



# ROLE OF VARIOUS SEGMENTATION AND CLASSIFICATION ALGORITHMS IN PLANT DISEASE DETECTION

Yatendra Kashyap, Dr. Shivshakti Shrivastava  
Rabindra Nath Tagore University Bhopal

**Abstract-** Using deep learning approaches, convolution neural network models proven best and well constructed to detect and diagnose the soya bean plant disease by using the foliar images of healthy and diseased soya bean plants. The models were trained using our own collection of almost 848 pictures exhibiting various soya bean plants foliar diseases in combinations by covering wide range of commonly found diseases .in this paper Several model and their designs were tested and studied, But out of all our proposed system perform well the accuracy of almost 95.73 percent. The proposed model's high success rate makes it an excellent advising or early warning tool for the farmers in order to avoiding the unnecessary use of pesticide on plants that are very harmful for human body and also this system helpful in enhancing the production rates of good quality soya bean crop. .

**Keywords-**Convolution Neural Network, Soybean Plant Disease, Data Augmentation, Transfer Learning.

## I. INTRODUCTION

Agriculture has a significant role in India's economic development. Agriculture accounts for over 70% of the Indian economy. As a result, crop damage would result in a significant loss of output, affecting the economy. Plants' leaves, which are the most vulnerable, show disease symptoms first[1]. From the beginning of their life cycle until they are ready to be harvested, the crops must be monitored for illnesses. Initially, the traditional naked eye observation approach was employed to keep a watch on the plants for illnesses. Various strategy which requires experts to manually monitor the crop fields [2]. A variety of strategies have been used to produce autonomous and semi-automatic plant disease detection systems in recent years. These methods have so far resulted to be fast, affordable and more accurate than the old way of manual observation by farmers [3].

As a result, researchers are being urged to develop more intelligent technology systems for detecting plant diseases that do not require human interaction. The goal of this study is to go over several plant disease detection techniques and discuss them in terms of various factors. The following are the portions of the paper. The first section provides a quick

overview of the significance of plant disease detection. The second section analyses and discusses recent work in this field, as well as the methodologies used. The third section covers the basic methods for designing a disease detection system. Finally, the fourth section brings this article to a close by outlining future directions.

For the past few years, Deep Learning has completely dominated the field of image classification, so the proposed Convolution Neural Network is used to detect coffee leaf diseases and categories them into five classes: healthy, diseased leaves with brown spots, frog eye on soy plant leaf, and frog eye on soy plant leaf with a high success rate.

## II. LITERATURE REVIEW

Machine learning algorithms for disease identification and classification have been compared in a large survey. We investigated the effectiveness of the Support Vector Machine (SVM) Classification Technique, the Artificial Neural Network (ANN) Classification Technique, the K-Nearest Neighbour Classification Technique, the Fuzzy C-Means Classifier, and the Convolution Neural Network Classification techniques for detecting plant diseases.

### A. Svm classifier

In Machine Learning, SVM Classifier is a supervised learning method in which analysed data is used for classification. SVM Classifier was used by the following authors to detect illness in various crops. [1] Detection of illnesses on citrus trees, such as canker and anthracnose infections on grapefruit, lemons, limes, and oranges. The trial result received a true acceptance rate of 95%. [2] The grape plant diseases Downy Mildew and Powdery Mildew were detected with an average accuracy of 88.89 percent for both infections. [3] Chimaera and Anthracnose, two oil palm leaf diseases, have a detection accuracy of 97 percent and 95 percent, respectively. [4] Potato plant diseases Late blight and Early blight are detected with 95% accuracy across 300 publicly available pictures.[5] Using data from both the LAB and HSI colour models, the grape leaf diseases Black Rot, Esca, and Leaf Blight are accurately diagnosed. [6] A method for identifying illnesses in tea plants was developed. SVM classifiers are used to detect three different types of diseases using fewer features. The



established approach correctly classified diseases 90% of the time. [7] Downy Mildew, Frog Eye, and Septories Leaf Blight are three diseases that can be detected using soybean culture. Using a large dataset, they reported an average classification accuracy of around 90%.

**B. Ann classifier**

In machine learning and pattern recognition, an Artificial Neural Network is a computational model. The following is a list of related work on plant disease detection using an ANN classifier. [8] A proposed method for plant disease recognition utilising a feed forward back propagation algorithm was evaluated, and it worked well with a precision of roughly 93 percent. They tested the treatment on plant diseases such as early scorch, cottony mould, late scorch, and small whitening. [9] Developed a model to improve the accuracy in identifying two forms of fungus-caused diseases in cucumber plants: Downy Mildew and Powdery Mildew. [10] Using a back-propagation algorithm, a system was developed to recognise and categorise diseases such as leaf spot, bacterial blight, fruit spot, and fruit rot in pomegranate plants, and the experimental results suggest that it works.[11] Using the neural network Back propagation approach, proposed a work on identifying the groundnut plant disease cercospora (leaf spot). The experimental results and observations demonstrate that they correctly recognised four types of diseases out of 100 sample diseased leaf photos with a 97.41 percent accuracy rate. [12] Proposed a method for detecting pomegranate plant disease, which was tested using 40 photos and found to be 90% accurate.

**C. Knn classifier**

In Machine Learning, K-Nearest Neighbours has been used for pattern recognition, statistical estimation, and classification. As shown below, we conducted a survey on plant disease identification using the KNN classifier. [13] An algorithm for detecting illness in sugarcane culture was proposed. For feature extraction, image processing methods are utilised. It detected Leaf scorch disease in sugarcane leaves with a 95% accuracy rate. [14] Using 40 photos,

developed a method to determine the severity and detection of the cotton plant disease Grey Mildew disease with an accuracy of 82.5 percent. [15] Using the GLCM feature extraction approach and the KNN classifier, an algorithm for plant disease identification was proposed. To classify data into several classes, the KNN classifier is proposed rather than the SVM classifier. When it comes to accuracy, the performance is put to the test.

**D. Fuzzy classifier**

An author [16] presented a method to determine the presence of infection in wheat crop photos using Fuzzy Classifier in a related paper on Fuzzy Classifier in plant disease detection. The dataset of healthy and ill leaves is used to evaluate this method. The accuracy of the classification of healthy and unhealthy leaves was 88 percent, and the recognition of disease was 56 percent.

**E. Deep learning**

Deep learning is a type of ANN learning that is also a type of machine learning method. [17] Using the CNN classification technique, a model was proposed to detect healthy leaves and 13 different diseased leaves of peach, cherry, pear, Apple, and grapevine. More than 30000 photos were included in the dataset, with separate class test accuracy ranging from 91% to 98 percent and average accuracy of 96.3 percent. [18] Developed a system for detecting plant illnesses using a public dataset of 54306 photos of 14 crops and 26 diseases, which performed with 99.35 percent accuracy using 20% of testing data and 98.2 percent accuracy using 80% of testing data.[29] Using a CNN classifier, created a model to recognise the soybean plant diseases Septoria, Frogeye, and Downy Mildew. A dataset with 12673 leaf photos divided into four classifications yielded a 99.32 percent accuracy. [30] Developed a CNN classification system for detecting plant illnesses. The dataset contains 87848 photos of 25 different plants in a set of 58 diseases, with a 99.53 percent accuracy. The comparison of different type of Machine Learning classifiers used in plant disease detection is summarized and is given in Table 1.

Table 1. Comparison of classification techniques.

Classification Technique	Crop	No. of Diseases	Result
SVM Classifier	Grapes[1]	2 Diseases	95%
	Citrus [2]	2 Diseases	88.9%
	Oil Palm[3]	2 Diseases	97%
	Potato[4]	2 Diseases	90%
	Tea[5]	2 Diseases	93%
	Soybean[6]	3 Diseases	90%
ANN classifier	Not Mentioned[8]	5 Diseases	93%
	Cucumber[9]	2 Diseases	Increased accuracy
	Pomegranate[10]	4 Diseases	90%



	Groundnut[11]	4 Diseases	97.41%
KNN Classifier	Sugarcane[13]	1 Diseases	95%
	Cotton[24]	1 Diseases	82.5%
Fuzzy Classifier	Wheat [16]	1 Diseases	88%
CNN Classifier	Peach, cherry, apple	13 Disease	96%
	Rice[28]	26 Disease	99.35
	Soybean[19]	3 Disease	99.32

### III. SEGMENTATION METHODS

This technique is useful in dividing the images into several segments or sets of pixels. The idea is that at least one segment will have the region of interest (ROI) which can be utilized for further sophisticated algorithms.

Segmentation Algorithms have been developed to segment the images and it can be classified into following [22]

- Segmentation by Clustering
- Segmentation by Edge Detection
- Segmentation by Fuzzy Logic
- Segmentation by Neural Network
- Segmentation by Region Based
- Segmentation by watershed

Algorithm	Description	Advantage	Limitations
Segmentation by clustering	Grouping of pixels having similar properties and defines the cluster values based on their visible intensities	Works actually well on tiny datasets and generates admirable clusters	Computation time is excessively large and also expensive..k-means is a distance-based algorithm. It is not suitable sometimes.
Segmentation by edge detection	Segmentation is done from end to end by identifying the boundaries.	Helps to retain gray tones in Edges and for good contrast images	Difficult for low contrast images .It is not suitable if edges are many.
Region based segmentation techniques	Separates the objects into different regions based on Morphological operations	Calculations are simple and operations are fast. It works well high contrast images.	Sometimes overlap of the greyscale pixel values faces difficult. It is good with implementation of Marker based.
Clustering techniques	It is a method to perform Image Segmentation of pixel-wise segmentation. In this type of segmentation, we try to cluster the pixels that are together.	Clustering algorithm helps to better understand customers, in terms of both static demographics and dynamic behaviours.	Disadvantages of clustering are complexity and inability to recover from database corruption.
Watershed segmentation	Here, pixel and region similarity is found. For each pixel identified, the region to which the pixel should belong is computed.	The result of watershed algorithm is global segmentation, border closure and high accuracy	it is highly sensitive to local minima, since at each minima.

#### IV. PROPOSED WORK

##### (a) Input image and dataset

We have collected the dataset consisting about 300 images to train and test our system from various farms of vidisha district of Madhya Pradesh, India. Various Diseases are

found during the database collection out of which we found 4 deadly disease that cause damage of soybean foliar are namely Bacterial blight, Downy Mildew, Frog Eye, Septoria Leaf Blight are shown in the below figure

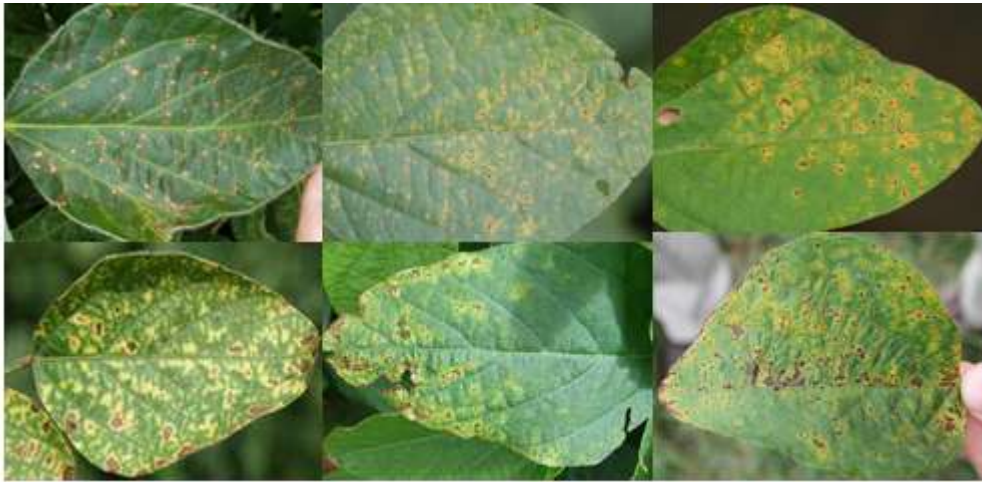


Fig. 1. Common Disease of soybean leaf (1) Bacterial blight (2) Downy Mildew (3) Frog Eye (4) Septoria Leaf Blight

##### (a) Pre Processing and masking

The proposed framework consists of 4 stages – pre-processing, segmentation, feature extraction and classification.–Bacterial blight, frog eye, downy mildew, septoria leaf blight .First the input leaf image is pre-processed to remove the unnecessary background and crop our region of interest. The pre-processed image is segmented using watershed segmentation method in which one or more clusters may contain the diseased region. A watershed clustering algorithm classifies pixels based on particular set of features into various number of classes. The classification is done by minimizing the sum of squares of distances between the pixels and the corresponding cluster

centroid for which Euclidian distance is considered. Then feature extraction process will take place using GCML and various entropy, mean, skiwness, etc. The classification happens in 2 stages. In first stage color and texture features are extracted from both infected and healthy clusters to train classifier in order to differentiate the healthy and infected clusters.

The complex image is pre-processed to remove the unwanted background and it is processed further. An illustration of pre-processing step is represented in Fig.2. The region of interest alone is selected to ease the process this is important in case of real time images. The binary mask is created with respect to ROI as shown in Fig2.

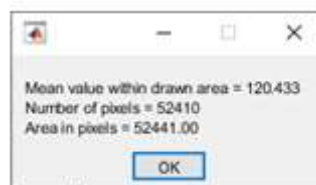


Fig.2 Area in Pixel

The color space conversion is essential to make the image into device independent and so RGB is transformed into  $L^*a^*b^*$  (a device independent model where  $L^*$  signifies the lightness,  $a^*$  and  $b^*$  are the chromaticity layers). This color space model closely resembles human perception and also splits information about chrominance better than other models. Algorithm.1.represents the way in which pre-

processing is done and Fig.5 shows the leaf images obtained after pre-processing The color space converted image then enters segmentation process in which only  $a^*$  and  $b^*$  channel are considered in order to reduce the time consumed for processing.



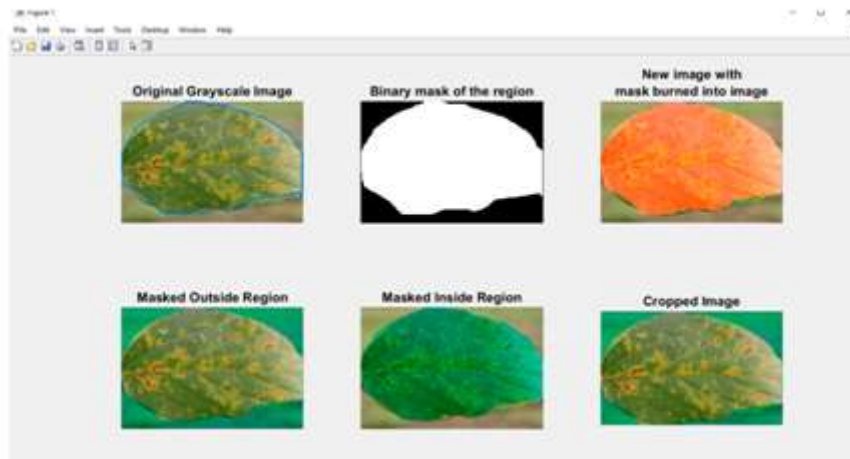


Fig.3 Binary Masking Image helpful in cropped form.

**(b) Segmentation**

Segmentation involves partitioning of image into various parts of same features or having some nearest similarity. It can be done using various methods like Otsu’ method, K-means clustering, Watershed segmentation. Our project uses watershed segmentation which extracts and places information in the cluster. The algorithm introduced by Luc Vincent and Pierre Soils is based on the concept of “immersion”. Each local minima of a gray-scale image  $I$  which can be regarded as a surface has a hole and the surface is immersed out into water. Then, starting from the minima of lowest intensity value, the water will progressively fill up different catchment basins of image

(surface)  $I$ . Conceptually; the algorithm then builds a dam to avoid a situation that the water coming from two or more different local minima would be merged. At the end of this immersion process, each local minimum is totally enclosed by dams corresponding to watersheds of image (surface)  $I$ . The watershed transform has been widely used in many fields of image processing, including image segmentation, due to the number of advantages that it possesses: it is a simple intuitive method, it is fast and can be parallelized and an almost linear speedup was reported for a number of processors up to 64) and it produces a complete division of the image in separated regions even if the contrast is poor.

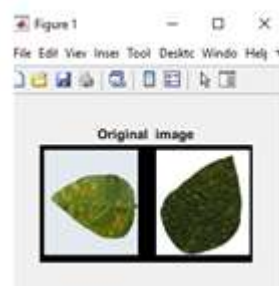


Fig 4(a) Input Image

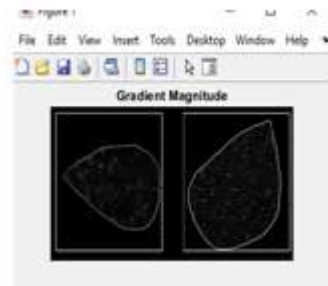


Fig.4 (b) Gradient Segmentation

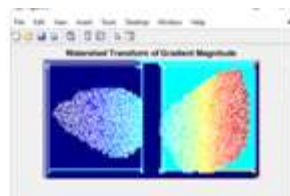


Fig 4(c) Colored Watershed



Fig.4 (d) Watershed segmentation



**(c) Feature extraction**

Feature extraction refers transforming input data into a set of features. Features are used to discriminate one input pattern from other pattern. The features from their color and texture and the combinations of these two are used to

evaluate the classification and to estimate the performance. For that various parameters are used to classify like Dissilarity, Homogeneity, Entropy, Energy, mean, variance and correlation is calculated that are as follows :-

```
>> glcm1

CG =

"Contrast Group:
Contrast:1212.470473
Dissimilarity:10.067938
Homogeneity:0.559686
"

SG =

"Statistics Group:
GLCM Mean:125.648695
GLCM Variance:10112.389302
GLCM Correlation:0.940050
"

OG =

"Orderliness Group:
Angular Second Moment :0.068523
Max Probability:0.166631
Entropy:2.373159
Energy:0.261769
"
```

Fig 5 Feature extraction with various parameters

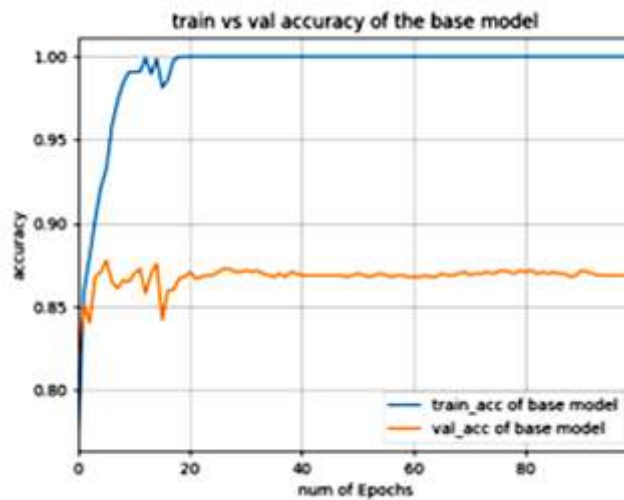


Fig. 6. Training vs validation accuracy of the base model

**V. RESULT**

CNN architectures vary with the type of the problem at hand. The proposed model consists of three convolution layers each followed by a maxpooling layer. The final layer is fully connected. After maxpooling is applied, the output is given as an input for the second convolution layer with 64 kernels of size 4x4. The last convolution layer has 128

kernels of size 1x1 followed by a fully connected layer of 512 neurons. The output of this layer is given to softmax function which produces a probability distribution of the four output classes. The architecture of the proposed model is shown in Table II



Layer	Type Filter	Size	Stride	Output size
L1	<b>Conv</b>	<b>3x3</b>	<b>1</b>	<b>128x128x32</b>
	<b>Pool</b>	<b>2x2</b>	<b>2</b>	<b>64x64x32</b>
L2	Conv	<b>4x4</b>	<b>1</b>	<b>61x61x64</b>
	<b>Pool</b>	<b>2x2</b>	<b>2</b>	<b>30x30x128</b>
L3	Conv	<b>1x1</b>	<b>1</b>	<b>30x30x128</b>
	<b>Pool</b>	<b>2x2</b>	<b>2</b>	<b>15x15x128</b>

Table II Architecture of proposed system

The dataset is split into three parts: 70% for training, 10% for validation, and 20% for testing. Various models with various designs and learning rates are put to the test. The network's parameters, such as the kernel size, filter size, learning parameter, and activation function, were chosen through trial and error.

As can be seen from the results, colour images have higher classification accuracy than grayscale and segmented images. This demonstrates the importance of the colour characteristic in extracting crucial features for categorization. Three convolution layers are followed by a max pooling layer in the model that delivers good classification accuracy.

An 95.8% accuracy rate was achieved using 75 epochs during the training of the model. The model also achieved a maximum accuracy rate of 100% when testing random images of plant varieties and diseases. The visualization of plots of train and test accuracy is described in fig 6. Shows the model is effective in detecting and recognizing plant diseases.

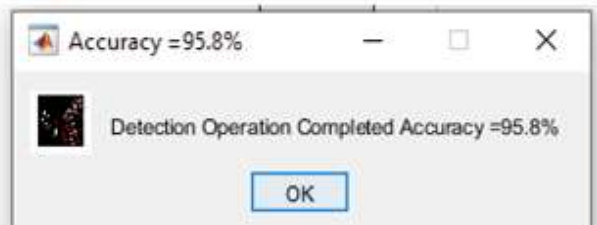


Fig.6 Accuracy of disease detection

## VI. CONCLUSION

The people around the world rely on the agricultural sector as one of the most important sectors where crops are the basic need for food. Early recognition and detection of these diseases are crucial to the agricultural industry. This paper has achieved its goal to detect and recognize 32 different plant varieties and plant diseases using convolution neural network. The trained model can be used to test real-time images to detect and recognize plant diseases. For the future work, additional plant varieties and different types of plant diseases may be included in the existing dataset to increase

the trained models. Other CNN architectures may also use different learning rates and optimizers for experimenting the accuracy of 95.8%, the proposed model can assist farmers to detect and recognize plant diseases.

## REFERENCES

- [1]. 'Etymology and classification' available at <https://en.wikipedia.org/wiki/Soybean> accessed December 2018
- [2]. Soybean diseases', available at <https://alliedcooperative.files.wordpress.com/2014/07/soybeandiseases.pdf>, accessed January 2017
- [3]. Shrivastava, S., Hooda, D.S "Automatic Automatic brown spot and frog eye detection system from the image captured in the field", Am. J. Syst., 2014
- [4]. Dandawate, Y., Kokare, R "An automated approach for classification of plant diseases towards development of futuristic decision support system" Proc. IEEE Int. Conf. Advances in Computing, Communications and Informatics , Kerala, India, August 2015
- [5]. Shrivastava, S., Singh, S.K., Hooda, D.S."Color sensing and image processing-based automatic soybean plant foliar disease severity detection and estimation" Multimedia Tools Appl., 2015
- [6]. Jadhav, S.B., Patil, S.B " Grading of soybean leaf disease based on segmented image K-means clustering" Int. J. Adv. Res. Electron, June 2015.
- [7]. Barbedo, J.G.A., Godoy, C.V. "Automatic classification of soybean diseases based on digital images of leaf" . SBI AGRO, October 2015
- [8]. Kaur, S., Pandey, S., Goel, S.: 'An automatic leaf disease detection system for legume species', J. Biol. Today's World, 2017
- [9]. Sanjay B. Patil1, Dr. Shrikant K. Bodhe, "Leaf disease severity measurement using image processin " Sanjay C 'International Journal of Engineering, October 2011.
- [10]. Cai, J., Miklavcic, S.: 'Automated extraction of threedimensional cereal plant structures from twodimensional orthographic images', IET Image Process.,2012..



- [11]. Sukhvair Kaur , Shreelekha and Shivani Goel, "Semiautomatic leaf disease detection and classification system for soybean culture", journal on IET Image processing, Vol. 12, Issue 6, 2018, pp. 1038- 1048.
- [12]. Kiran R. Gavhale, Ujwalla Gawande and Kamal O. Hajari, "Unhealthy region of citrus leaf detection using image processing techniques", IEEE International Conference on Convergence of Technology (I2CT), Pune 2014, pp. 1-6.
- [13]. Pranjali B. Padol; Anjali A. Yadav, "SVM Classifier Based Grape Leaf Disease Detection", IEEE Conference on Advances in Signal Processing (CASP), Pune 2016, pp. 175-179.
- [14]. Ahmad Nor Ikhwan Masazhar and Mahanijah Md Kamal, "Digital Image Processing Technique for Palm Oil Leaf Disease Detection using Multiclass SVM", IEEE 4th International Conference on Smart Instrumentation, Measurement and Applications ,Malaysia 2017, pp. 1-6.
- [15]. Monzurul Islam, Anh Dinh and Khan Wahid, "Detection of potato Diseases Using Image Segmentation and Multiclass Support Vector Machine", IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Canada 2017, pp. 1-4.
- [16]. Nithesh Agarwal, Jyothi Singhai and Dheeraj K. Agarwal, "Grape Leaf Disease Detection and Classification Using Multi- Class Support Vector Machine", proceeding of IEEE International conference on Recent Innovations in Signal Processing and Embedded Systems (RISE), Bhopal 2017, pp. 238-244.
- [17]. Md. Selim Hossain, Rokeya Mumtahana Mou, Mohammed Mahedi Hasan, Sajib Chakraborty and M. Abdur Razzak, "Recognition and Detection of Tea Leaf's Diseases Using Support Vector Machine" International Colloquium on Signal Processing (CSPA), Malaysia 2018, pp. 150-154.
- [18]. tanuja et al "plant disease detection system using image processing technique"International Journal of Computer Application in 2017.
- [19]. Dheeb Al Bashish, Malik Braik and Sulieman Bani-Ahmad, "A Framework for Detection and Classification of Plant Leaf and Stem Diseases", IEEE International Conference on Signal and Image Processing (ICSIP), Chennai 2010, pp. 113-118.
- [20]. Keyvan Asefpour Vakilian and Jafar Massah, "An artificial neural network approach to identify fungal diseases of cucumber" .Archives of Phytopathologyand Plant Protection, Vol. 46, Issue 13, Taylor & Francis 2013, pp. 1580-1588.
- [21]. Mrunmayee Dhakate and Ingole A. B. , "Diagnosis of Pomegranate Plant Diseases using Neural Network", IEEE 5th National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Patna 2015.
- [22]. Ramakrishnan M. and Sahaya Anselin Nisha A., "Groundnut Leaf Disease Detection and Classification by using Back Probagation Algorithm". IEEE International Conference on Communications and Signal Processing (ICCSP), Melmaruvathur 2015, pp. 0964 – 0968.
- [23]. Rashmi Pawar and Ambaji Jadhav, "Pomegranate Disease Detection and classification",IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, Chennai 2017, pp. 2475-2479.
- [24]. Umapathy Eaganathan, Jothi Sophia, Vinukumar Lackose, Feroze Jacob Benjamin, "Identification of Sugarcane Leaf Scorch Disease using K-means Clustering Segmentation and KNN based Classification", International Journal of Advances in Computer Science and Technology (IJACST), Vol. 3, No. 12, Special Issue of ICCEeT, Dubai , 2014, pp. 11- 16.
- [25]. Aditya Parikh, Mehul S. Raval, Chandrasinh Parmar and Sanjay Chaudhry, "Disease Detection and Severity Estimation in Cotton Plant from Unconstrained Images", IEEE International Conference on Data Science and Advanced Analytic, Canada 2016, pp. 594- 601.
- [26]. Gautham Kaushal, Rajini Bala, "GLCM and KNN based Algorithm for Plant Disease Detec-tion", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 6, Issue 7, 2017, pp. 5845-5852.
- [27]. Diptesh Majumdar, Arya Ghosh, Dipak Kumar Kole, Aruna Chakraborty and Dwijesh Dutta Majumder, "Application of Fuzzy CMeans Clustering Method to Classify Wheat Leaf Images based on the presence of rust" , Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications, Vol. 327, 2015, pp. 277-284.
- [28]. Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk and Darko Stefa-novic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification", Computational Intelligence and Neuroscience 2016.
- [29]. Sharada P. Mohanty, David P. Hughes and Marcel Salathe, " Using Deep Learning for Image- Based Plant Disease Detection", Frontiers in Plant Vol. 7, Article 1419, 2016.
- [30]. Serawork Wallelign, Mihai Polceanu and Cedric Buche, "Soybean Plant Disease Identification Using Convolutional Neural Network",International Florida Artificial Intelligence Research Society Conference , Melbourne, United States 2018, pp. 146- 151.





- [31]. Konstantinos P. Ferentinos, “Deep learning models for plant disease detection diagnosis” journal of Computers and Electronics in Agricultur science Vol. 145, Elsevier 2018, pp. 311-318
- [32]. Shruthi A Review system on Machine Learning Classification Techniques for plant Disease Detection” 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS),IEEE
- [33]. Vinod Kumar “A Study and Comparison of Different Image Segmentation Algorithms” International Conference on Advances in Computer, Bareilly, India IEEE 2016