



# Prediction of Skin Diseases using Big Data Techniques

<sup>1</sup>Mrs.K.SHUNMUGA PRIYA, <sup>2</sup>Dr.Mrs.V.SELVI

<sup>1</sup>Research Scholar, <sup>2</sup>Assistant Professor

Department of Computer Science,

Mother Teresa Women's University, Kodaikanal, Tamilnadu, India

## Abstract

Skin Disease prediction has become considerable in a range of applications such as health assurance, tailored health statement and public health. Due to the expenditure for dermatologists to observe every patient, there is a need for an automated system to assess a patient's threat of melanoma using images of their skin lesions captured using a standard digital camera. The established diagnosis technique aims at civilizing the quality of accessible diagnostic systems by proposing higher feature withdrawal and classification methods. In the Proposed method, 40 digital images composed from AOCD unit record and another 40 digital images from MIT unit database. These images are subjected to pre-processing the descriptions using Gaussian Filter technique. Then images are undergone image segmentation using K-means clustering algorithm to partition the disease reproduction area and non-affected area. Feature extraction is performed using Grey Level Co-occurrence Matrix (GLCM) for tentative texture which gave the statistical parameters, for better organization efficiency. The multi-SVM (Support Vector Machine) classifier is supervised learning model with connected algorithms that believe database images for classification analysis. The diagnosis system involves two stages of process such as training and testing, the skin tone values of the training data set are compared to the difficult data set of each type. Finally the performance study compared three algorithms such as Multi-SVM classifier, K-NN and Naïve Bayesian classifier. The overall correctness of using Multi-SVM classifier is 97% to 98%.

**Keywords:** Skin disease prediction, GLCM, Gaussian Filter, Multi SVM, K-NN, Naïve Bayesian, K-means.

## INTRODUCTION

Skin is the major organ in our body. Skin is that, which helps to protect the body from disease, heat, injury, and any type of damage which is caused by ultraviolet (UV) energy. It helps to build vitamin D. So production of skin from diseases is the important and intricate job in medicine.

Skin does many different things,

- i) Covers the interior organs and help protect them from wound
- ii) Serves as a difficulty to germs such as bacteria
- iii) Prevents the defeat of too much water and other fluids
- iv) Helps control body's high temperature
- v) To make the vitamin D.

So protect our body from skin diseases is to be required lead a healthy life. The main aim of this learn is to perform dissimilar types of skin disease images are collection in data set and pre-processing the contribution image using Gaussian filter and affect separate the subgroups affected area and non affected area using k-means clustering algorithm and affect feature removal using GLCM method and locate the diseases. Image processing operations can be in the region of separated into three major categories, Image Compression, Image Enhancement and renovation, and Measurement Extraction. It involves reducing the quantity of memory desired to store a digital image. Image desert which could be cause by the digitization procedure or by responsibility in the imaging set-up (for example, bad lighting) can be corrected using Image Enhancement technique.

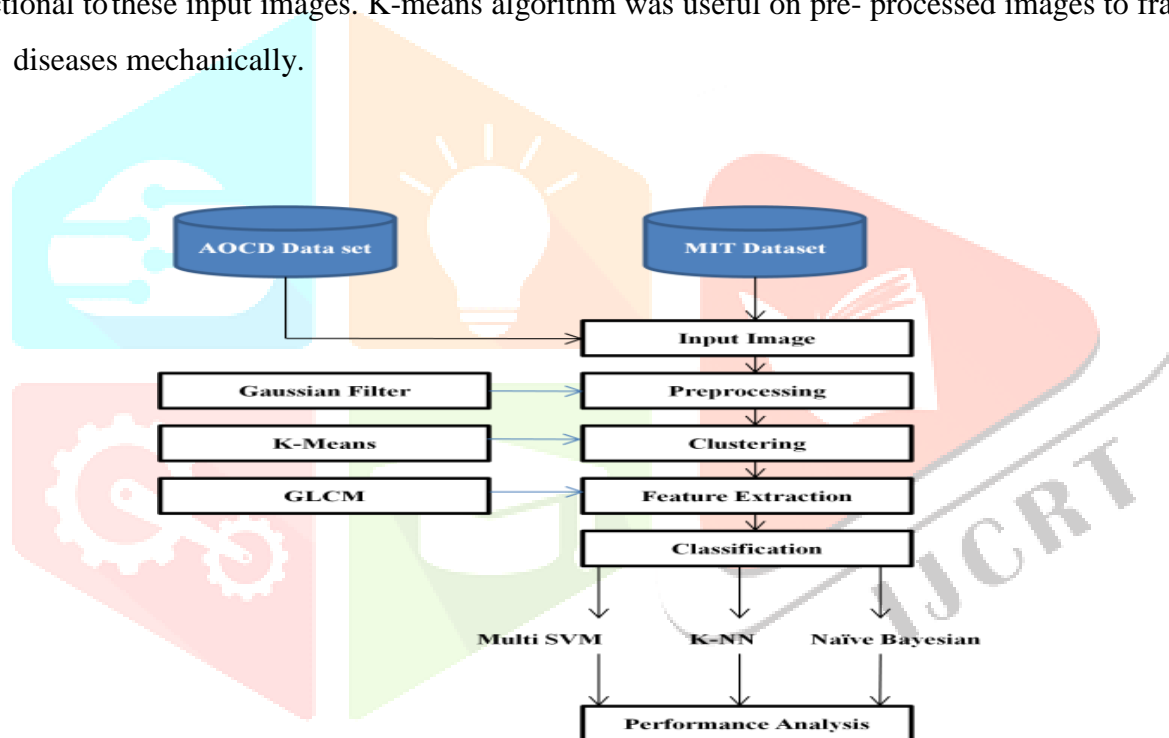
## LITERATURE REVIEW

Mugdha S Manerkar et al.,[1] using C means and Watershed algorithms for figure segmentation. Feature extraction is performed using Gray Level Co-occurrence Matrix (GLCM) and Image Quality Assessment (IQA) methods for texture which gave the numerical parameters of each algorithm, Features values of training data and testing. G.Ramya and J Rajeshkumar [2] used GLCM method for extracting features from the segmented diseased and classified the skin cancers based on fuzzy categorization, advanced accuracy compared to accessible one. B.Gohila vani et al.,[3] used a novel texture based skin lesion segmentation algorithm to categorize stages of cancer by Probabilistic Neural Network(PNN) on source of knowledge and instruction samples of data. Kawsar Ahmed et al.,[4] used pre-processing data is clustered using k-means clustering algorithm for extrication related and non relevant data to skin cancer. Many patterns are revealed using MAFIA algorithm. AprioriTid and decision tree algorithms of extract the repeated patterns from clustered dataset. I.Vijaya et al.,[5] focus on non-melanoma skin cancers and organize the types, calculate the type of illness correctly using support vector machine. Color rate and surface features are extract preprocessed instruction dataset. Y.P.Gowaramma et al., [6] used marker restricted watershed segmentation technique k-nn classifier along with curvelet filter. J.Priyadharshini et al., [7] using

organization algorithms to estimate the dermatological diseases in a variety of dimensions. Naive bayes algorithm technique for constructing classifiers. E.Barati et al., [8] emphasized highlighted and provided the variety of mining methodologies, they also found that apriori algorithm is best for quality extraction. Discussion result categorization method applied for good results. A.A.L.C Amarathunga et al.,[9] calculate the atmosphere and living situations of skin diseases. Image of skin disease captured to eliminate noises, pre-processing completed by Gaussian smoothing process. Divide the region of the disease by means of image segmentation algorithm. Feature extraction carried out using classification algorithm. Madhura Rambhajani et al.,[10] using Bayesian techniques with best first search feature selection used in order to categorize the dermatology diseases algorithm.

## METHODOLOGY

The digital skin illness images were full as from AOCD, MIT datasets and pre-processing technique were functional to these input images. K-means algorithm was useful on pre-processed images to fragment the skin diseases mechanically.



*Fig. 1 Data Flow Diagram*

### A. Input Image

The input images for this development are the digital similes of dissimilar skin diseases. These descriptions were taken as input from the data set.

### B. Pre-processing

In pre-processing concern Gaussian filtering to our input image. Gaussian filtering is repeatedly used to eliminate the noise from the image. Here we used wiener task to our input image. **Gaussian filter** is windowed clean of linear class; by its character is weighted mean. Named behind celebrated scientist Carl Gauss since weights in the filter designed according to Gaussian distribution.

The Gaussian Smoothing Operator perform a biased regular of instant pixels based on the Gaussian division. It is used to eliminate Gaussian noise and is a practical model of defocused lens. Sigma defines the amount of blurring. The radius slider is used to manage how large the pattern is. Large values for sigma will only give large blurring for larger pattern sizes. Noise can be further using the sliders.

### C. Clustering

Clustering is a development of unraveling dataset into subgroups according to the exclusive feature. Clustering separated the dataset into important and non-relevant dataset. Clustering is another tiresome term of data mining. The clustering difficulty has been addressed in frequent besides being proven helpful in many applications. The aim of clustering is to classify objects or data into a number of categories or classes where each class contains the same characteristic. The main settlement of clustering are that the data object is assigned to an unknown class that have unique feature and reduce the memory.

The k-means clustering is a widely established clustering tool that is used for robotics, diseases and artificial intelligence application principle. Here k is a positive integer in place of the number of clusters. The pre-processed data is cluster using the k-means clustering algorithm with the worth of k equal to 2. This represent there is two clusters where one is related data and another contains non-relevant information.

### D. Feature Extraction

The features were extracting from the images based on Gray Level Co-occurrence matrix (GLCM). GLCM statistics like energy, connection, entropy, and homogeneity were extracting from the midpoint growth and the whole skin region. The development calculation of the GLCM data was applied in different distance and in dissimilar angles. In each case of the computation of the GLCM statistics min, mean, max, and difference (i.e., max-min) values were extracting from the image region. The extracted values represent the surface of the input images in an efficient method. The significance can be extracted from the images was then saved as the skin texture. Few of the common figures applied to co-occurrence probability are given the following table with connected formula.

### E. Classification

Classification is a data mining function that assign items in a gathering to target categories or classes. The objective of classification is to completely estimate the goal class for each case in the data. Classification divide data sample into objective classes. The classification techniques predict the target class for each data point.

### F. Multi-SVM Classifier

This defines a group of the entire program in two disjoint groups of classes. This arrangement is then used to train a SVM classifier in the origin join of the consequence tree, using the samples of the first collection as supportive examples and the samples of the next group as negative examples. The classes from the first

cluster group are being allocate to the first (left) sub tree, while the instruction of the second clustering group are being assigned to the (right) second sub tree. The procedure continues recursively until there is simply one class per compilation which defines a leaf in the decision tree.

### G. K-NN classifier

Taxonomy has been done by KNN classifier. The k-Nearest Neighbors algorithm (or k-NN for short) is a non- parametric process used for classification and failure. The contribution consists of the k closest teaching examples in the quality space. The value depends on whether k-NN is used for group. k-NN is a type of instance-based data, or lazy knowledge, where the purpose is only approximated nearby and all division is delayed until classification. It can be caring to weight the help of the neighbors, so that the nearer neighbors offer more to the ordinary than the more distant ones.

### H. Naïve Bayesian Classifier

In machine learning, naïve Bayesian classifiers are a family of simple probabilistic classifiers based on apply Bayes theorem with muscular (naive) liberty assumption between the features. Naïve Bayes classifiers are extremely scalable, require a figure of parameters linear in the figure of variables (features/predictors) in a learning problem. Maximum- training can be done by evaluate a closed-form appearance, which take linear time, rather than by luxurious iterative estimate as used for several other type of classifiers.

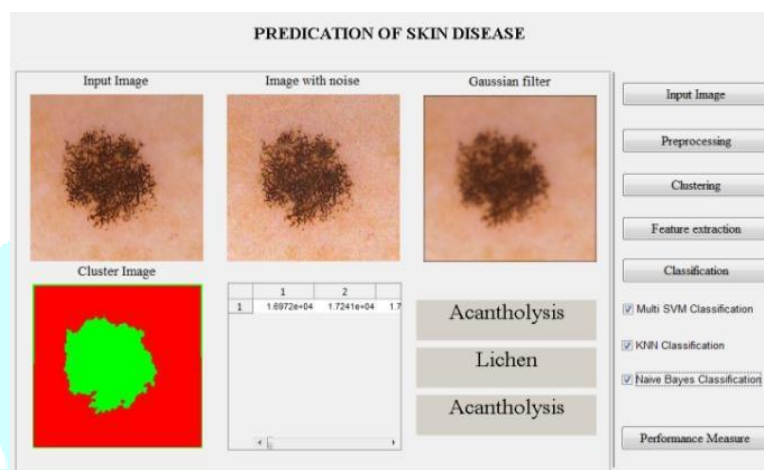
### PERFORMANCE ANALYSIS

The accurateness, understanding and specificity of the classifier are calculated. The accuracy represent the competence of the process. The understanding show how the algorithm gives correct categorization. The specificity show how the algorithm discards the accidentally categorization results. We calculated a spatial constancy constraint in a graphical model to get better the finding performance. The presentation of the procedure is calculated based on the calculation of Accuracy, Area beneath curve of the method.

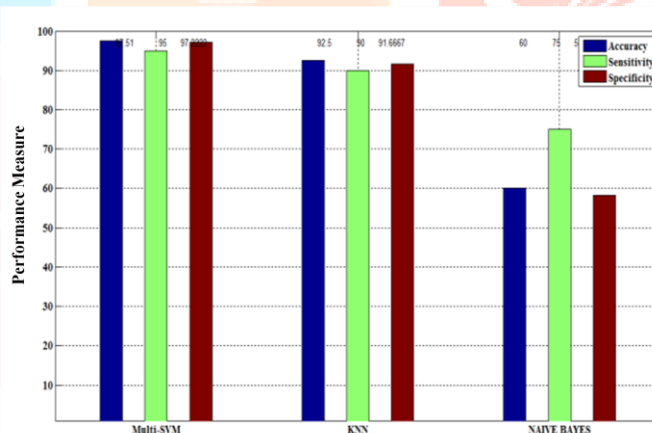
### EXPERIMENTAL RESULTS

The exceeding stated algorithms are implement in MATLAB. Around 40 digital skin diseases images are taken from two dissimilar data sets MIT and AOCD these are subjected to pre-processing technique such as image resizing, image format exchange, difference enhancement.

The pre-processed descriptions are given as input for smoothened using Gaussian filter for improved correctness images partition into affected area and non-affected area using k-means clustering algorithm. Features are extract by GLCM method. Features of test images are compare with the guidance data set by the disease categorization into specific category using multi SVM classifier. In Experimental results the concert measure of AOCD data set are accessible in Fig.3 and MIT dataset are presented in Fig.5 that instead of the performance measure of SVM, k-NN, Naïve Bayesian classifiers. On the whole accuracy of this study range from 97% to 98%.

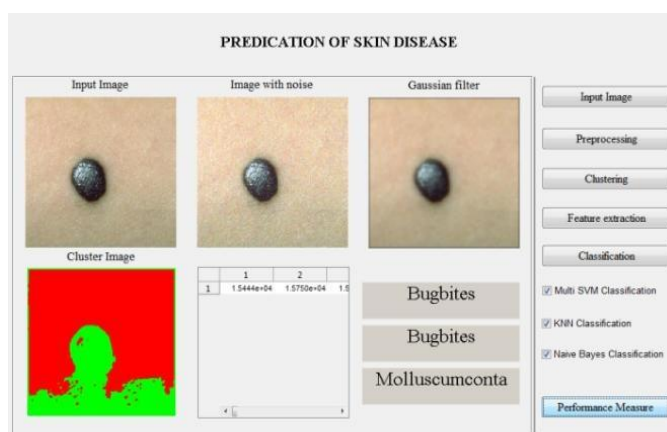


**Fig.2 . The GUI Form of Skin Disease Prediction AOCD dataset**



**Fig. 3. Performance measure of Dataset1**

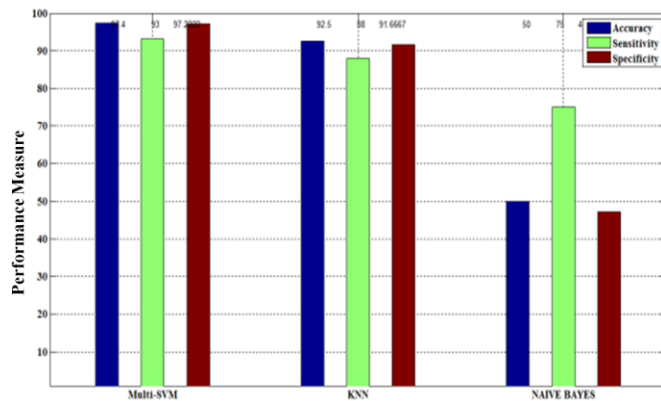
**TABLE I THE PERFORMANCE OF ACOD DATA SET**





Classification Algorithms	The Performance of ACOD dataset		
	Accuracy	Sensitivity	Specificity
Multi-SVM	97.5	95	97.2
K-NN	90	70	91.4
Naive Bayesian	60	75	58.3

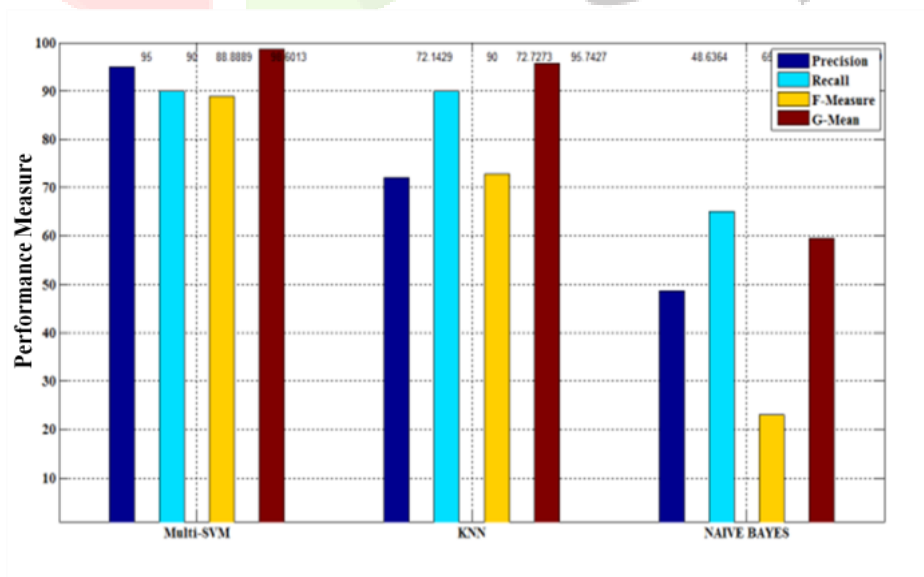
*Fig.4 . The GUI Form of Skin Disease Prediction MIT dataset*



*Fig. 5. Performance measure of Dataset2*

**TABLE II THE PERFORMANCE OF MIT DATA SET**

Classification Algorithms	The Performance of MIT dataset		
	Accuracy	Sensitivity	Specificity
Multi-SVM	97.4	93	97.2
K-NN	90	68	91.4
Naive Bayesian	55	75	52



*Fig. 5. Performance measures*

**TABLE III THE PERFORMANCE MEASURES**

<b>Classification Algorithms</b>	<b>The Performance Measures</b>			
	<b>Precision</b>	<b>Recall</b>	<b>F-Measures</b>	<b>G-Mean</b>
Multi-SVM	95	90	88.8	98.6
K-NN	72.1	70	66.6	85.5
Naive Bayesian	51.6	65	27.2	66.1

## CONCLUSION

In proposed system we compare one different recognized feature. We also experiment with dissimilar classifier architectures. The characteristic sets and most of the classifier architectures we experienced, offer a similar performance. To improve the concert the results of the experiment specify that this technique has much faster training and testing times than the extensively used multi SVM methods. Maintain vector machine has been used to train the model. The accurateness, compassion and specificity of the classifier are calculated. The result of the experiments proves that maintain vector machine is effective. As prospect improvement of this investigate work, more skin texture that help to boost the classification correctness can be recognized, extracted and used for knowledge. In future to implement multi SVM classifier With some other extract skin texture. A system framework is obtainable to distinguish multiple kinds of activities from videos by an SVM multi-class classifier with binary tree construction. The thought of hierarchical classification is introduce and multiple SVMs are aggregate to achieve the respect of actions. Each SVM in the multi-class classifier is trained unconnectedly to realize its best categorization performance by choosing proper skin texture before they are aggregate.

## REFERENCES

- [1] Mugdha Smanerkar, shashwata harsh, juhi saxena, simanta p sarma, dr. u. snekhalatha, dr.m. anburajan, "classification of skin disease using multi svm classifier" 3rd international conference on electrical, electronics, engineering trends, communication, optimization and sciences (eeecos)-2016.
- [2] G.Ramya, J.Rajeshkumar "Novel method for segmentation of skin lesions from digital images", international research journal of engineering and technology vol:02 issue:08 November 2015.
- [3] B.Gohila vani et al. "Segmentation and Classification of Skin Lesions Based on Texture Features" Int. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 4, Issue 12( Part 6), December 2014, pp.197-203.
- [4] Kawsar Ahmed, Tasnuba Jesmin, Md. Zamilur Rahman "Early Prevention and Detection of Skin Cancer Risk using Data Mining" International Journal of Computer Applications (0975 – 8887) Volume 62– No.4, January 2013.
- [5] I.Vijaya M. S "Categorization of Non-Melanoma Skin Lesion Diseases Using Support Vector



Machine and Its Variants”. International Journal of Medical Imaging. Vol. 3, No. 2, 2015, pp. 34-40. doi: 10.11648/j.ijmi.20150302.15

- [6] Y.P.Gowaramma et al., used marker controlled watershed segmentation method k-nn classifier along with curvelet filter.
- [7] J. Priyadharshini “A Classification via Clustering Approach for Enhancing the Prediction Accuracy of Erythematous-squamous (Dermatology) Diseases” IJSRD - International Journal for Scientific Research & Development| Vol. 3, Issue 06, 2015 | ISSN (online): 2321-0613.
- [8] E.Barati et al., “A survey on utilization of data mining approach for dermatological skin diseases prediction” Journal of selected areas in health informatics march 2011.
- [9] A.A.L.C. Amarathunga, et al.,”Expert system for diagnosis of skin diseases” International journal of scientific & technology research volume 4,issue 01, january 2015 issn 2277-8616 174 ijstr©2015.
- [10] MadhuraRambhajani “Classification of Dermatology Diseases through Bayes net and Best First Search” International Journal of Advanced Research in Computer and Communication Engineering” Vol. 4, Issue 5, May 2015.

