2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence) | 978-1-6654-3701-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/Confluence52989.2022.9734162

Detection of Unhealthy citrus leaves using Machine Learning Technique

1st Nidhi Goyal
Amity University
Noida, Uttar Pradesh, India
nidhigoyal1991@gmail.com[1]

2nd Sumit Kumar
Amity University
Noida, Uttar Pradesh, India
Sumitkumarbsr19@gmail.com[2]

3rd Mukesh Saraswat
Jaypee Institute Of Information Technology
Noida, Uttar Pradesh, India
saraswatmukesh@gmail.com[3]

Abstract* - Precision agriculture increases the productivity with fewer resources and contributes in decision making process. The production of plant gets widely affected by the diseases which degrades the quality of the plants and causes financial loss to the farmers. In past decades, image processing methods are widely used to detect and classify the disease of the plant. In tropical areas of the world, citrus plants are the grown. To prevent plants from the diseases, detection of diseases in early stages is helpful for farmers. Therefore, this paper detects the healthy and unhealthy citrus leaves using different techniques. The Otsu thresholding is performed for segmenting the image and extracting the features. Further, optimized set of features are selected using various metaheuristics algorithm such as WOA, LOA, CSO and DE. The classification is performed using Support vector machine (SVM) and Multilayer perceptron (MLP). The different performance quality measures are evaluated such as recall, precision. F1-score, sensitivity and accuracy. The overall results shows that MLP provides accuracy of 74.8% and SVM gives 61.6% by utilizing WOA. The experimental analysis shows that MLP provides more accuracy than SVM.

Keywords- Precision Agriculture, Citrus Leaves, Metaheuristic Algorithm, Machine Learning

I. INTRODUCTION

Agriculture is one of the main sources of human sapiens but its salvation from diseases is more concerned nowadays. Some farmers take the help of agriculturists and some of them are not aware of contacting these agriculturists hence take a lot of time and money, which leads to unhealthy cultivation. Usually, detection of unhealthy crops by farmers takes time therefore there is need of automated system for prevention of disease at early stage. There are various method introduced for detection of unhealthy and healthy crop. Thus, continued detection of disease in crops is important for attaining the demand of human sapiens. For redemption of the crops, different techniques found to be very useful. However, detecting the diseases is still a big challenge due to variation in the crop images. The paper proposed the efficient methodology for citrus leaves analysis and detection of unhealthy leaves. The different technique are employed for appropriate feature selection methods for classifying the healthy and unhealthy plants. Supervised technique is when the data is known while unsupervised technique is used when data is unknown. Moreover, feature selection helps to find the best set of features that allow one to make a useful model [10]. Mostly, feature selection is used for doing best classification. Classification is a process of grouping the things in a cluster based on their common features. The citrus leave representation of healthy and unhealthy is shown in Fig.1.

(a)

(b)

Fig.1. (a) Healthy Citrus Leaves (b) Unhealthy Citrus Leaves

Many researchers have used different classification techniques for classifying the leaves. In Agriculture field, the different classification algorithm can be used are Random Forest, K nearest Neighbours, support vector machines, etc. Once the classification

is performed, it becomes easy to detect the diseases of leaves in plants. Kumari et al. has used machine learning and hyperspectral imaging for wilt virus disease on capsicum leaves [10]. Thereafter used an unsupervised k-means algorithm to separate the background from the leaves. The study of techniques that can be used in detection and image processing steps and used Otsu's algorithm and K - means clustering for image segmentation. Arivazhagan et al. discusses the plant leaves recognition and analysis using machine learning technique [11].

Kulkarni et al. proposed a classification based system for plant disease [12]. Gabor filter and artificial neural network are used for segmentation and classification respectively. Cui et al has explained the technique for detecting and classifying the plant disease easily [13]. Researchers have used Otsu algorithm for segmentation, back propagation feed forward neural network for classification of diseases that leads to diseases in plant leaves [14].

II. Literature Review

Disease detection from leaves using different feature selections and choosing the best classification algorithm are being proposed by many authors but accuracy in results is still a challenge. Chaudhary et al shown algorithm to detect the disease spot segmentation [1]. The comparative analysis has been performed for depicting the effect of HSI, YCbCr and CIELAB color space for detecting the spot on leaves. The CIELAB method allows the calculation of the difference between color measurements. HIS is another method to measure the color mainly hue, intensity and saturation.

The pre-processing technique referred to as median filter is utilized for image smoothing. Further, the Otsu method was used for calculating the threshold. The results show that background noise can be reduced by applying CIELAB method. Islam et al. proposed metholodgy for detecting disease in leaves using image processing [5]. Further, Cany edge method is used for detection of shape and texture. BPNN is used for classification and 93.33% of accuracy is achieved. Badage et al. presented the paper focusing on the clustering the remote sensing images from satellites [6]. Model was trained by many diseased leaves and healthy leaves. Image was compared with the threshold to differentiate between disease and ageing. VijayaLakshmi et al. showed feature extraction using local binary patterns (LBPs) technique and for classification to explain the way of identifying the crop diseases from various leaf samples [2]. LBPs is a descriptor used for texture classification. The experimental analysis is performed on vine leaves and success rate of 95% is achieved.

Feature extraction is useful for classification but one should use the data set which is optimized and give better results. Hence there are techniques to optimise the feature selection, one of them is Particle swarm optimization (PSO). There are numerous papers which proposed the technique PSO for optimization. Kumar et al. et al has extracted the feature using the Gabor based techniques and then for identifying the best set of features the author has subjected them to PSO-CFS [3].

After the extraction four classification techniques KNN, J48, CART and RF have been used. Pantazi et al. proposes the Kernel based PSO for minimizing the overhead of selecting the best features [4]. Fuzzy relevance vector machine (FRVM) was used to classify the different types of leaves. FRVM helps to predict the type of leaf from a given dataset. Islam et al. uses artificial intelligence techniques (AIT) like artificial neural network, naïve bayes algorithm, k-Nearest neighbourhood identifier and SVM are implemented for designing an automatic identifier for plant leaves [7]. The AIT models are trained well, they can give better results and they can be successfully applied on leaves for detecting the diseases. Amrita et al. has used the algorithm for improving the different established techniques for detecting the diseases in leaves [8]. Proposed technique tested different diseases and concluded that the KNN algorithm was better than other algorithms such as SVM. Singh et al. presented a survey on leaves diseases detection and different types of classification techniques used. The paper also presents an algorithm for image segmentation technique [9]. Apart from being used for classification of plant leaves, this was used for automatic detection. The proposed algorithm was tested upon 10 different plant species like bananas, mango, potato, sapota etc. It has been observed that accuracy of 97.6 has been achieved.

III. PROPOSED METHODOLOGY

The simulated steps of proposed methodology are shown and depicted in Fig.2.

(a) Image acquisition

Image acquisition is a process of extracting the images from the given source such as sensors, DSLR etc. This method is considered as the first step in image processing. The dataset of healthy and unhealthy 100 leaves is acquired from mendeley repository [14].

(b) Image pre-processing

The image pre-processing method is to improve the quality of the image data by removing the background noise, suppressing the undesired distortion. Median filter technique is used for enhancing the images and removing the noise as shown in Eq.1.

$$Med[X(a)+Y(b)] \neq med[X(a)]+med[Y(b)] \quad (1)$$

Here, med represents the median point or value of image pixels

(c) Segmentation

Image segmentation is a technique to divide the image into many segments. In this paper, Otsu thresholding is applied which returns the single intensity threshold value that separate two different classes one is foreground and other is background as presented in Eq.2.

$$\alpha^2(t) = \omega_1(t)\, \alpha 1^2(t) + \omega_2(t)\, \alpha 2^2(t) \quad (2)$$

Here, $\omega_1(t)$ *and* $\omega_2(t)$ are the probabilities of the classes which are divided by the thresholding t, whose value range can be from 0 to 255.

(d) Feature Extraction and Selection

The different feature is extracted using gray level co-occurrence Matrix (GLCM) and histogram of oriented (HOG) technique. Here, 25 features are extracted namely. GLCM, HOG is a technique of extracting second order statistical texture features. GLCM is a method for examine the texture. Basically, it covert RGB to Gray level co-occurrence matrix. HOG helps to count occurrence of gradient orientation in portion of the image. Further, optimal feature are selected using metaheuristic algorithm such as GA, PSO, APSO, WOA, CSO, etc. Genetic algorithm is mainly focus on optimization to solve complex problems. Best solution is identified from the given population. Particle swarm optimization iteratively try to improve a candidate solution by considering given quality measures. WOA uses bubble net attacking technique. For encircling the prey, best agent solution is assumed to be near to destination and other agent to update their position considering best agent. Cat swarm-based optimization is another approach that tackle the complex problems. Mainly algorithm is based on two modes, namely tracing and seeking modes. Tracing mode is when cat is active and searching for prey while in seeking mode cat becomes aware of the prey. These algorithms help to generate high quality solutions.

(e) Classification

Classification is done by using the SVM and MLP methods [15]. SVM is used for classification which transforms the data and find the optimal boundaries between the possible outputs. Further, it helps to solve linear as well as non-linear problems. Artificial neural network (ANN) is a subfield of computing systems which helps to imitate like human brain to process the information. It has the capabilities of learning by itself by analysing the dataset.

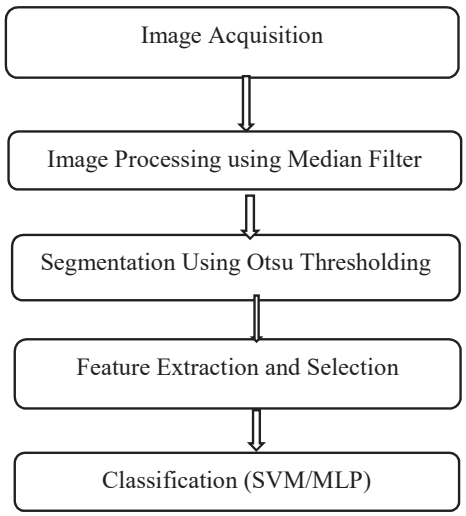

Fig.2. Flow Process of Proposed Methodology

IV. EXPERIMENT RESULT

The spots on leaves are identified by segmenting the unhealthy citrus leaves. Fig.3. shows the (a) healthy (b) unhealthy leaves (c) segmented citrus leaves.

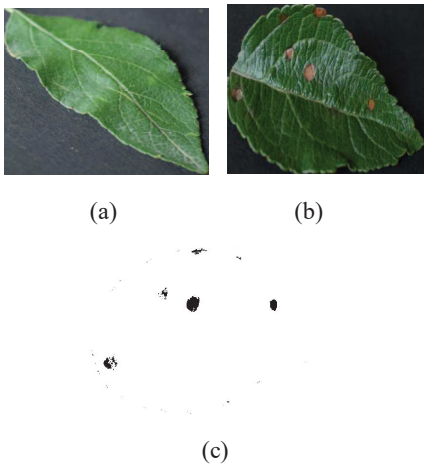

Fig.3. (a) Healthy Citrus leaf (b) Unhealthy Citrus leaf (c) Segmented Citrus leaf

The selected number of features is shown in Table 1. The performance quality matrices are represented using Table II & Table III respectively. The different measures are computed such as recall, precision, F1-score, specificity and accuracy by utilizing models of SVM and MLP. Further, Fig.4 and Fig.5 depicts the visualization analysis of overall results of SVM and MLP applied in dataset

TABLE 1. NUMBER OF SELECTED FEATURES

Algorithms	Total No. of Extracted features	Total number of selected features
WOA	25	20
PSO		18
CSO		16
DE		18

Table II. Analysis of different techniques on considered dataset in terms of various performance measures using SVM

Measures		WOA	PSO	CSO	DE
Healthy	Recall	0.396	0.957	0.693	0.627
	Precision	0.66	0.715	0.693	0.583
	F1-measure	0.682	0.572	0.693	0.627
	Specificity	0.704	0.594	0.715	0.638
	Accuracy	0.66	0.583	0.693	0.627
Unhealthy	Recall	0.572	0.715	0.66	0.704
	Precision	0.572	0.704	0.66	0.704
	F1-measure	0.561	0.704	0.66	0.704
	Specificity	0.583	0.726	0.682	0.726
	Accuracy	0.572	0.704	0.671	0.704
Overall	Recall	0.484	0.836	0.6765	0.6655
	Precision	0.616	0.7095	0.6765	0.6435
	F1-measure	0.6215	0.638	0.6765	0.6655
	Specificity	0.6435	0.66	0.6985	0.682
	Accuracy	0.616	0.6435	0.682	0.6655

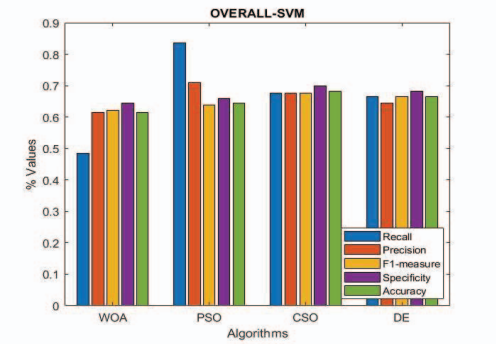

Fig 4: The overall results analysis using SVM

Table III. Analysis of different techniques on considered dataset in terms of various performance measures using MLP

Measures		WOA	PSO	CSO	DE
Healthy	Recall	0.748	0.803	0.726	0.671
	Precision	0.693	0.594	0.726	0.649
	F1-measure	0.693	0.594	0.726	0.649
	Specificity	0.715	0.605	0.748	0.671
	Accuracy	0.726	0.627	0.726	0.66
Unhealthy	Recall	0.759	0.759	0.671	0.715
	Precision	0.781	0.715	0.814	0.759
	F1-measure	0.77	0.726	0.748	0.715
	Specificity	0.792	0.748	0.77	0.737
	Accuracy	0.77	0.737	0.737	0.737
Overall	Recall	0.7535	0.781	0.6985	0.693
	Precision	0.737	0.6545	0.77	0.704
	F1-measure	0.7315	0.66	0.737	0.682
	Specificity	0.7535	0.6765	0.759	0.704
	Accuracy	0.748	0.682	0.7315	0.6985

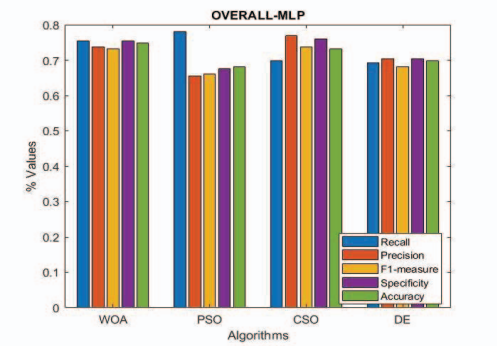

Fig 5: The overall results analysis using MLP

V. CONCLUSION AND FUTURE WORK

Detection of unhealthy leaves in early stage helps farmers to protect from heavy loss. In this paper, comparative analysis is done between classification technique by taking healthy and unhealthy citrus leaves. For this, the image is pre-processed by median filter and 25 features is extracted using GLCM technique. Metaheuristic algorithms helps to optimize the features which allows to extract some important features. After this, SVM classification technique is considered and compared the results with MLP. Table 2 and Table 3 shows the results of an analysis of different techniques in terms of various performance measures using SVM and MLP respectively. By observing the experimental

analysis MLP gave the better results as compared to SVM. In future, the analysis can be done by extending the large dataset. Also, identification of unhealthy stem and roots can be considered which can help to identify the severity level of diseases in plants.

REFERENCES

[1] Chaudhary, P., Chaudhari, A. K., Cheeran, A. N., & Godara, S. (2012). Color transform based approach for disease spot detection on plant leaf. *International journal of computer science and telecommunications*, *3*(6), 65-70.

[2] VijayaLakshmi, B., & Mohan, V. (2016). Kernel-based PSO and FRVM: An automatic plant leaf type detection using texture, shape, and color features. *Computers and Electronics in Agriculture*, *125*, 99-112.

[3] Kumar, A., Patidar, V., Khazanchi, D., & Saini, P. (2016). Optimizing feature selection using particle swarm optimization and utilizing ventral sides of leaves for plant leaf classification. *Procedia Computer Science*, *89*, 324-332.

[4] Pantazi, X. E., Moshou, D., & Tamouridou, A. A. (2019). Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers. *Computers and electronics in agriculture*, *156*, 96-104.

[5] Yigit, E., Sabanci, K., Toktas, A., & Kayabasi, A. (2019). A study on visual features of leaves in plant identification using artificial intelligence techniques. *Computers and electronics in agriculture*, *156*, 369-377.

[6] Badage, A. (2018). Crop disease detection using machine learning: Indian agriculture. *Int. Res. J. Eng. Technol.(IRJET)*, *5*(9), 866-869.

[7] Islam, S. T., Masud, M. A., Rahaman, M. A. U., & Rabbi, M. M. H. (2019, December). Plant leaf disease detection using mean value of pixels and canny edge detector. In *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)* (pp. 1-6). IEEE.

[8] Trivedi, J., Shamnani, Y., & Gajjar, R. (2020, February). Plant leaf disease detection using machine learning. In *International conference on emerging technology trends in electronics communication and networking* (pp. 267-276). Springer, Singapore.

[9] Singh, V., & Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information processing in Agriculture*, *4*(1), 41-49.

[10] Kumari, C. U., Prasad, S. J., & Mounika, G. (2019, March). Leaf disease detection: feature extraction with K-means clustering and classification with ANN. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1095-1098). IEEE.

[11] Arivazhagan, S., Shebiah, R. N., Ananthi, S., & Varthini, S. V. (2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *AgriculturalEngineering International: CIGR Journal*, *15*(1), 211-217.

[12] Kulkarni, A. H., & Patil, A. (2012). Applying image processing technique to detect plant diseases. *International Journal of Modern Engineering Research*, *2*(5), 3661-3664.

[13] Cui, D., Zhang, Q., Li, M., Zhao, Y., & Hartman, G. L. (2009). Detection of soybean rust using a multispectral image sensor. *Sensing and Instrumentation for Food Quality and Safety*, *3*(1), 49-56.

[14] Chouhan, S. S., Singh, U. P., Kaul, A., & Jain, S. (2019, November). A data repository of leaf images: Practice towards plant conservation with plant pathology. In *2019 4th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 700-707). IEEE.

[15] Vijh, S., Gaur, D., & Kumar, S. (2020). An intelligent lung tumor diagnosis system using whale optimization algorithm and support vector machine. *International Journal of System Assurance Engineering and Management*, *11*(2), 374-384.