PLANT LEAF DISEASE DETECTION USING IMAGE SEGMENTATION AND MACHINE LEARNING TECHNIQUES

Project report submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology

in

Computer Science and Engineering/Information Technology

By

Nitika Sharma (141258)
Shaswat Sood(141260)

Under the supervision of

Dr.Pradeep Kumar Singh

to

Department of Computer Science & Engineering and Information Technology
Jaypee University of Information Technology Waknaghat, Solan-173234, Himachal Pradesh

Certificate

Candidate's Declaration

I hereby declare that the work presented in this report entitled " Plant leaf disease detection using image segmentation and machine learning techniques" in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering/Information Technology** submitted in the department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology Waknaghat is an authentic record of my own work carried out over a period from August 2015 to December 2015 under the supervision of **(Supervisor name)** (Designation and Department name).

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Nitika Sharma, 141258

Shaswat Sood,141260

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

(Supervisor Signature)

Dr. Pradeep Kumar Singh
Assistant Professor (Senior Grade)
Computer Science & Engineering and Information Technology
Dated:

ACKNOWLEDGEMENT

On the submission of our thesis report on “Plant leaf disease detection using image segmentation and machine learning techniques”, we would like to extend our gratitude and sincere thanks to our supervisor **Dr. Pradeep Kumar Singh, Assistant Professor (Senior Grade), Department of Computer Science & Engineering and Information Technology** for his constant motivation and support during the course of our work . We truly appreciate and value his esteemed guidance and encouragement from the beginning. We are indebted to him for having helped us shape the problem and providing insights towards the solution. And for providing a solid background for our studies and research thereafter. He has been a great source of inspiration to us and we thank him from the bottom of our heart. Above all, we would like to thank all our friends whose direct and indirect support helped us.

Nitika Sharma,141258

Shaswat Sood,141260

INDEX

LIST OF ABBREVIATIONS

DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
GIF	Graphics Interchange Format
TIFF	Tagged Image File Format
DES	Data Encryption Standard Algorithm
LSB	Least Significant Bit Algorithm
RSA	Rivest-Shamir-Adleman
WMSW	Wireless Multimedia Sensory Networks
DPCM	Differential Pulse Code Modulation
CR	Compression Ratio
MSE	Mean Square Error
PSNR	Peak Signal to Noise Ratio
CS	Compressed Sensing

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

A huge development has been made in the field of image processing and machine learning and its application in various branches of engineering. We have entered the era of digitization. We have captured images with the help of digital camera. More clear the images are better and efficient the results. In this report we have done the classification of disease free, partially diseased and completely diseased leaves. We have used HSI color model for classification of our attributes and further we have used Neural Network Toolbox in Matlab for machine learning and analyzing the results.

CHAPTER-1

INTRODUCTION

1.1 Image Basic Concept

Image is a collection of rectangular array of dots which are known as pixels. The size of an image can be determined by the number of pixels present in it. It can simple be calculated by width x height. Each and every pixel present in an image is a certain type of color. While working with the black and white image in which the pixels are totally white or totally black, the options are really restricted as only a single bit is needed for every pixel. Such type of images are useful for line art like cartoons in newspaper. An additional type of colorless image is a grayscale image. Grayscale images are very often incorrectly termed as "black and white" images. They use 8 bits per pixel, which are sufficient enough to depict every shade of gray color that a human eye can recognize.

While we are working with the color images, things tend to become slightly complicated. The number of bits per pixel is referred as the depth of an image or the bit plane of an image. A bit plane which is having n bits can have 2n colors present in it. The human can identify around 224 colors and some claim the larger number as well. Very commonly found color depths are 8,16 and 24.

There are two basic methods to store the information of various colors present in an image.
The most direct way is to use the RGB (red, green, blue) color composition in which color of each pixel is represented by giving an order triple of numbers.
The second way to store information regarding the color is by using the table in order to store the information of triples and use a reference in the table for every pixel present. This helps in the betterment of storage requirements of an image.

1.2 Neural networks overview

Neural network are made out of basic components . These components are handled by the nervous system of the human body. As in nature, the components are associated in such a way that they operate the working of the system to large extent. You can prepare a neural system to play out a specific capacity by changing the estimations of the associations (weights) between components.

Regularly, neural network are balanced, or prepared, with the goal that a specific info prompts a particular target yield. The following figure shows such a circumstance. Here, the system is balanced, in light of an examination of the yield and the objective, until the point that the system yield coordinates the objective. In order to train a network the input/target are necessary.

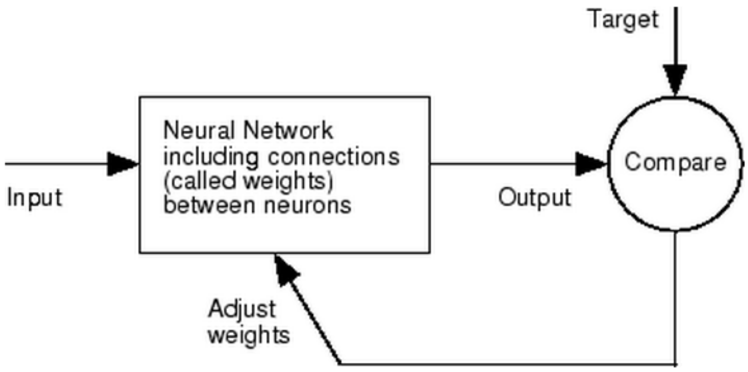


Figure 1.1

Neural network systems are designed in such a way that they solve difficult problems very easily containing design acknowledgment, ID, grouping, discourse, vision, and control frameworks.

Neural systems can likewise be prepared to take care of issues that are troublesome for ordinary PCs or individuals. The tool compartment stresses the utilization of neural system standards that development to—or are themselves utilized as a part of—building, money related, and other down to earth applications.

The accompanying points disclose how to utilize graphical instruments for preparing neural systems to tackle issues in work fitting, design acknowledgment, bunching, and time arrangement. Utilizing these instruments can give you a great prologue to the utilization of the neural system toolbox programming:

1.2.1 Procedure of the working of NN toolbox:

Four methods for operating NN toolbox.
Apparatuses are applied by principal route. Functionalities of these tools can be accessed from an ace device began by the command nnstart.

These apparatuses give a helpful method to get to the capacities of the tool compartment for the accompanying assignments:

- Function used for fitting is: nftool
- Function for recognizing of pattern: nprtool
- Function for clustering of data: nctool
- Function for analyzing the series of time: ntstool

The consecutive method for the utilization of NN toolbox is with the help of essential order linear tasks. Summon linear activities provide excessive adaptability as compared apparatuses. On the off chance that this is your first involvement with the tool compartment, the instruments give the best presentation. What's more, the devices can produce contents of recorded matlab code in order to give layouts to make tweaked order linear capacities. The way toward utilizing the instruments to begin with, and afterward producing and altering Matlab contents, is a magnificent method to find out the usefulness of NN toolbox.

The next method for utilizing the NN tool kit by customization. Such propelled ability enables us for making our particular custom NN systems, as yet approaching the full usefulness of the tool compartment. You can make systems with subjective associations, regardless. You have the capacity to prepare them utilizing existing tool kit preparing capacities .

The fourth method to utilize the tool kit is through the capacity to change any of the capacities contained in the tool compartment. Each computational part is composed in matlab code and is completely open.

All subsequent levels of tool stash utilization traverse the fledgling to the master: basic apparatuses control the new client through particular applications, and system customization enables analysts to attempt novel models with insignificant exertion.

1.2.2 Neural Network Design Steps

- Firstly do the collection of the data
- Then a network generation takes place.
- Composition of network takes place.
- Declaration of weights and biases takes place.
- Network is being trained.
- Validation of the network takes place.
- Then the network is ready for application.

1.3 Pattern Recognition

Pattern recognition is a fast growing technique now a days. It is playing a very significant role in various other techniques as well. It is a process with the help of which a pattern is recognized by computer or machine. It helps in putting patterns in to various categories with its reliable and efficient methods. The demand of such technique is increasing at a rapid pace. They can be deployed in real life applications like agriculture, weather forecast ,automatic disease detection and speech recognition etc.

Pattern recognition is a part of artificial intelligence which helps us in finding the regularities or a particular pattern present in the data. It is firmly related to technique called machine learning. Pattern recognition are basically used in vision of computer. It is an essential part of artificial intelligence that intends t provide human intelligence to the machines or the computer. It is having the capability to solving complex problems, efficient classification of the data and solve various other real world problems as well. It supports many versatile fields like computer science, mathematics and cognitive science etc.

It includes three basic elements which are features, pattern and classifier. Feature means the prominent attribute or a characteristic of data of an image. Features can include the numeric data like height or the color. Classifiers help in the division of the feature space into various decision regions. A decision region is separated by a decision boundary. A general pattern recognition system is composed of a sensor, a mechanism for preprocessing , a mechanism for feature extraction, classification algorithm and a training set.

Sensor:
It is an equipment which is used to recognize the actual physical object. If gives output usually in digital form so that in can be easily processed by the machine. The sensor is usually chosen from the sensors which are already existing.

Pre-processing:

It helps in the production of an efficient set of data by doing noise filtering, smoothing and normalization. It usually processes larger amount of data and reduces the various variations present in it. It helps in safe keeping of an image from various errors.

Feature Extraction:

It is basically used to collect the required information from the data input used by the sensor so that the classification can be done easily. It is usually done with the help of a software which can be modified according to the sensor.

Classification:

It is a technique which is used to do the classification of the object based upon its properties. It uses the various features which have been classified by the feature extraction and assign it to various classes according to its attributes. There are various categories of classification like nearest mean classification, classification using feed forward artificial neural network etc which are being used according to the requirement.

1.3.1 Pattern Recognition Technique:

Pattern recognition comprises of the following three things:

- Preprocessing
- Feature Extraction
- Classification

The foremost priority is to get a database which totally depends on the application. After the acquisition of the appropriate database it is pre-processed so that in can be efficiently used for the further steps. Features which have been carefully extracted from the database are then converted to the feature vectors. These feature vectors help in the statistical representation of the data. So, according to the application domain the classification of these features is done.

1.3.2 Preprocessing:

It is the initial step which is being performed when we do pattern recognition. It is quite effective for the further stages of the pattern recognition. It helps in improving the performance and efficiency of the system. It helps in the production of the consistent set of data by reducing various inconsistencies present in it. It protects the image from various errors .It helps in the doing the image segmentation i.e. dividing the image into smaller parts as per requirements. Segmentation plays a very critical role when it comes to the infection detection in the agricultural products.

1.3.3 Feature extraction

Feature extraction helps in solving the problem of high dimensionality of the input set which we are providing in pattern recognition. It helps in a decreasing the representation of set of features as it transforms the data into the feature vector. It only extracts the useful information from the input which is being provided so that we can get the desired results. The computation process of the extracted features is quite simple and it doesn't respond to various distortions and variations occurring in the images. Then the features which provide most exact and favorable outcome are selected. feature extraction can be implemented by various techniques like Fourier transform, Fuzzy invariant vector, Gabor transform and radon transform etc.

a) Fourier transform:

Fourier transform is basically used to examine the signal in order to check its frequency. It has properties like rotational property and translation property. It takes into account the spectrum magnitude and excludes the circular shift effect.

b) Fuzzy invariant vector:

After the feature has been extracted, it is then converted to a fuzzy invariant vector which helps in lowering down the effect of the noise which is occurring due to lower frequency. It also increases the discrimination. The computation of the fuzzy invariant vector is done using the fuzzy numbers. In this each and every harmonic of a pattern of input has a identical distribution.

c) Gabor wavelets transform:

This transform helps in the analysis time and frequency parameters. This transformation is based on the wavelets which is useful for the feature extraction. It also provides effective resolution. for doing effective pattern recognition it extracts the features of the input data locally. It works in three domains. These three domains are biological, empirical and mathematical. Its resemblance to the human vision system provides excellent results.

d) Radon transformation:

In this type of transformation the mapping is done by the coordinate system. Mapping is done with the help of the Cartesian coordinates and the polar coordinates. It helps in doing the projection of the image with the help of certain angles. The final result of the projection is the sum of the various intensities of the pixels in the certain direction. By projecting the image into different orientation slices, the transform does the capturing of the features quite effectively. This type of transform can be very well implemented in the Fourier domain.

1.3.4 Classification

The versatile set of required features extracted from various patterns in the past stages are utilized here. Here the classification and recognition of the features is done and they are mapped to their respective classes. Learning procedures are categorized into two parts. One is the supervised learning and other one is unsupervised learning. In the case of supervised learning, the classifiers are very well aware of the each and every pattern category among various pattern classes. Where as in the case of the unsupervised learning , the various attributes of the system are modified on the basis of input given to the system. It searches for the similar patterns in the data and find the correct output values.

a) Fuzzy ART:
ART stands for Adaptive Resonance Theory. It very well adaptable with the brain of human beings while doing the processing of the information. It can very well remember the

information without forgetting the information which is previously stored. Here the distribution of the various patterns is done on the basis of the previously stored patterns. A new pattern is formed , if no resemblance is found from the existing patterns.

b) Neural Networks:
Neural networks are mainly based on the biological concept in order to do the recognition of the patterns. It is very effective and powerful tool in order to achieve higher performance. It gathers the information from the human brain. The classification if done by mapping the feature space with the resultant classes. It is acts as a link between the input as well as the output sets. For better results multiple set of neural networks can be used. By using various combining methods input pattern is classified.

c) Markov random field:
MRF is used to combine the statistical data and the information of a particular structure. It extracts the states which are highly effective and with the help of this they design the statistical data.

d) Support Vector Machine
The Support Vector Machine classifier can deal with directly divisible information and in addition non-straightly distinct information utilizing part works. The piece capacity, for example, polynomial, Gaussian outspread premise work, exponential spiral premise work, spline , wavelet and autocorrelation wavelet bit, can outline preparing cases in input space into a highlight space.

e) Multi-class SVM
Multi-class SVM can be favored as a meta-level student. Multi-Classifier framework is higher order precision, in view of SVM for design acknowledgment. This combinational technique classifier depends on stacked speculation which consolidate classifiers from various students, having a two-level structure.

Example acknowledgment strategies are utilized as a part of horticulture applications. A product model framework for rice sickness discovery is produced by utilizing design acknowledgment strategies . Its database comprises of the contaminated pictures of different rice plants which handled by picture developing, picture division methods to distinguish contaminated parts of the plants. At that point the tainted piece of the leaf has been utilized for the arrangement reason utilizing neural system.

CHAPTER-2

LITERATURE REVIEW:

2.1 Paper Review: S.Phadiar ,J.Sil, Rice Disease Identification using Pattern Recognition Techniques, Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008), 25-27 December, pp.420 - 423, 2008.

Objective:
The aim of this paper is to describe a software prototype system for the detection of disease in rice plant on the basis of various images of the rice plants. Images of the infected part of the rice plant are taken using digital camera. In order to detect the defected part of the plant various techniques like image segmentation, image growing etc. have been used. By using neural network the infected part f the leaf is classified. Image processing and soft computing techniques have been applied on infected rice plant.

Techniques/Methodology adopted in paper:

- Image processing and pattern analysis methods.
- Hue Intensity Saturation (HIS) model
- Bi-level thresholding method
- Boundary detection algorithm using 8- connectivity method
- Self organizing map (SOM)

Findings/Results:

- In this research paper , the infected part of the rice plant is being classified using SOM(Self Organizing Map)neural network where the images are being obtained by doing the extraction of the infected part while four different types of images are being used for the testing purposes.
- Usage of zooming algorithm is also there for feature extraction of the image. Zooming algorithm by the usage of computationally efficient technique extracts the features of the image.

Limitation of the Proposed schemes:

- Results of zooming algorithm are not up to the mark and require some improvement.. Results shown by this are not that much extraordinary infact they are just satisfactory.
- It has also been observed that transformation of image in the frequency domain does not give a better classification when compared with the original image.

Future Scope:
By using efficient pattern recognition techniques, the system will be able to do the timely diagnosis of the field problem and the suggestion will help the farmers to take the appropriate measure to increase the quality of the crop .It will not only reduce the development cost in the future but also save the environment also.

2.2 Paper Review: A. Meunkaewjinda, P. Kumsawat and K. Attakitmongcol, Grape leaf disease detection from color imagery using hybrid intelligent system, 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Volume: 1,pp. 513 - 516,2008

Objective:
The aim of this paper is to do automatic plant disease diagnosis with the usage of multiple artificial intelligent techniques. In this paper the main focus is on the grape leaf disease. Once the system is trained, it can diagnose the plant leaf disease without doing its maintenance again and again from the beginning.

Techniques/Methodology adopted in paper:

- Genetic algorithm for optimization.
- Support vector machines for classification.
- Artificial neural network.
- Back-propagation neural network(BPNN).
- Anisotropic diffusion technique.
- Modified self-organizing feature map(MSOFM)

Findings/Results:

- Grape leaf disease extraction and classification system using color imagery, the system gives the desirable results. Back-propagation neural network provides efficient grape leaf extraction with complex background where as Modified self-organizing feature map and Genetic algorithm provides automatic adjustment for the colour extraction of the diseased grape leaf.

- This system has tested images of 426x568 pixels. There were 497 scab disease samples, 489 rust disease sample and 492 non disease samples used to train the SVMs. For testing stage there were 39 scab disease images,41 rust disease images and 35 non-disease images. The results show that the system provides desirable performance.

Limitation of the Proposed schemes:

- There were some problems for doing extraction of ambiguous colour pixels from the background of the image.
- Neural network do not allow better segmentation of the grape leaf disease pixels.

Future Scope:
The system will demonstrate automatic diagnosis capability with very effective performance for the further agricultural product analysis/inspection system development.
More computational effort will definitely help in the classification and recognition of the experimental results.

2.3 Paper review: E.Omrani, B.Khoshnevisan, S.Shamshirband, H.Saboohi, N.B.Anuar, M.H.N.M.Nasir, 'Potential of radial basis function-based support vector regression for apple disease detection', Department of Biosystem Engineering, pp.2-19,2014.

Objective:_The aim of this paper is to classify disease using soft computing approaches, Artificial Neural Networks(ANNs) , Support Vector Machines(SVM) in apple.

Techniques/Methodology adopted in paper:

- K-means clustering
- Artificial neural networks (ANNs)
- Support vector machines (SVMs)
- Gray level co-occurrence matrix (GLCM)
- Polynomial based (SVR_Poly)
- RBF-based SVR (SVR_rbf)
- Wavelet transform
- Principal component analysis (PCA)
- Back-propagation neural network.

Findings/Results:

- The usage of K-means and ANNs for clustering and classifying diseases affecting the leaves of plant. The outcome is 94% correct and swifter by 20% .
- K-means cluster divided the images into two groups: infected and healthy leaf areas. The diseases were classified based on the extracted features.
- The SVR_rbf model had a very small RMSE (0.13) during training and the value was 0.2 in testing.
- The SVR_poly had RMSE of 0.39 in training and RMSE of 0.42 during testing. It was seen that SVR_rbf model showed consistently good correlation throughout training and testing.
- A comparison of SVR_rbf results with SVR_poly and ANN reveals that SVR_rbf outperforms the POLY model in terms of prediction accuracy.

Limitation of the Proposed schemes:

- The Artificial neural networks (ANNs) approach didn't provide accurate results for the disease classification.
- The results showed by SVR-polt in detecting apple leaf diseases were not exact.

Future Scope:

- Usage of SVR algorithm should be increased as this algorithm contains the quadratic programming function resolution which is a work function that leads to a unique, optimum, and comprehensive solution.
- The SVR approaches against the ANN results demonstrated interesting improvement in the prediction system. It is potentially a promising alternative to existing prediction models.

2.4 Paper Review: A.Singh,B.Ganapathysubramanian, A.K.Singh,S.Sarkar, Machine Learning for High-Throughput Stress Phenotyping in Plants,Trends in Plant Science, Volume 21, Issue 2, February 2016, Pages 110-124.

Objective:
The aim of this paper is to give us an overview regarding the work done in the field of plant stress phenotyping using Machine Learning, classification, quantification and prediction. It will also tell about the general issues in Machine Learning strategy.

Techniques/Methodology adopted in paper:

- High-throughput phenotyping (HTP)
- High-throughput stress phenotyping (HTSP).
- ML algorithms
- Support vector machines (SVM)
- Artificial Neural Networks(ANN)

Findings/Results:
This review gave us an overview of Machine Learning and with the various advantages of machine learning in the future. The concepts discussed here can be applied to data collected across the spectrum of complexity and sophistication. We have identified several future avenues for using ML techniques that show tremendous promise but remain currently unutilized by the phenotyping community.

Limitation of the Proposed schemes:

- The features identified in the unsupervised process may not be meaningful to a human user.
- Generative model does not give better results as compared to the discriminative model.
- Discriminative model is less robust when it comes to over fitting issues.

Future Scope:

- Machine Learning approaches are scalable and also can provide modular strategy for the data analysis especially for the new domain of 'plant stress analysis'.
- It will also help in the gene discovery process as well as the introduction of novel selections protocols for the complex competitive traits like biotic and abiotic stress and yield.

2.5 Paper Review: A.Camargo, J.S.Smith, An image-processing based algorithm to automatically identify plant disease visual symptoms, Biosystems Engineering ,Volume 102, Issue 1, January 2009, pp. 9-21.

Objective:

The aim of this paper is to do the automatic identification of the plant disease by image processing from the visual symptoms by analyzing the colored images.

Techniques/Methodology adopted in paper:

- Image pre-processing
- Image enhancement
- Image segmentation
- Image post-processing

Findings/Results:

- The test set consisted of 20 images which were showing symptoms of plant disease in different crops used in the study. To create the manually segmented set of images, a grid was overlaid on the image and each position was then evaluated the white colour and black colour. White colour (1)depicted the pixel having diseased symptoms whereas the black (0) for non-diseased region.

- To evaluate the algorithm, original images were automatically segmented. The output which was produced was a binary image where 1 represented a pixel classified as diseased and 0 as non diseased.

Limitation of the Proposed schemes:

- Huge variation was there in results. To develop such detection system is a very difficult and challenging task.

- Accuracy is a major criteria in disease detection system.

Future Scope:

- The strength of this algorithm is its ability to identify the correct target (disea sed region) which is shown in the images with different range of intensities distribution which will definitely help in the future.

2.6 Paper review: S.Bashir,N.Sharma, Remote Area Plant disease detection using Image Processing', IOSR Journal of Electronics and Communication Engineering , Volume 2, Issue 6 ,pp.31-34,2012.

Objective of Paper: Disease can be recognized by using color and texture features. Disease detection in Malus Domestica is using K mean clustering ,color and texture analysis.

Techniques/Methodology adopted in paper:

- Otsu Segmentation
- K-mean clustering
- Back propagation
- Neural Network
- Co-occurrence matrix method

Findings/Results:

- Appropriate enhancement of images uses histograms. Image segmentation is used for presence of adequate symptoms for detection of disease.
- Spot on the image can be detected by texture segmentation. Rough, silky, bumpy texture of image can be identified by texture analysis. Texture analysis uses co-occurrence matrix method ,which uses Hue Saturation Intensity color space representation.
- Colorfulness in HIS space is given by the saturation component and transformation of color space can be done easily.

Limitation of the Proposed schemes:

- It is only limited to symptoms of particular disease not all.
- Large training sets required to recognize various leaves with pest or damaged leaves due to insects or diseases.

Future Scope:

Conventional method of disease detection in plants using naked eye was cumbersome and not much effective but by using computer vision toolbox the disease detection in plants is less time consuming and more efficient .

2.7 Paper review:S.Arivazhagan, R.N.Shebiah, S.Ananthi ,S.V.Varthini ,Using texture detection features, recoginizing unhealthy region of plant leaves and classification of plant leaf disease, Agric Eng Int: CIGR Journal, Vol. 15, No.1,pp.211-217,2013

Objective of Paper: The aim of the paper is to do detection of unhealthy region of plant leaves and classification of plant leaf disease using texture features.

Technique adopted in paper:

- RGB image acquisition
- Color transformation
- Masking and removal
- Mapping of RGB
- Segment the components
- Obtain the useful segments
- Computation of texture
- Classifier

Findings /Results:

- Green color pixels identified based on specified value of threshold . If green component is less than the threshold value ,the red ,green and blue component of the pixel is assigned zero and pixels are completely removed.
- Patch size of 32X32pixels chosen such that significant information is not lost.
- Various formulas for Contrast ,Energy,Local homogeneity,Cluster shade and cluster prominence .
- Minimum distance criteria is used for classification phase, the c-occcurence features for the leaves are compared with the corresponding values in feature library.
- Finite set of elements was drawn by using Support Vector Machine(SVM).
- For training the system 5% of the leaf images from each group are used and remaining serves as the set for testing.

Limitation of the proposed schemes:

- Large number of training samples are needed for disease identification rate.
- Image of the diseased leaf should be properly captured.

Future Scope:

Little computational effort can be used to classify and recognize the experiment results.
Training sample can be increased to improve disease identification rate .

2.8 Paper Review: J.Behmann ,A.K.Mahlein ,T.Rumpf , C. Romer ,L.Plumer ,A review of advanced machine learning methods for the detection of biotic stress in precision crop protection, Springer Media New York, Precision Agriculture,Volume 16, Issue 3, pp 239–260,June 2015

Objective of Paper: Biotic stress detection in precision crop protection using advanced machine learning methods.

Techniques/Methodlogy adopted in paper:

- Support vector machine
- Support vector regression
- Neural networks
- HSV-color space,(Hue ,Saturation,Value)
- Image Segmentation
- K-Means clustering
- RGB Color

Findings /Results:

- Kernel functions used to map the data requiring non-linear discrimination functions and in this feature space the data is linearly separable. It is used by support vector machine.
- Non-Linear SVM and Neural Networks gives good prediction accuracies than linear approaches. On the other hand, best classification performance in the study was given by the SVM compared to Neural Network.
- Grey scale image is used for extraction of texture features.
- High dimensional data is analysed with unknown statistical characteristics for precision crop protection by using machine learning methods.

Limitation of the proposed schemes:

- Stress effect of weeds and nitrogen in maize Neural Network gives accuracy upto (69%to 58%)
- Non-Linear Support Vector Machine gives accuracy upto 85% and is superiror to a linear SVM(Support Vector Machine)
- Algorithms used are very specific and generalizations of the prediction accuracies are not justified.

Future Scope:

Non-relevant or redundant features lead to decrease significantly the prediction accuracy and reducing these number of features is an important step in data analysis.

2.9 Paper Review:J.G.A.Barbedo,Digital image processing techniques for detecting, quantifying and classifying plant diseases, pp.1-12,Barbedo SpringerPlus ,2013

Objective of Paper: Detecting, quantifying and classifying plant diseases using digital image processing techniques.

Techniques/Methodology adopted in paper:

- Neural Networks
- Thresholding
- Dual-segmented regression analysis
- Quantification
- Colour analysis
- Fuzzy Logic
- Knowledge-based system
- Sobel operator
- Chlorosis algorithm

Findings/ Results:

- Background is discriminated from the leaf and then damaged regions is separated from healthy surface and the ratio between the number of pixel in damage by the total number of pixel of the leaf gives the final estimate.
- Subsequent steps uses 2 modification versions of I3 and only one modification of H. Sepration of diseased and healthy region is done using thresholding.
- Red and green component of image are combined using chlorosis algorithm for determining the yellowness of leaf. To discriminate leaves from background blue component is used.To identify and quantify the necrotic region is done Necrosis algorithm.
- Area occupied by the spots is estimated using thresholding the blue component of the image and algorithm to implement using white spots algorithm.

Limitation of the proposed schemes:

- Real-time monitoring was not used.
- Over fitting, overtraining, undersized sample sets and sample sets with low representativeness are the major problems.

Future Scope:

Crops are continuously monitored by the Real-time monitoring and alarm willbe issued as soon as disease is detected.

2.10 F.Qin, D.Liu, B.Sun, L.Ruan, Z.Ma, H.Wang,Identification of Alfalfa Leaf Diseases Using ImageRecognitionTechnology,p.p.1-15,Plos Journals,December 15, 2016

Objective of Paper:Using Image recognition technology identification of Alfalfa leaf diseases.

Techniques/Methodology adopted in paper:

- Fuzzy *C*-MEANS CLUSTERING
- *K*-MEDIAN CLUSTERING
- Euclidean DISTANCE
- Logistic REGRESSION ANALYSIS
- Naive BAYES ALGORITHM
- Cart
- LINEAR DISCRIMINANT ANALYSIS

Findings/ Results:

- Arithmetic square root of total number of features was randomly selected by each decision tree .For eg. If arithmetic square root is decimal ,then rounding up the decimal gives the number of features randomly selected by each decision tree.
- Disease recognition models built after feature selection gives the satisfactory recognition results. This indicates that features extracted from lesion images were efficient.
- Relief method gives the top 45 features for importance ranking based on the SVM model.
- For implementing K-median clustering algorithm linear discriminant was used and the highest score of median and mean are used for implementation.

Limitation of the proposed schemes:

- To get accurate results there is need of large datasets.
- Accuracy and efficiency depends on the experience and it is time consuming and subjective work.

Future Scope:

Optimal image recognition model of alfalfa leaf diseases can be done by developing mobile application.

2.11 Paper Review: **A.Meunkaewjinda**, **P.Kumsawat** and **K.Attakitmongcol**,Grape leaf disease detection from color imagery using hybrid intelligent system, 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Volume: 1,pp. **513 - 516,2008**

Objective: The aim of this paper is to do automatic plant disease diagnosis with the usage of multiple artificial intelligent techniques. In this paper the main focusis on the grape leaf disease. Once the system is trained, it can diagnose the plant leaf disease without doing its maintenance again and again from the beginning.

Techniques/Methodology adopted in paper:

- Genetic algorithm for optimization.
- Support vector machines for classification.
- Artificial neural network.
- Back-propagation neural network(BPNN).
- Anisotropic diffusion technique.
- Modified self-organizing feature map(MSOFM)

Findings/Results:

- Grape leaf disease extraction and classification system using color imagery, the system gives the desirable results. Back-propagation neural network provides efficient grape leaf extraction with complex background where as Modified self-organizing feature map and Genetic algorithm provides automatic adjustment for the colour extraction of the diseased grape leaf.

- This system has tested images of 426x568 pixels. There were 497 scab disease samples, 489 rust disease sample and 492 non disease samples used to train the SVMs. For testing stage there were 39 scab disease images,41 rust disease images and 35 non-disease images. The results show that the system provides desirable performance.

Limitation of the Proposed schemes:

- There were some problems for doing extraction of ambiguous colour pixels from the background of the image.
- Neural network do not allow better segmentation of the grape leaf disease pixels.

Future Scope:

- The system will demonstrate automatic diagnosis capability with very effective performance for the further agricultural product analysis/inspection system development.
- More computational effort will definitely help in the classification and recognition of the experimental results.

2.12 Paper review: E.Omrani, B.Khoshnevisan, S.Shamshirband, H.Saboohi, N.B.Anuar, M.H.N.M.Nasir, 'Potential of radial basis function-based support vector regression for apple disease detection', Department of Biosystem Engineering, pp.2-19,2014.

Objective: The aim of this paper is to classify disease using soft computing approaches, Artificial Neural Networks(ANNs) , Support Vector Machines(SVM) in apple.

Techniques/Methodology adopted in paper:

- K-means clustering
- Artificial neural networks (ANNs)
- Support vector machines (SVMs)
- Gray level co-occurrence matrix (GLCM)
- Polynomial based (SVR_Poly)
- RBF-based SVR (SVR_rbf)
- Wavelet transform
- Principal component analysis (PCA)
- Back-propagation neural network.

Findings/Results:

- The usage of K-means and ANNs for clustering and classifying diseases affecting the leaves of plant. The experimental results revealed that proposed approach can successfully detect and classify the inspects disease with 94% precision and 20% faster.

- K-means cluster divided the images into two groups: infected and healthy leaf areas. The diseases were classified based on the extracted features.

- The SVR_rbf model had a very small RMSE (0.13) during training and the value was 0.2 in testing.

- The SVR_poly had RMSE of 0.39 in training and RMSE of 0.42 during testing. It was seen that SVR_rbf model showed consistently good correlation throughout training and testing.

- A comparison of SVR_rbf results with SVR_poly and ANN reveals that SVR_rbf outperforms the POLY model in terms of prediction accuracy.

Limitation of the Proposed schemes:

- The Artificial neural networks (ANNs) approach didn't provide accurate results for the disease classification.
- The results showed by SVR-polt in detecting apple leaf diseases were not exact.

Future Scope:

- Usage of SVR algorithm should be increased as this algorithm contains the quadratic programming function resolution which is a work function that leads to a unique, optimum, and comprehensive solution.

- The SVR approaches against the ANN results demonstrated interesting improvement in the prediction system. It is potentially a promising alternative to existing prediction models.

2.13 Paper Review:A.Singh,B.Ganapathysubramanian, A.K.Singh,S.Sarkar, Machine Learning for High-Throughput Stress Phenotyping in Plants,Trends in Plant Science, Volume 21, Issue 2, February 2016, Pages 110-124.

Objective:

The aim of this paper is to give us an overview regarding the work done in the field of plant stress phenotyping using Machine Learning, classification, quantification and prediction. It will also tell about the general issues in Machine Learning strategy.

Techniques/Methodology adopted in paper:

- High-throughput phenotyping (HTP)
- High-throughput stress phenotyping (HTSP).
- ML algorithms
- Support vector machines (SVM)
- Artificial Neural Networks(ANN)

Findings/Results:

This review gave us an overview of Machine Learning and with the various advantages of machine learning in the future. The concepts discussed here can be applied to data collected across the spectrum of complexity and sophistication. We have identified several future avenues for using ML techniques that show tremendous promise but remain currently unutilized by the phenotyping community.

Limitation of the Proposed schemes:

- The features identified in the unsupervised process may not be meaningful to a human user.
- Generative model does not give better results as compared to the discriminative model.
- Discriminative model is less robust when it comes to over fitting issues.

Future Scope:

- Machine Learning approaches are scalable and also can provide modular strategy for the data analysis especially for the new domain of 'plant stress analysis'.

- It will also help in the gene discovery process as well as the introduction of novel selections protocols for the complex competitive traits like biotic and abiotic stress and yield.

2.14 Paper Review:VijaiSingh, A.K.Misra, Detection ofxplant leaf diseases using image segmentation and soft computing techniques,Volume 4,Issue 1, March 2017, PP. 41-49.

Objective:

The aim of this paper we do the automatic detection and classification of plant leaf diseases using an algorithm for image segmentation technique. It also covers survey on different diseases classification techniques that can be used for plant leaf disease detection.

Techniques adopted in paper:

- Genetic algorithm
- k-nearest-neighbor method
- Machine learning based recognition
- Color Co-occurrence Method
- Artificial neural network (ANN)

Findings/Results:

- By using Minimum Distance Criterion with K-Mean Clustering we did the first classification which showed its efficiency and accuracy of 86.54% . The detection accuracy was later improved to 93.63% by the proposed algorithm.
- The second phase of classification was done using SVM classifier and shows efficiency and accuracy of 95.71%.But with the help of proposed algorithm we were able to improve the detection accuracy to 95.71%.
- From the results it was clearly visible that only few sample from the Frog eye leaf spot and bacterial leaf spot leaves were misclassified. Only two leafs with bacterial leaf spot disease were classified as frog eye leaf spot and one frog eye leaf spot was classified as the bacterial leaf spot. The average accuracy of proposed algorithm is 97.6% .

Limitation of the Proposed schemes:

- If the training data is not linearly separable then it becomes quite difficult to determine the optimal parameters in the Support Vector Machine, which appears to be its major drawback.
- There is a need to do improvement in the recognition rate of the classification process.

Future Scope:

- Genetic algorithm optimizes continuous and discrete variables effectively. It searches for large data samples of the cost surface with large variables being processed at the same time.
- It helps in the optimization of variables with highly complex cost surfaces.

- Neural Networks can be used for the recognition rate of the classification process.

2.15 Paper Review: A.Camargo, J.S.Smith, Anximage-processing based algorithm to automatically identify plant disease visual symptoms, Biosystems Engineering ,Volume 102, Issue 1, January 2009, pp. 9-21.

Objective:

The aim of this paper is to do the automatic identification of the plant disease by image processing from the visual symptoms by analysing the coloured images.

Techniques/Methodology adopted in paper:

- Image pre-processing
- Image enhancement
- Image segmentation
- Image post-processing

Findings/Results:

- The test set consisted of 20 images which were showing symptoms of plant disease in different crops used in the study. To create the manually segmented set of images, a grid was overlaid on the image and each position was then evaluated the white colour and black colour. White colour (1)depicted the pixel having diseased symptoms whereas the black (0) for non-diseased region.

- To evaluate the algorithm, original images were automatically segmented. The output which was produced was a binary image where 1 represented a pixel classified as diseased and 0 as non diseased.

Limitation of the Proposed schemes:

- Huge variation was there in results. To develop such detection system is a very difficult and challenging task.

- Accuracy is a major criteria in disease detection system.

Future Scope:

- The strength of this algorithm is its ability to identify the correct target (diseased region) which is shown in the images with different range of intensities distribution which will definitely help in the future.
- Due to the higher complexity of the images used in this study, the strategy proposed here will be suitable for other type of images as well whose targets are different to that of images which are showing diseased plants.

CHAPTER-3

DATA SET

A dataset is an accumulation of related, discrete things of related information that might be gotten to exclusively or in blend or oversaw all in all element.

An informational collection is sorted out into some kind of information structure. In a database, for instance, an informational collection may contain a gathering of business information (names, pay rates, contact data, deals figures, et cetera). The database itself can be viewed as an informational collection, as can assemblages of information inside it identified with a specific kind of data, for example, deals information for a specific corporate office.

Data set used in the project is stored in an excel file and further used for getting results in Matlab software.
Our data set is a collection of four attributes which are as follows:

- The first component of our dataset comprises of Histogram of the leaf which shows the characteristics of either of diseased leaf or disease free leaf which further is analyzed and outcome is generated accordingly.
- The second component consists of the Hue component which is generally shows the dominant color.
- The third component is S component it shows the purity of the image or addition of white light.
- The last component is the I-Component it states the amplitude of light present in our image.

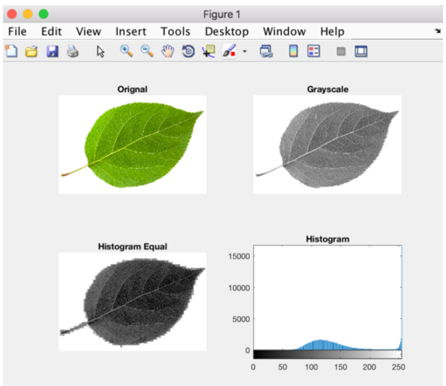


Figure3.1 shows the normal image ,grayscale image, Histogram Equal and the histogram.

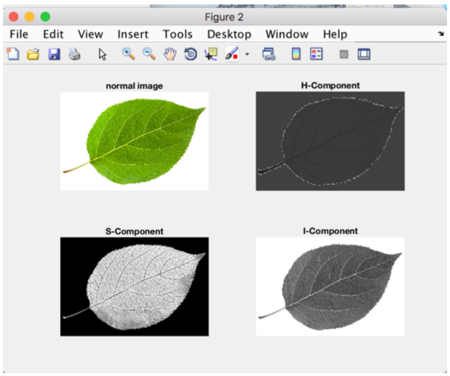


Figure3.2 shows the normal image the H,S and I (Hue,Saturation,Intensity)component of the good leaf.

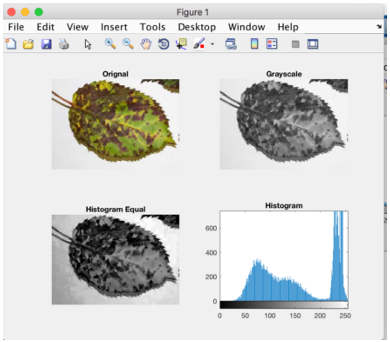


Figure3.3 The Histogram of the grayscale ,Histogram Equal image, Histogram, Diseased leaf.

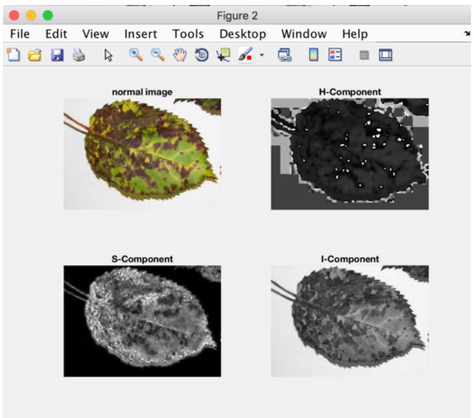
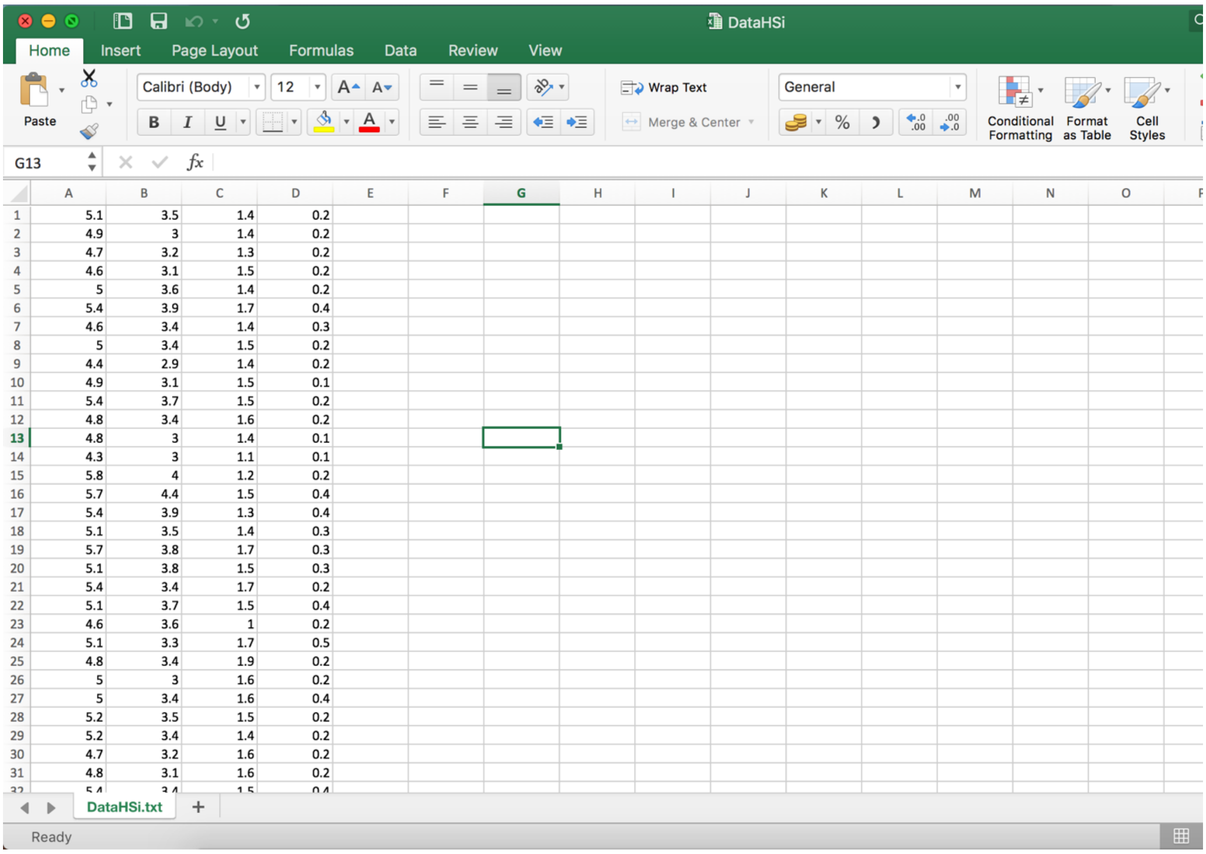


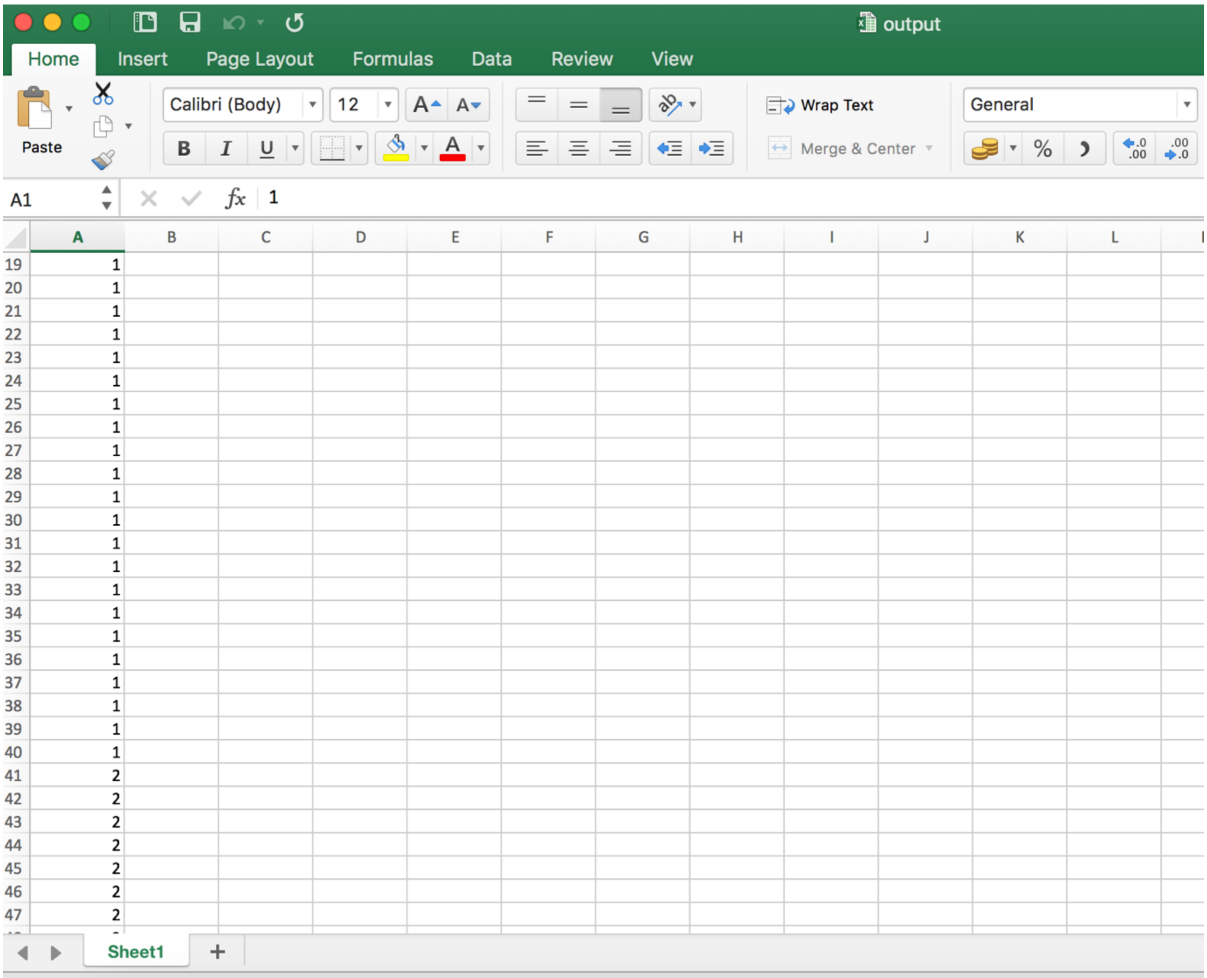
Figure3.4 It shows the normal image, H-component, S-component and I-Component of the diseased leaf.



	A	B	C	D
1	5.1	3.5	1.4	0.2
2	4.9	3	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1
11	5.4	3.7	1.5	0.2
12	4.8	3.4	1.6	0.2
13	4.8	3	1.4	0.1
14	4.3	3	1.1	0.1
15	5.8	4	1.2	0.2
16	5.7	4.4	1.5	0.4
17	5.4	3.9	1.3	0.4
18	5.1	3.5	1.4	0.3
19	5.7	3.8	1.7	0.3
20	5.1	3.8	1.5	0.3
21	5.4	3.4	1.7	0.2
22	5.1	3.7	1.5	0.4
23	4.6	3.6	1	0.2
24	5.1	3.3	1.7	0.5
25	4.8	3.4	1.9	0.2
26	5	3	1.6	0.2
27	5	3.4	1.6	0.4
28	5.2	3.5	1.5	0.2
29	5.2	3.4	1.4	0.2
30	4.7	3.2	1.6	0.2
31	4.8	3.1	1.6	0.2

Table3.1 This is our data set which consists of Histogram, H-Component ,S-Component and I-Component respectively. It is fed as input to the neural network.

Our dataset is not redundant and every time whenever the input is fed to the network a unique result is generated depending upon above stated characteristics.



	A
19	1
20	1
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	1
30	1
31	1
32	1
33	1
34	1
35	1
36	1
37	1
38	1
39	1
40	1
41	2
42	2
43	2
44	2
45	2
46	2
47	2

Table3.2 This figure depicts the target of previously given input in order to get the desired result whether the leaf is prone to disease or not.

Our result is divided into 3 categories that are 1,2 and 3 .They are explained as follows:

- If the value of target is 1 and less than 1.5 then it means that is is a Good leaf which is free from disease.
- Secondly if the value is 2 and less than 2.5 than it semi diseased leaf in which some portion is exposed to some kind of infection.
- Lastly if the target value if 3 and greater than 3.5 than it is completely infected.

6	4.8	3	1.4	0.3	1
7	5.1	3.8	1.6	0.2	1
8	4.6	3.2	1.4	0.2	1
9	5.3	3.7	1.5	0.2	1
10	5	3.3	1.4	0.2	1
11	7	3.2	4.7	1.4	2
12	6.4	3.2	4.5	1.5	2
13	6.9	3.1	4.9	1.5	2
14	5.5	2.3	4	1.3	2
15	6.5	2.8	4.6	1.5	2
16	5.7	2.8	4.5	1.3	2
17	6.3	3.3	4.7	1.6	2
18	4.9	2.4	3.3	1	2
19	6.6	2.9	4.6	1.3	2
20	5.2	2.7	3.9	1.4	2
21	5	2	3.5	1	2
22	6	3	4.8	1.8	3
23	6.9	3.1	5.4	2.1	3
24	6.7	3.1	5.6	2.4	3
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table3.3 These are some of the test inputs which are used for analyzing our NNtool
The first column if the histogram followed by H,S, I-Component followed by the Resultant data.

CHAPTER-4

Data Fitting

Neural systems are great at fitting capacities. Actually, there is evidence that a genuinely basic NN system can adjust itself in viable capacity.
Fit information utilizing bends, surfaces, and nonparametric techniques

Information fitting is the way toward fitting models to information and breaking down the exactness of the fit. Specialists and researchers utilize information fitting procedures, including numerical conditions and nonparametric techniques, to display gained information.

MATLAB gives you a chance to import and envision your information, and perform fundamental fitting systems, for example, polynomial and spline interjection. You can perform information fitting intelligently utilizing the MATLAB Basic Fitting device, or automatically utilizing MATLAB capacities for fitting.

4.1 Data Fitting Tool

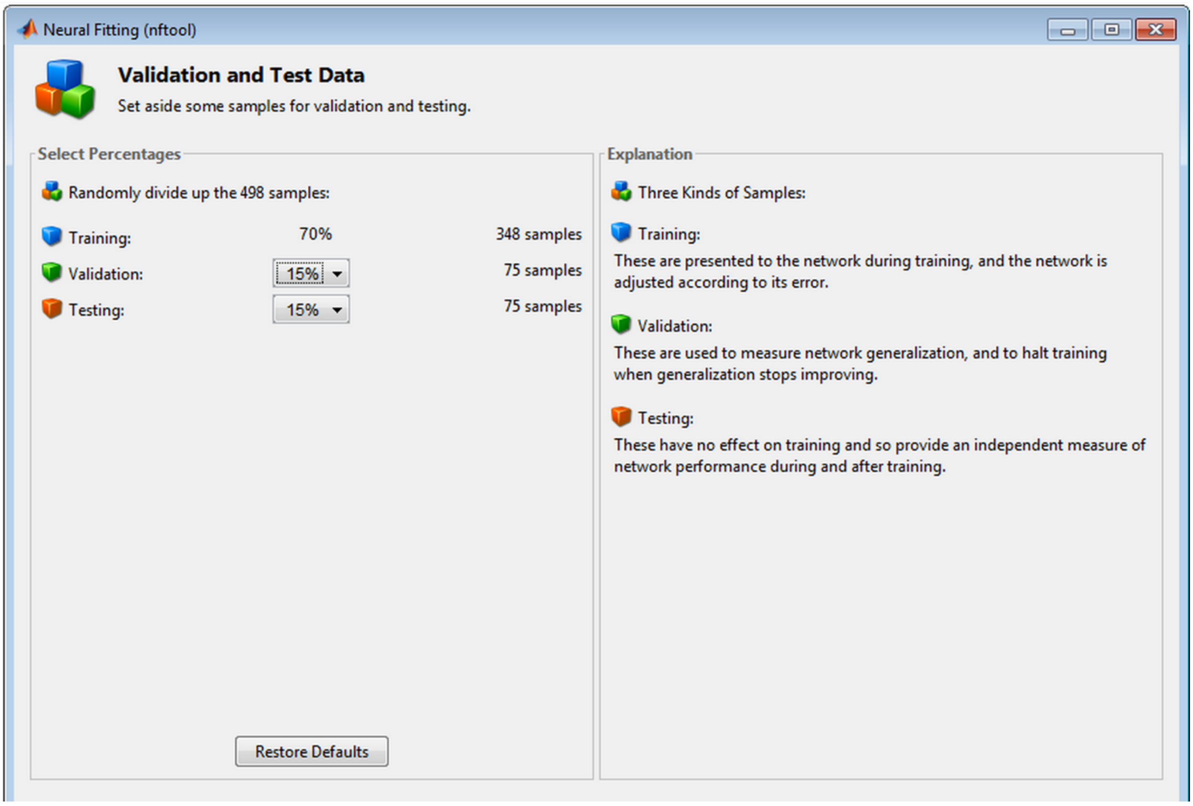


Figure4.1 nftool

Analysis: The info vectors and target vectors will be arbitrarily separated into three sets as takes after:

- 70% of leaf dataset will be utilized for preparing.

- 15% of leaf dataset will be utilized to approve that the system is summing up and to quit preparing before overfitting.

- The last 15% of leaf will be utilized as a totally autonomous trial of system speculation.

The standard framework that is used for work fitting is a two-layer feedforward sort out, with a sigmoid move work in the hid layer and a straight capacity work in the yield layer. The default number of covered neurons is set to 10.

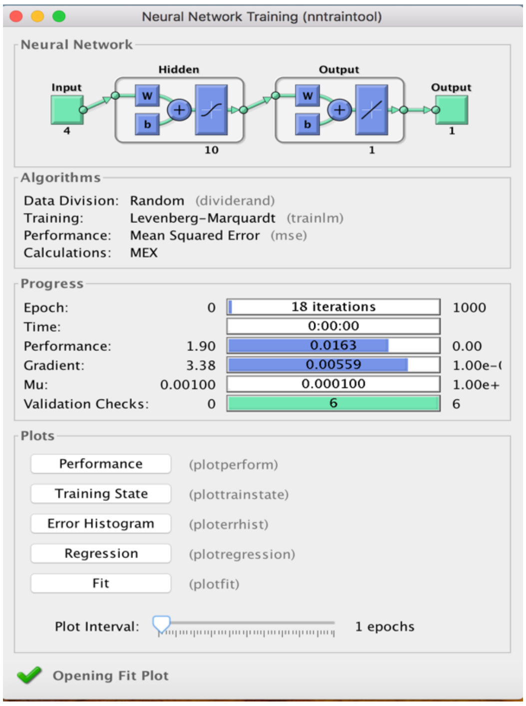


Figure4.2 nntraintool

Analysis: In this number of Epoch are having 18 iteration, Performance of the diseased and diseased free leaf dataset is equivalent to 0.0163,Gradient is equivalent to 0.00559 and the number of validation checks are equal to 6.

4.1.1Performance:

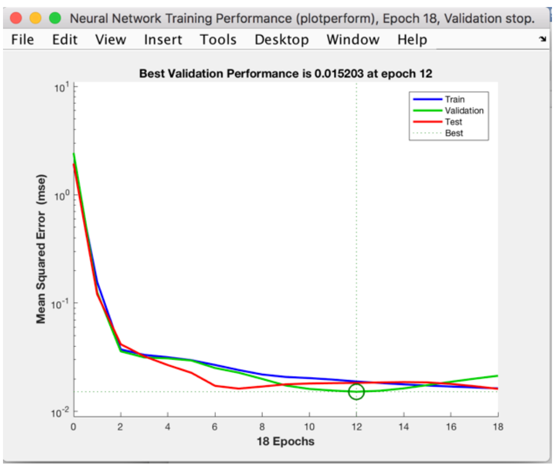


Figure4.3 plotperform

Analysis: After analysis of the data set consisting of diseased, semi-diseased and completely diseased leaves we are getting result at 12 epochs.
plotperform(TR) are basically used to plot the error vs. epoch. It is useful for doing training, validating and testing the performances exhibitions in preparation storage TR evaluated by function train.
By and large, the error diminishes after more epochs of preparing, yet may begin to increment on the approval informational collection as the system begins overfitting the preparation information.

Performance Plot demonstrates you mean square error flow for all your datasets in logarithmic scale. Preparing MSE is continually diminishing, so its approval and test MSE you ought to be occupied with. Your plot demonstrates an impeccable training.
Mean Square Error (MSE) is the mean (average) extent of the squares of the error: i.e., the separation between the model's approx of your test esteems and the genuine test esteem. (squaring just changes over things to a flat out esteem as opposed to fiddling with under or overshooting).

4.1.2Training State:

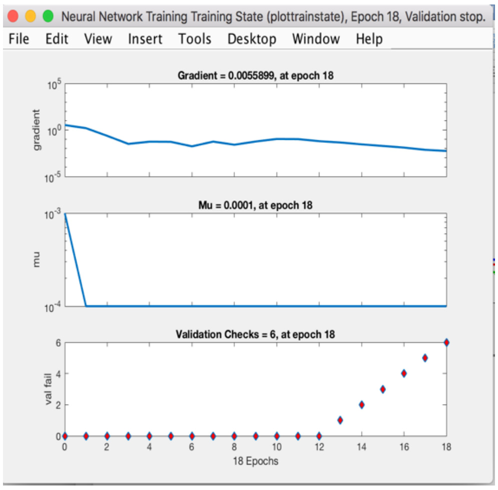


Figure4.4 plottrainstate

Analysis: After training the data set of leaves we get that Matlab naturally quits preparing after 6 flops consecutively.
the figure shows the Gradient =0.055899, at epoch 18 and Mu=0.0001,at epoch 18 and validation checks =6, at epoch at 18.

Training State demonstrates to you some other preparing insights. Gradient is an estimation of back propagation slope on every cycle in logarithmic scale. Validation falls flat are emphases when approval MSE expanded its esteem. A considerable measure of comes up short means overtraining.

4.1.3 Error Histogram:

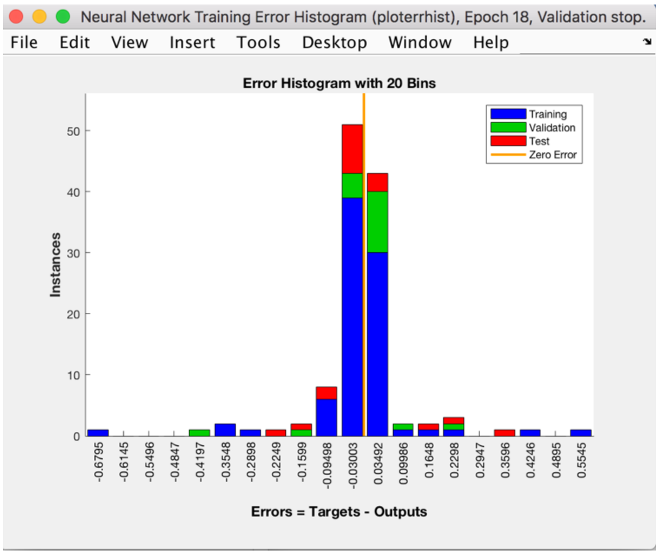


Figure4.5 ploterrhist

Analysis: You can see that while most bungles fall between - 0.4197 and 0.2298, there is a readiness point with a slip-up of 8 and endorsement centers with goofs of 09 and 11. These special cases are furthermore evident on the testing relapse plot. The important contrasts to the point and a goal of 30 and yield near 15.

The blue bars address getting ready data, the green bars address endorsement data, and the red bars address testing data. The histogram can give you an indication of inconsistencies, which are data centers where the fit is by and large more repulsive than the bigger piece of data. It is a brilliant idea to check the irregularities to choose whether the data is terrible, or if those data centers are not the same as whatever is left of the educational accumulation. If the irregularities are considerable data concentrates, yet are not in the least like whatever is left of the data, by then the framework is extrapolating for these core interests. You should assemble more data that looks like the inconsistency centers, and retrain the framework.

As the quantity of the error occurring in the dataset results is very less hence it shows that the classification of diseased ,partially diseased and completely diseased leaves have been done right.

4.1.4 Regression:

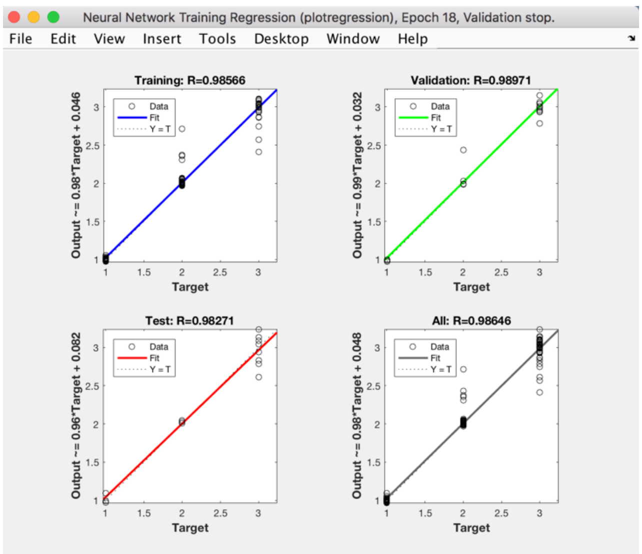


Figure4.6 plotregression

Analysis: The going with relapse plots show the framework yields concerning centers for planning, endorsement, and test sets. For a flawless fit, the data should fall along a 45 degree line, where the framework yields are comparable to the destinations. For this issue, the fit is sensibly helpful for every single educational accumulation, with R regards for every circumstance of 0.97 or above. If fundamentally more correct results were required, you could retrain the framework by clicking Retrain in nftool. This will change the hidden weights and slants of the framework, and may make an improved framework in the wake of retraining. Distinctive options are given on the going with sheet.

CHAPTER-5

PATTERN RECOGNITION USING NEURAL NETWORKS

5.1 Introduction

In pattern recognition the inputs are related with various classes. Neural systems are great at design acknowledgment issues. A neural system with enough components (called neurons) can order any information with discretionary precision. They are especially appropriate for complex choice limit issues over numerous variables. The system will be composed by utilizing the credits of neighborhoods to prepare the system to deliver the right target classes.
Neural systems have demonstrated themselves as capable classifiers and are especially appropriate for tending to non-linear issues. They are likewise great at perceiving patterns.

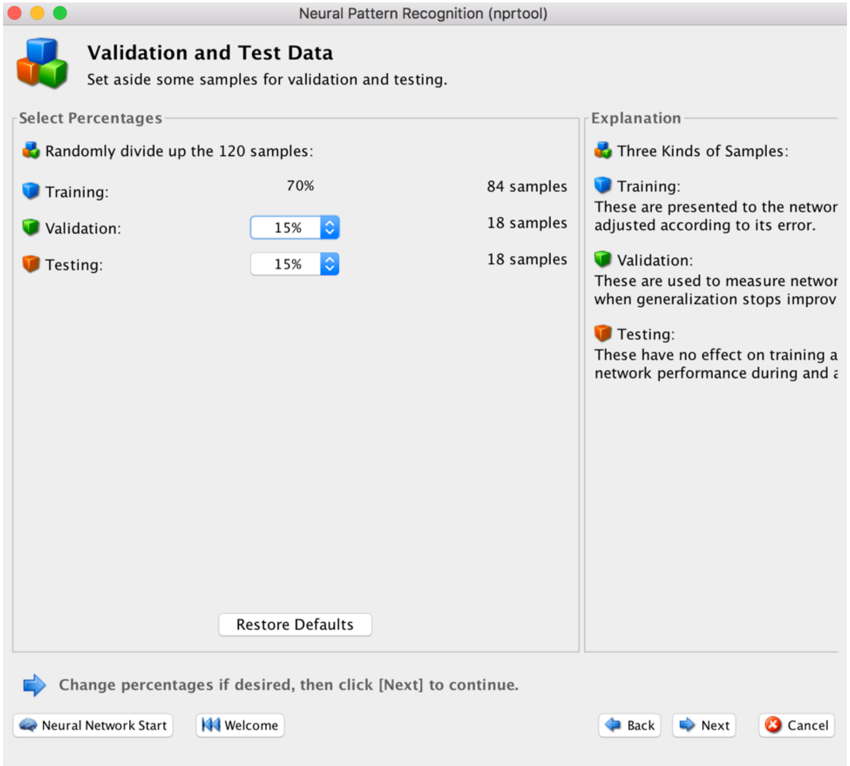


Figure5.1 nprtool

Analysis: Here the validation and test data takes place. Here total 120 samples have been taken into consideration in which 70% of the samples (84 samples) are used for training ,18 samples out of 120 which is 15% of the total sample are used for validation of the sample whereas rest 15% that is 18 samples are been used for testing of the data. Data is trained till the point it doesn't give the proper result during the validation and testing is an independent process.

5.2 Neural Network

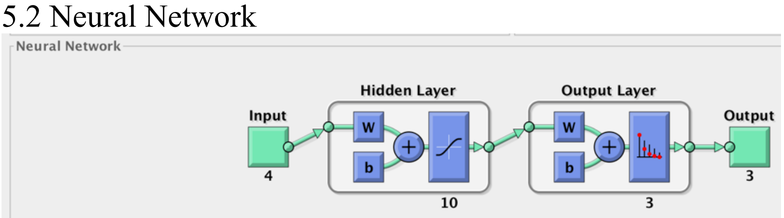


Figure5.2 NN layers

Analysis: It consists of Input which are 4 in number and the 10 hidden layers and the output layer which are 3 in numbers

Feed forward systems comprise of a progression of layers. The primary layer has an association from the system input. Each resulting layer has an association from the past layer. The last layer creates the system's yield.

Feed forward systems can be utilized for any sort of contribution to yield mapping. A feed forward connect with one hidden layer and enough neurons in the hidden layers, can fit any limited information yield mapping issue.

Specific variants of the feed forward organize incorporate fitting (fitnet) and pattern recognition (patternnet) systems. A minor departure from the feed forward organize is the cascade forward system (cascadeforwardnet) which has extra associations from the contribution to each layer, and from each layer to every after layer.

feedforwardnet(hiddenSizes,trainFcn) takes these contentions,

hiddenSizes -Row vector of at least one concealed layer sizes (default = 10)

trainFcn -Preparing capacity (default = 'trainlm')

also, restores a feedforward neural system

5.3 DataSet:

5.3.1 Input Dataset

The input data set consists of 4 x120 matrix representing the static data:

120 samples of 4 elements.

outpu3

4x120 double

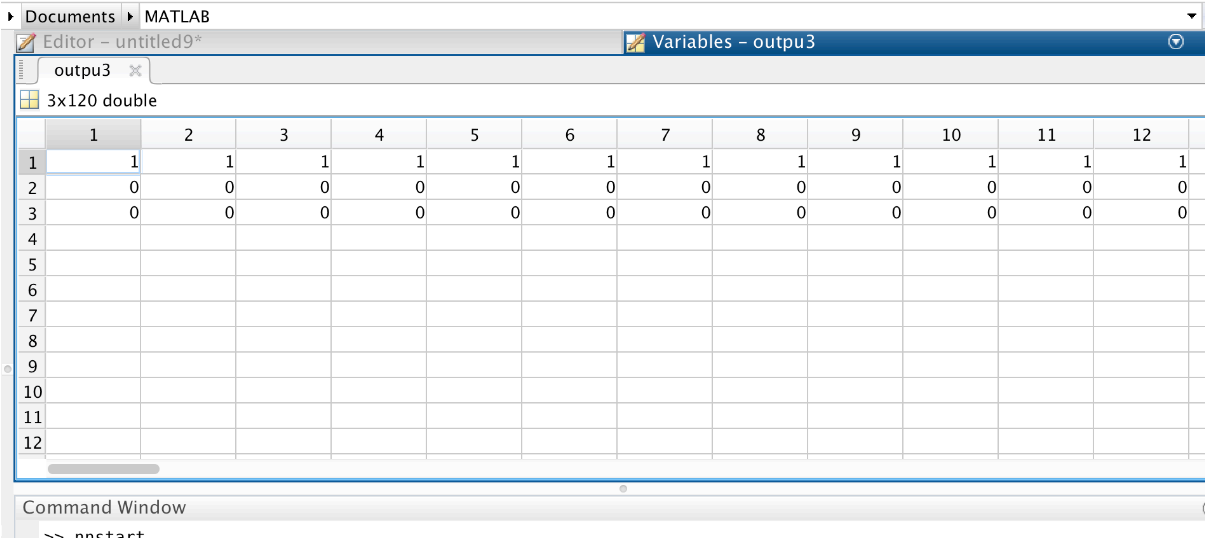
	47	48	49	50	51	52	53	54	55	56	57	58
1	5.8000	6.2000	5.6000	5.9000	6.1000	6.3000	6.1000	6.4000	6.6000	6.8000	6.7000	6
2	2.7000	2.2000	2.5000	3.2000	2.8000	2.5000	2.8000	2.9000	3	2.8000	3	2.9000
3	4.1000	4.5000	3.9000	4.8000	4	4.9000	4.7000	4.3000	4.4000	4.8000	5	4.5000
4	1	1.5000	1.1000	1.8000	1.3000	1.5000	1.2000	1.3000	1.4000	1.4000	1.7000	1.5000
5												
6												
7												
8												
9												
10												
11												
12												

Command Window

Table5.1 Input for recognition of the pattern.

5.3.2 Target Dataset

The target data set consists of 3 x 120 matrix, representing the static data: 120 samples of 3 elements.



	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	1	1	1	1	1	1	1	1	1
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4												
5												
6												
7												
8												
9												
10												
11												
12												

Table5.2 Target for pattern tool.

5.4 Train Data

Neural network training tool it consist of Epoch,time,Performance,Gradient,Valiation Checks. From this we can plot performance, Training State, Error Histogram, Confusion matrix and receiver operating characteristics.

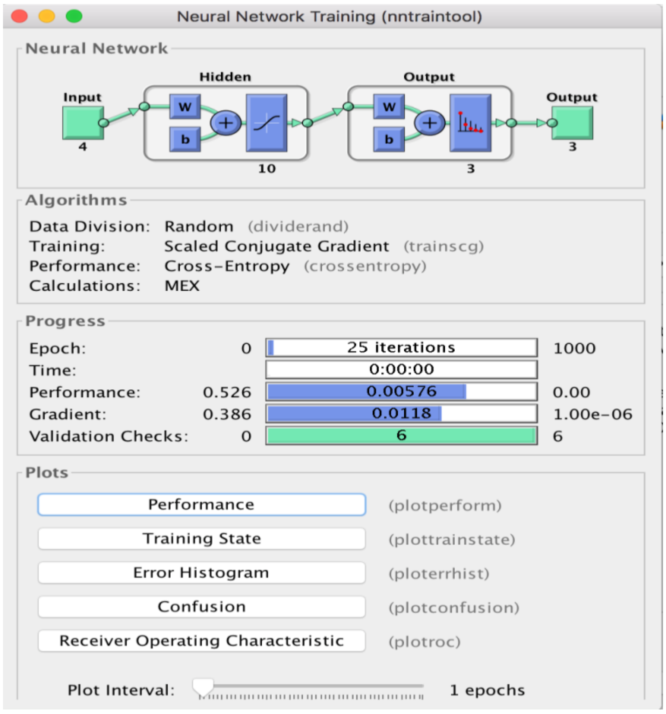


Figure5.3 nntrain tool for pattern

Analysis: In it epoch has 25 iterations ,performance is 0.00576 and the gradient is 0.0118 and the number of validation checks are 6.

5.4.1 Performance:

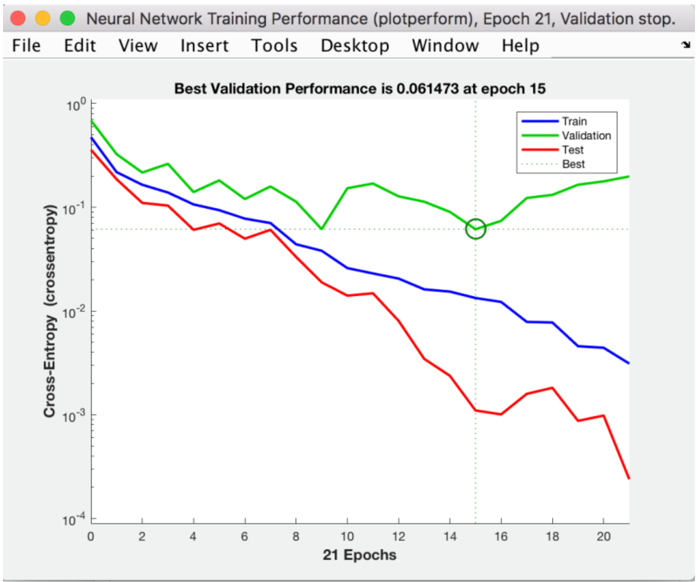


Figure5.4 plotperform for pattern.

Analysis: Performance Plot demonstrates you mean square error flow for all your datasets in logarithmic scale. Preparing MSE is continually diminishing, so its approval and test MSE you ought to be occupied with. Your plot demonstrates an impeccable training.
Mean Square Error (MSE) is the mean (average) extent of the squares of the error: i.e., the separation between the model's approx of your test esteems and the genuine test esteem. (squaring just changes over things to a flat out esteem as opposed to fiddling with under or overshooting).It is giving optimal result at 15 epochs. It is efficiently classifying the diseased and diseased free leaves.

5.4.2 Training state:

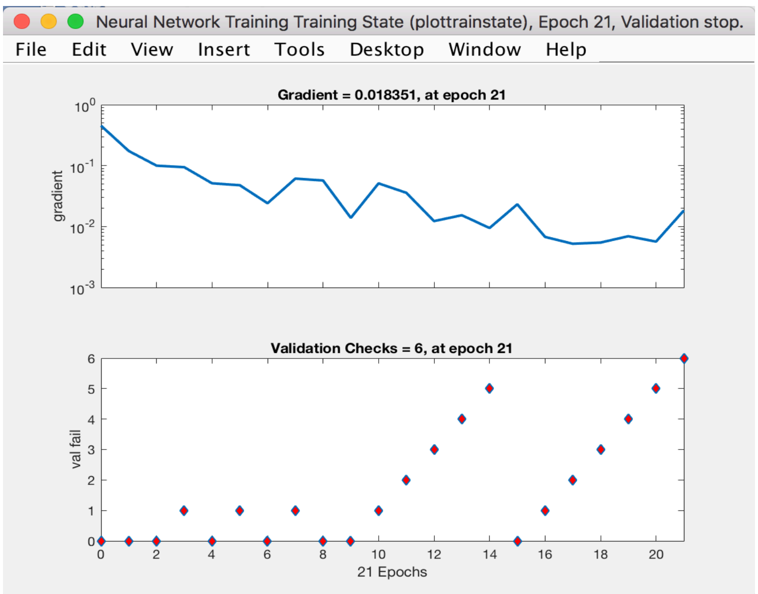


Figure5.5 plottrainstate for pattern.

Analysis: The figure shows the Gradient =0.18351 at epoch 21and validation checks =6, at epoch at 21.

5.4.3 Error Histogram:

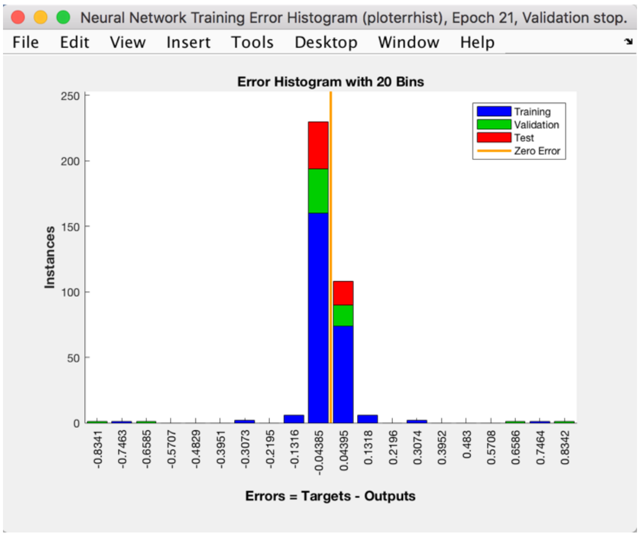


Figure5.6 ploterrhist in pattern.

Analysis: You can see that while most bungles fall between - 0.1316 and 0.1318.It has zero error at nearly 0.1.It works with the help of floor and ceiling function if the value is negative that it shows the lower value and on the other hand if the value is to the positive side it raises the value. The computation of the diseased and diseased free data has been done very accurately.

5.4.4 Confusion Matrix:

Figure5.7 plotconfusion in pattern.

Analysis: It demonstrates the rate of right and of base orders. The green squares refer to the correct classification of the results of the diseased, partially diseased leaves and completely diseased leaves where as on the other hand the red squares refer to the misclassification.
In case of training the confusion matrix the sun of the diagonal elements which are highlighted in the green color sun up to the value of 98.8% which means that it is showing correct results of the dataset of the diseased and diseased free leaves to large extent whereas the value in the red square is mere 1.2% which is a very small error value.
In the case of validation of confusion matrix the sun of the diagonal elements which are highlighted in the green color comprises of the 88.9% of the total value where as the value highlighted in the red color is 11.1%.
In the case of testing of the confusion matrix the value highlighted in the green color comprises of 100% and the values highlighted in the red colored box 0% which is a great sign and highly effective results.
So, the overall confusion matrix the values which are correctly tested for the diseased , partially diseased and completely diseased as it comprises of 97.5% are the values which are having little bit error are approximately 2.5% which is quite less.

5.4.5 Receiver Operating Characteristics:

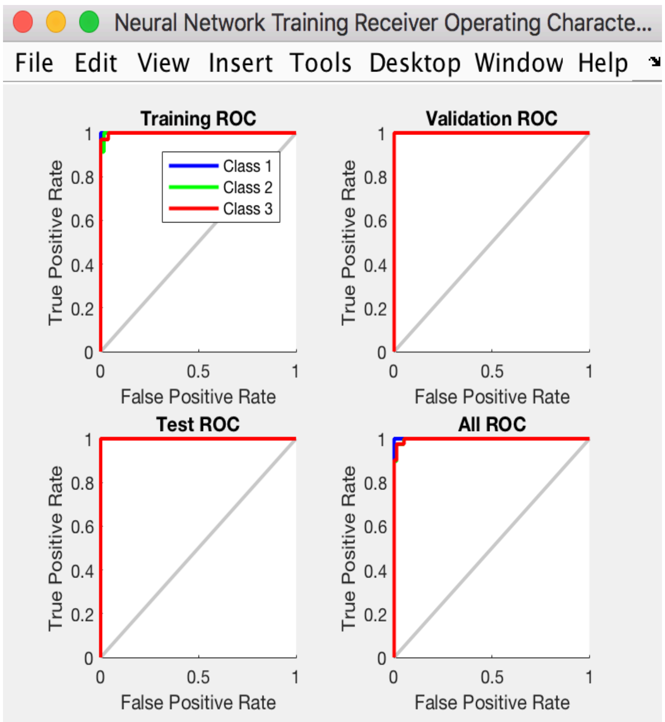


Figure5.8 ROC characteristics.
Analysis: The shaded region in every axis represent the ROC bends. The plot is being made between false positive rate and true positive rate.

CHAPTER-6

CLUSTERING TOOL

Neural systems is phenomenal utilization of grouping information. Gathering information by similarity. For eg. Performing market division by gathering individuals as indicated by their purchasing designs. Apportioning information identifies with their information mining. Grouping qualities with related articulation designs.

Cluster analysis includes applying at least one clustering calculations with the objective of finding concealed examples or groupings in a dataset. Clustering calculations frame groupings or groups such that information inside a group have a higher measure of closeness than information in some other group. The measure of comparability on which the bunches are displayed can be characterized by Euclidean separation, probabilistic separation, or another metric.

Cluster analysis is an unsupervised learning technique and a vital undertaking in exploratory information investigation. Well known bunching calculations include:

- Progressive grouping: assembles a multilevel chain of command of bunches by making a bunch tree
- k-Means bunching: allotments information into k unmistakable groups in view of separation to the centroid of a bunch
- Gaussian blend models: models groups as a blend of multivariate typical thickness segments
- Self-sorting out maps: utilizes neural systems that take in the topology and appropriation of the data.

6.1 Neural Clustering:

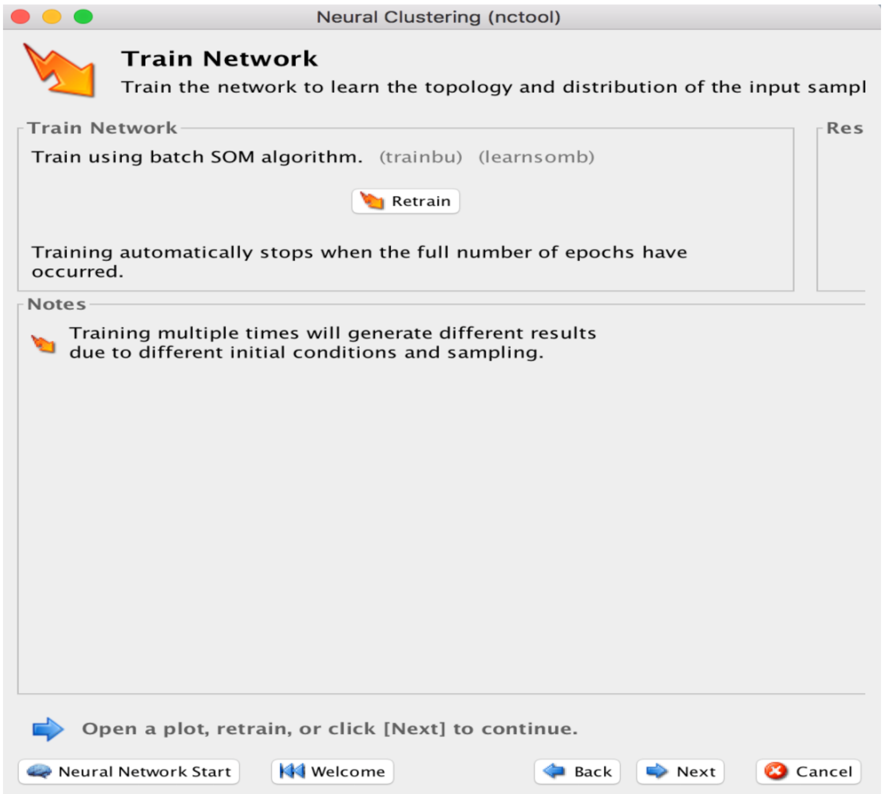


Figure6.1 nctool

Analysis: The data is being trained using nctool. The data is being accurately trained in order to understand the methodology and separation between the input samples proved to the network.

6.2 Train Data:

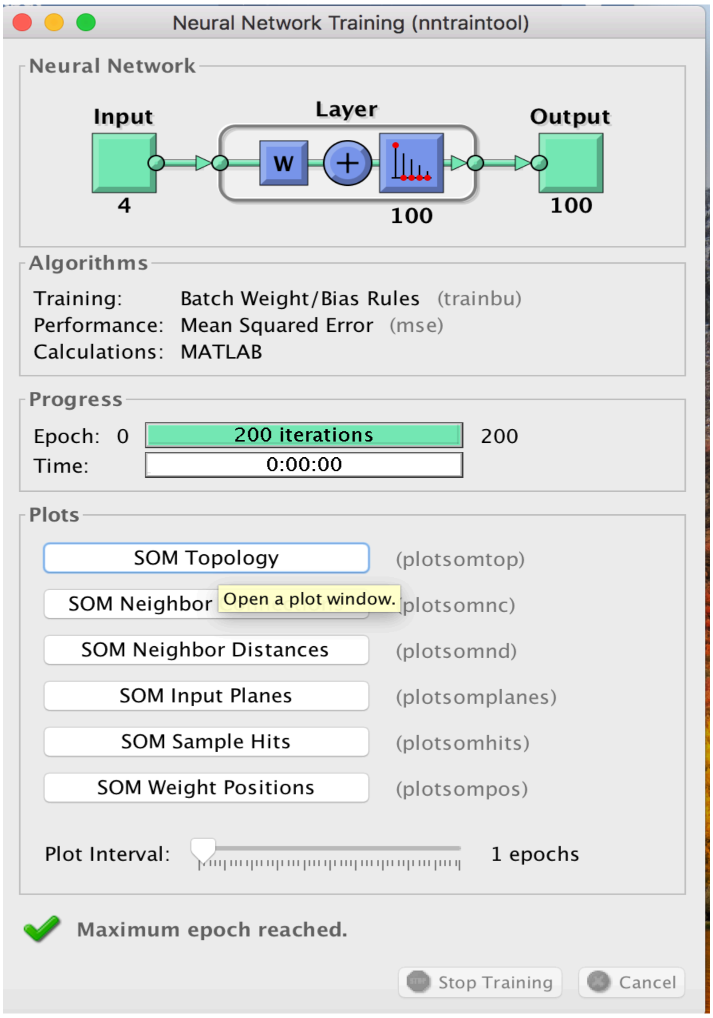


Figure6.2 nntraintool for clustering.

Analysis: nntraintool opens the neural system preparing GUI.
This capacity can be called to make the preparation GUI obvious before preparing has happened, in the wake of preparing if the window has been shut, or just to convey the preparation GUI to the front.
System preparing capacities handle all movement inside the preparation window.
To get to extra helpful plots, identified with the present or last system prepared, amid or subsequent to preparing, click their individual catches in the preparation window.

6.2.1 SOM Topology:

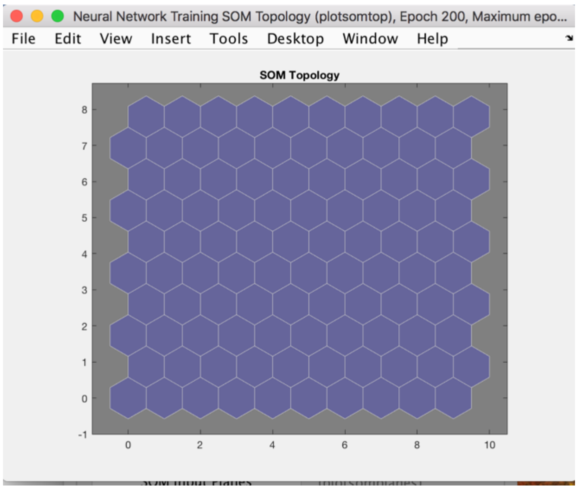


Figure6.3 plotsomtop

Analysis: It is hexagonal. In the abovementioned figure each and every hexagon builds communication network with neurons. The frame is 10x10 so it leads to 100 neurons in the system. It consists of four attributes in which the information vector and the information space is four dimensional.

6.2.2 SOM Neighbor Connection:

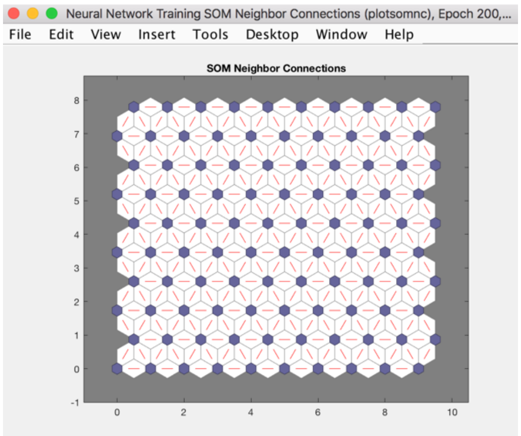


Figure6.4 plotsomnc

Analysis: plotsomnc(net) plots a SOM layer indicating neurons as dim blue patches and their immediate neighbor relations with red lines.
This plot bolsters SOM systems with hextop and gridtop topologies, however not tritop or randtop.

6.2.3 SOM Neighbor distance:

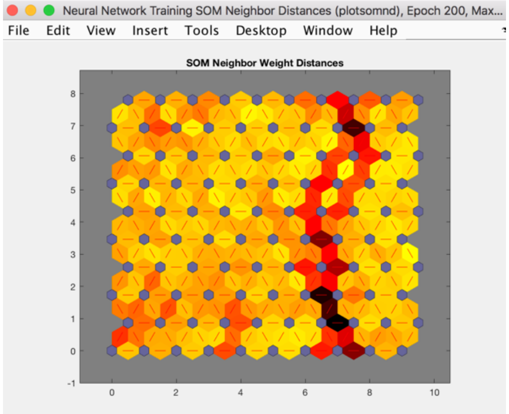


Figure6.5 plotsomnd

Analysis: In this figure the blue colored hexagons builds communication network with neurons. On the other hand red colored lines act as an association with the adjacent neurons. The pitch dark colors are related to bigger separations and lighter color speak to the littler separations.

6.2.4 SOM input Planes:

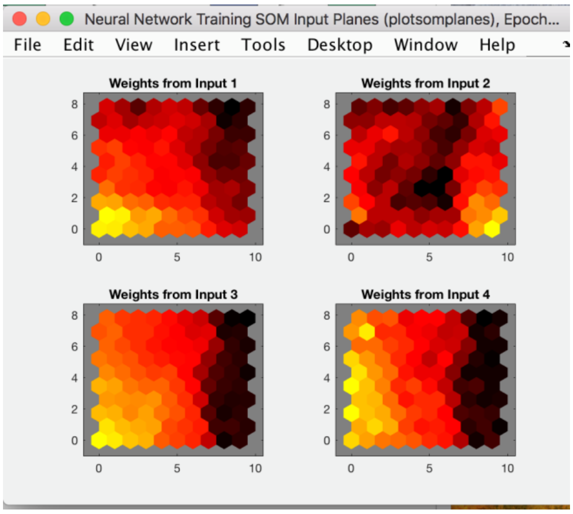


Figure6.6 plotsomplanes
Analysis: plotsomplanes(net) creates an arrangement of subplots. Each ith subplot demonstrates the weights from the ith contribution to the layer's neurons, with the most negative associations appeared as blue, zero associations as dark, and the most grounded positive associations as red.

6.2.5 SOM Sample Hits:

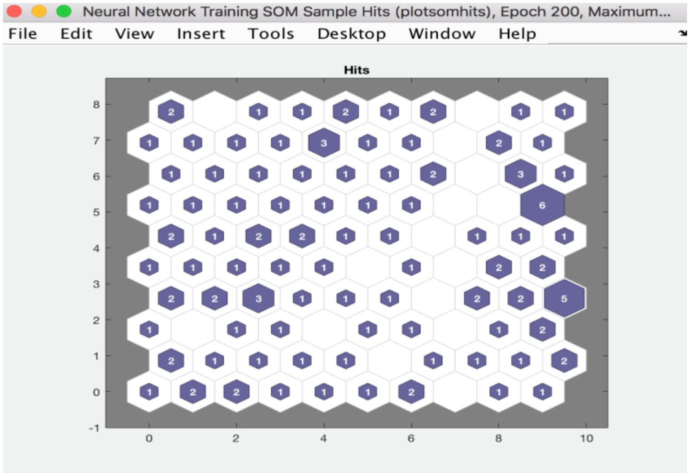


Figure6.7 plotsomhits

Analysis: plotsomhits(net,inputs) plots a SOM layer, with every neuron demonstrating the quantity of information vectors that it groups. The relative number of vectors for every neuron is indicated by means of the measure of a shaded fix.

This plot underpins SOM systems with hextop and gridtop topologies, however not tritop or randtop.

6.2.6 SOM Weight Positions:

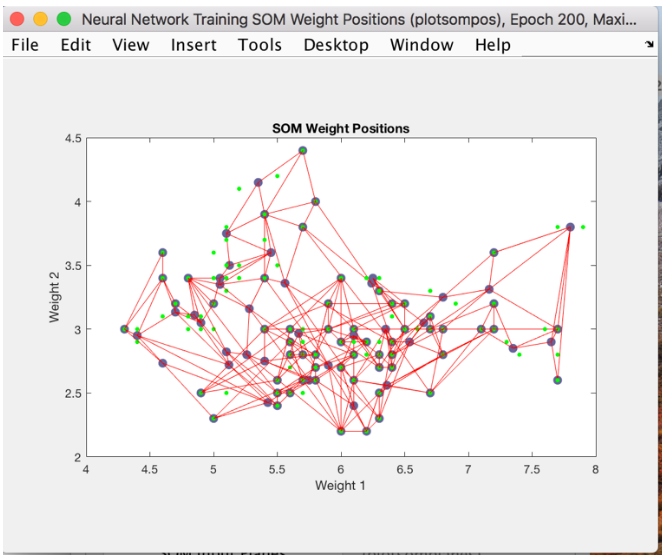


Figure6.8 plotsompos

Analysis: plotsompos(net) plots the info vectors as green specks and shows how the SOM arranges the information space by indicating blue-gray spots for every neuron's weight vector and interfacing neighboring neurons with red lines. plotsompos(net, inputs) plots the information close by the weights.

The value of weight1 is till 8 whereas the value mentioned in weight 2 is till 4.5. The used in the clustering is a four dimensional input as the number of input parameters used are four.

REFERENCES

[1]S.Bashir,N.Sharma,Remote Area Plant disease detection using Image Processing, IOSR Journal of Electronics and Communication Engineering , Volume 2, Issue 6 ,pp.31-34,2012.

[2]S.Arivazhagan,R.N.Shebiah, S.Ananthi,S.V.Varthini ,Using texture detection features, recoginizing unhealthy region of plant leaves and classification of plant leaf disease , AgricEngInt: CIGR Journal, Vol. 15, No.1,pp.211-217,2013.

[3]J.Behmann,A.K.Mahlein,T.Rumpf,C.Romer,L.Plumer, A review of advanced machine learning methods for the detection of biotic stress in precision crop protection, Springer Media New Y ork, Precision Agriculture,Volume 16, Issue 3, pp. 239–260, 2015.

[4]J.G.A.Barbedo,Digital image processing techniques for detecting, quantifying and classifying plant diseases, ArnalBarbedoSpringerPlus , pp.1- 12,2013.

[5] F.Qin, D.Liu, B.Sun, L.Ruan, Z.Ma, H.Wang,Identification of Alfalfa Leaf Diseases Using ImageRecognitionTechnology,Plos Journals, pp.1-15, 2016.

[6]A.Meunkaewjinda, P.Kumsawat and K.Attakitmongcol,Grape leaf disease detection from color imagery using hybrid intelligent system, 5th International Conference on Electrical Engineering/Electronics,Computer,Telecommunications and Information Technology, Volume: 1,pp. 513 - 516,2008.

[7]E.Omrani, B.Khoshnevisan, S.Shamshirband, H.Saboohi, N.B.Anuar, M.H.N.M.Nasir, 'Potential of radial basis function-based support vector regression for apple disease detection', Department of Biosystem Engineering, pp.2-19,2014.

[8]A.Singh,B.G.Subramanian,A.K.Singh,S.Sarkar, Machine Learning for High-Throughput Stress Phenotyping in Plants,Trends in Plant Science, Volume 21, Issue 2, pp. 110-124, 2016.

[9]V .Singh, segmentation 49,2017. [10]A.Camargo, J.S.Smith, An image-processing based algorithm to automatically identify plant disease visual symptoms,Biosystems Engineering ,Volume 102, Issue 1, pp. 9-21, 2009.

[11] ,AnInvestigat -ion Into Machine Learning Regression Techniques for the Leaf Rust Disease Detection Using Hyperspectral Measurement, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, 2016.

[12]M.K.Tripathi, Recent Machine Learning Based Approaches for Disease Detection and Classification of Agricultural Products,ComputingCommunication Control andautomation(ICCUBEA),2016 International Conference , 2016.

[13]M.Jhuria, A.Kumar, R.Borse, Image processing forSmart Farming : Detection of disease and food grading, Institute of Electrical and Electronics Engineering, 2013. [14] N. Bhardwaj, G. Kaur, P.K.Singh, A Systematic Review on Image Enhancement Techniques, Sensors and Image Processing, Advances in Intelligent Systems and Computing (AISC),

Springer, pp. 227-235, 2018.

[15] G. Kaur, N. Bhardwaj, P.K.Singh, An Analytic Review on Image Enhancement Techniques Based on Soft Computing Approach, Sensors and Image Processing, Advances in Intelligent Systems and Computing(AISC), Springer, pp. 255-265, 2018.

[16] K. Vasudeva, P.K. Singh,Y. Singh, A Methodical Review on Issues of Medical Image Management System with Watermarking Approach, Indian Journal of Science and Technology, Vol 9(32), DOI: 10.17485/ijst/2016/v9i32/100188, ISSN: 0974-5645, 2016.

[17] D. Agarwal, A. Gupta, P. K.Singh, A systematic review on Artificial Bee Colony Optimization Technique, International Journal of Control Theory and Application, Vol. 9(11), pp. 5487-5500, 2016,

[18] A. Sharma, P.K. Singh, P. Khurana, Analytical Review on Object Segmentation and Recognition, published in proceedings of 6thInternational Conference - Cloud System and Big Data Engineering (Confluence), 14th - 15thJanuary, 2016, Noida, India, IEEE, pp. 524 – 530,2016.

[19] R. Bhardwaj, P .K. Singh, Analytical Review on Human Activity Recognition in Video, published in proceedings of 6thInternational Conference - Cloud System and Big Data Engineering (Confluence), 14th-15thJanuary, 2016, Noida, India, IEEE, pp. 531 – 536,2016.

APPENDICES

Code for Neural fitting:

```
% Solve an Input-Output Fitting problem with a Neural Network
% Script generated by Neural Fitting app
% Created 10-May-2018 13:04:59
%
% This script assumes these variables are defined:
%
%   input - input data.
%   output - target data.

x = input;
t = output;

% Choose a Training Function
% For a list of all training functions type: help nntrain
% 'trainlm' is usually fastest.
% 'trainbr' takes longer but may be better for challenging problems.
% 'trainscg' uses less memory. Suitable in low memory situations.
trainFcn = 'trainlm';  % Levenberg-Marquardt backpropagation.

% Create a Fitting Network
hiddenLayerSize = 10;
net = fitnet(hiddenLayerSize,trainFcn);

% Choose Input and Output Pre/Post-Processing Functions
% For a list of all processing functions type: help nnprocess
net.input.processFcns = {'removeconstantrows','mapminmax'};
net.output.processFcns = {'removeconstantrows','mapminmax'};
```

```matlab

% Setup Division of Data for Training, Validation, Testing
% For a list of all data division functions type: help nndivide
net.divideFcn = 'dividerand';  % Divide data randomly
net.divideMode = 'sample';  % Divide up every sample
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;

% Choose a Performance Function
% For a list of all performance functions type: help nnperformance
net.performFcn = 'mse';  % Mean Squared Error

% Choose Plot Functions
% For a list of all plot functions type: help nnplot
net.plotFcns = {'plotperform','plottrainstate','ploterrhist', ...
    'plotregression', 'plotfit'};

% Train the Network
[net,tr] = train(net,x,t);

% Test the Network
y = net(x);
e = gsubtract(t,y);
performance = perform(net,t,y)

% Recalculate Training, Validation and Test Performance
```

```matlab
% Recalculate Training, Validation and Test Performance
trainTargets = t .* tr.trainMask{1};
valTargets = t .* tr.valMask{1};
testTargets = t .* tr.testMask{1};
trainPerformance = perform(net,trainTargets,y)
valPerformance = perform(net,valTargets,y)
testPerformance = perform(net,testTargets,y)

% View the Network
view(net)

% Plots
% Uncomment these lines to enable various plots.
%figure, plotperform(tr)
%figure, plottrainstate(tr)
%figure, ploterrhist(e)
%figure, plotregression(t,y)
%figure, plotfit(net,x,t)

% Deployment
% Change the (false) values to (true) to enable the following code blocks.
% See the help for each generation function for more information.
if (false)
    % Generate MATLAB function for neural network for application
    % deployment in MATLAB scripts or with MATLAB Compiler and Builder
    % tools, or simply to examine the calculations your trained neural
    % network performs.
```

```
        % network performs.
        genFunction(net,'myNeuralNetworkFunction');
        y = myNeuralNetworkFunction(x);
    end
    if (false)
        % Generate a matrix-only MATLAB function for neural network code
        % generation with MATLAB Coder tools.
        genFunction(net,'myNeuralNetworkFunction','MatrixOnly','yes');
        y = myNeuralNetworkFunction(x);
    end
    if (false)
        % Generate a Simulink diagram for simulation or deployment with.
        % Simulink Coder tools.
        gensim(net);
    end

```

Code for pattern recognition:

```
% Solve a Pattern Recognition Problem with a Neural Network
% Script generated by Neural Pattern Recognition app
% Created 10-May-2018 13:10:04
%
% This script assumes these variables are defined:
%
%   input - input data.
%   outpu3 - target data.

x = input;
t = outpu3;

% Choose a Training Function
% For a list of all training functions type: help nntrain
% 'trainlm' is usually fastest.
% 'trainbr' takes longer but may be better for challenging problems.
% 'trainscg' uses less memory. Suitable in low memory situations.
trainFcn = 'trainscg';  % Scaled conjugate gradient backpropagation.

% Create a Pattern Recognition Network
hiddenLayerSize = 10;
net = patternnet(hiddenLayerSize, trainFcn);

% Choose Input and Output Pre/Post-Processing Functions
% For a list of all processing functions type: help nnprocess
net.input.processFcns = {'removeconstantrows','mapminmax'};
net.output.processFcns = {'removeconstantrows','mapminmax'};
```

```
net.output.processFcns = {'removeconstantrows','mapminmax'};

% Setup Division of Data for Training, Validation, Testing
% For a list of all data division functions type: help nndivide
net.divideFcn = 'dividerand';  % Divide data randomly
net.divideMode = 'sample';  % Divide up every sample
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;

% Choose a Performance Function
% For a list of all performance functions type: help nnperformance
net.performFcn = 'crossentropy';  % Cross-Entropy

% Choose Plot Functions
% For a list of all plot functions type: help nnplot
net.plotFcns = {'plotperform','plottrainstate','ploterrhist', ...
    'plotconfusion', 'plotroc'};

% Train the Network
[net,tr] = train(net,x,t);

% Test the Network
y = net(x);
e = gsubtract(t,y);
performance = perform(net,t,y)
tind = vec2ind(t);
tind = vec2ind(t);
yind = vec2ind(y);
percentErrors = sum(tind ~= yind)/numel(tind);

% Recalculate Training, Validation and Test Performance
trainTargets = t .* tr.trainMask{1};
valTargets = t .* tr.valMask{1};
testTargets = t .* tr.testMask{1};
trainPerformance = perform(net,trainTargets,y)
valPerformance = perform(net,valTargets,y)
testPerformance = perform(net,testTargets,y)

% View the Network
view(net)

% Plots
% Uncomment these lines to enable various plots.
%figure, plotperform(tr)
%figure, plottrainstate(tr)
%figure, ploterrhist(e)
%figure, plotconfusion(t,y)
%figure, plotroc(t,y)

% Deployment
% Change the (false) values to (true) to enable the following code blocks.
% See the help for each generation function for more information.
if (false)
```

```
if (false)
    % Generate MATLAB function for neural network for application
    % deployment in MATLAB scripts or with MATLAB Compiler and Builder
    % tools, or simply to examine the calculations your trained neural
    % network performs.
    genFunction(net,'myNeuralNetworkFunction');
    y = myNeuralNetworkFunction(x);
end
if (false)
    % Generate a matrix-only MATLAB function for neural network code
    % generation with MATLAB Coder tools.
    genFunction(net,'myNeuralNetworkFunction','MatrixOnly','yes');
    y = myNeuralNetworkFunction(x);
end
if (false)
    % Generate a Simulink diagram for simulation or deployment with.
    % Simulink Coder tools.
    gensim(net);
end

```

Code For clustering:

```
% Solve a Clustering Problem with a Self-Organizing Map
% Script generated by Neural Clustering app
% Created 10-May-2018 13:14:50
%
% This script assumes these variables are defined:
%
%   input - input data.

x = input;

% Create a Self-Organizing Map
dimension1 = 10;
dimension2 = 10;
net = selforgmap([dimension1 dimension2]);

% Choose Plot Functions
% For a list of all plot functions type: help nnplot
net.plotFcns = {'plotsomtop','plotsomnc','plotsomnd', ...
    'plotsomplanes', 'plotsomhits', 'plotsompos'};

% Train the Network
[net,tr] = train(net,x);

% Test the Network
y = net(x);

% View the Network
```

```matlab
view(net)

% Plots
% Uncomment these lines to enable various plots.
%figure, plotsomtop(net)
%figure, plotsomnc(net)
%figure, plotsomnd(net)
%figure, plotsomplanes(net)
%figure, plotsomhits(net,x)
%figure, plotsompos(net,x)

% Deployment
% Change the (false) values to (true) to enable the following code blocks.
% See the help for each generation function for more information.
if (false)
    % Generate MATLAB function for neural network for application
    % deployment in MATLAB scripts or with MATLAB Compiler and Builder
    % tools, or simply to examine the calculations your trained neural
    % network performs.
    genFunction(net,'myNeuralNetworkFunction');
    y = myNeuralNetworkFunction(x);
end
if (false)
    % Generate a matrix-only MATLAB function for neural network code
    % generation with MATLAB Coder tools.
    genFunction(net,'myNeuralNetworkFunction','MatrixOnly','yes');
    y = myNeuralNetworkFunction(x);
```

```matlab
    y = myNeuralNetworkFunction(x);
end
if (false)
    % Generate a Simulink diagram for simulation or deployment with.
    % Simulink Coder tools.
    gensim(net);
end
```