *International Journal of Advanced Research in Computer and Communication Engineering*
20. Ramjeet Singh Yadav1,*, P. Ahmed1, A. K. Soni1 and Saurabh Pal2," Academic performance evaluation using soft computing techniques"