Review on Leaf Plant Disease Classification Using Machine Learning Techniques

U. I. Ismail[1*], M. K. Ahmed[2]

[1]*Department of Computer Science Federal University of Kashere, Gombe Nigeria*
[2]*Department of Computer Science Gombe state University, Gombe Nigeria*
**Corresponding Author*

Abstract: **Agriculture plays a vital role in the world economy. It basically provides job opportunities for the teaming population, eradicates poverty and contributes to the growth of the economy. Hence the need for improved effort for classifying diseases in plant from its leaf is important as it leads to increase in crop yield. Machine learning methods had being used in leaves plant diseases classification. This paper reviews various techniques used for plant leaf disease classification, and found that Most of the researchers used Support Vector Machine (SVM) algorithms for plant disease classification which they concluded that (SVM) is not suitable for large dataset and it does not perform very well when the dataset has more noise, also the target class will be overlapping. To overcome this difficulties a proposed methodology with different approaches to Machine learning was suggested; Deep learning is a sort of machine learning in which a model figures out how to accomplish classification tasks in a direct way from pictures, Neural network will be train using Fine-tuning techniques on different neural networks architectures and at the end comparisons will be done to find out the best neural networks that will be the best for providing an improved solution for leaf plant disease classification by checking their performance best on their accuracy and confusion matrix.**

***Keywords*: Plant disease, Machine learning, classification, Neural Network, Fine-tuning**

I. INTRODUCTION

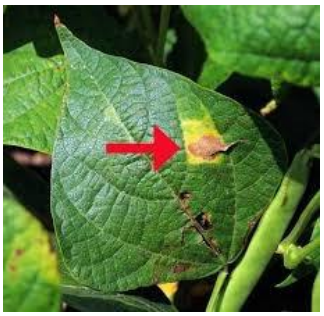
Agriculture is the prime income source in various countries of the world. Grounded on the significance of agriculture, farmers select their crops, paddies, and the related pesticide to restructure the development of the plant in the limited time [1]. In Nigeria, Agriculture is the basis of the economy and employs 75% of the work force. Despite the importance of petroleum as a major contributor to gross domestic product (GDP), the role of agriculture remains most significant in Nigerian economy since independence. Agriculture provides employment for most rural dwellers and it accounts for more than one third of total gross domestic product (GDP) and labour force for the majority of rural Nigerians [2]. Diseases of plants are major causes of plant damage and consequently agriculture and economic loses. Timely identification of plant disease is a critical factor to make harvest healthy and fruitful. The most common approach for identification of diseases of plants is visual observation by experts. But this approach can be time consuming or difficult due to lack of experts at the sites of cultivation [3].Support vector machines (SVM) was normally used for classification and identification of plant diseases via image properties. In machine learning, SVM is a supervised learning associated with learning algorithms to analyze and classify data [4]. The original SVM algorithm was invented by Vladimir N. Vapnik and Alexey Ya Chervonenkis in 1963, while the current version of the algorithm is published in 1995 by Corinna Cortes and Vapnik [5]. The earlier version of SVM has the restriction that training data can be separated without errors while the new version of the algorithm extends it to non-separable training data [5]. A semi-supervised support vector machine classifier (S3VM) based on active learning and context information is presented by [6] First a semi-supervised learning method uses active learning to select unlabeled samples as the semi-label samples. Then the context information is exploited to further expand the selected samples and re-label them, along with the labeled samples train S3VM classifier. This paper presents a review on different machine learning techniques for detection and classification of plant disease using leaf recognition and proposed an improved method for the classification of plant disease by it leaves. Diseases in plant create a certain patterns on the leaves of the affected plant. This review proposed a methodology with different approaches to Machine learning, Fine-tuning techniques will be use on different neural networks architectures and at the end it will be compared to find the best neural networks that will be the best for providing an improved solution to this problem by checking their performance in accuracy and confusion matrix.

Some Leaf Plant Diseases Analysis And Their Symptoms

Some common symptoms of leaf disease were discussed by [7] such as fungal, bacterial and viral:

(1) Bacterial disease symptoms: The disease is characterized by tiny pale green spots which soon come into view as water-soaked. The lesions enlarge and then appear as dry dead spots as shown in image (a) below.

(a) Bacterial disease on leaf



(b) Viral disease on leaf

(c) Fungal disease on leaf

(2) Viral disease symptoms: Among all plant leaf diseases, those caused by viruses are the most difficult to diagnose. Viruses produce no telltale signs that can be readily observed and often easily confused with nutrient deficiencies and herbicide injury. Aphids, leafhoppers, whiteflies and cucumber beetles insects are common carriers of this disease, e.g. Mosaic Virus, Look for yellow or green stripes or spots on foliage, as shown in image (b) above. Leaves might be wrinkled, curled and growth may be stunted.

(3) Fungal disease symptoms: Plant leaf diseases, that affected by fungus is shown in the image (c) above. Late blight caused by the fungus Phytophthora infester. It first appears on lower, older leaves like water-soaked, gray-green spots. When fungal disease matures, these spots darken and then white fungal growth forms on the undersides. Early blight is caused by the fungus Alternaria solani. It appears on the lower, older leaves like small brown spots with concentric rings that form a bulls eye pattern. When disease matures, it spreads outward on the leaf surface causing it to turn yellow. In downy mildew yellow to white patches on the upper surfaces of older leaves occurs. These areas are covered with white to greyish on the undersides as shown in the image (c) above.

General Structure of Plant Disease Classification Using Machine Learning

Several researchers explained the general structure for the classification of plant diseases by observing leaf of the plant. [8] Explained the digital image processing can be used to detect diseased leaf, shape and colour of affected area. Figure 1 explained the general flow of Image processing technique with five steps:

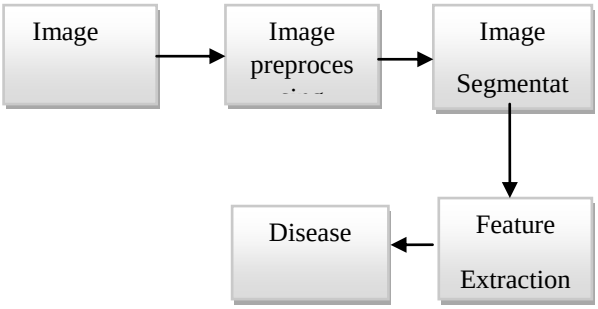


Figure 1. Structure of Leaf Plant Disease Classification

Image Acquisition: can be done by using a digital camera, digital scanner or smart phone camera. The image of leaf plant with the disease and the one without the disease were captured using that equipment to create the dataset.

Image preprocessing: Noise and the unwanted object are removed from the image and the techniques used at this stage are image resizing, image smoothing, increase the contrast and image enhancement etc.

Image Segmentation: This is a process of dividing the image into various part of the same feature; techniques used are include K – means, Converting RGB to HSV and other techniques

Feature extraction: Extraction of Colour, shape and texture feature of disease part of the plant can be done using Grey Level Co-occurrence Matrix(GLCM), Blend vision and machine intelligence etc.

Classification: Feature analysis, Machine learning model and fuzzy rule-based classification are applied to extracted features above for detecting and classifying the disease. Machine learning model mostly used are Artificial Neural Networks (ANN), Support vector machine (SVM), Random forest, K – nearest neighbor, and another type of machine learning model

Testing: The model is tested using a sample of feature left for testing the model, or by applying the image with disease and undergo all the step above hope the classifier will classify the disease. Checking the accuracy of the model is also done at this stage.

II. LITERATURE REVIEW

This stage talks about the most recent research works in plant leaf disease classification from different plant species [9] came with a cloud based system that determined if the crop is

healthy or unhealthy. The aim is to perform a real time classification using supervised learning. They concluded that Convolutional Neural Networks (CNNs) achieved better classification scores than Support Vector Machine (SVM) due to the size of the large data set they used where even an RBF kernel, so it does not help the case of the SVM as the data is nonlinear in nature and is a huge data set. They used the dataset of 1030 plant disease images in the project and are splited using the train_test_split module of sklearn library into 600 training images and 430 validation images. In this phase, the CNN model is tested using the validation images and obtain an accuracy score of 93.4% depicting the efficiency of the model. This accuracy score during the development of the model is used to improvise it by tweaking the hyper parameters, adding or removing layers and changing non-linearities [10] enhance classification of 20 different diseases from 10 distinct plant species from the dataset obtain from XDB plant disease database. This processing encompassed two pre-processing activities (image selection and resizing) and the application of two modified VGG architectures, VGG16 and VGG19, along with pre trained weights from ImageNet database. The classification performances of VGG16 and VGG19 on training and validation data were 94.3% and 87.1%, and 84.3%and 80.4%, respectively. A comparison study was carried by [10] based on classification metrics, such as accuracy, precision, recall and F1-score. The obtained results demonstrated that for this particular database, a pre-trained CNN with depth equal or smaller than VGG16 can compute disease-sensitive features which aggregate more refinement to recognize plant pathologies, which justified the better performance of VGG16 against VGG19 on training, validation, and test data.

Research conducted by [10] aim at presenting a study deals with automatic disease detection of plant leaf of *Phaseolus vulgaris* (beans) and *Camellia assamica* (tea) using image processing techniques. It involves image acquisition, image preprocessing, image segmentation, feature extraction and classification. Development of automatic detection system using advanced computer technology such as image processing help to support the farmers in the identification of diseases at an early or initial stage and provide useful information for its control. Classification is used in the interpretation of the extracted diseased region in an image which helps in the identification of the type of disease infection in leaves. In their analysis back propagation neural network (BPNN) is used which build association between known pattern of input and specific output. The input layer analyzes the diseased region while the output layer specifies the disease outcome of the affected region. [11] Their study has presented a reduced feature set based approach for recognition and classification of images of plant diseases. Algorithms for extraction of color and texture features have been developed, which are in turn used to train support vector machine (SVM) and artificial neural network (ANN) classifiers. The color and texture features have been used in order to work with the sample images of plant diseases**.** The study considers 900 sample images which obtained from department of plant pathology, at the University of Agricultural Sciences, Dharwad, INDIA with 6 classes of disease as fungal, bacterial, viral, nematodes, deficiency and normal (not affected) The results reveal that SVM classifier is more suitable for identification and classification of plant diseases affecting plants with 92% accuracy.

[12] Have worked on Decision Support System (DSS) that gives advice to the farmers as and when require over mobile internet. It focuses on the approach based on image processing for detection of diseases of soybean plants. 120 soybean images are captured directly from the farm using mobile camera having resolution greater than 2 mega pixels classify as 75 healthy and 45 diseased leaves and uses Support Vector Machine (SVM) to classify the images of soybean leaves as healthy and diseased. The algorithm comprises of four major steps: image acquisition, extracting the leaf from complex background, statistical analysis and classification. The pre-processing step includes conversion from RGB to HSV (Hue Saturation Value) color space. For extracting the region of interest (ROI) from the original image, multi-thresholding is used. The color based and cluster based methods are used for segmentation. The algorithm uses Scale Invariant Feature Transform (SIFT) technique which automatically recognizes the plant species based on the leaf shape. The SVM classifier proves its ability in automatic and accurate classification of images. Finally, [13] concluded from their experimental results that this approach can classify the leaves with an average accuracy of 93.79%. The restriction of their work is that they don't classify the diseases as they only detect the leaves as unhealthy or healthy. [14] This research focuses on the approach based on image processing for detection of diseases of soybean plants. The soybean images are captured using mobile camera having resolution greater than 2 mega pixels. The purpose of the project is to provide inputs for the Decision Support System (DSS), which is developed for providing advice to the farmers as and when require over mobile internet. Their work classifies the images of soybean leaves as healthy and diseased using Support Vector Machine (SVM). The algorithm comprises of four major steps: image acquisition, extracting the leaf from complex background, statistical analysis and classification. The pre-processing step includes conversion from RGB to HSV (Hue Saturation Value) color space. For extracting the region of interest (ROI) from the original image, multi-thresholding is used. The color based and cluster based methods are used for segmentation. The algorithm uses Scale Invariant Feature Transform (SIFT) technique which automatically recognizes the plant species based on the leaf shape [14] used SVM classifier proves its ability in automatic and accurate classification of images. Finally, it can be concluded from the experimental results that this approach can classify the leaves with an average accuracy of 93.79%. The system will enable the farmers to get advice from the agricultural experts with minimal efforts while [15] recommended a model that does not need pre-processing and can perform a successful classification. In this study,

EfficientNet deep learning architecture was proposed in plant leaf disease classification and the performance of this model was compared with other state-of-the-art deep learning models. The PlantVillage dataset was used to train models. All the models were trained with original and augmented datasets having 55,448 and 61,486 images, respectively. EfficientNet architecture and other deep learning models were trained using transfer learning approach. In the transfer learning, all layers of the models were set to be trainable. The results obtained in the test dataset showed that B5 and B4 models of EfficientNet architecture achieved the highest values compared to other deep learning models in original and augmented datasets with 99.91% and 99.97% respectively for accuracy and 98.42% and 99.39% respectively for precision.

[16] Recommend a methodology for analyzing and detecting of plant leaf diseases using digital image processing techniques. The experimental results demonstrate that the proposed system can successfully detect and classify four major plant leaves diseases: Bacterial Blight and Cercospora Leaf Spot, Powdery Mildew and Rust [17]. This research was aims to detect blast disease and reduce the crop loss and hence increase the rice agriculture production in an effective manner. In modern agriculture field, pest and disease identification is a major role of rice cultivation. Image classification by the use of deep convolutional neural networks of training and methodology used the facilitate a quick and easy system implementation. Pests and diseases are a threat to paddy production, especially in India, but identification remains to be a challenge in massive scale, and automatically collecting images from Image Net dataset. The results presented by [17] shows that we can effectively detect and recognize the rice diseases and pests including healthy plant class using a deep convolutional neural network, with the best accuracy of 96.50%. The significantly high success rate makes the model a really useful advisory or early warning tool, and an approach that would be further expanded to support an integrated plant disease identification system to work in real cultivation conditions.

Table 1: A glance on Tomato plant disease classification using different techniques

Authors	Classification Techniques	Accuracy %	Number of Disease	Dataset used	Number of images
Hiteshwari and Satish, 2016 [18]	Multi-Layer Feed Forward Back Propagation Neural Network	87.2%	6	Own images	180
Aravind k. 2018 *et .al* [19]	AlexNet and VGG16 net	97.29% and 97.49%	7	Plant Village	13,262
Sumair Aziz, 2019 *et.al* [20]	Support Vector Machine (SVM)	94%	5	Plant Village	1,882
Sue Han,2020 *et.al* [21]	VGG16 and IPM	44.54% and 28.13%	2	PV dataset	43,810

The table above talks about the distinctive research works, which has been completed in tomato leaf disease classification using different techniques with different number of images that used to classify different disease from tomato leaf.

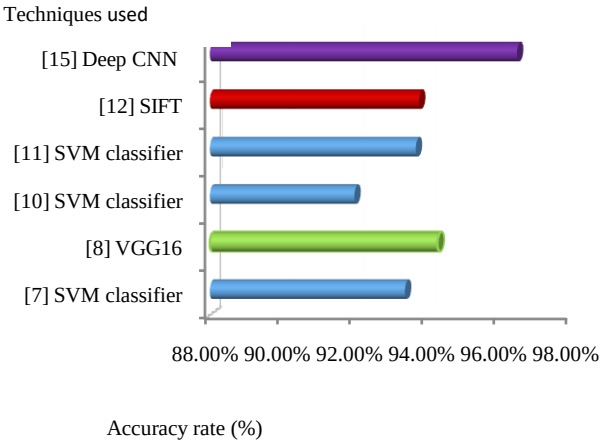


Figure 2: Accuracy Level Comparison of Different Techniques

Fig.2 above shows the accuracy level comparisons of some different techniques used in this review in classifying disease from the leaf of plant from different species using different techniques such as Deep Convolutional Neural Network (DCNN), Scale invariant Future Transform (SIFT), Support Vector Machine (SVM),VGG16.

III. PROBLEMS IDENTIFIED

Most of the researchers [7, 10, 11, 14] used supervised learning classification algorithms for plant disease classification they concluded that it is not suitable for large dataset as it stop learning when the dataset is getting too large and it does not perform very well when the dataset has more noise, the target class will be overlapping, to overcome these challenges the suggested approach will be discuss below.

IV. PROPOSED APPROACH

Deep learning with convolutional neural network will be applied on a small leaf dataset with fine-tuning techniques. An accuracy score might be increase. High overfitting will be note, also since the dataset is small, data augmentation will also be applied to the dataset with the images augmented

before it will be input into the model, differing from the many cases where augmentation is applied on-the-fly. This may increase the success rate value and may reduce overfitting. It will also show that augmentation that will be applied in this way may improve the performance of the model. Transfer learning with the same augmentation method will be applied, that may result in higher test accuracy and stable results with low overfitting. The proposed methodology may provide a good result on the plant leaf dataset that will be use in the experiments. Finally, an outline of a framework for improved plant disease classification using leaf recognition will be present. The deep learning framework will be developed with Python, Open CV, and Anaconda with Jupyter Notebook and Keras Deep learning model implementing Tensorflow in the backend.

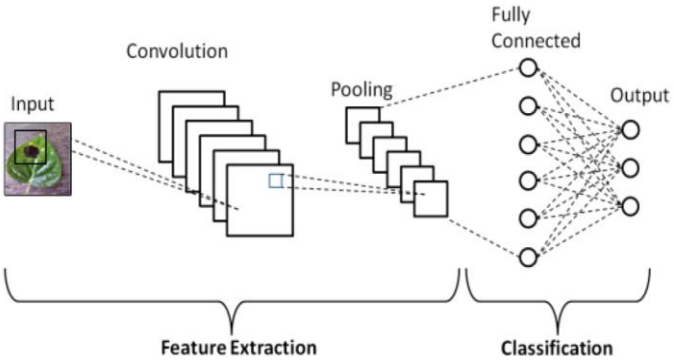


Figure 3. Classic Convolution Neural Network (CNN) architecture

V. CONCLUSION

Leaf plant disease classification stands as an essential area in plant pathology, contributing enormous percentage in incrementing the crop production in a country. In this review paper, leaf plant disease classification using images of the plant was suggested by fine-tuning with convolutional neural networks on different architectures like AlexNet, VGGNet, ResNet (50,101) and Inception (V3,V4) which at the end a comparison will be done to find out the best architecture that will provide an improve solution to the problem by evaluating and comparing their performance by checking their accuracy and confusion matrix.

REFERENCES

[1] Gayathri, Devi T. and Neelamegam, P. "Paddy leaf disease detection using SVM with RBFN classifier", International Journal of Pure and Applied Mathematics, vol. 117, no. 15, pp. 699- 710, 2017.

[2] Adams, Oluwadamilola, Kemi, "Challenges of Rice Production in Nigeria: A Case Study of Kogi State", Food Science and Quality Management, ISSN 2224-6088 (Paper) ISSN 2225-0557 (Online) Vol.74, 2018

[3] Murtaza, Ali Khan."Detection and Classification of Plant Diseases Using Image Processing and Multiclass Support Vector Machine", International Journal of Computer Trends and Technology (IJCTT) – Volume 68 Issue 4 – April 2020

[4] Stoean, C. Stoean, R. "Support vector machine and evolutionary Algorithms for Classification", Springer International Publishing https://doi.org/10.1007/978-3-319- 06941-8 2014

[5] Cortes, C. and Vapnik, V. "Support-vector networks". Machine Learning. URL https://doi.org/10.1007/BF00994018. 1995

[6] Gao, F. L., Zhang, Y., Sun, J., Wang, J., Yang, E. "A novel semi-supervised support vector machine classifier based on active learning and context information". Multidimensional Systems and Signal Processing; 27(4): 969-988. URL https://doi.org/10.1007/s11045-016-0396-1. 2016

[7] Vishnu, S. and Ranjith, A. "Plant Disease Detection Using Leaf Pattern", International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 6, June 2015.

[8] Saranya, G. M., Meenakshi, K. and Nithya, M. "Plant Leaf Disease Classification and Detection System Using Machine Learning". Article in Journal of Physics Conference Series 2020.

[9] Sue Han Lee, Hervé Goëaua, Pierre Bonnet, Alexis Joly, "New perspectives on plant disease characterization based on deep learning". Computers and Electronics in Agriculture 2020.

[10] Aravind Krishnaswamy Rangarajan, Raja Purushothaman, Aniirudh Ramesh. "Tomato crop disease classification using pre-trained deep learning algorithm".International Conference on Robotics and Smart Manufacturing (RoSMa2018).

[11] Hiteshwari Sabrol and Satish Kumar, "Fuzzy and Neural Network based Tomato Plant Disease Classification using Natural Outdoor Images". Indian Journal of Science and Technology, Vol 9(44), DOI: 10.17485/ijst/2016/v9i44/92825, November 2016.

[12] Sumair Aziz, Mudassar Bashir, Ovais Mughal, Muhammad Umar Khan, Arsalan Khan, "Image Pattern Classification for Plant Disease Identification using Local Tri-directionalFeatures". IEEE 978-1-7281-2530-5/19/$31.00 2019

[13] Jie Hang, Dexiang Zhang, Peng Chen, Jun Zhang and Bing Wang, "Classification of Plant Leaf Diseases Based on Improved Convolutional Neural Network". Article in Sensors September 2019.

[14] Lorick Jain, Harsha Vardhan, M.A, Nishanth, M.L and Shylaja,S.S. "Cloud-based system for Supervised Classification of Plant diseases using Convolutional Neural Networks".IEEE International Conference on Cloud Computing in Emerging Markets 2017.

[15] Vanessa Rezende, Michel Costa, Adam Santos and Roberto, C. L. de Oliveira. "Image Processing with Convolutional Neural Networks for Classification of Plant Diseases".Brazilian Conference on Intelligent Systems (BRACIS) 2019.

[16] Malti K. Singh and Subrat Chetia, "Detection and Classification of Plant Leaf Diseases in Image Processing using MATLAB"International Journal of Life Sciences Research Vol. 5, Issue 4, pp: (120-124), 2017.

[17] Jagadeesh D.Pujari, Rajesh Yakkundimath and Abdulmunaf Syedhusain Byadgi. "SVM and ANN Based Classification of Plant Diseases Using Feature Reduction Technique" International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 3, Nº7.

[18] Yogesh Dandawate and Radha Kokare. "An Automated Approach for Classification of Plant Diseases Towards Development of Futuristic Decision Support System in Indian Perspective", International Conference on Advances in Computing, Communications and Informatics (ICACCI) 2015.

[19] Ümit Atila, Murat Uçar, Kemal Akyol, Emine Uçar. "Plant leaf disease classification using EfficientNet deep learning model."Contents lists available at Science Direct Ecological Informatics (2021)

[20] R.Jeya Bharathi, "Paddy Plant Disease Identification and Classification of Image Using AlexNet Model" The International journal of analytical and experimental modal analysis Volume XII, Issue III,ISSN NO: 0886-9367. 2020

[21] Yin Min, Oo and Nay Chi Htun," Plant Leaf Disease Detection and Classification using Image Processing" International Journal of Research and Engineering ISSN: 2348-7860 Vol. 5 No. 9. 2018