

Automated System for Prediction of Disease of the Skin using Image Processing and Machine Learning

Chaitra T C
Dept. of CSE
GSSSIETW, Mysuru

Nisarga R
Dept. of CSE
GSSSIETW, Mysuru

Srushti N
Dept. of CSE
GSSSIETW, Mysuru

Sunku Vineela
Dept. of CSECSE GSSSIETW,
Mysuru

Rajath A N
Assistant Professor, Dept. of CSE
GSSSIETW, Mysuru

Abstract:- Human skin is that the largest organ in our body that provides protection against heat, light, infections and injury. It conjointly stores water, fat, and nourishment. Cancer is that the leading reason behind death in economically developed countries and also the second leading reason behind death in developing countries. carcinoma is that the most typically diagnosed kind of cancer among men and ladies. Exposure to ultraviolet illumination rays, modernize diets, smoking, alcohol and phytotoxin area unit the most cause. Cancer is more and more recognized as a important public ill health in African nation. There area unit 3 kind of carcinoma and that they area unit recognized supported their own properties. seeable of this, a digital image process technique is projected to acknowledge and predict the various kinds of skin cancers victimisation digital image process techniques. Sample carcinoma image were taken from yank cancer society center and DERMOFIT that area unit standard and wide focuses on carcinoma analysis. The arrangement was supervised love the predefined categories of the kind of carcinoma. Combining Self organizing map (SOM) and radial basis perform (RBF) for recognition and diagnosing of carcinoma is far and away higher than KNN, Naive Thomas Bayes and ANN classifier. it absolutely was conjointly showed that the discrimination power of morphology and color options was higher than texture options however once morphology, texture and color options were used along the classification accuracy was magnified. the most effective classification accuracy (88%, 96.15% and 95.45% for Basal cell malignant neoplastic disease, malignant melanoma and epithelial cell malignant neoplastic disease respectively) were obtained victimisation combining Kyrgyzstani monetary unit and RBF. the general classification accuracy was ninety three.15%.

Keywords— Skin diseases, pre-processing techniques, k-Mean, Gray Level Co-Occurrence Matrix(GLCM), Multi-Support Vector Matrix(SVM).

I. INTRODUCTION

Human malignant neoplastic disease may be a health problem that appears on the outer layers of the skin that unit caused once the skin cells unit dead or broken attributable to over exposure to Sun's actinic ray. but malignant neoplastic disease may occur on areas of one's skin not usually exposed to sunlight. The human skin is that the most important organ in our body that gives protection against heat, light, infections and injury. It to boot stores water, fat, and

ergocalciferol. The Human skin consists of two major layers cited as stratum and dermis. the best or the outer layer of the skin that {is cited as [is named] is termed} the stratum composed of three types of cells flat and scaly cells on the surface referred to as SQUAMOUS cells, spherical cells cited as BASAL cells and MELANOCYTES, cells that provide skin its color and protect against skin damage. The inner layer of the skin cited because the dermis is that the layer that contains the nerves, blood vessels, and sweat glands. There unit three style of malignant neoplastic disease melanoma, Basal cell malignant growth illness and vegetative cell malignant growth illness. malignant neoplastic disease is diagnosed by physical examination and diagnostic assay. diagnostic assay may be a quick and easy procedure where 0.5 or all of the spot is removed and sent to a laboratory. it ought to be done by doctor otherwise you're typically cited a specialist or sawbones. Results would possibly take a number of week to be ready. drugs imaging analysis believes that designation of malignant neoplastic disease area unit typically automatic supported an exact physical feature and color information that's that the characteristics of assorted category of malignant neoplastic disease.

Now a day's people unit laid low with skin diseases, over 100 twenty 5 million people laid low with disease to boot malignant neoplastic disease rate is quickly increasing over last few decades notably melanoma is most diversifying malignant neoplastic disease. mycosis rate is high notably at rural areas. If skin diseases are not treated at earlier stage, then it ought to cause complications inside the body in conjunction with spreading of the infection from one individual to the alternative. The skin diseases area unit typically prevented by investigation the infected region at associate early stage. The characteristic of the skin footage unit varied, thus it's troublesome job to arrange associate economical and powerful formula for automatic detection of the illness of the skin and its severity. Skin tone and color plays a awfully necessary role in illness of the skin detection. Color and coarseness of skin unit visually whole completely different. Key think about skin diseases treatment is early detection extra treatment reliable on the

primary detection.

Machine learning (ML) is that the study of laptop computer algorithms that improve automatically through experience. it's seen as a collection of AI. Machine learning algorithms build a mathematical model supported sample data, cited as "training data", thus on type predictions or picks whereas not

being expressly programmed to do and do thus. Machine learning algorithms unit utilised during a massive kind of applications, like email filtering and laptop computer vision, where it's difficult or not possible to develop normal algorithms to perform the specified tasks.

Machine learning is closely related to procedure statistics, that focuses on making predictions using computers. The study of mathematical improvement delivers ways in which, theory and application domains to the world of machine learning. processing may be a connected field of study, that focus on wildcat data analysis through unattended learning. In its application across business problems, machine learning is to boot cited as prophetic analytics.

The discipline of machine learning employs various approaches to help computers learn to accomplish tasks where no completely satisfactory formula is accessible. In cases where Brobdingnagian numbers of potential answers exist, one approach is to label variety of the correct answers as valid. this could then be used as employment data for the laptop to reinforce the algorithm(s) it uses to examine correct answers. as an example, to teach a system for the task of digital character recognition, the MNIST dataset has sometimes been used. There unit four machine learning ways in which supervised machine learning formula, Un-supervised machine learning formula, Semi supervised machine learning formula, reinforcement machine learning formula.

Image method may be a way to perform some operations on an image, thus on urge associate accumulated image or to extract some useful information from it. could be a} variety of signal method throughout that input is a image and output is additionally image or characteristics/features associated with that image. Nowadays, image method is among quickly growing technologies. It forms core analysis house among engineering and field of study disciplines too.

Image method primarily includes the following three steps unit commerce the image via image acquisition tools, Analysing and manipulating the image, Output throughout that result area unit typically altered image or report that is supported image analysis. There unit two types of ways in which used for image method notably, analogue and digital image method. Analogue image method area unit typically used for the arduous copies like printouts and footage. Image analysts use various fundamentals of interpretation whereas using these visual techniques. Digital image method techniques facilitate in manipulation of the digital footage by using computers. the three general phases that ever one types of data have to be compelled to endure whereas using digital technique unit pre-processing, improvement, and show, information extraction.

An image may be defined as a two-dimensional function, $f(x, y)$, where x and y are plane coordinates. The amplitude of f at any pair of coordinates (x, y) is called the intensity or grey

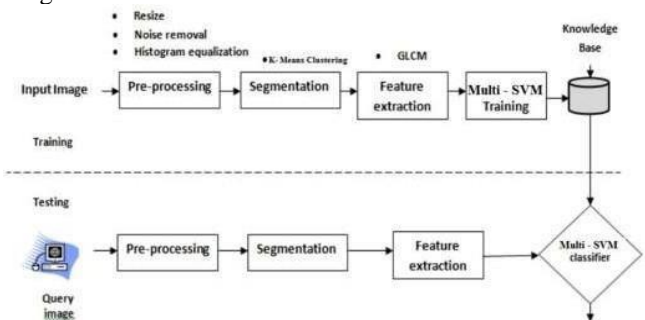
level of the image at that point. When x, y , and the amplitude values of f are all finite, discrete quantities, we call the image a digital image. The field of digital image processing refers to processing digital images by means of a digital computer. A digital image is composed of a finite number of elements, each of which has a particular location and value. These elements are referred to as picture elements, image elements, and pixels. Pixel is the term most widely used to denote the elements of a digital image.

II. REVIEW OF OTHER METHODS

In the past two decades, strong impulse has been given to developing automated systems capable of assisting physicians in medical imaging task [4].The paper proposed by Muhammad Zubair Asghar et al [5] presents an online expert system to detect certain skin diseases that make use of forward-chaining with depth-first search method. However, using a system based on set rules and symptoms is not feasible due to the various manifestations of a single skin disease. The system created by us would be superior in performance in such a manner as the issue we are addressing is probabilistic in itself and thus we require a system which has the ability to adapt and learn from the underlying pattern that exists in the skin disease which can be inferred by the image. Another system that was proposed by A.A.L.C. Amarathunga et al [6] introduces a data mining unit to the system of skin detection. However, it lacks the choice of attributes essential for detection. Neither the data source nor the attributes used for learning/testing have been mentioned. M. SHAMSUL et al [7] proposed a system that uses an image processing system with pre-processing algorithms and a feed forward neural network. Florence et al [8] categorize the image as a bacterial or viral skin infection using image processing techniques. Damilola et al [9] developed a system that assembles pigmented skin lesions image results, analysis, comparing observation and conclusions by medical experts using prototyping methodology. Our system makes use of image processing with pre-processing algorithms and feed forward back propagation method in artificial neural networks that are discussed in the following section.

III. PROPOSED ALGORITHM TOWARDS THE EXTRACTION OF TEXT

The system amalgamates 2 approaches that square measure image process and machine learning algorithms delineated in figure one.



Figure(1) : Proposed System

The first stage makes use of various image process unit to isolate and extract the region of interest that's the morbid

region from the image.

Initially those pre-processing algorithms were applied on color skin pictures and more image process algorithms were applied. The Second stage makes use of machine learning unit, the input pictures area unit fed to a man-made neural network to coach network associated predict an correct result. For all the method to require place we would have liked information within the kind of skin disorder image, notably of skin problem, psoriasis, impetigo, melanoma, and scleroderma. The coaching information set was obtained from.

LOAD THE IMAGE:

When you save your GUI layout, GUIDE mechanically generates associate M-file that you simply will use to manage however the GUI works. This M-file provides code to initialize the GUI and contains a framework for the GUI callbacks the routines that execute in response to user-generated events like a click. victimisation the M-file editor, you'll add code to the callbacks to perform the functions you would like. The system displays the house page with some choices. The user can browse for the input image. User clicks on LOAD IMAGE Button to transfer the image.

Consider the subsequent formula to load the image:

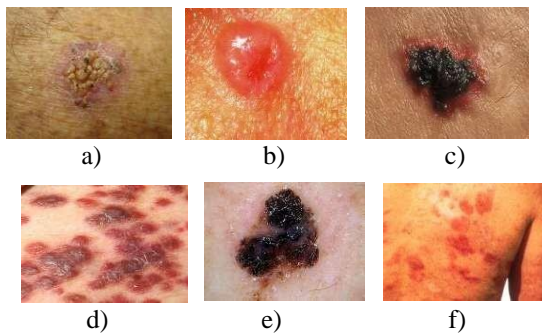
Algorithm: Load Image

Input: Path of the image wherever it's hold on in memory.

Output: Image in such path is loaded for more process.

1. Use "imread()" operate by providing path name as parameter to load the image.
2. Show the loaded image by victimisation "imshow()" operate as Figure one.

The sample information set that is employed in our project implementation is shown in below figure 2:



Figure(2) : Sample Datasets

PREPROCESSING:

The Dermoscopic Image in Digital format is subjected to varied Digital Image process Techniques. The standard image size is taken as 360x360 pixels. sometimes the image consists of noises within the kind of hairs, bubbles etc. These noises cause quality in classification. so as to avoid that, pictures area unit subjected to varied image process techniques. one in every of the key part in image process is filtering of image preprocessing is completed to removes the noise, fine hair and bubbles within the image. For smoothing image from noise, median filtering is employed. Median filtering may be a common step in image process. Median filtering is employed for minimizing the influence of tiny structures like skinny hairs and isolated islands of pixels like

tiny air bubbles.

SEGMENTATION:

K-mean is that the preferred partitioning technique of bunch. Kmean may be a unattended, non-deterministic, numerical, repetitive technique of bunch. In k-mean every cluster is described by the norm of objects within the cluster. Here we tend to partition a group of n object into k cluster so intercluster similarity is low and intracluster similarity is high. Similarity is measured in term of norm of objects in an exceedingly cluster. The formula consists of 2 separate phases.

1. Choose k center of mass at random, wherever the worth k is fastened earlier.
2. Every object in information set is associated to the closest center of mass. geometrician distance is employed tolive the gap between every information object and cluster center of mass.

The complete k-means formula of 2 phases is as follows: Algorithm:k-means

Input K: Variety of cluster (for dynamic bunch initialize the worth of k by two) fastened variety of cluster = affirmative or no (Boolean). D : an information set containing n objects. Output a group of k clusters.

Method:

1. At random opt for k information item from D dataset because the initial cluster centers.
2. Repeat
3. Assign every information item di to the cluster to that object is most similar supported the norm of the thing in cluster;
4. Calculate the new norm of the info things for every cluster and update the mean value;
5. Till no amendment.
6. If fastened variety of cluster = affirmative (Go to step 12)
7. Reason the inter-cluster distance victimisation equivalent weight (1)
8. Reason the intra-cluster distance.
9. If new intra-cluster distance previous inter-cluster distance goto step ten else goto step eleven.
10. k=k+1

11. Stop This formula offer optimum variety of cluster for unknown information set. The time taken by this formula for little information set is nearly same as customary k-mean formula however the time taken by dynamic bunch formula for giant data set is a lot of as compare to straightforward k- mean formula.

Output: a group of k cluster as per input technique

1. Randomly opt for k information item from D dataset as initial cluster center of mass
2. Repeat
3. Assign every information item d i to the cluster to that object is most similar supported the norm of the thing in cluster
4. Calculate the new norm of the info things for every cluster and update the norm
5. Till no amendment.

FEATURE EXTRACTION:

In feature extraction ways we tend to about to extract the

options supported the subsequent ways extraction: form Feature Extraction, Texture Feature Extraction, Color Feature Extraction.

1. Form Feature Extraction: form feature extraction utilized in this project is solidity, extent, axis length and eccentricity. This feature taken form feature extraction utilized in this paper are solidity, extent, axis length and eccentricity. These options taken from analysis so as to extract form feature in unhealthy region. Eccentricity is employed to acknowledge whether or not the rust form could be a circle or line section. axis length is employed to live length of axis of the unhealthy region. Extent is employed to live space of unhealthy region that's divided by the realm of the bounding box. Solidity is employed to live space of unhealthy region divided by pixels within the umbellate hull. it's computed by dividing {the space |the world |the realm} by umbellate area.

2. Texture Feature Extraction: grey Level Co-occurrence Matrix (GLCM) extract second order applied mathematics texture options. Texture feature extraction utilized in this analysis are distinction, correlation, energy and homogeneity. This options taken from analysis to extract texture feature in skin unhealthy region.

Contrast of the constituent and its neighbors is calculated over all of the image pixels. Contrast is employed to live distinction between neighborhood constituent.

$$f1 = \sum_{i=1}^k \sum_{j=1}^k (i - j)^2 P_{ij} \quad (1)$$

Correlation could be a live of correlation of a constituent with its neighbors over all of the image.

$$f2 = \sum_{i=1}^k \sum_{j=1}^k \frac{(i-mr)(j-mc)P_{ij}}{\sigma_r \sigma_c} \quad (2)$$

Energy could be a total of G (grey level co-occurrence matrix components).

$$f3 = \sum_{i=1}^k \sum_{j=1}^k P_{ij}^2 \quad (3)$$

Homogeneity computes similarity of G to the square matrix.

$$f4 = \sum_{i=1}^k \sum_{j=1}^k \frac{P_{ij}}{1+|i-j|} \quad (4)$$

All of All of the four options delineated during this section represent texture of the pictures of unhealthy region as compared with the conventional one.

3. Color Feature Extraction: Color could be a delicacy for image illustration that's invariant with relation to scaling, translation and rotation of a picture. Mean, lopsidedness and kurtosis are wont to represent color as options. To do this, we tend to remodel RGB to research laboratory.

Mean wont to represent average worth of every color channel.

$$\mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N P_{ij} \quad (5)$$

Skewness and kurtosis wont to live the distribution of every color channel. Skewness may be delineated as:

$$S = \frac{\sum_{i=1}^M \sum_{j=1}^N (P_{ij} - \mu)^3}{MN\sigma^3} \quad (6)$$

Skewness could be a live of symmetry. If a distribution or knowledge is parallel, it's an equivalent to the left and right of the middle purpose. Kurtosis will represent whether or not the info are peaked or flat relative to a standard distribution. Kurtosis may be delineated as follows:

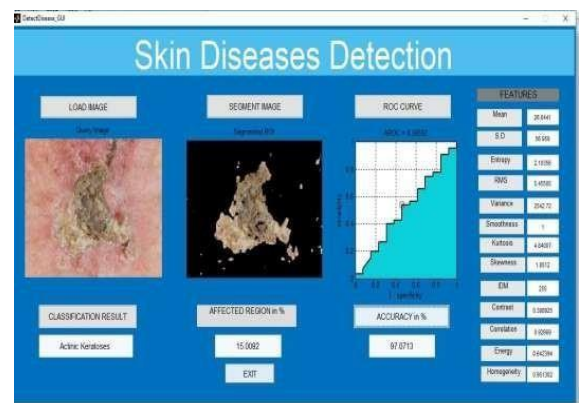
Combination of mean, lopsidedness and kurtosis is employed to represent color feature of traditional and unhealthy image of skin.

$$k = \frac{\sum_{i=1}^M \sum_{j=1}^N (P_{ij} - \mu)^4}{MN\sigma^4} \quad (7)$$

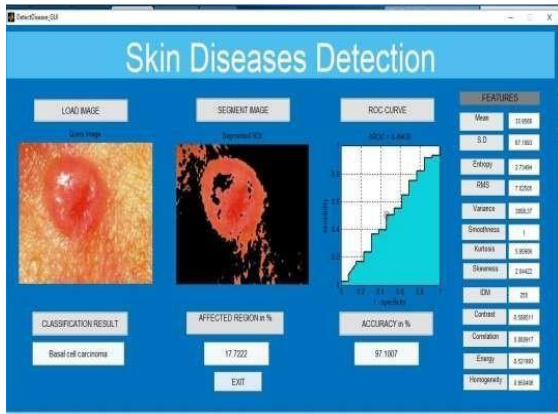
CLASSIFICATION:

Multiclass SVM essentially consists of the educational module and therefore the classification module, wherever the classification model is applied to new knowledge. It may be enforced by changing single category SVM into multiples of the binary classifications which may be done by identifying the classifiers on the idea of the actual label vs the rest(one-versus-all) or between each try of classes(one-vs-one).The provided MATLAB functions may be wont to train and perform multiclass classification on an information set employing a dendrogram-based support vector machine (D-SVM).The two main performs are: Train DSVM: this can be the function to be used for coaching, Classify DSVM: this can be the perform to be used for D-SVM classification.

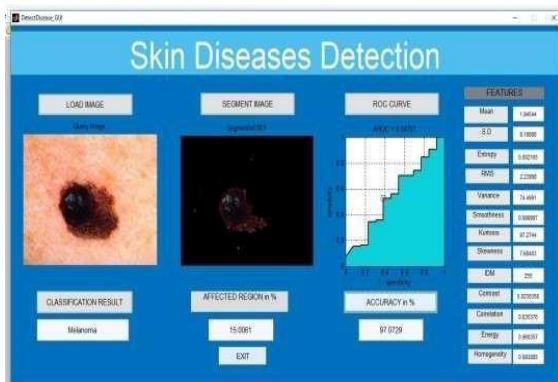
IV. EXPERIMENTAL RESULTS



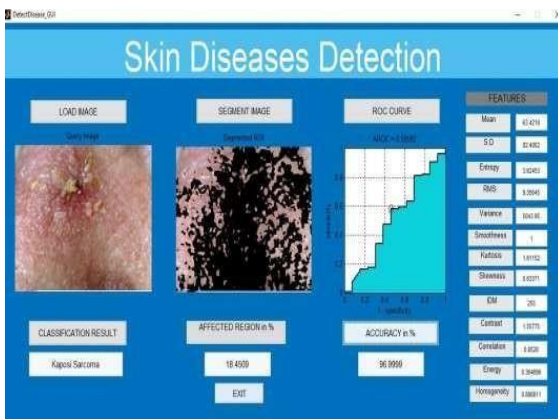
Figure(3) : Resultant Image 1 of Skin Disease Detection



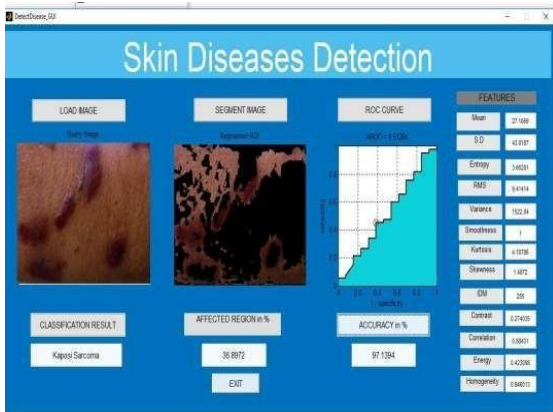
Figure(4) : Resultant Image 2 of Skin Disease Detection



Figure(5) : Resultant Image 3 of Skin Disease Detection



Figure(6) : Resultant Image 4 of Skin Disease Detection



Figure(7) : Resultant Image 5 of Skin Disease Detection

V. CONCLUSION

The main task of this project was to implement a system which might be able to predict disease of the skin. It may be simply finished that the projected system of carcinoma detection may be enforced victimization grey level co-occurrence matrix and support vector machine to classify simply whether or not image is cancerous or non-cancerous. Accuracy of projected system is ninety fifth. it's painless and dateless method than diagnostic test technique. it's additional advantageous to patients.

REFERENCES

1. J Abdul Jaleel, Sibi Salim, Aswin.R.B, "Computer Aided Detection 01 Skin Cancer", International Conference on Circuits, Power and Computing Technologies, 2013.
2. C.Nageswara Rao, S.Sreehari Sastry and K.B.Mahalakshmi "Co-Occurrence Matrix and Its Statistical Features an Approach for Identification Of Phase Transitions Of Mesogens", International Journal of Innovative Research in Engineering and Technology, Vol. 2, Issue 9, September 2013.
3. Santosh Achakanalli & G. Sadashivappa , "Statistical Analysis Of Skin Cancer Image –A Case Study " , International Journal of Electronics and Communication Engineering (IJECE), Vol. 3, Issue 3, May 2014.
4. "Digital image processing" by jayaraman. Page 244,254-247,270-273. (gray level, median filter).
5. Algorithm For Image Processing And Computer Vision .Page 142-145.(Thresholding)
6. Kawsar Ahmed, TasnubaJesmin, "Early Prevention and Detection of Skin Cancer Risk using Data Mining", International Journal of Computer Applications, Volume 62– No.4, January 2013.
7. M.Chaithanya Krishna, S.Ranganayakulu, "Skin Cancer Detection and Feature Extraction through Clustering Technique", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 3, March 2016.
8. A.A.L.C. Amarathunga, "Expert System For Diagnosis Of Skin Diseases", International Journal Of Scientific & Technology Research, Volume 4, Issue 01, 2015.
9. Mariam A.Sheha , "Automatic Detection of Melanoma Skin Cancer", International Journal of Computer Applications, 2012.
10. Anshubharadwaj, "Support Vector Machine", Indian Agriculture Statistics Research Institute.