Lecture Notes in Networks and Systems 1111

Jyoti Choudrie
Parikshit N. Mahalle
Thinagaran Perumal
Amit Joshi *Editors*

ICT for Intelligent Systems

Proceedings of ICTIS 2024, Volume 5

Springer

Machine-Learning Methods for Plant Leaf Disease for Improving Agricultural Production

Ashish Nagila and Abhishek K. Mishra

Abstract Agricultural production is an essential component in India's economy. Agricultural production ensures that everyone can eat, even in cases of rapid population increase. In particular, tomatoes have emerged as India's most valuable crop. It is recommended to anticipate plant diseases when they are in their early phases of growth so that food can be provided for all residents. Predicting disease in immature crops is a sad reality, though. The research aims to inform farmers about state-of-the-art methods for decreasing tomato leaf diseases. When machine-learning techniques applied, it can detect leaf diseases in tomato plants. The samples of sick tomato leaves are considered in this study. By analysing these diseased tomato leaf samples, farmers would be able to identify infections based on their early symptoms. After down sampling the tomato leaf samples to 256 × 256 pixels, histogram equalisation is applied to boost the tomato section's quality. In order to partition the data into Voronoi cells, the K-means clustering algorithm is employed. The procedure of contour tracing is used to determine the boundaries of leaf samples. Discrete wavelet transforms, principal component analysis, and greyscale co-occurrence matrices are among the descriptors that can be used to extract the informative features of a leaf sample. A model is proposed, which is combination of SVM and CNN for classify the disease for tomato leaves. With an accuracy of over 99% in both the training and test datasets, the results demonstrate that the suggested model performs admirably when it comes to detecting and classifying illnesses in tomato leaves.

Keywords Discrete wavelet transform · Convolutional neural network · Nearest neighbour · Principal component analysis

A. Nagila (✉) · A. K. Mishra
IFTM University, Moradabad, Uttar Pradesh, India
e-mail: ashishnagila01@gmail.com

J. Choudrie et al. (eds.), *ICT for Intelligent Systems*, Lecture Notes in Networks and Systems 1111, https://doi.org/10.1007/978-981-97-6681-9_26

1 Introduction

We can now meet the food needs of our society with the help of modern technology. However, we still haven't gotten around to protecting our crops and food. Plant diseases, declining pollinator populations, and climate change are just a few of the challenges that farmers face today. Establishing a solid groundwork for these components should take precedence [1, 2]. By employing technologically advanced analysis and detection processes, farmers may put an end to these issues. When faced with an epidemic like COVID-19, the nation looks to modern technologies to address concerns and contain the disease [2–5]. Famines and droughts are two ways in which plant diseases put human lives in jeopardy. Consequently, when farming is done for profit, it leads to significant losses. "Machine learning and computer vision are two examples of how technology might be applied to help avoid diseases [6–8]. Our goal in this research was to find a way to treat plant diseases using machine learning. This approach consists of three steps: identify, examine and verify with a database [9]. Scientists and researchers analyse plant leaf diseases to determine the main problems and technical hitches [10, 11]. Below are only some of them: (1) Leaf image needs to be of outstanding quality. (2) The need for publicly accessible datasets. (3) Noisy data influencing leaf sample results. (4) While segmentation can aid in disease diagnosis, training and testing are necessary steps for the samples.

(5) Categorization presents yet an additional difficulty for the detection of leaf diseases. (6) Depending on the surroundings, the colour of the leaves may change. (7) Different type of plant show a range of illnesses, making disease identification challenging. Taken together, the aforementioned challenges and the proposed model's use of IP and ML techniques yield higher accuracy. Taking everything into account, this work proposes a method for detecting leaf illnesses using IP and ML technology that does not require human participation.

There are three levels of support for the mentioned structure. The initial step in optimising the superiority component of the leaf section is to engage K-means clustering and histogram equalisation. The K-means clustering reaction allows for early prediction of an infected leaf's state. Secondly, the features and informative regions of the samples are extracted using the PCA, GLCM, and DWT. The properties are subsequently classified with machine learning algorithms such as CNN, SVM, and KNN. In Sect. 2, you will find a detailed explanation of the methods currently used to combat leaf diseases. The final portion delves into the model that can be used to detect illnesses in leaves. In Sect. 4, we examine the outcomes of the suggested model using six samples of sick tomato leaves. Chap. 5 concludes the model for leaf detection.

2 Literature Review

Disease identification in leaves has been a primary focus of image processing for a long time and shows no signs of going away. The automated detection of agricultural illnesses through the application of machine learning image processing methods has witnessed a remarkable surge in the past few years. Krisika P. et al. Methods such as picture scaling, contrast enhancement, and colour space conversion are employed in the pre-processing phase by Dataset Kaggle [9]. The segmentation and feature extraction procedure makes use of K-Means clustering in conjunction with the Grey Level Co-occurrence Matrix (GLCM). R. Meena et al. employed a multi-class support vector machine (SVM) for classification in this work. Following the colour space conversion, a complex method was employed in [10]. Changing the leaf colour is the main way to get LAB production. The K-means clustering technique is utilised for segmentation. The feature extraction or classification processes make use of Support Vector Machines (SVMs) as well as Grey Level Co-occurrence matrix (GLCMs). Researchers utilised a digital camera to take pictures, which were subsequently enhanced using a median filter. One common method for segmentation is K-means clustering. The Support Vector Machine (SVM) classification procedure. Data segmentation was used to identify target locations, particularly contaminated areas, in the study conducted by Pooja et al. [12]. After utilising a use of Otsu's analysis and the k-means clustering method allowed for the conversion of RGB to HSI. The point method and the border detection method are used to carry out the classification procedure. Rukaiyya et al. employed the same methods in their study such as standardisation and variance correction [13]. Following the installation of two-level thresholds, the colour change is transformed into YCBCR. The two most popular methods for extraction and classification are Hidden Markov Models (HMMs) as well as Grey-Level Co-occurrence Matrixes (GLCMs) [14]. According to what Chaitali et al. indicated, image segmentation was utilised for the goal of posterior dissection. [15]. We use ANN, Support Vector Machines (SVM), and K-Nearest Neighbours (KNN) methods to classify the data. To sort the data into categories, we utilise the K-Nearest Neighbours (KNN) algorithm, which compares the distances of the training and test sets [16]. With Otsu's detection and the k-Mean clustering method. Utilising a border and spot detection technology, the segmentation process is thereafter carried out. The investigation conducted by Rukaiyya et al. utilised pre-processing processes like as normalisation and contrast correction [13]. Before bi-level thresholding is used, the colour transformation is changed to YCBCR. When it comes to feature extraction and classification, two often used algorithms are the Hidden Markov Model (HMM) and the Grey-Level Co-occurrence Matrix (GLCM) [14]. According to Chaitali et al. [15], image segmentation is used to remove backgrounds. K-Nearest Neighbours (KNN), Artificial Neural Networks (ANN), and Support Vector Machines (SVM) are the methods used to execute the classification procedure. By comparing the locations of training and test subjects, the K-nearest neighbours (KNN) method is used to categorise samples [16]. Models for the extraction thresholding method and morphological operation have been suggested. Then, a Support Vector Machine (SVM) classifier with several classes is used. To accomplish segmentation, LAB colour spaces are utilised. These spaces are generated by analysing the colour and brightness components of several picture regions to establish a set of markers. In order to extract characteristics, the GLCM is used. In their study, looked at a number of digitally photographed plant leaf samples. Among the samples were roses and beans, lemons and sunburn, early scorch and beans, respectively, caused by bacterial disorder, sunburn, sunburn, and fungal disorder. The verdant areas serve as a backdrop for the whole thresholding procedure. The segmented image is then obtained by utilising the genetic algorithm. By modifying the colour co-occurrence, relevant information can be more easily extracted from segmented images. The SVM classifier is used after the Minimum Distance Criterion, which is the first step in the classification process. The accuracy rate is at 97.6%. Abed et al. [17] used the min–max linear method, a scaling and stretching technique, to improve the raw data quality in their investigation. All work on the HIS model has been finished; now we may move on to the segmentation phase. For data partitioning, we use K-means clustering techniques in conjunction with mixed Euclidean distance. For the purpose of feature extraction and classification, the Grey Level Co-occurrence Matrix (GLCM) and Support Vector Machine (SVM) approaches are employed. Arya et al. [18, 19] alters the colours of the input RGB picture before transforming it to HIS format. In the end, Otsu's methodology is used to separate the components. After adding the images from the 1981 study by Nema et al. [20] to the database, analysis was carried out using the LAB colour space. Using disease classification and then K-means clustering with Support Vector Machines (SVM), leaf disease segmentation was accomplished. Statistical measurements including median, mean, mode, and standard deviation were employed by the writers to support their conclusions. Kanbur et al. [21] used a number of attributes to build the disease identification model for leaves. When tested using a local leaf database, the model performed better; however, it can also be tested with publicly available datasets. To identify the affected area of the leaf, Pushpa et al. [22] use an Indices Based Histogram technique. The difficulties of mean-shift segmentation, polygon approximation, and slice segmentation have been adequately handled by the writers. Following the steps outlined by Kaleem et al., the images were preprocessed [23]. Reducing the photos to 300×300 pixels, eliminating noise from the background, making them brighter, and tweaking the contrast were all part of this process. An SVM classifier is utilised for categorising illnesses in leaves, while K-means clustering is employed for segmentation and data extraction.

3 Proposed Model

This section explains how to build up a model for detecting leaf diseases using ML and IP techniques. The proposed model for detecting leaf diseases using computer vision and machine learning is shown in Fig. 1. It consists of DWT, PCA, GLCM, and CNN. The correctness of the evaluation and the healthiness of a leaf ailment can be evaluated by looking for six abnormalities in the tomato samples used for the test. The tomato samples are shrunk to 256 × 256 pixels as part of the image processing approach so that their dimensions remain constant throughout the experiment. To improve the quality and efficiently segment the sets of leaf samples, HE and K-means clustering techniques are being used. In the early phases of a leaf's functioning, the infection status can be predicted using the K-means clustering technique. Using the contour tracing method, you may find out where the leaf sample borders. To extract the relevant sample regions and attributes, the DWT, PCA, and GLCM techniques are used. Following this, the features are categorised using machine-learning approaches such as Support Vector Machines (SVM), K-Nearest Neighbours (KNN), and Convolutional Neural Networks (CNN), while the model's performance is tracked.

3.1 Dataset

The analysis includes the plants affected by several illnesses, as recorded in the tomato leaf village database [14]. Tomato leaves with six different diseases are photographed so as to carry out the tests for identifying leaf diseases. Figure 2 shows examples of leaf images from the database.

3.2 Pre-processing

In order to identify the diseased region, pictures of the leaves are fed into the K-means clustering algorithm [16–18, 24, 25]. Thanks to the K-mean clustering algorithm, we can find the picture's data centre, image clusters, and the distance between the centre and neighbouring clusters. Figure 3 shows the result of the leaf sample after applying the k-mean clustering method [17]. Scanned leaf samples can have their general shape information recovered via contour tracing [18, 19]. After the contour has been obtained, its characteristics are examined and patterns are subsequently classified using them. It is useful for determining how effective the feature extraction method was [17, 25]. Figure 4 displays the images that were produced via contour tracing.

3.3 Feature Extraction

Unstructured Wavelet Improved tomato samples are subjected to transform the DWT [20] to extract data on the pertinent traits. You can separate a signal into sub-bands having lower frequency (LH, LL, and HL) and higher frequency (HH) components using the discrete wavelet transform (DWT). Figure 5 shows that compared to the

Fig. 1 Proposed model for plant leaf

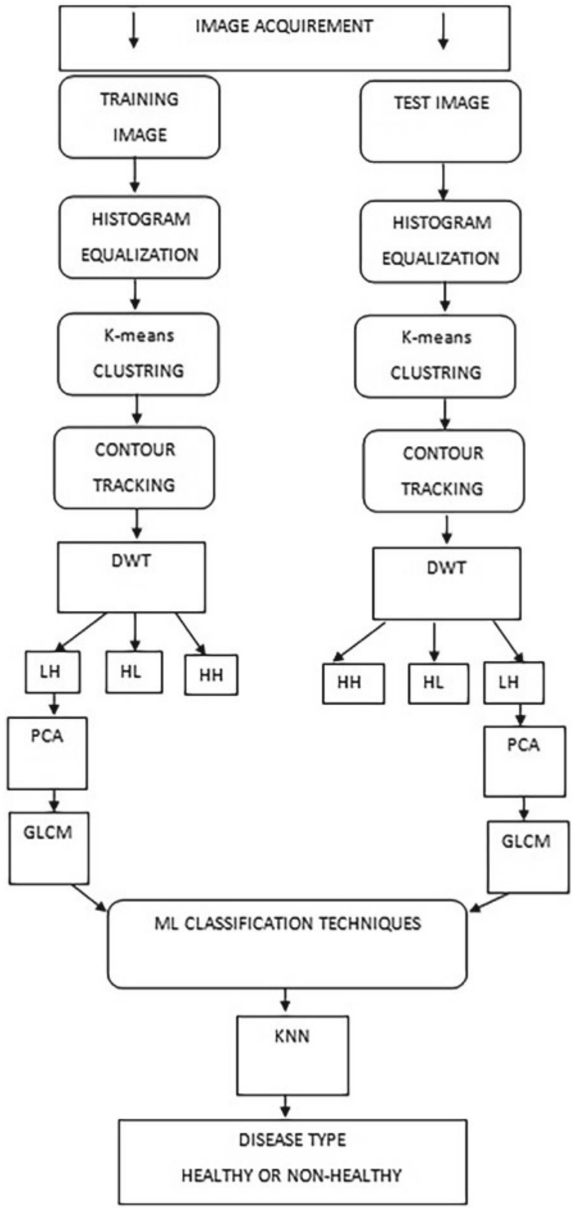


Fig. 2 Tomato sample with diseases

Fig. 3 K-mean clustering

higher frequency components of DWT, the LL component carries the most readily available information. Greyscale Matrix Illustrating Their Presence: To select the most valuable features, principal component analysis [21, 22] is used to the features that are generated by wavelet decomposition. In order to find the distribution of higher order grey values, the GLCM employs the neighbourhood criteria [23, 26]. These qualities were built on top of the GLCM approach, which is used to extract leaf features. Here is a rundown of the most popular texture-based features. Terms like "uniformity auto" are defined in [27, 28]. We have entropy, differences, and correlation. A feature vector is constructed by combining the features obtained by DWT, GLCM, and PCA. This vector is then supplied to a classifier in order to identify and categorise images.

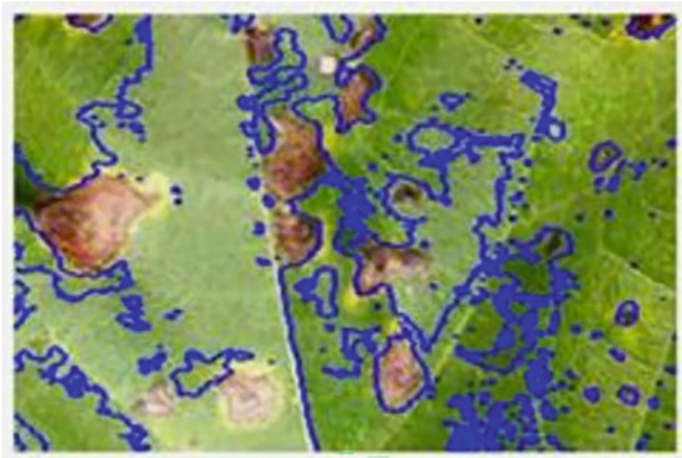
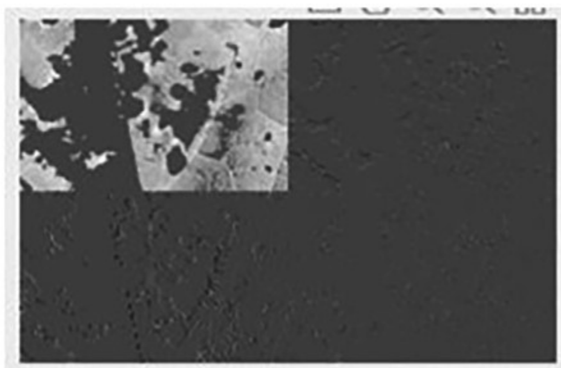


Fig. 4 Contour tracing

Fig. 5 DWT decomposition



3.4 Evaluation Parameter

The factors associated with the proposed model, namely Precession, Recall, F-measure and accuracy, have been established and are outlined in Eqs. 1, 2, 3, and 4.

$$\text{Precision Measure}(\%) = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \tag{1}$$

$$\text{Recall Measure}(\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \tag{2}$$

$$\text{F} - \text{measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} * \text{Recall}} \times 100 \tag{3}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} = \text{FP} = \text{FN}} \times 100 \tag{4}$$

(TP: True Positives) (FP: False Positives) (FN: False Negatives) (TN: True Negatives).

4 Experimental Results

We are contemplating using the tomato leaf samples from the village dataset to do the model evaluation. Using the 100 healthy leaf samples obtained, the suggested system is tested. The programme achieved an impressive success rate of 99% when it came to properly identifying 99 different samples. If there are 100 tomato samples infected with the Mosaic virus, the model can accurately identify 100 of them. When it comes to the type of mould known as leaf mould, the model's predictions are spot on. Using 100 examples of yellow curls, the model achieved a 99% success rate. The findings of the Target Spot test were 100% and the spotted spider mite test were 99%, in a similar vein. A thorough evaluation is required to determine the suggested model's validity and effectiveness; thus, 600 samples from the tomato village dataset are tested on it. The model's accuracy improves to 99.5% because of this. Model validation occurs during training and testing using dataset samples. It is recommended to complete the assignment with the specified software and hardware requirements. Dataset: Plant village dataset with samples of tomatoes in six distinct disorders; Windows 10 operating system; GPU-NVIDIA core; Python language libraries: Tensor flow, OpenCV, Keras, NumPy, Matplotlib. Variables such as $F1$ score, Precision, and Recall are used to evaluate the performance of the proposed model (Fig. 6).

The suggested model is evaluated on a dataset comprising 600 total samples pertaining to tomato leaf disease. We compare the results of the proposed model to those of previous models. Comparing the proposed method (DWT + PCA + GLCM + RF) to the current models in use, it has shown a higher level of accuracy, precisely 99.09%.

5 Result Comparison

We are contemplating using tomato leaf samples from village dataset to do the model evaluation. Using the 100 healthy leaf samples obtained, the suggested system is tested. The programme achieved an impressive success rate of 99% when it came to properly identifying 99 different samples. If there are 100 tomato samples infected with the Mosaic virus, the model can accurately identify 100 of them. When it comes to the type of mould known as leaf mould, the model's predictions are spot on. A 99% success rate was attained by the model when fed 100 samples of yellow curls.

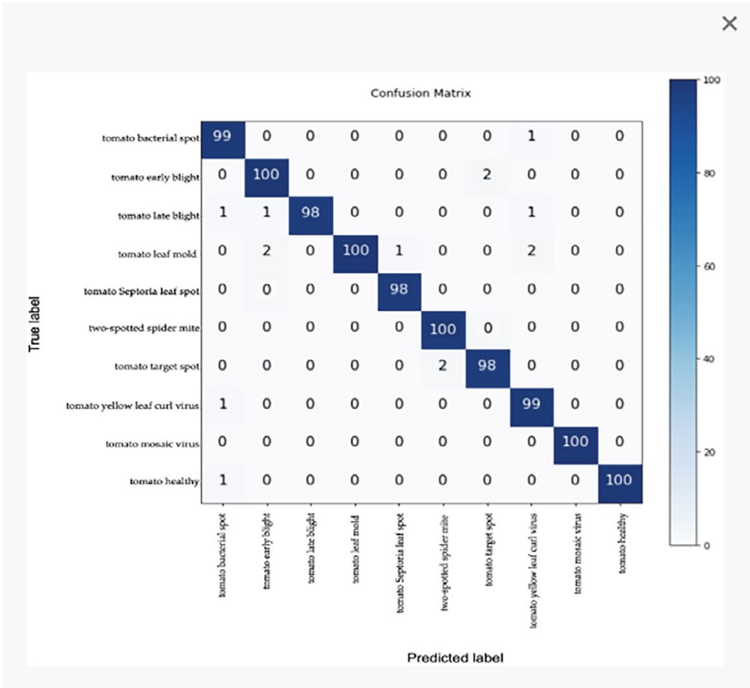


Fig. 6 Confusion matrix of the proposed model

The findings of the Target Spot test were 100% and the spotted spider mite test were 99%, in a similar vein. Six hundred samples from the tomato village dataset are run through the suggested model to get an estimate. Hence, the model attains a superior level of accuracy, reaching 99.5% (Table 1).

Table 1 The suggested model is compared to current approaches

Authors	Methodologies/Description	Accuracy in percentage
Hossain et al. [18]	Based on univariate statistical features test + SVM	(90)
Vidyashree et al. [29]	K-mean clustering + GLCM + SVM	(90)
Vadivel et al. [30]	Fast enhanced learning method	(99)
Harakannanavar et al. [31]	DWT + PCA + GLCM + CNN	(99.09)
Proposed model	DWT + SVM + GLCM + CNN	(99.5)

6 Conclusions

Through the use of the suggested model, this study has demonstrated a method for disease classification and identification in tomato leaves. When it comes to accuracy, the proposed method achieves promising results close to 99%. With the potential for an increase in the number of diseases in the future, this experiment does have certain restrictions. Future work on the model can focus on enhancing its feature extraction capabilities and expanding its testing to include additional dataset's leaf samples.

References

1. Krithika P, Veni S (2017) Leaf disease detection on cucumber leaves using multiclass support vector machine. In: IEEE international conference on wireless communications, signal processing and networking, pp 1276–1281
2. Prakash R, Saraswathy GP, Ramalakshmi G (2017) Detection of leaf diseases and classification using digital image processing. In: IEEE international conference on innovations in information, embedded and communication systems, pp 1–4
3. Mishra B, Nema S, Lambert M, Nema S (2017) Recent technologies of leaf disease detection using image processing approach-review. In: IEEE international conference on innovations in information, embedded and communication systems, pp 1–5
4. Puttamadappa C, Parameshachari BD (2019) Demand side management of small scale loads in a smart grid using glow-worm swarm optimization technique. Microprocessors Microsyst 71:102886
5. Pooja V, Das R, Kanchana V (2017) Identification of plant leaf diseases using image processing techniques. In: IEEE international conference on technological innovations in ICT for agriculture and rural development, pp 130–133
6. Vu DL, Nguyen TK, Nguyen TV, Nguyen TN, Massacci F, Phung PH (2020) HIT4Mal: hybrid image transformation for malware classification. Trans Emerg Telecommun Technol 31(11):e3789
7. Shaikh RP, Dhole SA (2017) Citrus leaf unhealthy region detection by using image processing technique. In: IEEE international conference on electronics, communication and aerospace technology, pp 420–423
8. Yu K, Lin L, Alazab M, Tan L, Gu B (2020) Deep learning-based traffic safety solution for a mixture of autonomous and manual vehicles in a 5G-enabled intelligent transportation system. IEEE Trans Intell Transp Syst 22(7):4337–4347
9. Dataset Kaggle. https://www.kaggle.com/thanjaivadivelm
10. Dhaware CG, Wanjale KH (2017) A modern approach for plant leaf disease classification which depends on leaf image processing. In: IEEE international conference on computer communication and informatics, pp 12–16
11. Nayak JP, Anitha K, Parameshachari BD, Banu R, Rashmi P (2017) PCB fault detection using image processing. IOP Conf Ser Mater Sci Eng 225. IOP Publishing
12. Nguyen CH, Pham TL, Nguyen TN, Ho CH, Nguyen TA (2021) The linguistic summarization and the interpretability, scalability of fuzzy representations of multilevel semantic structures of word-domains. Microprocessors Microsyst 81:103641
13. Gupta V, Sengar N, Dutta M, Travieso C, Alonso J (2017) Automated segmentation of powdery mildew disease from cherry leaves using image processing. In: IEEE international conference and workshop on bioinspired intelligence, pp 1–4
14. Yu K, Tan L, Lin L, Cheng X, Yi Z, Sato T (2021) Deep-learning-empowered breast cancer auxiliary diagnosis for 5GB remote E-health. IEEE Wirel Commun 28:54–61

15. Vijai Singh, A.K. Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques. Inf Process Agric 4 (1):41–49
16. Abed S, Esmaeel A (2018) A novel approach to classify and detect bean diseases based on image processing. In: IEEE symposium on computer applications and industrial electronics, pp 297–302
17. Chandrasekaran G, Nguyen TN, Hemanth D (2021) Multimodal sentimental analysis for social media applications: a comprehensive review. Wiley Interdisc Rev 11(5):e1415
18. Hossain S, Mou R, Hasan M, Chakraborty S, Razzak A (2018) Recognition and detection of tea leaf's diseases using support vector machine. IEEE Int Colloquium Signal Process Appl 150–154
19. Zhang J, Yu K, Wen Z, Qi X, Paul AK (2021) 3D reconstruction for motion blurred images using deep learning-based intelligent systems. CMC Comput Mater Continua 66(2):2087–2104
20. Saleem MH, Khanchi S, Potgieter J, Arif KM (2020) Image-based plant disease identification by deep learning meta-architectures. Plants (Basel) 9(11):1–23
21. Pushpa SH, Ashok A (2021) Diseased leaf segmentation from complex background using indices based histogram. In: IEEE International conference on communication and electronics systems, pp 1502–1507
22. Kaleem MK, Purohit N, Azezew K, Asemie S (2021) A modern approach for detection of leaf diseases using image processing and ML based SVM classifier. Turkish J Comput Math Educ 12(13):3340–3347
23. Vadivel T, Suguna R (2021) Automatic recognition of tomato leaf disease using fast enhanced learning with image processing, Taylor Francis. Acta Agricult Scandinavica Sect B Soil Plant Sci 71(1):1–13
24. Nagashetti SM, Biradar S, Dambal SD, Raghavendra CG, Parameshachari BD (2021) Detection of disease in Bombyx Mori Silkworm by using image analysis approach. In: 2021 IEEE Mysore sub section international conference (MysuruCon). IEEE, pp 440–444
25. Arya M, Anjali K, Unni D (2018) Detection of unhealthy plant leaves using image processing and genetic algorithm with Arduino. In: IEEE international conference on power, signals, control and computation, pp 1–5
26. M TS, S SK, S SM, Devi PR (2021) Leaf disease detection using deep learning. In: IEEE international conference on electronics and sustainable communication systems, pp 1797–1804
27. Bhagwat R, Dandawate Y (2021) Comprehensive multilayer convolutional neural network for plant disease detection. Int J Adv Comput Sci Appl 12(1):204–211
28. Khan RU, Khan K, Albattah W, Qamar AM, Image-based detection of plant diseases: from classical machine learning to deep learning journey. Wirel Commun Mob Comput 1–13
29. Kanabur V, Harakannanavar SS, Puranikmath VI, Torse D (2019) Detection of leaf disease using hybrid feature extraction techniques and CNN classifier. Springer Comput 1213–1212
30. Vadivel T, Suguna R (2021) Automatic recognition of tomato leaf disease using fast enhanced learning with image processing, Taylor Francis. Acta Agricult Scandinav Jica Sect B Soil Plant Sci 71(1):1–13
31. Harakannanavar SS, Rudagi JM, Puranikmath VI (2022) Plant leaf disease detection using computer vision and machine learning algorithms. Global Trans Proc 3:305–310
32. Narmadha RP, Arulvadivu G (2017) Detection and measurement of paddy leaf disease symptoms using image processing. In: IEEE international conference on computer communication and informatics, pp 1–4
33. Gokula KV, Deepa JR, Rao PV, D V, K S (2021) An automated segmentation and classification model for banana leaf disease detection. J Appl Biol Biotechnol 1–12