INTERNATIONAL JOURNAL OF TECHO-ENGINEERING

ISSN : 2057-5688

EARLY DISEASE CLASSIFICATION OF MANGO LEAVES USING FEED-FORWARD NEURAL NETWORK MODEL

BALASUBRAMANYAM MEDA[1]

S KALYAN[2]

[1]PG Scholar, [2]Assistant Professor, Department of Computer Science and Engineering, Sarada Institute of Science Technology and Management, Ampolu Road, Srikakulam -532 404 Andhra Pradesh, India

Abstract: *Plant disease is described by a physiological disorder that alters the normal structure of the plant development, growth, and function. Disease can affect the quality and quantity of the crops that can affect the economics of countries like Bangladesh in which agriculture is the principal job. Rice being the primary crop, identifying the disease of mango is important to avoid loss of quantity and yields. The classification of rice diseases includes visual patterns and colors of the affected part. The manual observation of color and pattern used to categorize the disease is a lot of work and may be less effective when dealing with non-native illnesses. This paper introduces a novel method of detecting and classifying diseases by percent of RGB value of the affected region by using image processing. After the amount of RGB of the affected area is determined, and then grouped into various categories and fed to a simple classifier known as Naive Bayes which categorizes the illness into different types. This research has developed an app for mobile use to identify mango plant diseases system that uses fuzzy entropy and classifier probabilistic neural networks (PNN) which operates using Android smartphone's OS. Mango disease is a significant cause of lower yield and lower profits in rice production. Rice is among the main foods consumed by people around the world. However, the rice production is hindered by various kinds of mango disease. One of the most prevalent illnesses associated with mango is called leaf diseases. In general, it can be long and tiring for farmers in remote regions to detect the symptoms of mango leaf diseases because of the absence of experts. While experts are readily available in certain areas, detection is carried out by the naked eyes, which leads to inaccurate recognition at times. This is the most significant issue in the entire world in the agriculture sector. It is important to detect this disease can prevent enormous economic loss to the farmer. This paper proposes a machine-learning method to identify the signs for the disease within the rice plants. The detection of disease in plants is achieved using a machine* *learning algorithms. Pictures of healthy as well as blast disease afflicted leaves are taken to be used in the system proposed.*

Keywords: Neural network, image classification, plant disease, feature selection, precision agriculture.

I. Introduction

The plant diseases that plague the world are the primary source of quality and quantity losses in the agricultural production. These losses have a negative impact on the cost of production and the profits of all the agricultural stakeholders. However, equipment for rapid and accurate identification are scarce. The well-being and livelihood of farmers, along with the supply of food and nutritional security of a country are in danger if any type of epidemic occur. Traditionally the farmers and plant pathologists use their eyes to identify diseases and take their decisions based upon their experiences but this is not always reliable and can be biased as at the beginning, several types of illnesses seem to be the same. Furthermore, their experiences have to be passed on generation through generations. This leads to ineffective use of pesticides which results in more expensive production costs. Based on these points of evidence, the requirement for a reliable disease detection system and an accurate database that can assist farmers is essential, particularly for novices and young farmers. Computer vision advances open the way to achieve this through modern Deep learning (DL) or machine learning (ML) algorithms. It is also necessary for a prompt disease detection system that can protect the crop from disease before it becomes. There are numerous studies done for this purpose. The majority of them use the so-called "Plant village""" dataset, which is a well-known dataset that is accessible on the internet and with CNNs being the most well-known models. But, CNNs require a lot of data to train [22. In this study we have proposed two strategies: CNN models enhanced with transfer learning (TL) and Artificial Neural Network (ANN) using the feature Selection (FS) for address the multi-class classification problem for three kinds of diseases that are Anthracnose, Gall Midge, and Powdery Mildew. The proposed frameworks are aimed to improve the accuracy of models in the event that data is not available.

II. Related review

CNNs are a kind of hierarchical model in which the features of an object are taught by training numerous instances. They are composed of layers that are later constructed on top of the previously learned features [33. Saleem et al. [4] carried out a review of plant diseases detection and classification with Deep Learning techniques. The author concluded that "PlantVillage data set features a simple or plain background , and that more realistic scenarios must be taken into consideration. Additionally, multispectral or hyperspectral imaging should be utilized in conjunction with DL models to create early disease detection systems as well as a wider selection of training data must be obtained from multiple sources across different geographical regions cultivation conditions, as well as the modes of image capture. Konstantinos et al. [5] have implemented the Visual Geometry Group (VGG) model to detect diseases in plants The network was able to achieve 99.53 percent accuracy over the data provided in [11. Rangarajan et al. [6] employed AlexNet as well as VGG16 to identify tomato leaf diseases. VGG16 achieved 97.29 percent, and AlexNet was able to reach 97.49 percent accuracy. Mohanty et al. [7] have implemented a Transfer learning (TL) method using an already trained AlexNet to recognize crop diseases. The model is able to classify 26 distinct diseases across 14 species of crop with a size of 54,306 photos and 99.35 percent accuracy. Too et al. [8] carried out a study of the effectiveness of several deep learning models , namely Visual Geometry Group (VGG), Inception V4, ResNet, and DenseNet in the classification of diseases with the "Plant Village"" dataset and found that DenseNet is the one that was most effective with 99.75 percent accuracy. VGG16 was also used by Shijie et al. [9] to categorize tomato-related diseases, and had a 98 percent accuracy. There are many other applications for CNN are also covered in [10[10] to classify tomato diseases.

Khirade et al. [11] reviewed a variety of the various methods used to divide the infected portion of the plant. The study also looked at methods for feature extraction and classification to determine the features of affected leaves as well as the classification of diseases of plants. A variety of approaches employ ANN methods to determine the type of diseases in plants, like self-organizing feature maps backpropagation algorithm as well as SVMs, support vector machines (SVMs) and others. Singh et al. (12) used ANN in conjunction with image segmentation to identify illnesses on different varieties of plants, including beans, bananas, jackfruit and mango, as well as lemon potato, tomato and the sabota. First using it is used to determine the minimum distance Criterion using K-Mean Clustering is employed, and the classification is performed using SVM. The method proposed reached an average precision of

97.6 percent. Kulkarni et al. [13] have proposed a technique employing ANN along with Gabor filter to extract features to detect early signs of plant disease and the possibility of recognizing as high as 91 percent. The ANN utilized the combination of color and texture characteristics to classify.

In sum, the previous research findings are important. But, there are areas of uncertainty. For one, "Plant Valley"" is the most commonly utilized dataset in earlier studies. This dataset contains 54.303 healthful and diseased leaf photos belonging to 38 different types of plants and diseases. The majority of them are at the final stages of disease spread, in which the affected area is vast compared to the size on the leaf which means it is easy to identify the kind of disease. Also, when testing with images that were taken in conditions that differ from the ones used for training and conditions, it was observed that the CNN algorithms' efficiency dropped dramatically, as noted by Mohanty and co. [7]. Also, he suggested additional training videos that are more realistic situations should be taken into consideration. Another thing to note is that the "Plant Village does not have a mango leaf samples that is the primary issue in this study. A brand-new study [14] utilized in [10] in order to determine the diseases of mango leaves but they concentrate on healthy leaves and those that are infected by "Anthracnose'. In the study involving ANN researchers, they used various processing methods to extract desired characteristics from the photos before putting them into the ANN. But there was no mention of a step to select features to choose the most beneficial features.

III. Proposed Approach

Our data set includes more than 450 photos of mango leaves, belonging to four distinct types (three ailments and one health) Anthracnose Gall Midge, Powdery Mildew and Healthy. They are also classified into four categories of our classification, as illustrated in Figure. 1. The samples come from different locations within An Giang province, which is widely regarded as one of the regions with the highest production of mangoes in Vietnam. The leaves were taken from the tree when blobs began appearing on the tree, and their photos were captured on one day. The images are taken with cameras with a Resolution of 3096 x 3096, with no background. The photos are captured under various lighting conditions in the chamber in Fig. 2. The chamber is comprised consisting of an aluminium frame, and one model camera model CANON60D is positioned on top of it to point directly to the frame on the bottom. Additionally lighting sources, they are set around the camera in order to minimize shadows from an object. The intensity of the light source can be altered to mimic realistic lighting conditions. The flowchart of this process is illustrated in Fig. 3. In the beginning, Mango images are pre-processed by rescaling to an lower resolution, and comparing to their original dimensions. The next step is the centre alignment process is responsible for ensuring the figure 1. Four leaf diseases are studied during this research: The region of a leaf at the center of the image, and it should be exactly to the upper and lower edges of an image. Because there are a variety of shades of leaf images so we employ the method of enhancing contrast to alter the pixel intensity, which can are beneficial in the case of providing more information in specific regions of the image. The values of the image are normalized within the range [0,1] so that loss functions will achieve the optimal value for the entire image quickly. Normalization can also speed the process of convergence of this back-propagation method. The image data is split into three sets: Training, Validation, and Testing sets The first, 20 percent of the entire dataset is set to be test set, then the remainder of the dataset is divided into the same fashion, 80/20, with 20% representing an validation set, and the other 80 percent being the training set. The validation and training set is used to train the model, and tests set serves for evaluating how well the model. The data is randomly split by using Python's "train_test_split" built-in function in the "scikitlearn' library" [17The data is split randomly using the "scikitlearn'. The method of shuffle in Python was employed in conjunction with "train_test_split' to create random subsets. This method is based upon the random number generator, and reorders images in order to ensure the process of sampling is impartial. We also divide the data in the "'PlantVillage collection" similarly to develop already-trained CNN models.

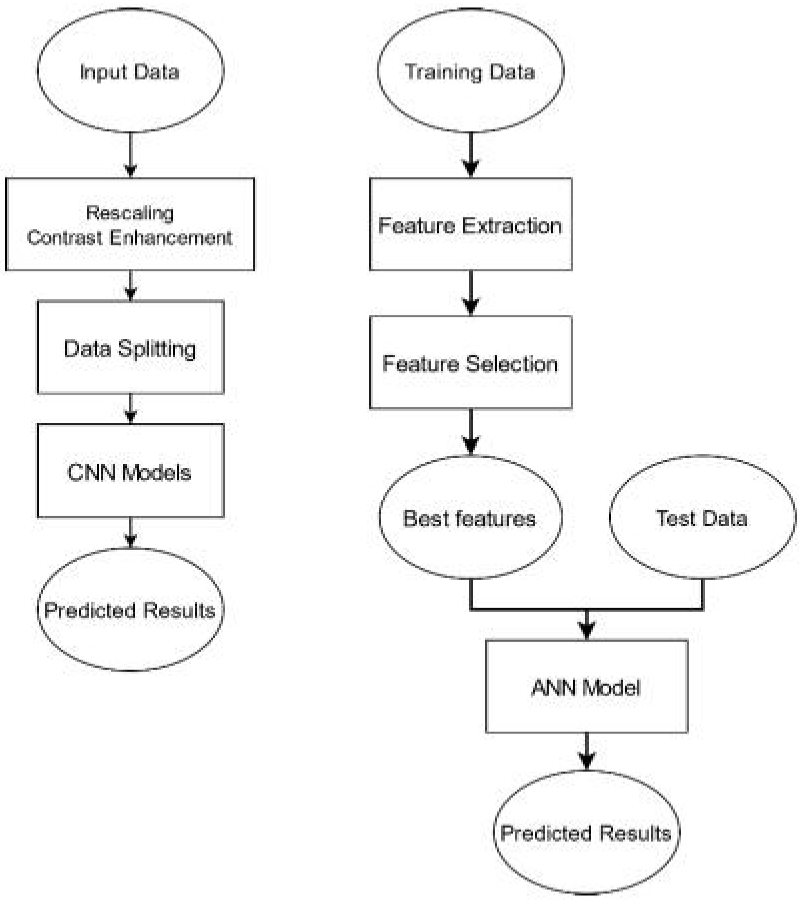

Figure 1: Flowchart of the two approaches in this study

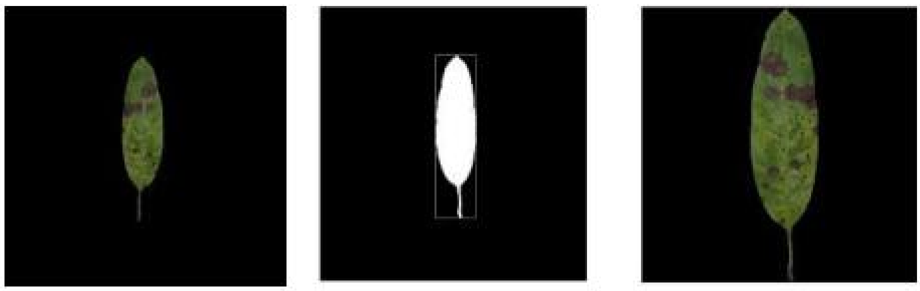

Figure 2: An example of rescaling and center alignment for a leaf image

IV. IMAGE PRE-PROCESSING

Since leaves come in various sizes, it is important to rescaling the images to ensure that the training image and test image share the same size. Rescaling can be used to reduce the original images into lower resolution , which is which is 256 x 256 pixels, to be precise. In the beginning, the original image is divided and then converted into binary to determine the minimal bounding box. The vertical dimension of this bounding box is used to adjust the resizing up to 256 pixels, ensuring that each leaves correspond exactly to each other on the image that was scaled. The horizontal dimension that the bounding boxes have is utilized to move the leaf image to the exact middle of the scaled image. The result can be seen in Fig. 2. Because of the various contrasts within the leaf region the method of contrast enhancement can be used to alter the intensity of pixels, which is beneficial by giving more details in certain regions of an image. A variety of techniques to enhance contrast have been used extensively to enhance the quality of an image [18and 19. This paper aims in order to enhance features that have low in contrast and to increase in terms of quality of contrast using the contrast enhancement method that is modeled on the one used in [19] prior to further analysis. The principal idea behind this approach is to keep the brightness of an input image while adjusting contrast within local regions. First, the image input with RGB color channels are converted to HSI ones. This method focuses only upon the parameter intensity but keeps additional hues and values of saturation. After that, the intensity is divided in two sub-parameters that include low and high groups. This is accomplished using the search algorithm shown in the following equation.

$\gamma hi = \{\gamma (i) | i > \gamma m\}, \gamma lo = \{\gamma (j) | j \leq \gamma m\}$ (1)

where γhi and γlo are intensity high and low groups respectively, γm is a trial threshold intensity value which is defined to divide the image into two sub-images. After obtaining estimates of the two sub-parameters of intensity, a combination of them is performed to achieve the enhanced intensity. The enhanced intensity is calculated by the following equation as follows:

$\gamma enhance (i) = \gamma lo + (\gamma hi - \gamma lo) \times \chi (i)$ (2)

After the process of enhancing contrast We then look into four of the most popular CNN models, which are AlexNet [15], VGG16 [18], ResNet [12] (ResNet-50 variant). In this study, we maintained the original design of the CNN and only altered the final layer that is fully connected of them to have four nodes in accordance with the four classes we have identified in our study. We also use transfer learning to refine the models and enhance their performance. Transfer learning (TL) is an answer to the shortage of data for training within deep learning13. The term "TL" refers to using the information from a specific task in order to solve a different task. When it comes to deep-learning, TL aids the models in learning the attributes of an extensive dataset, so that it can perform better with a relevant dataset however, it could be smaller in size and this technique has demonstrated efficiency in image classification tasks [7, 14and [15and [15]. In our research in the beginning, the models are developed with an existing dataset called the Plant Village dataset. The Plant Village dataset contains a vast amount of data that lets the convolutional layers of the models to be trained to understand similar features efficiently. On the basis of the Plant Village dataset, pre-trained models are developed, and then they are then trained again on our data set to test the models. The cross-entropy method is employed as a loss-function to calculate the error prediction following an initial classification. Additionally the optimization algorithm that is used for the process of training is performed by the Adam optimizer. The maximum number of epochs needed to learn the proposed model is 30, with an initial rate of 0.0005.

The Feature Extraction method for the Ann Approach is a Contrast-Limited adaptive histogram equalization (CLAHE) was developed by K. Zuiderveld in 1994 [1717. The method analyzes a graph of intensities within a context area that is centered around every pixel, and then sets the intensity of the pixels as the rank of the pixel's intensity on its histogram. This histogram is an altered version of the normal histogram, where the enhancement in contrast caused by the method at every intensity level is restricted to a limit that is set by the user. In this research, CLAHE is utilized to accomplish the thresholding on the images. This image was transformed to HSV format then CLAHE applies to the H channel to increase intensity of deficient regions. The regions that are defective are separated and returned on to the initial image. The procedure is described in Fig. 6. In this method instead of identifying the kind of disease leaves are infected similar to the method used by CNNs we take the leaf and learn about the features and categorize the different types of blobs that are infested. This approach is more precise in the event that the leaf is affected by multiple types of disease.

Proposed Algorithm 1 Adaptive Particle – Grey Wolf Optimization

---

Initialize the particle population

Initialize parameters

while (t < Max number of iteration)

for each particle with position xp

calculate fitness value f(xp)

if f(xp) is better than pbestp then

pbestp ← xp endif if f(pbestp) is better than

gbest then gbest ← pbestp

endif

end for

update w according to equation (5)

for each particle with position xp

update c1, c2 according to equation (3), (4)

calculate velocity of each particle by equation (6)

update position of each particle by equation (7)

end for

if rand (0,1) < prob

run GWO

xp = position of the best wolf

endif

t = t + 1

end while

return g

They have a variety of local optimizations. The problem gets worse when you consider larger dimensions and is often employed to test meta-heuristic algorithms. "Range" indicates the limit of the search area, while fmin represents the best. The APGWO algorithm that is proposed APGWO produces competitive results when as compared to the standard PSO or GWO algorithm. Fig. 7 shows some instances. The APGWO-wrapper solution will be a binary array having a size that is 1xn in which n represents the total amount of features. Certain features will be assigned the value 1 and zero otherwise. The parameters set for the algorithms are as follows: 20 search agents (for PSO main loop), 20 search agents (for nested GWO loop), 20 iterations for main PSO loop, 5 iterations for nested GWO, wMax = 0.9, wMin =0.2.

V. Conclusion and Future Enhancement

In this article, we have proposed the multi-class classifier for mango leaf diseases by using deep neural networks. In the beginning the wrapper-based feature-selection method employing an Adaptive Particle-Grey Wolf metaheuristic (APGWO) was applied to select 81 features from of the original 114 features. The features selected are used as inputs to the MLP to perform the task of classification. The method developed beat deep learning models like VGG, AlexNet, ResNet-50 that are already improved through transfer learning (89.41 percent vs 78.64 percent, 79.92%, and 84.88 percent in turn). Additionally this MLP network is smaller, leading to a faster performance. This is beneficial as we intend to apply this method on devices with limited resources like smartphones. In addition, we can fine tune the parameters of this MLP model, including the number of layers, amount of nodes hidden and also an activation method. Plantations are used to produce the most diverse data. Achieving a better control of the feature selection algorithm. Develop a complete health monitoring system that is used on a variety of platforms.

VI. References

[1] S. D. Khirade and A. B. Patil, ''Plant disease detection using image processing,'' in Proc. Int. Conf. Comput. Commun. Control Autom., Feb. 2015, pp. 768–771, doi: 10.1109/ICCUBEA.2015.153.

[2] V. Singh and A. K. Misra, ''Detection of plant leaf diseases using image segmentation and soft computing techniques,'' Inf. Process. Agricult., vol. 4, no. 1, pp. 41–49, Mar. 2017, doi: 10.1016/j.inpa.2016.10.005.

[3] A. H. Kulkarni and A. Patil, ''Applying image processing technique to detect plant diseases,'' Int. J. Mod. Eng. Res., vol. 2, no. 5, pp. 3661–3664, 2012.

[4] S. S. Chouhan, A. Kaul, U. P. Singh, and Madhav Institute of Technology Science, ''A database of leaf images: Practice towards plant conservation with plant pathology,'' Mendeley Data, V1, 2019, doi: 10.17632/hb74ynkjcn.1.

[5] L. Yu and H. Liu, ''Efficient feature selection via analysis of relevance and redundancy,'' J. Mach. Learn. Res., vol. 5, no. 10, pp. 1205–1224, 2004.

[6] D. P. Hughes and M. Salathe, ''An open access repository of images on plant health to enable the development of mobile disease diagnostics,'' 2015, arXiv:1511.08060. [Online]. Available: http://arxiv.org/abs/ 1511.08060

[7] J. G. A. Barbedo, ''Factors influencing the use of deep learning for plant disease recognition,'' Biosyst. Eng., vol. 172, pp. 84–91, Aug. 2018, doi: 10.1016/j.biosystemseng.2018.05.013.

[8] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, ''Convolutional neural networks: An overview and application in radiology,'' Insights Imag., vol. 9, no. 4, pp. 611–629, Aug. 2018, doi: 10.1007/s13244-018- 0639-9.

[9] M. H. Saleem, J. Potgieter, and K. M. Arif, ''Plant disease detection and classification by deep learning,'' Plants, vol. 8, no. 11, p. 468, Oct. 2019, doi: 10.3390/plants8110468.

[10] K. P. Ferentinos, ''Deep learning models for plant disease detection and diagnosis,'' Comput. Electron. Agricult., vol. 145, pp. 311–318, Feb. 2018, doi: 10.1016/j.compag.2018.01.009.

[11] A. K. Rangarajan, R. Purushothaman, and A. Ramesh, ''Tomato crop disease classification using pre-trained deep learning algorithm,'' Procedia Comput. Sci., vol. 133, pp. 1040–1047, Jan. 2018, doi: 10.1016/j.procs. 2018.07.070.

[12] S. P. Mohanty, D. P. Hughes, and M. Salathé, ''Using deep learning for image-based plant disease detection,'' Frontiers Plant Sci., vol. 7, p. 1419, Sep. 2016, doi: 10.3389/fpls.2016.01419.

[13] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, ''A comparative study of fine-tuning deep learning models for plant disease identification,'' Comput. Electron. Agricult., vol. 161, pp. 272–279, Jun. 2019, doi: 10.1016/j. compag.2018.03.032.

[14] J. Shijie, J. Peiyi, H. Siping, and S. Haibo, ''Automatic detection of tomato diseases and pests based on leaf images,'' in Proc. Chin. Autom. Congr. (CAC), Oct. 2017, pp. 2510–2537, doi: 10.1109/CAC.2017.8243388.

[15] U. P. Singh, S. S. Chouhan, S. Jain, and S. Jain, ‘‘Multilayer convolution neural network for the classification of mango leaves infected by anthracnose disease,’’ IEEE Access, vol. 7, pp. 43721–43729, 2019, doi: 10. 1109/ACCESS.2019.2907383.

[16] J. O. Pedersen and Y. Yang, ‘‘A comparative study on feature selection in text categorization,’’ in Proc. ICML. 14th Int. Conf. Mach. Learn., 1997, pp. 412–420. [Online]. Available: https://pdfs. semanticscholar.org/f87c/96a5f22a741fa1015089f15beb78cda95aa2.pdf

[17] K. Yan and D. Zhang, ‘‘Feature selection and analysis on correlated gas sensor data with recursive feature elimination,’’ Sens. Actuators B, Chem., vol. 212, pp. 353–363, Jun. 2015, doi: 10.1016/j.snb.2015.02.025.

[18] A. Jain and D. Zongker, ‘‘Feature selection: Evaluation, application, and small sample performance,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 2, pp. 153–158, Feb. 1997, doi: 10.1109/34.574797.

[19] C. De Stefano, F. Fontanella, C. Marrocco, and A. S. D. Freca, ‘‘A GAbased feature selection approach with an application to handwritten character recognition,’’ Pattern Recognit. Lett., vol. 35, pp. 130–141, Jan. 2014, doi: 10.1016/j.patrec.2013.01.026.

[20] E. Zorarpacı and S. A. Özel, ‘‘A hybrid approach of differential evolution and artificial bee colony for feature selection,’’ Expert Syst. Appl., vol. 62, pp. 91–103, Nov. 2016, doi: 10.1016/j.eswa.2016.06.004.

[21] J. Too, A. Abdullah, N. M. Saad, N. M. Ali, and W. Tee, ‘‘A new competitive binary grey wolf optimizer to solve the feature selection problem in EMG signals classification,’’ Computers, vol. 7, no. 4, p. 58, Nov. 2018, doi: 10.3390/computers7040058.