



COTTON CROP DISEASE DETECTION - TECHNOLOGIES AND METHODOLOGIES

RISHIKESH GADADE¹, SHIKHAR GUPTA², SANDESH GUPTA³, Dr. PARUL JADHAV⁴

Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra

^{1,2,3}School of Computer Science and Engineering, Dr. Vishwanath Karad MIT World Peace University, PUNE-38

¹Email: gadaderishikesh@gmail.com; ²guptashikhar031297@gmail.com;

³sandeshgupta1211@gmail.com

⁴School of Electrical and Telecommunications Engineering, Dr. Vishwanath Karad MIT World Peace University, PUNE-38

Email: parul.jadhav@mitwpu.edu.in

DOI: [10.33329/ijer.9.1.22](https://doi.org/10.33329/ijer.9.1.22)



ABSTRACT

Agriculture is one of the prime sources of livelihood in India and it provides employment for rural people on a large scale in developing countries like India. Agriculture in peninsular India is composed of many crops and according to survey nearly 58% population is dependent on agriculture in India [16]. Very few Indian farmers have adopted technically efficient cultivation due to severe lack of technical knowledge. Farmers find it difficult to keep an eye on each and every plant in the cultivation area to detect any developing disease or infection. Several image processing, machine learning and deep learning approaches and methodologies have been proposed in recent decades to serve the purpose of detecting the initiation of a disease, determining the type and severity. of the disease the leaf is contaminated with and recommending upon relevant measures to be implemented. This paper focuses on reviewing all the methodologies used in most of the research work that has been done in the field of Image Processing and Machine Learning for the detection and diagnosis of disease in a cotton plant leaf from its image and also will try to explore different advancements in plant disease detection using Machine Learning and Deep Learning.

Keywords— Deep Learning, Image Processing and Machine Learning

1. Introduction

Agriculture has always been the prime factor in the economy of a developing country like India. Unfortunately, agriculture is a very constrained occupation in India because of the unpredictable weather conditions, uncertain rain and pathogens.

The damage inflicted by adverse climatic conditions pests and pathogens to a vast amount of crops cannot be neglected.

Cotton is considered as the “king” of cash crops in India. Cotton is one of the major crops in agriculture which decides the economy of the nation. However, there are certain issues such as: identify deficiency of nutrition in plants, to identify various diseases, various pests which affect crops. Each issue has an importance. Among one is detection of pests so that proper action should be taken to control it leading to minimizing loss.

On occurrence of any such predicament, misidentification of any diseases can be possible and incorrect controls like non-affecting pesticides can be used leading to wasting of work and money and most importantly it may lead to serious problems to crops.

Farmers have the option to approach experts but then the availability and convenience of the expert comes into picture. An automated method for detection and diagnosis can be very fruitful and convenient to the farmers.

The paper contains the following sections:

Section 2 – Key Issues In Disease Detection Of A Cotton Plant

Section 3 – Diseases Affecting the Cotton Crop

Section 4 – Literature Survey (Methodologies Used)

Section 5 - Importance Of Automation

2. Key Issues in Disease Detection of a Cotton Plant

Many Researchers from various organizations have done research on cotton plants and the diseases affecting it and also various techniques have been proposed by them to identify the disease. Gathering input data from several sources is an inevitable part of the Mechanized procedure in identifying disease. In this review, based on the various research papers that are been considered and their matter has been talked about, below are the mentioned challenges on disease and methods:

- Quality image of cotton plant leaves
- Vast amount of data is required to achieve considerable accuracy.
- Acquired images are affected by background data, noise and probably need a lot of pre-processing.
- Segmenting the exact spot in a leaf into predicting meaningful disease. Preparation of training and testing samples from the input image.

- Classification plays a role in recognizing segmented diseased parts into meaningful disease.
- Colour of plant leaf, size and texture are varying when the climate changes.
- Constant observation is needed for particular plants.
- Research suggests that image processing and machine learning techniques, especially deep learning techniques imbibed with transfer learning, comparably hold a good capacity in identifying disease. Hence, existing research calls for an improvisation.

3. Diseases affecting the Cotton Crop

3.1 Alternaria leaf spot -Fungal disease:

These are the small, circular brown lesions on cotyledons and seedling leaves which expand and develop a concentric pattern necrotic areas coalesce and often have a purple margin. A brown bed-bug surfacing occurs on the leaves as a result of the lesions on the central part which dry and further drop out from the plant.

Cause: Nutrient deficiency, drought, and other pests are the main reasons that causes stress in plants and makes them more susceptible to disease. There is a prompt spread of fungus in dense canopies, especially during periods of warm, wet weather.



Fig.1 Alternaria leaf spot

3.2 Ascochyta blight - Blight:

These are the brown or gray spots on leaves surrounded by a red halo. Elongated red-purple cankers on stems cause the wilting and death of leaves above.

Cause: Favorable weather conditions for disease are cool and wet ones. This situation arises when the cultivation area receives rainfall for a prolonged period of time.



Fig.2 Ascochyta blight

3.3 Cercospora leaf spot - Fungus:

These spots look like circular red wounds on leaves which enlarge and turn white or gray in the center. Lesions often have a pattern of concentric rings and possess a red margin. Dark gray spore masses form in the centers of the lesions making them appear dark gray.

Cause: Fungus grows out since the winters in crop debris from the previous growing season. *Torula alternata* leaf spot and other foliar disease and this disease are found close at hand.



Fig3. Cercospora leaf spot

3.4 Fusarium wilt

This is the wilting of cotyledons and seedling leaves. Cotyledons become chlorotic at the edges and then necrotic. Older plants exhibit symptoms of wilting and leaf chlorosis. Wilting is generally a gradual process but it may get accelerated if a heavy summer rain occurs. Due to severe infection, plants even become diminutive and may be further terminated. The Vascular system of infected plants becomes discolored and can be seen by cutting the stem. One can see the discolouring changes by cutting the stem in the vascular system of plants due to infection

Cause: Disease emergence is favored by warm temperatures; fungus may be introduced to the field through infected seed or by contaminated equipment and human movement.

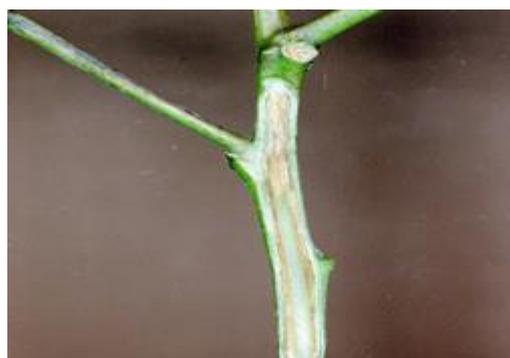


Fig.4 Fusarium wilt

3.5 Target spot

The symptoms are usually to be found on leaves, boll bracts and on bolls. They can be spotted by the initial appearance of small chocolate brown spots which later enlarge to become circular to irregular spots with target markings. These symptoms are mainly seen in the lower canopy. Typically, the infected leaves retain their green color or green yellow color. Under severe conditions leaf and flower drop may be possible.

Cause: The pathogen also infects cucumber, sweet potato, soybean and tomato. It will cause severe loss if pathogens infect at flowering stage. The disease is common in South-eastern cotton producing states of the USA. The target spot is confused with the leaf spot caused by *Cercospora* spp, *Stemphylium* spp, or *Alternaria* spp. But this

spot is essentially surrounded by reddish to purplish margin of the canopy.

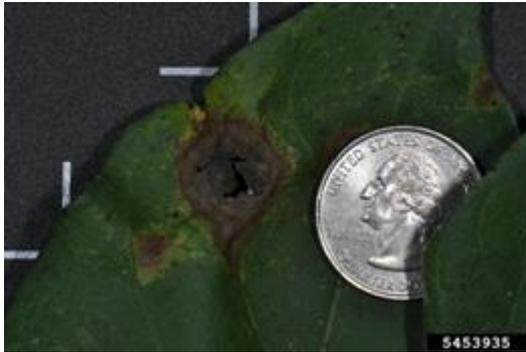


Fig.5 Target spot

3.6 Bacterial blight/angular leaf spot of cotton

The water-soaked spots which appear on leaves, are delimited by leaf veins, giving them an angular appearance. The lesions increase in size and turn black. Leaves then drop from the plant. Elongated gray-black lesions extending from the leaves to petioles and stem may be caused by the disease, which is known as the "blackarm" phase. The stem may get weakened by the highly severe blackarm symptoms. The water-soaked lesions may be present on bolls.

Cause: Disease is often introduced to cotton fields by infested seeds.



Fig.6 Bacterial blight

4. Literature Survey (Methodologies Used)

Farmers all over the country would be greatly benefited if this disease detection process is automated. Many researches and organizations have given multiple techniques and solutions for detection and classification of several diseases on cotton crop to get understanding of this research

area, we carry out a study on various types of plants with diseases. This survey will help to propose a novel way to make the disease detection more accurate and efficient.

Images from a mobile phone can also be used as a form of real time input for disease detection methodology[1].The image was converted into a grayscale image as it is easy to extract notable features from it .The color filter was applied and affected leaf spot color used RGB Pixel counting values (each and every disease) feature .Green pixels are the prominent indicators of healthy part in leaves. Green pixels are masked using color filtering and then they are separated. To get clarity on the edges Canny and Sobel edge detectors using Homogeneous operator techniques. The Homogeneity-based edge detector takes the result of any edge detector and divides it by the average range of the part. This division removes the effect of not level lighting in the image.

Color models play an important role in feature extraction. Three color models are prominent in measurement of the severity of damage in the leaf due to disease or pest, namely RGB color model, HSI color model and YCbCr color model. To measure the severity of the disease the degree of damage was calculated as the ratio between the area of damaged leaf to the total area of the leaf. The HSI color model is very important and attractive color model for image processing applications because it represents colors similarly how the human eye senses colors. The HSI color model represents every color with three components namely hue (H), saturation(S), intensity (I). The YCbCr color model has been defined in response to increasing demands for digital algorithms in handling video information, and has since become a widely used model in a digital video. These color models separate RGB (Red-Green-Blue pixels) into luminance and chrominance information and are useful in compression applications however the specification of colors is somewhat unintuitive. YCbCr model was considered as more intuitive than the other considered models.[2]

The use of transfer learning for classification processes can be very advantageous. The main advantage in using transfer learning is that instead of starting the learning process from scratch, the model starts from patterns that have been learned when solving a different problem which is similar in nature to the one being solved. This way the model leverages previous learnings and avoids starting from scratch.[3] uses a public dataset of 54,306 images of diseased and healthy plant leaves, a deep convolutional neural network is trained to classify crop species and disease status of 38 different classes containing 14 crop species and 26 diseases. This research compared five different architectures including VGG16, ResNet50, InceptionV3, InceptionResNet and DenseNet169. Stochastic Gradient Descent is used for optimization as it updates weights after every iteration. A further head model is built on top of these architectures. ResNet50 architecture stands out as the most accurate architecture.

Various Machine Learning algorithms provide very high accuracy in classification of diseases. Supervised Classifiers like Support Vector Machine (SVM), k-Nearest Neighbors(k-NN) or Naive Bayes classifier are the more favored for classification than the Deep Learning approaches, mostly because of their complexity and computation limitations. Likewise, [4] uses a Multiclass SVM for classification of diseased cotton leaves. This model was Trained for 3 diseases (Red Spot Disease, White Spot Disease, Crumple Leaf Disease) with a total of 103 images. Acquired images are converted from RGB format to YCbCr format. YCbCr gives the luminance and chrominance information of the input image. Using color co-occurrence method, the input image is quantized. The quantized images are then converted to Gray Level Co-occurrence Matrix (GLCM) by defining the angle and distance between two pixels. Finally, from this GLCM matrix, the texture feature values are extracted for further classification. After extracting color and texture features, the classification is performed by using Support Vector Machine (SVM). As the proposed system consists of 4 classes, the multiclass classifier is used for the classification.

Some works follow one-step procedure for classification i.e., to train the model directly on vast amounts of labeled data. This eliminates the need for segmenting out the leaf image from the background of the entire image as the Deep Learning models will learn iteratively on the vast amount of data and give a considerably good accuracy. But in the cases of limited amount of data, a two step models can be used in which the first part will segment out the leaf and the other part will work on classifying the disease. [5] proposed such a two cascade KNN classifier model to eliminate the problem of noisy and cluttered background in the leaf images, making segmentation difficult. The first classifier segments the leaf from its background using 15 statistical features (Means of RGB channels, Standard deviations of RGB channels, Contrast, Correlation, Homogeneity, Energy, Entropy, Hue component, Gradient, Mean Luminescence) Then using Hue and Saturation from converting it to HSV color space the second classifier detects the disease and its severity. Severity is taken as the ratio of total percent of disease intensity in observation to the total number of leaves under observation.

Features to be extracted should be studied well before training them on Neural Networks as the neurons in the network learn on these features repeatedly through back-propagation and different loss functions. The features with more information gain will give a huge impact on the fine tuning of the model. Also, there are multiple algorithms in play to extract these specific features. Similarly [6] used a new Neural Network called Cross Information gain forward Neural Network(CIG-DFNN) to identify and detect cotton leaf diseases. Three features are used for matching the train image features in database images. Color variance based on Skew divergence feature is calculated by color histogram and color descriptor. The shape feature also used Skew divergence to calculate the values using Sobel and Canny through the find out edge variance, edge location using Edge detection method. The skew divergence texture feature is calculated by Gober filter and texture descriptor. Performance evaluation of proposed CIG-DFNN system based on the overall accuracy is 95%. This method helps to

correctly classify the diseases Bacterial Blight, Fusarium wilt, Leaf Blight, Root rot, Micro Nutrient, Verticillium wilt.

An completely integrated mobile application for disease detection will be much better than clicking an image of the diseased part and uploading it on a laptop or a computer and then passing it through our classifier model. It will be also prove convenient if the application also performed more additional functionalities supporting irrigation or infrastructure management.[7] proposed such an mobile application which is a Support Vector Machine based regression system for identification and classification of five cotton leaf diseases i.e. Alternaria, Bacterial Blight, Fusarium wilt, Cereospra, and Gray Mildew. An Android application was developed which detected and identified diseases and provided users with their solutions. The App also displayed the information such as humidity, moisture and temperature along with the water level in the container. By using this app farmers can turn on or turn off the relay to control the motor functions and sprinkler assembly according to need. In this paper four different sensors that are temperature, moisture, humidity and water sensors are used and interfaced with raspberry pi for soil quality monitoring.

Performance of any model depends upon the features selected for training and testing. This features can be selected through manual supervision as in [6] or one can employ a dimensionality reduction algorithm on the given dataset to get the important features as the author did in [8].The author used Principal Component Analysis(PCA) and Nearest Neighborhood Classifier (KNN) for diagnosis of diseases on cotton leaf. Principal Component Analysis(PCA) is an unsupervised machine learning technique which PCA is used to identify features with maximum variance and contribution of each component is tested with other components to form an covariance matrix. The optimal number of principal components is determined by looking at the cumulatively calculated variance ratio as a function of the number of components. After implementing PCA/KNN multi-variable techniques, statistical

analysis of the data which is related to the green channel of the image is done. Here Green channel will be taken into consideration for feature collection because disease or deficiencies of elements are reflected most prominently by green channel.

Active contour model(snakes) is a framework introduced in 1988 which is essentially specialized in delineating(separating) an object from cluttered 2D images. This model is an energy minimizing curve, which is influenced by external constraint and image forces that fits it on the object contours In [9], Active contour model is used for image segmentation and Hu's moments are extracted as features for the training of adaptive neuro-fuzzy inference systems. The method of using invariant moments as features for training the neural network is understandable as it would make it easy to define the object by a set of measurable quantities called invariants that are insensitive to particular deformation and has high discrimination power to differentiate the objects belonging to dissimilar classes. The Author uses Hu's moments defined in terms of central moments which are calculated by using the function `cent_moment` which is available in Matlab. Backpropogation is later used to tune the neural network.

MATLAB is an interactive programming environment for scientific computing which is mainly used in the fields of machine learning and data analysis. It is basically a computational library with tons of statistical functions and algorithms which can be used in Deep Learning models. Use of this software is found quite frequently in such kind of classification studies for disease detection in plants like in [10], the Author proposes to detect the different diseases of cotton by applying artificial neural networks. Artificial Neural Network (ANN) tool of MATLAB is used in this process to classify the quality of cotton leaf diseases based on the RGB and HSV components of the image. Back propagation as a function also exists in MATLAB which helps compute the gradient of the loss function with respect to the weights of the network for a single input - single output example, and does it efficiently, unlike a naive direct computation of

the gradient with respect to each weight individually at every instance.

A robust automatic cotton crop disease recognition method has been proposed using the different invariant feature descriptors with the support vector machine [11]. The choice of filter used to remove noise plays an important role. The Gaussian filter is used here which removes the noise by defining and calculating the weighted mean for the filter box. Gaussian filter blurs the image to remove the missing pixels and pixels with noise to create the smooth image. Feature extraction is done by using multi-dimensional Histogram of oriented gradients (HoG), which has been obtained from all of the color planes in the horizontal and vertical form in order to get the multidimensional reflection of the infected region of the cotton crop.

An Ensemble learning methodology such as a Random Forest Classifier which is internally built from a number of decision trees can also provide results as comparable to the Support Vector Machine. Such a study is conducted in [12], which uses Random Forest Classifier for detection of diseases on cotton plant leaves. Feature extraction method used here is Histogram of Gradients. The histogram of oriented gradients (HOG) is a technique element descriptor utilized as a part of Computer vision and image processing for the sake of object detection. Three features considered here are Hu moments, Haralick texture and color histogram. The author also suggests that the accuracy can be increased when trained with vast number of images and by using other local features together with the global features which are also invariant in some aspects such as SIFT (Scale Invariant Feature Transform), SURF (Speed Up Robust Features) and DENSE along with BOVW (Bag Of Visual Word).

While the 2D Convolutional Neural Networks are the most obvious in classification tasks through imaging, a relatively 3D Convolutional Neural Networks, which also considers depth while training, can also provide fruitful results. In [13], the author proposed a novel 3D deep convolutional neural network (DCNN) that directly assimilates the

hyperspectral data to identify charcoal rot diseases in soybean stems. 3D CNN allows considering both spatial and spectral correlations simultaneously. Saliency maps in this methodology enabled interpretability to track the physiological insights of model predictions.

This paper suggests that saliency map visualization can be used to explain the importance of specific hyperspectral wavelengths in classification of diseased and healthy soybean stem tissue.

Though this study was performed for charcoal rot in soybean stems, a similar approach can also be used to detect the fusarium wilt in cotton plants.

The size of the dataset is also a factor which affects the decision making process of the classification models using deep learning or transfer learning. If the database is large enough to include a wide variety of backgrounds for all classes being considered, the influence of unimportant objects may become diluted, but given the difficulties involved in building a truly robust and integrated plant disease image database, this is generally not the case[14]. The study was carried out using a dataset that varied in terms of plant species, diseases and image capture conditions, but also had a number of samples that often was too small for the CNN to thoroughly capture the characteristics and variations associated with each class. While this prevents the trained networks to be used in practice, those limitations produce a wealth of information that can be used in future studies on the subject.

Finding or building a comprehensive dataset for further such studies is a herculean task. In[15], the author discusses the challenges deep-learning faces in detecting plant diseases. The author states that DL methods require a greater amount of data. This is a drawback since currently available datasets are usually small and do not contain enough images, which is a necessity for high-quality decisions. A comprehensive dataset must contain images captured in different conditions, as much as possible. One of them is that currently available datasets do not contain images gathered and

labeled from real-life situations. Therefore, the training was conducted with images taken in a staged(controlled) environment. Another limitation is that currently, the proposed methods cannot detect multiple diseases in one image or can only detect single occurrences of the same diseases in one image[17].

The author suggests the use of Generative Adversarial Networks (GANs) could be used for generating synthetic data.

The author proposes a two-stage architecture, the Plant Disease Net, which is motivated by advances in GANs and transfer learning.

5. Importance of Automation

As consulting an expert in the early phase of disease formation can be difficult or rather inconvenient, an automated system can overcome this need and provide at par results to a farmer at his convenience. Over decades many researchers have experimented lot of research on cotton plant leaves to detect and recognize a type of diseases, this automation can find early disease that helps to prevent damages for plants and list of some diseases and techniques are specified above in the literature review, which actually bring to a close the importance of continuing the research for the next level of competency. The gap in the identification of disease is growing day by day, as finding pathologists are difficult. Automation helps to prevent the spending a lot of money on actual experts and wrongly prescribed fertilizers and pesticides.

Conclusion

The Literature review has a detailed explanation of the importance of disease detection both to cotton plants and to farmers.

This research puts a light on various image processing and feature extraction techniques along with different classification algorithms for disease detection on cotton plants. The Literature review also speaks about the rising importance of GANs and how it can tackle the problem of inadequate dataset.3D -CNNs were also noted for further consideration in disease detection using Images.

This paper will better guide the future enthusiasts in helping farmers with more efficient automated techniques in cotton leaf disease detection. The paper depicts the development in image processing in agriculture field and considering the type of disease for further research work.

REFERENCES

- [1]. P. Revathi and M. Hemalatha, "Classification of cotton leaf spot diseases using image processing edge detection techniques," 2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET), Tiruchirappalli, 2012, pp. 169-173, doi: 10.1109/INCOSET.2012.6513900.
- [2]. He, Qinghai & Ma, Benxue & Qu, Duanyang & Zhang, Qiang & Hou, Xinmin & Zhao, Jing. (2013). Cotton Pests and Diseases Detection Based on Image Processing. TELKOMNIKA Indonesian Journal of Electrical Engineering. 11. 10.11591/telkomnika.v11i6.2721.
- [3]. Abhinav Sagar, J. Dheeba. (2020). On Using Transfer Learning for Plant Disease Detection. Cold Spring Harbor Laboratory. 10.1101/2020.05.22.110957
- [4]. Supriya S. Patki, Dr. G. S. Sable. (2016). Cotton Leaf Disease Detection & Classification using Multi SVM. International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), vol. 5, no. 10, pp. 165-168
- [5]. A. Parikh, M. S. Raval, C. Parmar and S. Chaudhary, "Disease Detection and Severity Estimation in Cotton Plant from Unconstrained Images," 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, 2016, pp. 594-601, doi: 10.1109/DSAA.2016.81.
- [6]. Revathi, P. & Hemalatha, M.. (2013). Identification of cotton diseases based on cross information gain_deep forward neural network classifier with PSO feature selection. 5. 4637-4642.

- [7]. A. A. Sarangdhar and V. R. Pawar, "Machine learning regression technique for cotton leaf disease detection and controlling using IoT," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2017, pp. 449-454, doi: 10.1109/ICECA.2017.8212855.
- [8]. V. A. Gulhane and M. H. Kolekar, "Diagnosis of diseases on cotton leaves using principal component analysis classifier," 2014 Annual IEEE India Conference (INDICON), Pune, 2014, pp. 1-5, doi: 10.1109/INDICON.2014.7030442.
- [9]. P. R. Rothe and R. V. Kshirsagar, "Cotton leaf disease identification using pattern recognition techniques," 2015 International Conference on Pervasive Computing (ICPC), Pune, 2015, pp. 1-6, doi: 10.1109/PERVASIVE.2015.7086983.
- [10]. N. Shah and S. Jain, "Detection of Disease in Cotton Leaf using Artificial Neural Network," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 473-476, doi: 10.1109/AICAI.2019.8701311.
- [11]. Prashar K, Talwar R, Kant C (2017) Robust automatic cotton crop disease recognition (ACDR) method using the hybrid feature descriptor with SVM. In: 4th International conference on computing on sustainable global development, vol INDIACom-2017, PaschimVihar, New Delhi
- [12]. S. Ramesh et al., "Plant Disease Detection Using Machine Learning," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, 2018, pp. 41-45, doi: 10.1109/ICDI3C.2018.00017.
- [13]. Nagasubramanian, K., Jones, S., Singh, A.K. et al. Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods* 15, 98 (2019). <https://doi.org/10.1186/s13007-019-0479-8>
- [14]. Jayme Garcia Arnal Barbedo,(2018),Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification,Computers and Electronics in Agriculture,Volume 153,2018,Pages 46-53,ISSN 0168-1699,
- [15]. Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., Stefanović, D. (2019). Solving Current Limitations of Deep Learning Based Approaches for Plant Disease Detection. *Symmetry* 11, 21. doi: 10.3390/sym11070939
- [16]. [https://www.ibef.org/industry/agriculture-india.aspx#:~:text=Agriculture%20is%20the%20primary%20source,\)%20in%20FY20%20\(PE\)](https://www.ibef.org/industry/agriculture-india.aspx#:~:text=Agriculture%20is%20the%20primary%20source,)%20in%20FY20%20(PE)).
- [17]. S. Begampure, P. M. Jadhav, "Comprehensive Review Of Generic Object Detection Frameworks Using Deep Learning Approach," IEEE International Conference on Contemporary Engineering and Technology, April, 26th-27th, 2019, Chennai.