MACHINE LEARNING CLASSIFICATION OF INFECTION IN *OCIMUM TENUIFLORUM* USING PREDICTIVE MODELLING

Manjot Kaur[1*], Someet Singh[2], Anita Gehlot[3]

Abstract

India's economy relies heavily on agricultural production and is a major source of employment. The early identification of plant leaf diseases is crucial for maximizing revenue and crop productivity. There are more tulsi (*Ocimum Tenuiflorum*) plant products produced in India than anywhere else in the world. Early methods of observing solely through visual inspection were time-consuming and inaccurate. The present work identifies and categorizes leaf diseases by using a variety of image-processing approaches. The present study demonstrates comprehensive methods for identifying and classifying infections in medicinal plants utilizing image processing and machine learning. The plant village input image dataset having three different types of infected *Ocimum Tenuiflorum* leaf and healthy leaf is amassed as the basis for this dataset. Through the use of a computer vision lab framework, the image datasets are augmented, pre-processed, segmented, extracted and validated with certain features. Five machine learning classifiers are evaluated using an optimized dataset of infected leaves of *Ocimum Tenuiflorum*, including logistic regression, linear discriminant analysis, k nearest neighbour, classification and regression trees, random forests, naive bayes, and support vector machines. According to the results, the random forest classifier outperforms the others with an accuracy of 99.86%, followed by the linear discriminant analysis with 98.59%, and the support vector machine with 97.42%.

Keywords: Augmentation, Image Processing, Machine Learning, *Ocimum tenuiflorum*, Predictive Modelling

[1*]School of Electrical & Electronics Engineering, Lovely Professional University, Phagwara-144411, India, manjotkaur.hpk@gmail.com
[2]School of Electrical & Electronics Engineering, Lovely Professional University, Phagwara-144411, India, someet.17380@lpu.co.in
[3]Uttaranchal Institute of Technology, Uttaranchal University, Dehradun-248007, India, anita.ri@uttaranchaluniversity.ac.in

***Corresponding Author:** Manjot Kaur
School of Electrical & Electronics Engineering, Lovely Professional University, Phagwara-144411, India, manjotkaur.hpk@gmail.com

DOI: 10.48047/ecb/2023.12.si10.00251

1 Introduction

It is possible to detect most of the infection signs on the roots, stems, and leaves of the herb *Ocimum Tenuiflorum*. A system for detecting and classifying leaf diseases can incorporate the use of image processing and artificial intelligence. It is normal for leaf diseases such as Bacterial Blight, Anthracnose, Alternaria Alternata, and others to occur frequently. There is a possibility that a leaf infection may reduce the production of the herb as well. A variety of pathogens, such as bacteria and fungi, may be responsible for causing these infections, and using classifier models, we can automatically identify and categorize these diseases. It is the primary objective of this study to create, implement, and evaluate a software solution based on image processing for classifying leaf diseases using a machine learning prediction model [1]. Several characteristics can be used to describe leaf diseases, among which is Alternaria Alternata, which results in small, round, reddish-brown patches on the leaves of this plant. Another example would be Anthracnose, which consists of these tiny, sporadic, dark violet or black leaf spots with yellowish haloes that can occur regularly or sporadically. The yellowing and falling of the leaves. Another variety of this disease is Bacterial blight, which is characterized by the appearance of an overwhelming number of tiny, irregularly shaped, water-soaked patches on leaves. There was a disease called Cercospora, which is a tiny, brown spot on the leaves with yellow halo dots. It is possible to see spots that are dispersed, rounded, irregular, or both, and as they mature, they become darker [2].

Research in Image Processing and Machine Learning for disease detection and recognition is one of the most active areas of research that shows immense promise for detecting pests and diseases early and accurately. The only way to create an infection detection system that is straightforward, reliable, and effective is by using image processing [3]. The use of machine learning (ML) is one of the most effective ways to construct a predictive system for images and medical data. The majority of the past research has presented a common workflow, which includes image procurement, image pre-processing, feature extraction, and classification techniques, all of which fall under the general category of image processing and machine learning. The following are a few of the goals and benefits that can be achieved through the use of image analysis in a variety of different ways, such as identifying and quantifying the impact of an infected leaf, determining the borders of the affected area, determining the gue of the impacted area and distinguishing the infection category appropriately [4]. As far as image processing is concerned, the most challenging aspect is the collection of databases, whose primary purpose is to provide basic information about the crop and its illnesses. As a result, it is necessary to conduct a thorough investigation into the numerous types of diseases, their effects on crops, and their patterns in order to identify and eradicate them. Based on the patterns of infection, a system will be developed that will be able to detect infections efficiently. It is documented in the Charaka Samhita, an old ayurvedic treatise, that *Ocimum Tenuiflorum* has been used in a variety of medicines for many years and even in its current form [5]. In addition to its significant medicinal properties, *Ocimum Tenuiflorum*, also known as Tulsi, possesses analgesic, antiemetic, and immunomodulatory properties. A number of previous studies have concluded that tulsi possesses a strong antibacterial property that can be used against a wide range of pathogenic Gram-negative bacteria [6, 7]. This plant's medicinal properties are also utilized in nanotechnology, which was used to synthesize nanoparticles from its leaves aqueous extract. The nanoparticles demonstrated strong antibacterial activity against Gram-positive, Gram-negative, and spore-forming bacteria. There are a number of ailments that can be treated by taking extracts of *Ocimum Tenuiflorum* such as the common cold, headaches, cardiovascular diseases, and stomach infections. *Ocimum Tenuiflorum* is an analgesic with antibacterial properties, and it helps decrease bronchitis, which is why it is a key component of most ayurvedic cough syrups [9].

It has evolved for plants to have unique defence mechanisms at both the molecular and cellular levels to deal with a variety of biotic and abiotic stress conditions. Various molecular processes activate the plant's immune system under stress conditions, including changes in gene expression related to transcription factors, signal transduction and kinase cascade pathways, hormone signaling, and heat shock proteins, thereby enhancing plant immunity [10]. It has been found that plants are one of the richest sources of drugs of traditional systems of medicine, modern systems of medicine, nutraceuticals, food supplements, folk medicine, pharmaceutical intermediates, and chemical entities for manufacturing synthetic drugs [11]. It has been documented since the beginning of human civilization that plants and plant products are used as medicine. As a matter of fact, the medicinal plant is widely considered to be one of the best sources for finding potential therapeutic benefits derived from its varying parts, such as seeds, roots, stems, and leaves. Identification of plants is based on their components, including seeds, flowers, fruits, leaves, and stems, but leaves play a major role in identifying plants when compared to other plant organs, as they are readily available throughout all seasons, and possess unique characteristics that aid in identifying any plant. In recent decades, diseases in leaves have been the leading cause of crop waste in many countries, which is why the crop needs to be treated with reliable and effective methods as soon as possible. An android application was developed for the identification of herbal plants via texture extraction techniques by Sana O et al. In this system, medicinal plants are identified based on a picture of a confirmed herbal plant. The use of co-accident matrices has been found to be an effective technique for extracting textures from images and for the classification of plants using image processing techniques [12]. An algorithm for fruit recognition with higher accuracy has been developed by Ruaa Adeeb et al., by comparing various machine learning algorithms [13]. It is necessary to classify the disease data and use a variety of classification and data analysis techniques in order to determine if we can predict certain diseases based on the data. The purpose of this research project is to examine a variety of prediction approaches in order to demonstrate how they compare to one another in terms of their accuracy and effectiveness.

2 Methodology

There are many diseases that cause leaves to deteriorate, such as bacteria, fungi, viral infections, and insect-borne diseases. The paper describes the implementation of the model using these disorders as the dataset. The procedure used for analyzing the examined data can be classified into four categories. These are image pre-handling, image segmentation, feature extraction, and detection and classification of diseases. As shown in figure 1, the block chart of the framework is presented in the form of a block diagram [14].

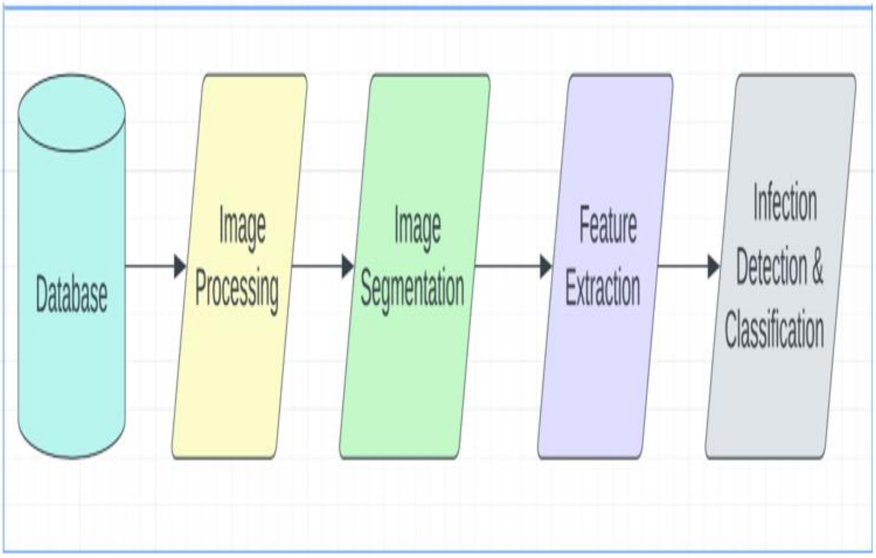

Figure 1: Block diagram of Machine Learning Prediction Model

In order to perform predictive analytics, predictive modeling is the key component. Predictive modeling is an approach rather than a procedure. The fact that predictive models frequently incorporate a machine learning algorithm makes it clear that predictive analytics and machine learning can be thought of as a combination. Over time, these models are capable of being trained to respond to changes in information or values and to produce the outcomes that the industry demands. This is the most crucial step in the ML-based classification process in terms of feature selection (FS). This method seeks to eliminate superfluous traits that serve no purpose in the classification process and instead choose the most reliable ones that can be used in the process. [15].

2.1 Data augmentation

The purpose of this section is to explain the recommendation for categorizing infections in plant leaves according to the recommended approach. A comprehensive analysis of the collected dataset will be carried out in the first phase of the project using the Python compiler in Visual Studio. Table 1 shows the requirements and calculations to be carried out. Table 2 shows the data that has been generated as a result of these calculations for further processing.

Table 1: Augmentation parameters

Augmentation Image Data Generator parameters	Selected Value/ Range
Rotation range	40
Shear range	0.2
Zoom range	0.2
Horizontal flip	True
Brightness range	(0.5,1.5)
Batch size	16

Table 2: Dataset after augmentation

Type of leaf	Actual dataset	Augmented dataset
Bacteria	4	204
Fungal	19	490
Healthy	15	765
Pests	16	815
Total	**54**	**2274**

The figure above shows the sample images of the left side prior to augmentation and the right side after the augmentation has been performed on both sides. As a result, the process was carried out by optimizing different input parameters for example the rotation range of 40 degrees in order to be able to view the infection area in the region with an optimal view. In addition to affecting the image's orientation, shearing also has the capability of changing its dimension. Through the enhancement of the images, we are able to ensure the reliability of our model and its ability to identify the infection under a variety of conditions. In order to achieve this, Keras is a sophisticated machine learning framework built on top of TensorFlow, which is used for the purpose of this process. The ImageDataGenerator class is used to augment the data in a variety of ways and includes a number of arguments, such as the rotation, zoom, shear, and brightness ranges, as well as various flips and batch sizes [16].

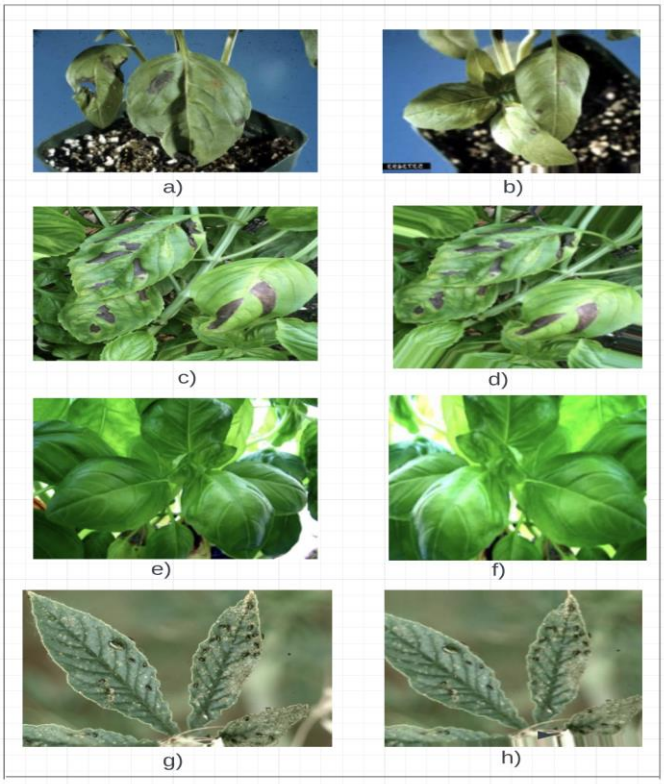

Figure 2: a) bacterial infected leaf, b) augmented bacterial leaf, c) fungal infected leaf, d) augmented fungal leaf, e) healthy leaf, f) augmented healthy leaf, g) pests infected leaf, h) augmented pests leaf

2.2Process Flow:

Using OpenCV, a computer vision software package, each image that was captured for the dataset was analyzed [17]. In order to prepare the data for the segmenting of the images, BGR, RGB, and HSV formats were used for the pre-processing of the data. The feature extraction of the image was carried out using Hu-moments, Harlick texture, and color histogram of the image in computer vision software. The feature labels were trained, and 10-fold cross-validation was carried out for classifier models. An AI model is evaluated on a restricted information test using cross-validation, which is a technique of resampling. In the final step, the accuracy results for the seven classifier models are visualized by using a boxplot in order to compare their accuracy. This figure shows the flowchart for the construction of the prediction model as illustrated in Figure 3.

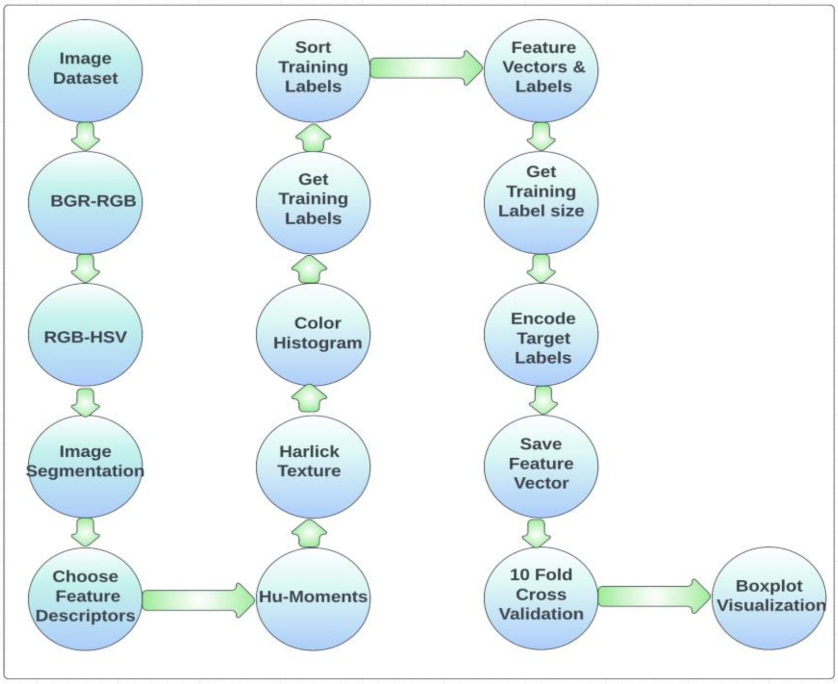

Figure 3: Process Flow for Machine Learning Prediction Model

Various parameters are configured and tuned to be used in order to perform the various data operations that can be found in table 3.

Table 3: Tunable parameters

Tunable Parameters	Value
Images_per_class	15
Fixed_size	Tuple (500,500)
Num_trees	100
Test_size	0.25

3 Experimental results and discussions

Using a variety of classifier models to categorize and detect leaf diseases, this part presents and analyzes the experiments and findings of an automated system for the categorization and detection of leaf diseases. The overall performance of the algorithm is analyzed and discussed. As a part of this research, tests were conducted to determine the optimal segmentation result, reduce measurement errors, and confirm the accuracy of detecting and categorizing leaf diseases. There have been experiments carried out to demonstrate that the system is accurate at detecting certain types of objects, but at the same time, it has also been tested by extracting attributes from the images in order to check the accuracy of the detection. The graphs in figures 4 and 5 show the comparison between the different machine learning algorithms that are used. On the basis of a Pareto chart comparison, it is evident that the RF classifier model outperforms the other classifier models both in terms of accuracy (99.862%) and execution time (8122 milliseconds), followed by the LDA and SVM models. Even though KNN is next to LDA and SVM in accuracy, the execution time is much longer.

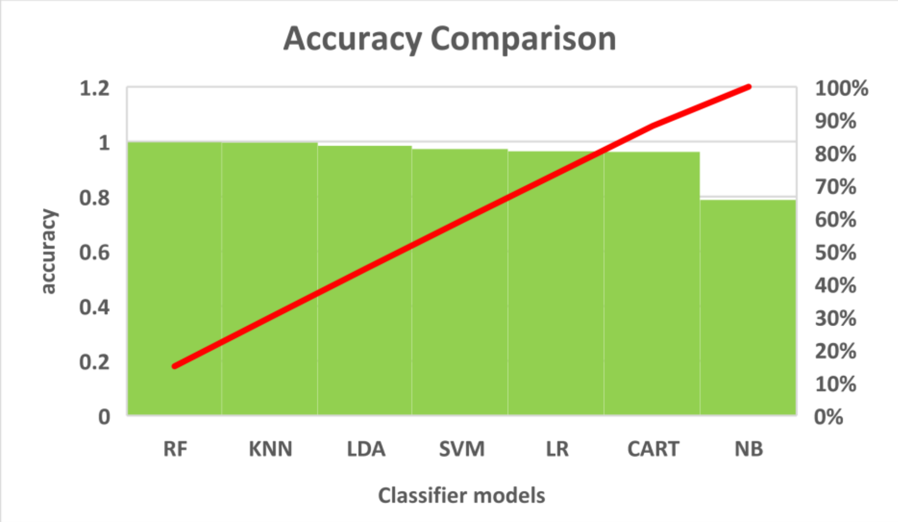

Figure 4: Machine learning algorithms accuracy comparison

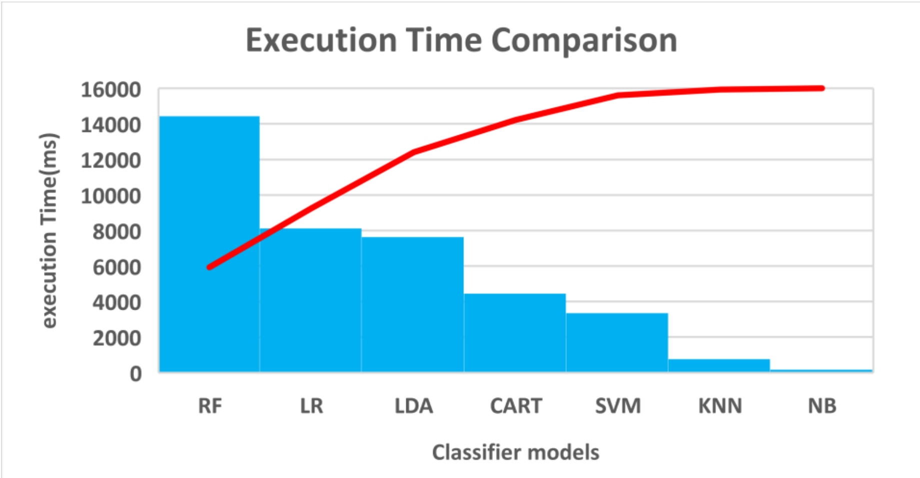

Figure 5: Machine learning algorithms execution time comparison

It can be said that in comparison to other reported methodologies, the one that is provided is more effective and reliable [18,19]. Furthermore, the categorization of plant leaves that are used in this procedure is superior to what is currently in place because of its accuracy, satisfaction, and superiority. As shown in table 4, a comparison of the current taxonomy of plant leaf infection can be made.

Table 4: Comparison Table with existing literature

Reference Paper	Dataset	Pre-processing	Features	Training Model	Tools	Average Accuracy
[20]	Medicinal Plant Leaves-`6	Cropping Leaf Region, RGB to GL	Multispectral, Texture, Runlength	MLP, LB, B, RF, SL	Computer Vision Laboratory Setup, Open CV	99.10% tulsi-AA
[21]	Medicinal Leaves	RGB-Grayscale-Binary	Morphological, Filtering	Multilayer Perception	Matlab, SPSS	Recognition Rate-70.87%
[22]	Medicinal Plants-900	HSV, YCbCr	Colour, Edge, Texture	SVM, RBENN	Matlab, ANN Tool	AA-90% Combined
Proposed methodology	Tulsi	BGR, HSV	Hu Moments, Harlick Texture, color histogram	SVM,KNN, RF	Computer Vision Laboratory Setup, Open CV	SVM-98.2408 KNN-99.6801 RF-99.89

4 Conclusion

In this paper, we develop a machine learning (ML) system to detect *Ocimum Tenuiflorum* infection on leaves using machine learning techniques. The aim of this study is to construct a machine learning classification system based on machine learning (ML) which is used to categorize *Ocimum Tenuiflorum* leaves as being either infected or healthy, using three feature extractors for the datasets. Gathering a clean, uniform dataset is the first step, followed by identifying edges and lines, extracting fused features, optimizing extracted features, selecting the most important feature, and selecting an effective machine learning algorithm. The categorization of data can be improved by incorporating five valuable traits that can help to enhance the accuracy of the categorization. Seven AI-based classifiers in all are taken into account. The results of this study contribute to a broader classification of therapeutic plant leaves. The ability to identify the right kind of infection in a plant can also be very useful for those in the fields of economy and agriculture if they are able to aid in the production of *Ocimum Tenuiflorum* leaves.

4.1 Future Suggestions

The study focuses only on the infection of the leaves of the *Ocimum Tenuiflorum* medicinal plant, despite the fact that there are countless varieties of medicinal plants and herbs throughout the world. The current approach uses pixels, but in the future, an object-based approach may be more appropriate. When this suggested method is applied to the infected leaves of various medicinal plants in the future, hyperspectral and 3D digital image datasets can be used to enhance it.

5 References

1. Azlah, M.A.F., Chua, L.S., Rahmad, F.R., Abdullah, F.I., & Wan Alwi, S.R. (2019). Review on Techniques for Plant Leaf Classification and Recognition. *Computers*, 8(4), 77.
2. Gowans M. Brownish Yellow Spots on a Holy Basil Plant. https://homeguides.sfgate.com/brownish-yellow-spots-holy-basil-plant-71343.html (April 2021).
3. Barbedo, J.G.A. (2013). Digital image processing techniques for detecting, quantifying, and classifying plant diseases. *SpringerPlus*, 2(1), 1-12.
4. Sakhamuri, S., & Kompalli, V.S. (2020). An overview on prediction of plant leaves disease using image processing techniques. *In IOP Conference Series: Materials Science and Engineering* (Vol. 981, No. 2, p. 022024). IOP Publishing.
5. Mishra, T., Rai, A. (2021). Antimicrobial potential of Ocimum sanctum and Adhatoda vasica: The Medicinal Herbs. *International Journal of Pharmacy and Biological Sciences* 1191, 17-28.
6. Singh, S., Malhotra, M., Majumdar, D.K. (2005). Antibacterial activity of Ocimum sanctum L. Fixed oil. *Indian J Exp Biol*; 43:835-7.
7. Singh, V., Verma, O. (2010). Ocimum sanctum (tulsi): Bio-pharmacological activities. WebmedCentral Pharmacol; 1:WMC001046.
8. Patil, R.S., Kokate, M.R., Kolekar, S.S. (2012). Bioinspired synthesis of highly stabilized silver nanoparticles using Ocimum tenuiflorum leaf extract and their antibacterial activity. Spectrochim Acta A Mol Biomol Spectrosc; 91:234-8.
9. Roopashree, S., Anitha, J. (2020). Enrich Ayurveda knowledge using machine learning techniques. *Indian Journal of Traditional Knowledge;* 19(4), 813-820.
10. Sinha, A.K., Jaggi, M., Raghuram, B., Tuteja, N. (2011). Mitogen-activated protein kinase signaling in plants under abiotic stress. Plant Signal Behav.; 6(2):196–203.
11. Hammer, K.A., Carson, C.F., Riley, T.V. (1999). Antimicrobial activity of essential oils and other plant extracts. *J. Appl. Microbiol*., 86(6): 985.
12. Sana, O.M., Jaya, R. (2015). "Ayurvedic Herb Detection Using Image Processing". *International Journal of Computer Science and Information Technology Research*, 3(4), 134139.
13. Ruaa Adeeb Abdulmunem Al-falluji (2016). "Color, Shape, and Texture based Fruit Recognition System". *International Journal of Advanced Research in Computer Engineering & Technology*, 5(7), 2278 – 1323.
14. Vishnoi, V.K., Kumar, K., & Kumar, B. (2021). Plant disease detection using computational intelligence and image processing. *Journal of Plant Diseases and Protection*, 128(1), 19-53.
15. [15] Ganatra, N., Patel, A. (2020). A Multiclass Plant Leaf Disease Detection using Image Processing and Machine Learning Techniques. *International Journal on Emerging Technologies* 11(2), 1082-1086.
16. Dhingra, G., Kumar, V., & Joshi, H. D. (2018). Study of digital image processing techniques for leaf disease detection and classification. *Multimedia Tools and Applications*, 77(15), 19951-20000.

17. Mingqiang, Ji., Yang, Yu., & Zheng, Y., Zhu, Q., & Huang, M., & Guo, Ya. (2020). In-field automatic detection of maize tassels using computer vision. *Information Processing in Agriculture* 8, 87-95.
18. Jasim, M. A., & Al-Tuwaijari, J. M. (2020). Plant leaf disease detection and classification using image processing and deep learning techniques. *In 2020 International Conference on Computer Science and Software Engineering (CSASE)* (pp. 259-265). IEEE.
19. Vijayashree, T., Gopal, A. (2015). Classification of Tulsi Leaves Based on Texture Analysis. *Middle-East Journal of Scientific Research 23 (Sensing, Signal Processing and Security).* 102-107.
20. Naeem, S., Ali, A., Chesneau, C., Tahir, M. H., Jamal, F., Sherwani, R. A. K., & Ul Hassan, M. (2021). The Classification of Medicinal Plant Leaves Based on Multispectral and Texture Feature Using Machine Learning Approach. *Agronomy*, 11(2), 263.
21. Ananthi, C., Periyasamy, A., & Muruganand, S. (2014). Pattern Recognition of Medicinal Leaves Using Image Processing Techniques. *Journal of Nanoscience and Nanotechnology*, 2(2), 214-218.
22. Anami, B.S., Nandyal, S.S., Govardhan, A. (2010). A combined color texture and edge features-based approach for identification and classification of Indian medicinal plants. *International Journal of Computer Applications* 6(12), 45-51.