JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 01
2024

Research Article
https://doi.org/10.56979/701/2024

Multifeature Analysis to Detect Cotton Leaf Curl Virus

Nazir Ahmad[1*], Salman Qadri[2], Nadeem Akhtar[3], and Syed Ali Nawaz[1]

[1]Department of Information Technology, IUB, Bahawalpur, 63100, Pakistan.
[2]Department of Computer Science, MNS-University of Agriculture, Multan, 61000, Pakistan.
[3]Department of Software Engineering, IUB, Bahawalpur, 63100, Pakistan.
[*]Corresponding Author: Nazir Ahmad. Email: nazeerrana@iub.edu.pk

Received: January 01, 2024 **Accepted:** May 09, 2024 **Published:** June 01, 2024

Abstract: Plant leaf diseases have devastating impacts on yield production, both in terms of quantity and quality. The cotton leaf curl virus (CLCuV) is one of the most destructive diseases that affect cotton crops worldwide. Disease detection based on symptoms is laborious and demands a great deal of experience and knowledge. The purpose of this research study is to design an automated system to detect CLCuV accurately. A dataset of healthy, mildly, and severely infected CLCuV is captured with a digital camera from cotton fields. An image enhancement tool is used to standardize the dataset for image analysis. Histogram, gray level co-occurrence matrix, and run length matrix features are extracted by the image analysis tool. Fisher, Probability of error plus average correlation and Mutual information feature optimization techniques are used to get the most optimal features to reduce computation costs. MultiClass, Bagging, Logistic Model Tree (LMT), and Radom Forest (RF) machine learning (ML) classifiers are deployed to observe the impact of CLCuV. True Positive (TP) rate, False Positive (FP) rate, Precision, Recall, F-measure, Mathews Correlation Coefficient (MCC), Receiver Operating Characteristic (ROC), and Precision Recall Curve (PRC) performance evaluation parameters are calculated to measure the effectiveness of ML classifiers. The RF classifier outperformed and demonstrated 87.542% accuracy, while other ML classifiers also achieved satisfactory results.

Keywords: Supervised Learning; Machine Learning ; Logistic Model Tree; Random Forest.

1. Introduction

One of the subareas of artificial intelligence (AI) is computer vision. Computer vision involves equipping computers with the ability to see and process visual information, enabling them to make quicker, more accurate decisions than humans can consistently achieve. Relatively, Computer vision is new and used in many other areas of life like agriculture, healthcare, transportation, sports, manufacturing, remote monitoring and automation [1]. For any community to feed its people, agriculture is necessary. It is essential to the expansion and development of every nation's economy. Agriculture is directly responsible for the existence of humans because it is seen as a source of sustenance. Several nations around the world rely entirely on agriculture for their economies, which illustrates the importance of this sector [2].

Understanding the effects of plant pests and diseases on crop performance is a crucial area of study in agricultural computational modeling. Historically, attempts were made to develop theoretical models that accounted for the influence of pests and diseases on yield independently, resulting from the interactions of genotype, environment, and management [3]. The concept of production situation was introduced in 1982. Pest and disease control is also an aspect of a production situation that includes farmer crop management. This generally accepted classification of yield levels includes crop genetics as a factor determining potential yield, and grouping water and nitrogen stress as limiting variables to achievable output. In a later study, there were three distinct types of yield. Solar radiation and temperature determine

the potential yield, which is the first type. The availability of water and nutrients constrains the second, attainable yield. Finally, factors such as diseases, pests, and environmental stressors influence the actual yield. According to this framework, the gap between the potential yield and the actual yield determines the decrease in crop yield due to biotic stresses [4].

Insect pests are a significant contributor to crop production and storage losses in agricultural systems. It is estimated that these parasites cause losses of 60–70% in tropical countries, primarily in stored products. Insects represent an incredibly vast array of animal species that inhabit our planet. Insects inhabit every habitat besides the open ocean, including swamps, forests, deserts, and extremely harsh environments. Insects are truly remarkable creatures, displaying an unparalleled level of adaptability that surpasses all other animal categories in terms of sheer numbers. Insects play a crucial role in our ecosystem and have a significant impact on both humans and the environment. Insect pests cause significant harm to humans, farm animals, and crops [5].

The cotton industry is a significant economic sector in all countries whose economies revolve around agriculture. Despite this, pathogens frequently threaten cotton production, resulting in significant economic losses. Researchers worldwide have detected a broad array of diseases in cotton. The main challenges to the production of cotton fiber are the leaf curl virus, bacterial blight, leaf spot, and seeding diseases. . Researchers pursue the overarching goal of ensuring a minimal incidence of diseases. To estimate the economic impact of diseases, it is critical to understand their causes; this knowledge ultimately facilitates the formulation of management strategies. CLCuV has become a significant threat to all cotton-growing regions due to the evolution of the viral disease complex [6].

Over 60% of Pakistan's foreign exchange earnings are derived from cotton, making it one of the country's most valuable commodities. The current CLCuV epidemic emerged in the Punjab region adjacent to Multan and was initially documented in 1985; however, it had been identified in this area since 1967. Early in the 1990s, CLCuV emerged as the most significant impediment to cotton production in Pakistan; it has since spread to India and, more recently, to other provinces in Pakistan to the south and west. The distinctive symptoms consist of vein enlargement, darkened veins, leaf curling, and enations that often transform into cup-shaped, leaf-like formations on the undersides of leaves. Rapidly following the identification of the vector of CLCuV as the whitefly Bemisia tabaci (Genn.), it was hypothesized that the disease-causing agent is a geminivirus [7].

In recent years, with the development of digital image processing technologies, a lot of methods have been used for identification and diagnosis of plant diseases. Color features have been extracted by using different color space models like HSV, RGB and YCbCr for disease detection. With the advancement in machine learning, many novel methods and models have also been used in the agricultural area. One of new category is deep learning which uses artificial neural network (ANN) with large number of processing layers. At present, many scholars and experts are using deep learning models like state vector machine (SVM) and convolution neural network (CNN) with efficiently which exhibits more accuracy in the areas of pattern recognition, voice recognition and the process where huge dataset is used for analysis purposes [8].

Convenient and technologically advanced plant diagnostics simplify the identification of damaged leaves while minimizing individual effort. Early detection of leaf maladies may also contribute to an increase in the profitability of production. Transfer learning (ResNet50) and K-Nearest Neighbor (KNN) machine learning algorithms used to the detection of cotton leaf diseases. After being trained with an adequate amount of data, RESNET50 achieves an accuracy of 95% in differentiating between healthy and unhealthy leaves. The KNN algorithm has an 86% accuracy rate in identifying diseases in leaves [9].

State Vector Machine (SVM) was deployed to classify the five cotton leaf diseases. Features were extracted with partial least-square regression (PLSR). The accuracy result for the SVM was obtained to 83.6 percent [10]. Machine learning approaches for crop disease detection range from the more conventional use of deep features to more recent approaches based on hand-crafted features (HCF). Novel hybrid architecture was used for detecting crop diseases by integrating many features. A convolutional neural network (CNN) is employed to extract high-level characteristics that are then combined with HCF [11].

Various diseases, including CLCuV, bacterial blight, and ball rot, have a significant impact on global cotton crop productivity. Image processing techniques, in conjunction with machine learning algorithms, are effectively utilized in several domains and have also been applied for the purpose of crop disease

identification. A deep learning approach was used to classify diseases such as bacterial blight and CLCuV affecting the cotton crop. The dataset of cotton leaves exhibiting disease symptoms was gathered from diverse places in Sindh, Pakistan. Inception v4 architecture was deployed as a convolutional neural network for detecting damaged plant leaves by bacterial blight and CLCuV. Inception v4 achieved an accuracy of 88.26% [12]. A deep learning model based on CNN was developed with the capability of identifying 15 distinct varieties of diseases in four prominent crops across the world. The proposed model was deployed in 19 classes, of which 15 were diseased and 4 were healthy, and achieved a high accuracy of 85.5% [13].

The principal aim of this study is to design an automated detection system for cotton leaf curl virus. This research also encompasses the subsequent aims:

- Determining the most pertinent texture features.
- Constructing an innovative classification framework that leverages soft computing to achieve the primary objective.

2. Materials and Methods

This study uses image processing techniques to design CLCuV automated detection system. Texture features were extracted from photographic data to accomplish the proposed work. The Conon IXUS 185 digital camera model with 20 megapixels was utilized to capture digital photographic data. Bahawalpur and Multan divisions of Punjab, Pakistan, were selected for the experimentation of the proposed study. Steps of data collection, data preprocessing, feature extraction, feature optimization and classification were performed. Figure 1 show Multifeature analysis framework to detect CLCuV.

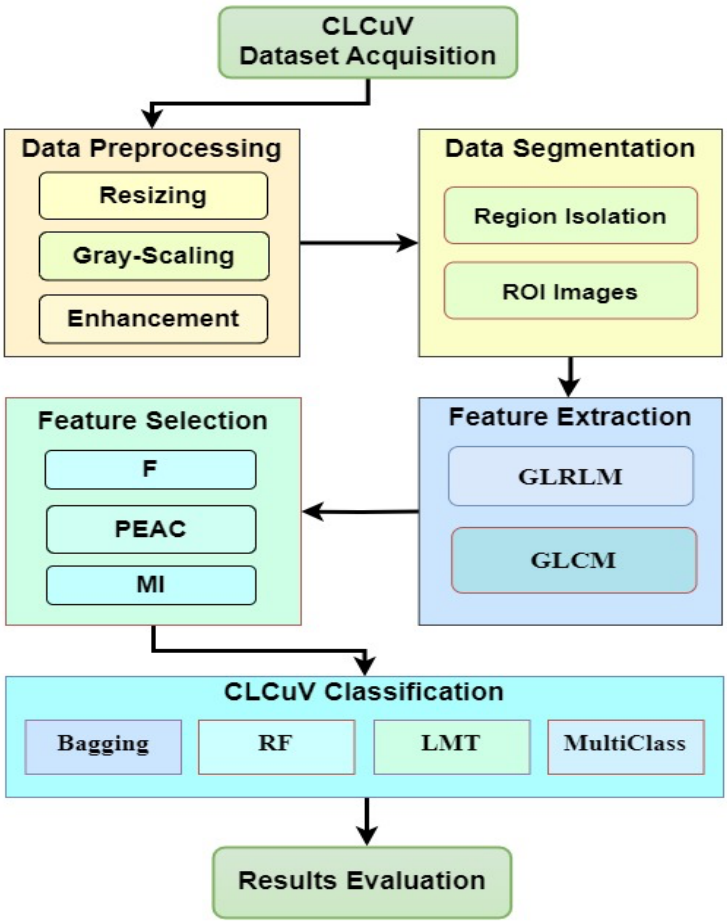


Figure 1. Multifeature analysis framework to detect CLCuV

2.1. CLCuV Dataset Acquisition

Digital photographic image dataset was collected from cotton fields using a Canon digital camera. The acquired dataset was divided into three categories: one is healthy, and the others are mildly and severely infected. The dataset was collected in an open environment with no experimental setup.

2.2. Data Pre-processing

Data preprocessing include the manipulation, filtration, or augmentation of data prior to analysis, and is frequently a crucial stage in the data mining process [14]. A total of six hundred colored digital images were obtained for each category. Experts and plant pathologists collaborated to select the top 100 pictures to get more accurate results. In order to obtain only prominent leaves, captured images were cropped to proportions ranging from 5152 × 3826 pixels to 3215 × 2776 pixels, as the selected images contained extraneous regions. The cropped photos were resized to 128 × 128, 256 × 256, and 512 × 512 pixels for each severity level and saved in bitmap (.bmp) format. Figure 2 shows cropped images of the healthy, mildly and severely infected cotton leaves.

a. healthy leaves b. mildly infected leaves c. severely infected leaves

Figure 2. Healthy, mildly and severely infected cotton leaves images

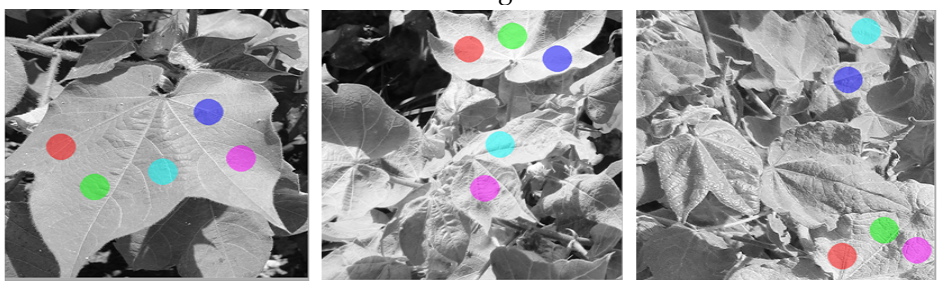
Cropped images were transformed to grayscale, so that minimum memory space would be used for the whole process. Enhancement filters were employed to restore the sharpness and shadow of grayscale photographs by increasing the contrast level.

2.3. Data Segmentation

After all the collected digital photographic images were resized and enhanced, the next step was to segment the images. During this process, the diseased regions of the curl virus were isolated by applying the automated clustering-based segmentation scheme. Next, ROIs were taken from these identified damaged regions and ROI image dataset was established.

2.4. Feature Extraction

In machine learning and data analysis, feature extraction is the process of finding important features in raw data and pulling them out. These features are then used to make a dataset with more information, which can then be used for different jobs like classification and prediction [15]. Image texture analysis software MaZda version 4.6 was used to get texture features. It works well and can be trusted to distinct segments of a picture, select features, extract features, and calculate texture features [16]. In images and videos, region of interest (ROI) is a specific area or region that holds information that is useful for the task at hand [17]. ROI's of various sizes, like 16 × 16, 20 × 20 for image size 128 × 128, 20 × 20, 24 × 24 for image size 256 × 256, and 28 × 28, 32 × 32 for image size 512 × 512, were used to accomplish the experimentation process. ROI's of size 32 × 32 for 512 × 512 are shown in Figure 3.



a. healthy leaves b. mildly infected leaves c. severely infected leaves

Figure 3. ROI's of healthy, mildly and severely infected leaves

The feature space is very large because it includes nine first-order histogram features, eleven second-order gray level co-occurrence matrix features with five inter-pixel distances in four directions, and five second-order run-length features in four directions.

2.5. Feature Selection

Machine learning employs feature reduction, also known as feature optimization, as a technique to reduce the number of input variables or features in a dataset. Our main objective is to streamline the dataset while preserving its fundamental attributes, which can enhance the effectiveness of machine learning models [18]. Working with a dataset that has hundreds of thousands of features can significantly impact the time and complexity required to compute results. Having an excessive number of features in the machine learning process can lead to a dimensional disaster. Each ROI's extracted features are not all equally important. Extracting the most significant features is essential for obtaining more accurate results. This can be achieved through various feature optimization techniques. MaZda implements the fisher co-efficient, probability of error plus average correlation coefficient, and mutual information coefficient feature reduction technique [19]. Most thirty prominent features were selected using MaZda. Table 1 shows the optimized features for each ROI.

Table 1. Optimized features for each ROI

F		PA		MI	
1	S(0,1)SumVarnc	11	Kurtosis	21	S(5,0)DifEntrp
2	S(1,-1)SumEntrp	12	WavEnLH_s-2	22	S(5,-5)DifEntrp
3	S(5,-4)DifEntrp	13	WavEnHL_s-1	23	S(4,-4)DifEntrp
4	S(1,-1)SumVarnc	14	WavEnHL_s-2	24	Horzl_GLevNonU
5	S(4,4)DifEntrp	15	Vertl_LngREmph	25	S(5,5)InvDfMom
6	S(0,1)SumEntrp	16	S(5,-5)Correlat	26	135dr_GLevNonU
7	S(1,1)SumEntrp	17	S(1,5)Correlat	27	S(0,5)DifEntrp
8	S(1,0)SumEntrp	18	Skewness	28	S(4,4)InvDfMom
9	S(0,2)SumEntrp	19	S(0,1)Correlat	29	S(4,0)DifEntrp
10	S(3,3)DifEntrp	20	WavEnLL_s-3	30	S(2,0)SumEntrp

2.6. CLCuV Classification

Classification is a method used in supervised machine learning to predict the correct label of given input data. During the classification process, the model is trained extensively using the provided training data. After completing the training, we test the model using separate test data to evaluate its performance [20]. Predictive modeling and data analysis are done using the Waikato Environment for Knowledge Analysis (WEKA). Several machine learning classifiers were deployed to detect CLCuV, but MultiClass, Bagging, Logistic Model Tree (LMT), and Random Forest (RF) proved to be the top performers for digital photographic data. In machine learning, model performance review checks how well a model is doing at the job it was designed for. TP rate, FP rate, Precision, Recall, F-measure, MCC, ROC, and PCR evaluation parameters are used to find out how well the suggested model works with a digital photographic dataset.

3. Results and Discussion

MultiClass, Bagging, LMT, and RF supervised ML models were used to detect CLCuV from healthy, mildly, and severely infected cotton leaves. The k-fold cross-validation set to 10 for digital photographic dataset. ML classifiers were implemented on 128 × 128 image sizes with ROI sizes of 16 × 16 and 20 × 20, 256 × 256 image sizes with ROI sizes of 20 × 20 and 24 × 24 feature datasets, but promising results were not found and the result accuracy was less than 80%. The same ML classifiers were deployed on 512 × 512 image size with 28 × 28 and 32 × 32 ROI size feature datasets. It was observed that RF outperformed with an accuracy of 87.542%, while the classification results accuracy for MultiClass, Bagging, and LMT were 85.175%, 85.393%, and 85.984%, respectively, as shown in Table 2. Figure 4 shows ML classifiers accuracy results for detecting CLCuV on a digital photographic dataset.

Table 2. ML classifiers detailed accuracy for an image size of 512 × 512 with ROI size 32 × 32

Classifier	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC	PRC	Accuracy

Multi Class	0.852	0.073	0.851	0.852	0.849	0.779	0.956	0.919	85.175
Bagging	0.854	0.073	0.853	0.854	0.854	0.781	0.951	0.91	85.393
LMT	0.86	0.069	0.859	0.86	0.859	0.791	0.967	0.94	86.013
RF	0.875	0.063	0.876	0.875	0.875	0.813	0.958	0.929	**87.542**

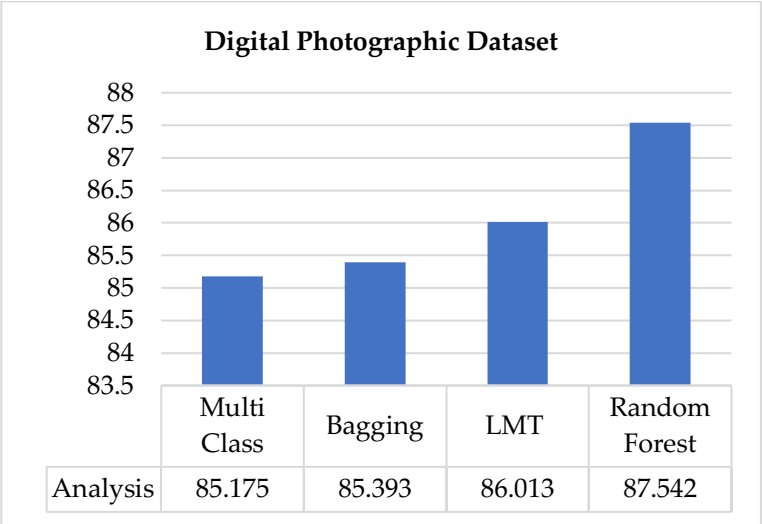


Figure 4. ML classifiers accuracy results

Confusion matrix (CM) tables 3, 4, 5, and 6 show the results of MultiClass, Bagging, MLT, and RF classifiers respectively, for a digital photographic dataset.

Table 3. CM showing results of MultiClass classifier for digital photographic dataset

Class	Healthy	Mildly infected	Severely infected	Total
Healthy	423	60	12	495
Mildly infected	65	405	25	495
Severely infected	9	32	454	495

Table 4. CM showing results of Bagging classifier for digital photographic dataset

Class	Healthy	Mildly infected	Severely infected	Total
Healthy	442	46	7	495
Mildly infected	37	419	39	495
Severely infected	5	24	466	495

Table 5. CM showing results of MLT classifier for digital photographic dataset

Class	Healthy	Mildly infected	Severely infected	Total
Healthy	457	27	11	495
Mildly infected	34	436	25	495
Severely infected	6	13	476	495

Table 6. CM showing results of RF classifier for digital photographic dataset

Class	Healthy	Mildly infected	Severely infected	Total
Healthy	421	62	12	495

Mildly infected	63	395	37	495
Severely infected	3	43	449	495

4. Conclusion

A Multifeature framework to detect various severity levels of CLCuV has been designed in this research. Once the preprocessing is completed, the feature extraction process extracts the most relevant features from the specified datasets. In feature optimization, the thirty most optimized features were selected from the digital photographic dataset. A digital photographic dataset is fed into the four specified ML classifiers, namely, MultiClass, Bagging, LMT, and RF, to detect CLCuV. It was found that RF is the most effective classifier to detect CLCuV, with an accuracy of 87.542%.

References

1. B. S. Rakhimov, F. B. Rakhimova, S. K. Sobirova, F. O. Kuryazov and D. B. Abdirimova, "Review And Analysis Of Computer Vision Algorithms," The American Journal of Applied sciences, vol. 3, no. 5, pp. 245-250, 2021.
2. E. Loizou, C. Karelakis, K. Galanopoulos and K. Mattas, "The role of agriculture as a development tool for a regional economy," Agricultural Systems, vol. 173, pp. 482-490, 2019.
3. T. B. Shahi, C. Y. Xu, A. Neupane and G. Willian, "Recent Advances in Crop Disease Detection using UAV and Deep Learning Techniques," Remote Sensing, vol. 15, no. 9, pp. 1-29, 2023.
4. M. Manosathiyadevan, V. Bhuvaneshwari and R. Latha, "Impact of Insects and Pests in loss of Crop production: A Review," Sustainable agriculture towards food security, pp. 57-67, 2017.
5. M. Ouhami, A. Hafiane, Y. Es-Saad, M. E. Hajji and R. Canals, "Computer Vision, IoT and Data Fusion for Crop Disease Detection Using Machine Learning: A Survey and Ongoing Research," Remote Sensing, vol. 13, no. 13, p. 2486, 2021.
6. S. Savary and L. Willocquet, "Modeling the Impact of Crop Diseases on Global Food Security," Annual Review of Phytopathology, vol. 58, pp. 313-341, 2020.
7. S. Chohan, R. Perveen, M. Abid, M. N. Tahir and S. Muhammad, "Cotton Diseases and Their Management," Cotton Production and Uses: Agronomy, Crop Protection, and Postharvest Technologies, pp. 239-270, 2020.
8. F. Jiang, Y. Lu, Y. Chen, D. Cai and G. Li, "Image recognition of four rice leaf diseases based on deep learning and support vector machine," Computers and Electronics in Agriculture, vol. 179, p. 105824, 2020.
9. S. Kotian, P. Ettan, S. Kharche, K. Saravanan and K. A. Kumar, "Cotton Leaf Desease Detection using Machine Learning," in 2nd Intrnational Conference on "Advancement in Electronics & Communication Engineering", India, 2022.
10. A. A. Sarangdhar and V. Pawar, "Machine learning regression technique for cotton leaf disease detection and controlling using IoT," in In 2017 international conference of electronics, communication and aerospace technology (ICECA), 2017.
11. R. Bhagwat and Y. Dandawate, "A Framework for Crop Disease Detection Using Feature Fusion Method," International Journal of Engineering and Technology Innovation, vol. 11, no. 3, pp. 216-228, 2021.
12. S. Anwar, A. R. Kolachi, S. K. Baloch and S. R. Soomro, "Bacterial Blight and Cotton Leaf Curl Virus Detection Using Inception V4 Based CNN Model for Cotton Crops," in In 2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS), 2022.
13. P. Bajpai, A. Pandey and P. Narang, "Plant Disease Detector using CNN," in In 2021 IEEE Bombay section signature conference (IBSSC), 2021.
14. F. Kamiran and T. Calders, "Data preprocessing techniques for classification without discrimination," Knowl Inf Syst, vol. 33, pp. 1-33, 2015.
15. G. Kumar and P. K. Bhatia, "A Detailed Review of Feature Extraction in Image Processing Systems," in International Conference on Advanced Computing & Communication Technologies (ACCT), 2014.
16. P. M. Szcxypinski, M. Strzelecki, A. Materka and A. Klepaczko, "MaZda- A software package for image texture analysis," Computer methods and programs in biomedicine, vol. 94, no. 1, pp. 66-76, 2009.
17. A. J. I. Barbhuiya and K. Hemachandran, "Wavelet Tranformations & Its Major Applications In Digital Image Processing," International Journal of Engineering Research & Technology, vol. 2, no. 3, pp. 1-5, 2013.
18. A. Jovic, K. Brkic and N. Bogunovic, "A review of feature selection methods with applications," in 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015.
19. B. S. Prajapati, V. K. Dabhi and H. B. Prajapati, "A Survey on Detection and Classification of Cotton Leaf Diseases," in International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016.

20. A. Ali, S. Qadri, W. K. Mashwani, S. B. Belhaouari, S. Naeem, S. Rafique, F. Jamal, C. Chesneau and S. Anam, "Machine learning approach for the classification of seed using hybrid features," International Journal of Food Properties, vol. 23, no. 1, pp. 1110-1124, 2020.