# Classifiers and Medical Image Processing: A Review

Shweta V. Marulkar[1], Bhawna Narain[2]
[1,2]*Research Scholar, MATS School of IT, MATS University, Raipur (C.G.), India*

***Abstract*** **- Digital image processing have wide scope in globe such as military, medical, robotics, forensic science etc. Now a day for such type of applications feature extraction of digital image is important part of processing. Classification of medical images is mandatory step in pattern recognition. Classification improves the efficiency and accuracy. Image pre-processing, segmentation, feature extraction can be carried out for classification process. Preliminary disease detection rate depends on all the steps but classification has its own importance in pattern recognition. Different important classifier such a support vector machine (SVM), artificial neural network (ANN), decision tree, KNN etc. Enlisted classifier has their importance in one or the other way. In this paper there is a discussion about many classifiers. Input to the proposed system is nail image from which nail texture, shape and colour features are extracted and then by considering these features analysis of nail is done which will then be used for the diagnosis of various diseases. This proposed system will surely help the medical practitioners in the early diagnosis of diseases.**

***Index Terms*** **- Digital image processing, Nail features Analysis, disease prediction, Supervised Learning, artificial neural network.**

## I.INTRODUCTION

Number of diseases can be easily determined by observing nails of both hand and leg. Digital image processing is an important concept in medical science and technology. In the modern era majority of doctors uses digital technology to diagnose the diseases. Image analysis techniques can be applied for various purposes such as to detect diseased area of specific organ, to find the boundaries and quantify disease infected area, to recognize colour, size and shape of infected area to diagnose the disease. Image acquisition, Image processing, Segmentation, Feature extraction, Comparison with database and result are the basic steps for the image processing technique to solve the problem. Various diseases affect the appearance of nail by colour, texture, shape and pliability. Nail Disorders can be classified into four categories – Congenital, Traumatic, Infectious, Tumours. Classification of digital image based on spectral information represented by digital numbers. The digital image classification process has two approaches: Per-pixel classification and Object-oriented classification. Input pixels can be categorizing into spectral feature class based on its individual multispectral vector, which does not involve any context or neighbourhood evaluation in per-pixel classification. Input pixels are grouped into spectral features using an image segmentation algorithm, In Object-oriented classification. These objects are characterized in both the raster and vector domains. Supervised classification, Unsupervised classification and Rule -based classification-based classification are the three common per-pixel methods.

*A. Supervised Classification*

From the knowledge of domain labelled data points are created. These labelled data points are used for accurate categorization of an image. Expertise knowledge is required for supervised classification.

*B. Unsupervised Classification*

A number of an unsystematically sampled data's groups, will be divided automatically into the same classes by using clustering techniques. These clustered classes later used for determining population measurements. This type of classification is called the unsupervised classification. This type of classification is very appropriate to classify large data in less time.

## II. REVIEWS OF LITERATURE

For object detection and digital image analysis image classification is important step. Image classification is either final output or intermediate output depending on application. Lot of research work have been already conducted for identifying best digital image classification technique. It is hard to decide any one technique as the best technique among all, because the results and its accuracy depend on a number of factors.

1. Digital Image Processing Technique for Palm Oil Leaf Disease Detection using Multiclass SVM Classifier proposed by Ahmad Nor Ikhwan Masazhar et al. The concept of k means clustering and SVM classifier to determine two palm oil diseases based on symptoms of the disease through its leaf for disease detection. Thirteen types of features are extracted from leaf image using k means clustering technique and multiclass SVM classifier classifies the disease. SVM achieves 97% of accuracy for chimaeras & 95% for Anthracnose. SVM also gives pattern of the disease infection symptoms on the leaf. Proposed system follows image acquisition, contrast enhancement, conversion from RGB colour space to L*a*b colour space, for segmentation k means clustering technique, for feature extraction gyrocomatrix and GLCM was created from image, for classification multiclass SVM classifier used, and finally feature extracted from the image will be classifies and compare by each row of dataset.
2. S. Arivazhagan et al. Proposed Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. In this paper creation of color transformation structure of RGB Image, Segmentation process is performed and the green pixels are masked and removed using specific threshold value, the interested segments are considered for computing the texture features using color co-occurrence method and finally the extracted feature are passed through classifier. SVM classifier is used to improve detection accuracy. The basic concept of two class problem is then extended to multiclass problem and detected leaf disease is classified into various categories. The preliminary disease detection can be made using this method.
3. A Review on Medical Image Processing proposed by Navnish Goel et al. A good image denoise model is that it should completely remove noise like Gaussian, Poisson, Rician and impulse noise such as salt and pepper noise as much as possible and preserve edges. A digital medical image consists of different types like x-ray, Computed tomography (CT-Scan), Medical resonance Imaging (MRI), Ultrasound, Positron Emission Tomography (PET) etc. Tumor detection, Craniofacial features, Congenital heart defects, Diagnosis heart volve diseases, Tuberculosis (TB), Pathological brain detection (PBD), Birth defects etc are applications of medical image analysis.
4. Classification of Cotton Leaf Spot Disease Using Support Vector Machine proposed by Prof. Sonal P. Patil et al. The work is based on segmentation techniques in which input images are pre-processed for quality improvement. Then texture and color feature extraction techniques are used for feature extraction such as boundary, shape, color, Texture for disease spots to detect the disease. To detect diseased leaf, stem and fruit, quantify affected area, find out the boundaries, determine the color, shape and size analysis of agricultural images can be done. Segmented image is considered for image analysis and important features are extracted and classification of diseases is performed using SVM classifier.
5. Mbarek Marwan et al. proposed Using Cloud Solution for Medical Image Processing: Issues and Implementation Efforts for security requirements they analyze and compare the current status of the cloud-based image processing implementations. Software, platform and infrastructure are the three main services offered by cloud providers to consumer. It also provides cost efficient computational resources. Several implementations are proposed to overcome security and privacy issues such as harmonic encryption, secrete sharing scheme and SOA. Extension of this work they propose hybrid solution based on segmentation in conjunction with watermarking techniques. For the above said technique the input image is divided into potions to increase data confidentiality, watermarking used to ensure authentication. To improve performance and availability the framework will be implemented over Hadoop system and to address interoperability issues use DICOM standerd.
6. A New Multimodal Brain Tumour Image Fusion for Pre-processing of Medical Image Database proposed by Denzil Bosco.T. et al. The objective of proposed technique is to get valuable reciprocal data from image combination. In this experiment first input image is separated into high rehash and low rehash. The two-scale representation of picture utilises ordinary channel which is fundamental and feasible. The Discrete wavelength Transform (DWT) Algorithm is used in proposed method. MRI and CT scan are the type of input images and

Discrete Wavelet Transform (DWT) and Integrated Discrete Wavelet Transform (IDWT) are fused (Output) images.

7. Disease Symptoms Identification In Paddy Leaf Using Image Processing proposed by R.P. Narmadha et al. The experiment consists of image acquisition of different type such as jpeg, png, gif, bmp consider as an input. During pre-processing RGB image is converted to gray scale image. Histogram equalisation and control management performed to enhance contrast of the image. K means clustering used for segmentation. SVM, ANN, Fuzzy classification is used for the classification of features like texture; structure and geometric feature are different type of feature extraction. To establish the classification system four characteristics of lesion type, boundary colour, spot colour, and broken paddy leaf colour were tested. For determining type of lesion the ratio of height and width of the injury spot gives unique shape characteristics.
8. A survey of image classification methods and technology for improving classification performance proposed by D.Lu and Q. Wend et. al. various factors examine current practices, problems, and prospects of image classification and affects complex process for image classification. The tactics and methods that used for improving classification accuracy plays important role.
9. A survey on image classification methods proposed by Jipsa Kurian, Vkarunakaran et. al. during the image processing image classification and method for classifying image is more complex task. Number of methods is available to classify blurry and noisy images. Supervised and unsupervised classifications are two main methods for image classification. It is more complex and difficult to classify if it contains blurry and noisy content.
10. Content based colour image classification using SVM proposed by Saurabh Agrawal, N.K. Varma, Prateek Tamrakar and Pradip Sircar et.al. In this paper histogram of color components produced from feature extraction for color image classification purpose and support vector machine is used to improve the classification. To achieve better efficiency and insensitivity to small changes in camera view-point i.e. translation and rotation color image histograms are produced.
11. SVM based fuzzy decision tree for classification of high spatial resolution remote sensing images proposed by S.Moustakidis, G. Mallinis, N. Koutsias, John B. Theocharis. Class grouping algorithm used for determining tree structure. fuzzy decision tree is proposed in this paper. binary svm are used for the implementation of node discrimination. Some advantages such as enhanced classification accuracy, interpretable hierarchy, and low model complexity shown by FDT-SVM. In addition, effective feature selection is incorporated within the tree building process, selecting suitable feature subsets required for the node discriminations individually. It provides hierarchical image segmentation and has reasonably low computational and data storage demands.

## III. COMPARATIVE STUDY OF DATA CLASSIFICATION TECHNIQUES

### *A. Artificial Neural Network (ANN)*

Any human learning process consist of acquiring information, practicing such things, and experience, to design human like behaviour in machine we use artificial intelligence. A neural network model is similar to human nervous system. To cultivate the human like actions in machine we use artificial intelligence and neural network, so it called as artificial neural network. The neurons contain in neural network that are correlated together to convert inputs into useful output. Datasets are used to train artificial neural network. If these datasets are previously known to us then ANN is trained in a supervised manner, and it learns precisely and quickly about the pattern suppressed in dataset. Character Recognition, Fault detection, Speech detection, Product detection are applications of ANN. ANN has basically two layers such as input layer and output layer but some systems also have hidden layers. Numbers of neurons are present in every layer and they connected through weighted links.

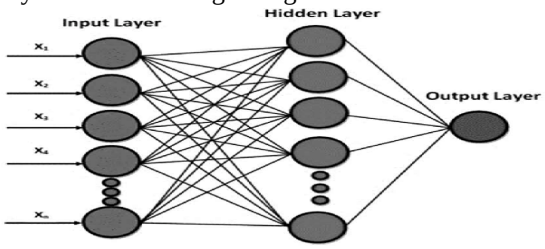


Fig. 1 Structure of ANN

*B. Support Vector Machine (SVM)*

The Support vector machine belongs in the supervised learning category. Regression and classification carried out by using SVM.A point in n-dimensional space is used to represent every object or item in this classifier. The value of each point feature, represented by the specific coordinate. Then the objects divided into classes are segregate by finding n-dimensional plane. The diagram shows support Vectors that represent the co-ordinates of each item. The SVM algorithm is a good choice to segregates the two classes.

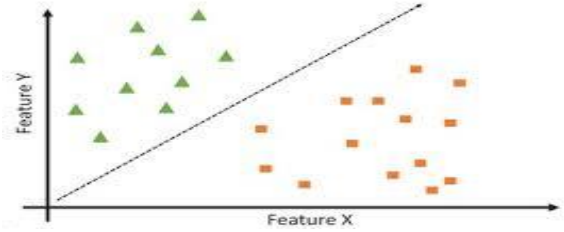


Fig. 2 Structure of SVM

*C. k-Nearest Neighbours (KNN)*

The k-nearest neighbour algorithm is a simple, easy-to-implement supervised machine learning algorithm. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. In this algorithm the task is to put newly recognized object into existing classes.

*1) K-Nearest Neighbour Classifier:*

For working on large amount of well-informed information various data mining techniques were used. K-Nearest Neighbour (KNN) is a very efficient, effective and simplest algorithm for object classification. Process of KNN is identify the class or set for newly recognized object of their nearest neighbour. Suppose Category x and Category y, and we have a new data point p1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. KNN is frequently used in many studies, especially when there are only few or no information about the data distribution. KNN is a non-parametric algorithm, i.e. it does not make presumptions about distribution of data used in analysis. KNN is also called as a lazy algorithm as it only uses quick training phase. KNN does not make generalization which implies that KNN maintains all training data.

*2) Classification using KNN:*

A method for K-nearest neighbour algorithm is categorizing data based on the adjoining training examples in the feature universe. The algorithm for KNN classifier is as follows:

1. First the collected database is divided into a training set and testing set.
2. For each row in the testing set, Select the number K of the neighbours then Calculate the Euclidean distance of K number of neighbours and Take the K nearest neighbours as per the calculated Euclidean distance.
3. Process of calculating distance between test data set and every element in training data set can be considered as follows. The Euclidean distance between two input vectors p and q is computed as the magnitude of difference in vectors i.e. p - q, Where both the data are having "m" dimensions i.e. p= (p1, p2… pm) and q= (q1, q2,…, qm).

The Euclidean distance between "p" and "q" is found to be:

$$D(p,q) = |p-q| = \sqrt{(p_1-q_1)^2 + (p_2-q_2)^2 + \ldots + (p_m-q_m)^2}$$

4. Among these k neighbours, count the number of the data points in each category.
5. Assign the new data points to that category for which the numbers of the neighbours are maximum.

*3) Decision Tree (DT)*

Supervised learning algorithm used by decision tree and used for classification purpose. Continuous and categorical input and output variables are the two cases for which the decision tree is applicable. In the DT algorithm the substantial discriminator in input variables is used to fragment the samples of dataset into homogeneous sets. According to their target variables decision trees are classified. Continuous Variable Decision Tree and Categorical Variable Decision Tree are the types of decision tree. Decision

trees comes with some merits as, it can be easily understandable and follows non parametric methods. Decision trees require minimum data cleaning.

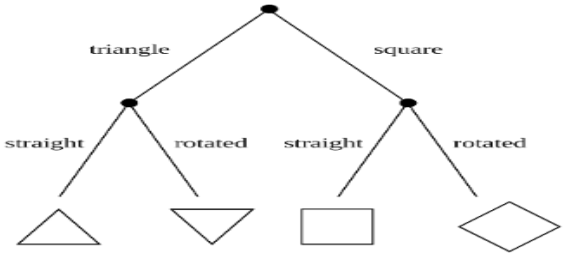


Fig.3. Structure of DT

## IV. THE PROPOSED PRELIMINARY DISEASE DIAGNOSIS SYSTEM

The complete methodology of our system is represented in following flowchart. The individual steps are modularized and are often autonomous and sometimes dependent on each other.

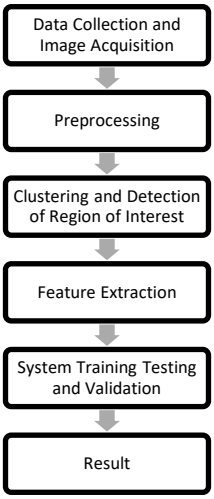


Fig.4. Methodology of system

### *A. Data Collection and Image Acquisition*

For system development collection of nail images is first and primary task. High quality data is collected through digital camera. Eight different pre-processing algorithms were used such as converting colour image to grey scale image, sharpening filter, median filter, smooth filter, binary mask, RGB extraction, histogram and sobel operator. For converting into gray scale image RGB values of input images were extracted. To sharpen the details of the infected region Sharpening filter is applied to the gray scale image. To remove the noise from the image Median filter is used after sharpening filter. In the next step smoothing filter which replaces each pixel with the mean values of its neighbours, including itself. Mean filter produces binary images and colour distribution of binary images shown by histogram. To extract average colour code of the infected area from the binary image YCbCr were used. To detect edge of the infected area of binary image Sobel operator were applied.

### *B. Pre-processing*

The image data is improved in the second step of pre-processing and by transforming input colour image to colour space which have intensity component the contrast enhancement is done.

### *C. Clustering and Detection of Region of Interest*

After pre-processing, a threshold was applied and k-means clustering was performed on the colour gradient. Then morphological closing was performed on the clusters to obtain the binary mask and by applying the mask we segmented the diseased part from the nail input image. All the consequent steps are illustrated below.

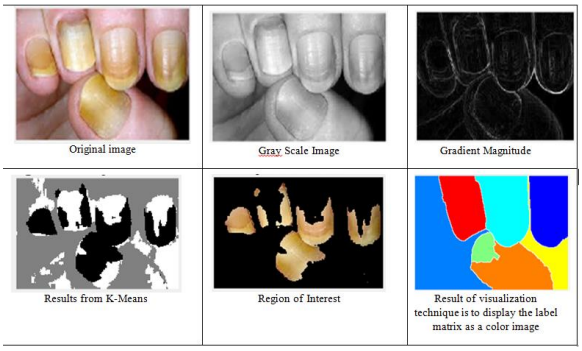


Fig.4. Clustering and Detection of Region of Interest

### *D. Feature Extraction*

Automated Visual feature extraction: The feature collections that were extracted from the images of the ROI and healthy nails are the following:

- Mean and Standard Deviations of Colours of 3 Colour Channels (R, G and B) of ROI
- Mean and Standard Deviations of Colours of 3 Colour Channels (R, G and B) of healthy nails
- Distribution (scattering of the ROI-s)
- Area (in mm2)
- Energy, Entropy, Contrast and Homogeneity from GLCM in each colour channel.

At this point, we have separated the healthy nail images and the individual ROI-s in structures that contain the images of the RGB components. From the RGB components, we have computed the mean,

variance (standard deviation) of the channels individually. Energy, entropy, contrast and homogeneity were computed from the regions by computing the gray level co-occurrence matrices (GLCM) in each channel. The separation of ROI and healthy skin was done using masks. Thus we can see how the healthy and normal nail images were separated to extract colours from them.

*E. System Training Testing and Validation*

We have used feed-forward back-propagation neural network training to perform this step. We validated and tested our system using the tenfold cross validation process. The virtue of using a cross validation technique was that there were no overlapping of the test data and training data, making the system testing results viable and dependable.

*F. Result*

This step gives us the name of preliminary disease from nail images.

## V. CONCLUSION

This paper tries to study different image classification techniques and merits and demerits of the techniques. Most common approaches for image classification can be categories as supervised and unsupervised, or parametric and nonparametric or object-oriented. Very popular classifier in the field of image processing such as ANN, SVM, KNN and DT explains in this paper. So, in this paper gives a theoretical knowledge of enlisted classifiers.

## VI.ACKNOWLEDGMENT

The authors would like to express their deep gratitude towards the Department of Animation Science, YCIS, Satara, Maharashtra and MATS School of IT, MATS University, Raipur for their support and encouragement during this work.

## REFERENCES

[1] Ahmad Nor Ikhwan Masazhar, Mahanijah Md Kamal, "Digital Image Processing Technique for Palm Oil Leaf Disease Detection using Multiclass SVM Classifier", Proc. of the 4th IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Putrajaya, Malaysia, 28-30 November 2017

[2] S. Arivazhagan, R. Newlin Shebiah*, S. Ananthi, S. Vishnu Varthini," Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features", Agric Eng Int: CIGR Journal, Vol. 15, No.1, Pp – 211-217, March 2013.

[3] NAVNISH GOel, DR. AKHILENDRA Yadav, DR. BRIJ MOHAN Singh, "Medical Image Processing: A Review", Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH-16), 978-1-4673-9080-4/16/$31.00 ©2016 IEEE, Pp – 58-62, 2016.

[4] Prof. Sonal P. Patil, Ms. Rupali S.Zambre, "Classification of Cotton Leaf Spot Disease Using Support Vector Machine", Int. Journal of Engineering Research an Applications, Vol. 4, Issue 5( Version 1), Pp – 92-97, ISSN : 2248-9622, May 2014.

[5] Mbarek Marwan, Ali Kartit and Hassan Ouahmane, "Using Cloud Solution for Medical Image Processing: Issues and Implementation Efforts".

[6] Denzil Bosco.T, Lakshmi Narayanan.A and Veeramuthu A, "A New Multimodal Brain Tumour Image Fusion for Pre-processing of Medical Image Database", International Conference on Communication and Signal Processing, 978-1-5090-3800-8/17/$31.00 ©2017 IEEE, Pp – 1724-1728, April 6-8, 2017.

[7] R.P. Narmadha, G.Arulvadivu, "Disease Symptoms Identification In Paddy Leaf Using Image Processing", Advances in Computational Sciences and Technology, Volume 10, pp. 2213-2223, ISSN 0973-6107, 8 November 2017.

[8] D.Lu and Q. Wend, "A survey of image classification methods and technology for improving classification performance", Department of Geography, Geology, and Anthropology, Indiana State University, USA, IJRS, Vol.28, No.5, 10 March 2007.

[9] Jipsa Kurian, V. Karunakaran, "A survey on image classification methods", Department of Computer Science, Karunya University, Coimbatore, India, IJARECE, Volume 1, Issue.4, Oct 2012.

[10] Saurabh Agrawal, Nishchal K Verma, Prateek Tamrakar, Pradip Sircar, “Content based color image classification using SVM”, Department of Electrical Engg., IIT, Kanpur, India, 2011 8th international conference on information technology.

[11] Shweta Marulkar, Rajivkumar Mente, "Nail Image Segmentation for Disease Detection", International Conference on Recent Trends in Image Processing and Pattern Recognition RTIP2R 2018: Recent Trends in Image Processing and Pattern Recognition pp 116-126

[12] Mente, R.S., Marulkar, S.V., "A review: fingernail images for disease detection." Int. J. Eng. Comput. Sci. 6(11), 22830–22835 (2017)