

Covariance Kalman Geometric Graph Based Feature Extraction And Bernoulli Kernel Classifier For Plant Leaf Disease Prediction

Mohammed Zabeeulla A N^a, Dr. Chandrasekar Shastry^b

^a Research Scholar, Jain (Deemed-to-be University)

^b Dean, PG Studies, FET, Jain (Deemed-to-be University)

^azabee225@gmail.com, ^bcsshastry2@gmail.com

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: As far as the agricultural domain is concerned, one of the most hot research areas of analysis is accurate prediction of leaf disease from the leaf images of a plant. The prediction of agricultural plant diseases by means of the image processing techniques will hence reduce the dependence on the farmers to safeguard their agricultural land and also their products. However, with the presence of noise, the leaf disease prediction is said to be hindered. To address this issue, in this paper, Covariance Kalman Geometric Graph-based Bernoulli Classifier (CKGG-BC) for Plant leaf disease prediction is proposed. The CKGG-BC method is split into three parts. To start with the plant leaf image provided as input, the Covariance Kalman Filtered Preprocessing model introduced for the image enhancement. Second, Geometric Graph-based Segmented Co-occurrence Feature Extraction model is applied to the preprocessed image to accurately segment the infected leaf areas and followed by which extracting the accurate infected leaf areas. Finally, Bernoulli Online Multiple Kernel Learning Classifier is applied for accurate plant leaf disease prediction with minimum classification error. The proposed method provides a significant refinement with respect to state-of-the-art methods. Even under complex background conditions, i.e., in the presence of noise, the average accuracy of the proposed method is said to be improved and hence paves mechanism for prediction of plant leaf disease in a significant manner. Experimental results exhibit the effectiveness of the proposed method in terms of computational overhead, accuracy, true positive rate and classification error respectively.

Keywords: Kalman Filtered Principal Component, Geometric Graph, Co-occurrence Feature Extraction, Bernoulli Trials, Online Multiple Kernel Learning Classifier

1. Introduction

The development of plant diseases has negative reactions on agricultural production, and if the plant diseases are not detected timely, it will result in a growth in food insecurity. In specific, the predominant crops like, rice, maize, etc., are indispensable for undertaking the food supply and agricultural production. The preliminary warning and forecast are the foundation of significant avoidance and control for plant diseases. They play critical parts in the agricultural production management and decision-making process. To the date till, visual observations of knowledgeable producers are still the predominant pattern for plant disease detection in rural areas of developing countries.

AlexNet and GoogleNet CNN was proposed in [1] to present a significant soybean diseases identification method on the basis of a transfer learning model using pretrained AlexNet and GoogleNet convolutional neural networks (CNNs). Here, five-fold cross-validation strategy was utilized. Initially, the preprocessed images were applied as input to the pre-trained GoogleNet CNN architecture. Here, retraining was performed on the preprocessed images with the purpose of classifying into four class species of objects from the input disease data set, therefore resulting in the accuracy improvement.

Despite improvement observed in the accuracy, with the presence of noise, the accuracy factor is said to be compromised. To address this issue in this work, Covariance Kalman Filtered Preprocessing model are applied as preprocessor that with the assistance of Covariance Kalman filter the principal components are obtained and only with this preprocessed images, further analysis is said to be carried out. In this way, the accuracy of leaf disease prediction is said to be improved even in the presence of noise.

An Inception model by Visual Geometry Group for deep convolutional network, called, INC-VGGNet was proposed in [2] for plant leaf disease identification. To start with a pre-trained model obtained from the massive dataset was considered as input and then transferred to the particular task trained by our own data. Here, VGGNet pre-trained on ImageNet and Inception module were selected, erstwhile of initiating the training from the starting

stage by initializing the weights in a random manner. Here, the weights were initialized with the aid of pre-trained networks, therefore contributing to validation accuracy.

Though validation accuracy was improved, geometric properties were not analyzed, therefore, contributing to classification error. To address this issue, in our work, Geometric Graph-based Segmented Co-occurrence Feature Extraction model is first applied then to the resultant segmented infected leaf areas, the prediction is made. In this manner, the classification error involved in prediction of leaf diseases is said to be reduced significantly.

Motivated by the above said issues, in this work, Covariance Kalman Geometric Graph-based Bernoulli Classifier (CKGG-BC) for Plant leaf disease prediction is proposed. This study aims to introduce Bernoulli Online Multiple Kernel Learning Classifier as an approach for classifying three types of leaves, i.e., pepper, tomato and potato according to sample leaf images. This study presents three main contributions in plant disease classification:

- (1) Implementation of the Covariance Kalman Filtered preprocessing model by using the covariance between the two vectors (i.e., state and observation vectors) on plant leaf dataset for enhancing the image even in the presence of noise.
- (2) To design a Geometric Graph-based Segmented Co-occurrence Feature Extraction algorithm by first obtaining the Geometric Pixel to accurately segment the infected areas and extract the geometric properties of the infected leaf areas by means of Graph-based Segmented Co-occurrence Feature Extraction.
- (3) Accurate and timely prediction of leaf disease using Bernoulli Online Multiple Kernel Learning Classification algorithm
- (4) The proposed method is evaluated on plant disease dataset using well-known metrics, such as, computational overhead, accuracy, true positive rate and classification error with the results compared using state-of-the-art methods.

The organization of the remaining paper is as follows. Section 2 provides highlights of state-of-the-art plant leaf disease prediction methods. Section 3 elaborates in detail the proposed method, Covariance Kalman Geometric Graph-based Bernoulli Classifier (CKGG-BC) for Plant leaf disease prediction. Section 4 showcases the experimental settings with a detailed discussion of the proposed method in Section 5. Section 6 concludes the work.

2. Related Work:

Plant disease has become a crucial warning food security globally. A significant amount of losses are said to occur due to plant disease globally each year. In the current situation, there is a requirement for efficient detection of the plant disease at an early stage so that plant diseases can be controlled significantly, therefore enhancing the agro ecosystem sustainability.

In [3], a novel automatic method for tassel detection was proposed using a color attenuation prior model that in turn removed the saturation present in the images via saturation graph. Followed by which the area of interest was detected by means of an Itti visual attention detection algorithm. Finally, the false positives present in the images were discarded via texture features and vegetation indices. However, measures were not taken to monitor large farms. To address this issue, an optimization method based on particle swarm optimization was proposed in [4] for detecting the disease in sunflower leaf. A review of neural network techniques via hyperspectral data for detecting disease in plant leaf was investigated in [5].

Swift and precise plant disease detection is evaluative to increasing agricultural productivity in a feasible manner. To be specific, speaking in a conventional manner, human experts have been depended upon to discover anomalies in plants generated by different factors, to name a few being, pesticides, deficiencies caused due to malnutrition or deteriorated weather conditions. However, it is found to be laborious process involving cost, time and in certain cases also proven to be impossible. In [6], a thorough review of current works conducted in the area of crop pesticide and recognition of disease with the aid of image processing and machine learning techniques were proposed. An elaborate review on deep learning techniques for leaf disease identification was investigated in [7].

Plant diseases are not only a warning to securitization of food globally, but can also prove to be involved catastrophic effects for small scale farmers whose day to day livelihoods heavily depend on healthy crops. In the developing country like India, more than 80 percent of the agricultural production is produced by small scale farmers. Hence, plant leaf disease identification in the early stage plays a main role for sustaining the productivity. In [8], multiple convolutional neural networks were applied to improve the validation accuracy involved in grape disease identification. Yet another novel classifier algorithm utilizing sine cosine algorithm based rider neural network was proposed in [9] to enhance the accuracy involved in classification. In [10], a deep convolutional neural network was applied to 14 crop species and the disease was diagnosed in a significant manner.

Owing to the high nutritional and medicinal aspect involved in apples it is said to be considered as the most productive fruit globally. Despite its familiarity and increased usage, numerous diseases are said to occur routinely on a large scale in the production of apple, thereby resulting in considerable economic losses. Hence, the timely and efficient apple leaf detection is said to be pivotal corroborating the healthy development of the apple industry and therefore has become a hot research area in the field of agricultural. A deep CNN model was proposed in [11] for faster detection of disease and also in an accurate manner. Gray level co-occurrence matrix was utilized in [12] to denote the diseased part and also classification of disease was performed in a precise manner utilizing random forest. Yet another method to minimize the convergence iterations utilizing rectified linear unit functions was proposed in [13].

With the high-speed evolution of the smart farming, plant disease identification is said to become digitalized and also data driven, therefore ensuring state-of-the-art decision support, analysis in a smart manner and ensuring prompt planning. A mathematical model of plant disease detection and recognition using deep learning was presented in [14] that in turn helped in enhancing the accuracy of detection and training efficiency. An investigation was performed in [15] by training the convolutional neural network for disease identification in plants based on segmentation, therefore ensuring timely detection.

In the recent years, there are escalating tendencies of utilizing deep learning for detecting disease present in plant leaf. However, their executions may be laborious and cumbersome process involving images with adequate resolutions. In developing countries like India however, with restricted internet connectivity, methods that would accomplish well even when data with low resolution are utilized are necessitated. A new residual network as branches to DCNN architecture was designed in [16] to minimize the issues related to gradient and therefore ensuring prompt detection. Yet another method to evaluate the detection and tracking performance utilizing Kalman filter with data association algorithm was proposed in [17].

Numerous visual computing based methods have been proposed in the recent years with the objective of predicting the plant leaf disease in an early manner. Despite early detection achieved, accuracy with which the detection was carried out is still said to be considered as a challenging issue. To address this issue, a novel hybrid method involving neuro fuzzy classifier was carried out in [18], therefore contributing to improved detection accuracy. Image processing technology was utilized in [19] for improving image recognition accuracy.

Motivated by the above research works conducted in the field of agriculture for early plant leaf detection even with the presence of noise, in this work a method called, Covariance Kalman Geometric Graph-based Bernoulli Classifier (CKGG-BC) is proposed to accurately predict the plant leaf disease. The proposed method is described in detail in the forthcoming section.

3. Methodology

In the recent years, digital signal processing methods have earned appeal for the plant leaf disease detection. These digital signal processing methods includes diverse steps such as image preprocessing, segmentation, feature extraction and classification of disease class. During the acquisition of data or plant leaf, undesired noise gets added to the original plant leaf resulting in the noisy plant leaf image collection. This noisy plant leaf image may deteriorate the leaf image analysis quality. To address this issue, Covariance Kalman Geometric Graph-based Bernoulli Classifier (CKGG-BC) method is proposed that accurately predicts the plant leaf disease even in the presence of noise in a computationally efficient manner. Figure 1 shows the block diagram of Covariance Kalman Geometric Graph-based Bernoulli Classifier (CKGG-BC) method.

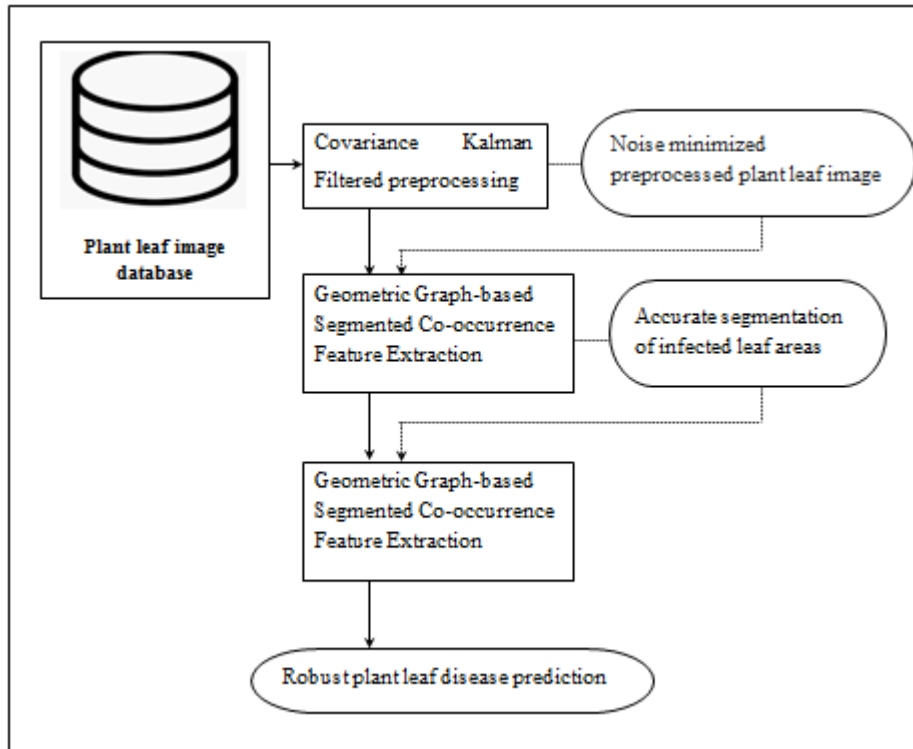


Figure 1 Covariance Kalman Geometric Graph-based Bernoulli Classifier (CKGG-BC) method

As shown in the above figure, three different steps are carried out. To start with, image pre-processing is applied as the preliminary step. This proposed Covariance Kalman Geometric Graph-based Bernoulli Classifier (CKGG-BC) method utilized Covariance Kalman Filter as noise reduction in the preprocessing step, which also tends to enhance the quality of image. Second, Geometric Graph-based Segmented Co-occurrence Feature Extraction model is applied to the preprocessed image with which accurate segmentation of infected leaf areas and geometric properties of leaf area are extracted in a computationally efficient manner. Finally, robust leaf disease prediction is made by means of Geometric Graph-based Segmented Co-occurrence Feature Extraction. The elaborate description of the proposed method is given below.

3.1 Covariance Kalman Filtered preprocessing model

The objective of image enhancement lies in enhancing the perception or viewpoint of information in leaf images for human viewers. In this paper, Covariance Kalman Filtered Preprocessing model are applied as preprocessor. Figure 2 shows the block diagram of Covariance Kalman Filtered preprocessing model.

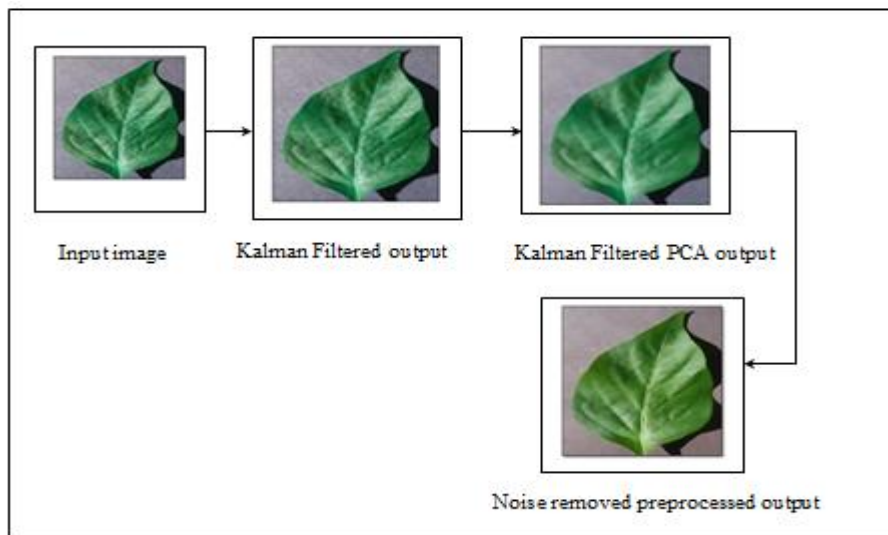


Figure 2 Block diagram of Covariance Kalman Filtered preprocessing model

As given in the above block diagram, the Covariance Kalman Filtered preprocessing model initially utilizes the state equation of the linear leaf image database to evaluate the state of the leaf image database in an optimal manner by means of the input and output of the system. As the observed leaf image database includes the effects of noise and interference, the optimal estimate is obtained by means of an advanced filtering mechanism. The Kalman filter utilizes the state equation 'S' and the observation equation 'Ob' to perform preprocessing. The recursive filtering model is utilized to evaluate the mean square error of the leaf image at the next moment, and then accurately predict the leaf disease even in the presence of noise. Let us assume the state equation 'S_t' and the observation equation 'Ob_t' of a linear leaf image system as given below.

$$S_t = STM_{t,t-1}S_{t-1} + SN_{t-1} \quad (1)$$

$$Ob_t = OM_t S_t + ON_t \quad (2)$$

From the above equations (1) and (2), 'S_t' and 'S_{t-1}' represents the state features of leaf images obtained at two different timestamps 't', 't - 1', 'STM_{t,t-1}' representing the state transition matrix, 'OM_t' denoting the observation matrix at timestamp 't' with 'SN_{t-1}' and 'ON_t' forming the state noise and observation noise respectively. Then, the state vector for the corresponding leaf image is mathematically expressed as given below.

$$SV_t = [a_{pt}, b_{pt}, a_{vt}, b_{vt}, w_t, h_t] \quad (3)$$

From the above equation (3), the state vector representation 'S_t' is obtained on the basis of the position 'p' and velocity 'v', width 'w' and height 'h' on the 'a' and 'b' axis of the leaf image. In a similar manner, the observation vector for the corresponding leaf image is mathematically expressed as given below.

$$ObV_t = [a_{pt}, b_{pt}, w_t, h_t] \quad (4)$$

Based on the observation vector, error covariance prediction equation is estimated in our work by means of analysis of principal components. The objective behind the application of analyzing the principal components for error covariance prediction is to remove highly correlated features and to enhance visualization. Let plant leaf 'PL = [PL₁, PL₂, ..., PL_n]' represent a column (i.e. pixel column) vector with a dimension of 'n', each column with mean and variance is represented as given below.

$$Mean[PL_i] = \mu_i \quad (5)$$

$$Var[PL_i] = Mean[(PL_i - \mu_i)^2] = \sigma_{ii} \quad (6)$$

Then, from the above equations (5) and (6), with state and observation vectors 'S_t' and 'Ob_t', the covariance between the two vectors is mathematically evaluated as given below.

$$COV[PL_S, PL_{Ob}] = Mean[(PL_S - \mu_i)((PL_{Ob} - \mu_j))] = \sigma_{ij} \quad (7)$$

From the above equation (7), the population (representing the plant leaf) mean vector 'μ' and population covariance matrix 'Σ' is formulated as given below.

$$\mu = Mean[PL] = \begin{Bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{Bmatrix}; \Sigma = COV[PL] = Mean[(PL_S - \mu)(PL_{Ob} - \mu)] = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix} \quad (8)$$

With the objective of designing a deterministic model so that to reduce the noise the state estimation, is updated as given below.

$$S'(t) = STM_{t,t-1}S'_{t-1} + \Sigma[Ob_t - OM_t S'_t] \quad (9)$$

The pseudo code representation of Kalman Filtered Principal Component Preprocessing is given below.

Input: Plant leaf image database, plant leaf images 'PI = pi ₁ , pi ₂ , pi ₃ , ..., pi _n '
Output: noise minimized preprocessed plant leaf image 'PPI'
<ol style="list-style-type: none"> 1: Begin 2: For each plant leaf images 'PI' 3: Estimate state equation 'S_t' and the observation equation 'Ob_t' using (1) and (2) 4: Estimate state vector 'SV_t' and the observation vector 'ObV_t' using (3) and (4) 5: Obtain principal component with mean and variance using (5) and (6)

- 6: Obtain covariance between the two vectors using (7)
- 7: Obtain overall covariance using (8)
- 8: Update state equation using (9)
- 9: **Return** error-minimized (noise-minimized) image
- 10: **End for**
- 11: **End**

Algorithm 1 Kalman Filtered Principal Component Preprocessing

As given in the above Kalman Filtered Principal Component Preprocessing algorithm, the objective remains in obtaining the noise minimized preprocessed plant leaf image by utilizing two different functions. First, a modified Kalman Filter utilizing error covariance prediction. Followed by which the Principal Components are utilized to enhance the image via preprocessing, therefore reducing the memory involved (i.e., computational overhead) in leaf disease prediction, owing to only the principal components utilized for further processing.

Geometric Graph-based Segmented Co-occurrence Feature Extraction model

Upon completion of the pre-processing, the second stage in accurate leaf disease prediction is segmentation. It is utilized in identifying the regions in the leaf image that are probably to qualify as diseased infected regions, therefore clarifying later stages. In this paper, a Geometric Graph-based Segmented Co-occurrence Feature Extraction model to segmentation is performed with the objective of identifying the regions in the leaf image infected with disease. In our work, accurate prediction of leaf disease is performed in two stages. The first stage does the task of region of interest (ROI) identification by eliminating the background using Geometric Pixel Graph-based optimized segmentation. On the other hand, the second stage performs with the diseased infected region recognition by means of Co-occurrence Feature Extraction.

Let ' $G = (V, E)$ ' with vertices ' $V \in PPI$ ', the preprocessed leaf images to be segmented, and the edges ' $(e_i, e_j) \in E$ ' representing the adjacent vertices pairs. As far as segmentation of preprocessed leaf images are concerned, the elements in ' V ', represent the pixels, whereas weight of an edge is measure of dissimilarity between any two pixels connected by that edge in terms of color and shape. To accelerate this, Geometric Pixel Graph-based optimized segmentation using feature extractors (color and spatial distribution) is implemented to produce a new region of interest image denoted by ' $SI_{a,b}^{Inf}$ ', as show in figure.

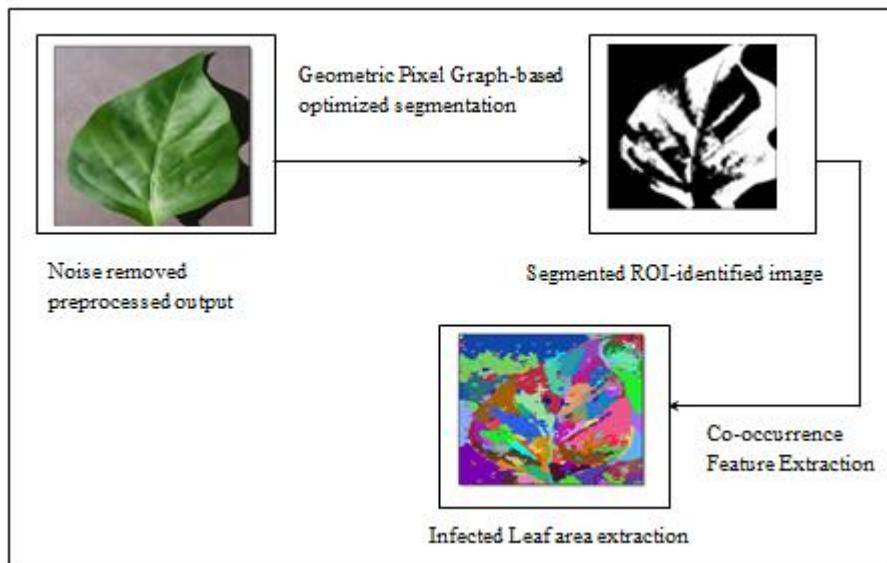


Figure 3 Block diagram of Geometric Graph-based Segmented Co-occurrence Feature Extraction model

As illustrated in the above Geometric Graph-based Segmented Co-occurrence Feature Extraction model, two different processes are carried out separately. In the Geometric Pixel Graph-based optimized segmentation, segmentation ' S ' is a separation of ' $V \in PPI$ ' into regions ' $PPI \in R_1, R_2, \dots, R_n$ ' such that each region corresponds to a connected region in a graph. The region of interest identification is made by means of Simple

Linear approach that utilizes channel information (i.e., color and spatial distribution) and is mathematically represented as given below.

$$S_c = \sqrt{(L_1 - L_2)^2 + (ca_1 - ca_2)^2 + (cb_1 - cb_2)^2} \quad (10)$$

$$S_{sd} = \sqrt{(a_1 - a_2)^2 + (b_1 - b_2)^2} \quad (11)$$

$$ROI = (S_c \cup S_{sd}) \quad (12)$$

From the above equations (10) and (11), the color channel segmented pixels for preprocessed leaf image ‘ S_c ’ is obtained on the basis of the lightness ‘ L_1, L_2 ’ and two color channels ‘ (ca_1, ca_2) ’ & ‘ (cb_1, cb_2) ’ respectively. In a similar manner, the spatial distributed segmented pixels for preprocessed leaf image ‘ S_{sd} ’ is obtained via two spatial axis ‘ (a_1, a_2) ’ and ‘ (b_1, b_2) ’ respectively. Finally, with the aid of color channel segmented pixels and spatial distributed segmented pixels, the actual region of interest is obtained from equation (12) by eliminating the background, considered to be of least significance. In the second stage, the diseased infected region recognition ‘ $SI_{a,b}^{Inf}$ ’ is identified by means of Graph-based Segmented Co-occurrence Feature Extraction. First Graph-based Segmentation is applied to the identified region of interest as given in equation (12). This is mathematically formulated as given below.

$$SI = D(S_c, S_{sd}) = \begin{cases} True, & If Diff(S_c, S_{sd}) > MID(S_c, S_{sd}) \\ False, & Otherwise \end{cases} \quad (13)$$

$$MID(S_c, S_{sd}) = MIN[Int(S_c) + \alpha(S_c), Int(S_{sd}) + \alpha(S_{sd})] \quad (14)$$

Next with the above segmented images as given in equation (13), feature extraction for extracting the geometric properties of the infected leaf areas is obtained using the co-occurrence matrix is mathematically formulated as given below.

$$SI_{a,b}^{Inf} = \begin{cases} 1, & If SI(S_c, S_{sd}) = i \text{ and } SI(S_c + \Delta S_c, S_{sd} + \Delta S_{sd}) = j \\ 0, & Otherwise \end{cases} \quad (15)$$

From the above equation (15), the offset ‘ S_c ’, ‘ S_{sd} ’ refers to the channel and spatial distribution operator that is being applied to the ‘ i, j th pixel values’ of the co-occurrence matrix. As a result of this, the geometric portions of the leaf images being infected are obtained in an accurate manner with improved true positive rate. The pseudo code representation of Geometric Graph-based Segmented Co-occurrence Feature Extraction is given below.

Input: preprocessed plant leaf image ‘ PPI ’,
Output: Accurate segmentation of infected leaf areas ‘ $SI_{a,b}^{Inf}$ ’,
<ol style="list-style-type: none"> 1: Begin 2: For each preprocessed plant leaf image ‘PPI’ 3: Estimate channel information (i.e., color and spatial distribution) using (10) and (11) 4: Identify region of interest using (12) 5: Perform graph-based segmentation based on color and spatial distribution using (13) 6: Evaluate infected leaf image using (15) 7: Return segmented infected image ($SI_{a,b}^{Inf}$) 8: End for 9: End

Algorithm 2 Geometric Graph-based Segmented Co-occurrence Feature Extraction

As given in the above Geometric Graph-based Segmented Co-occurrence Feature Extraction algorithm, the objective remains in extracting the geometric properties of the infected leaf areas via accurate segmentation of the infected leaf areas with maximum accuracy and true positive rate. To achieve this objective, first Geometric Pixel is applied to the preprocessed leaf image as input. Next, with the segmented infected leaf areas, relevant feature for leaf disease prediction are extracted by means of Graph-based Segmented Co-occurrence Feature Extraction, therefore attaining the required objective.

Bernoulli Online Multiple Kernel Learning Classifier

Finally, with the extracted geometric properties of the infected leaf areas Bernoulli Online Multiple Kernel Learning Classifier model is applied for leaf disease prediction with minimum classification error. The Bernoulli Online Multiple Kernel Learning Classifier learns a classifier for a given kernel (segmented infected image), combines classifiers by linear weights in such a manner limiting the number of segmented infected image in each single kernel classifier, therefore minimizing the classification error. Figure 4 shows the block diagram of Bernoulli Online Multiple Kernel Learning Classifier model.

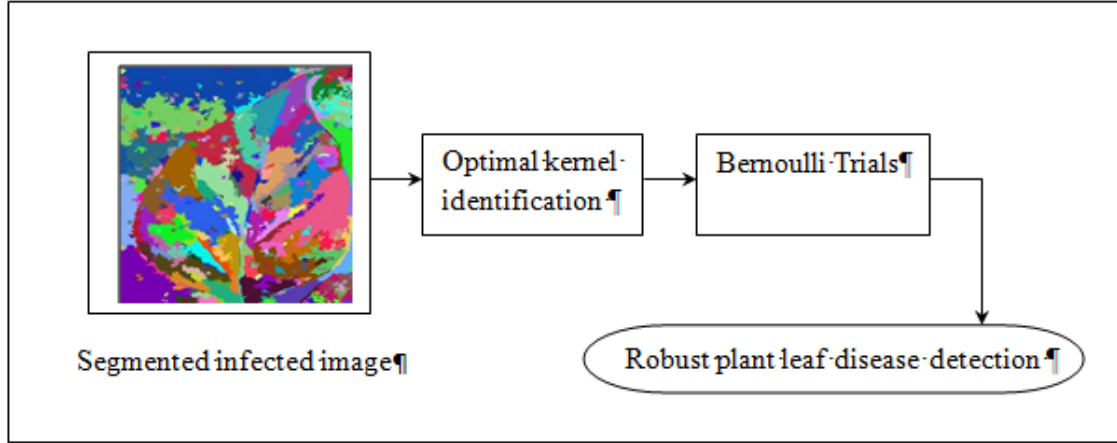


Figure 4 Block diagram of Bernoulli Online Multiple Kernel Learning Classifier

As shown in the above figure, with the segmented infected image considered as the input vector ' $p = p_1, p_2, \dots, p_n$ ' and the output vector being ' $q = q_1, q_2, \dots, q_n$ ', ' $q \in \{-1, +1\}$ ', weights ' $W = W_1, W_2, \dots, W_n$ ', ' $i = 1, 2, \dots, n$ ' and a collection of ' m ' kernel functions ' $K = (K_1, K_2, \dots, K_n)$ ' then, the objective of our work remains in learning a kernel-based prediction function by first obtaining the optimal collections of ' m ' kernels, represented by ' $\theta = (\theta_1, \theta_2, \dots, \theta_m)$ ' to minimize the classification error. This is formulated as given below.

$$\alpha = \text{Min} \frac{1}{n} \sum_{i=1}^n |p'_i - p_i| \quad (16)$$

$$p'_i = \begin{cases} -1, & \text{if } W \cdot p < Th \\ +1, & \text{if } W \cdot p \geq Th \end{cases} \quad (17)$$

To limit the number of segmented infected image to be included in each single kernel classifier, a sampling rule is proposed that determines if an incoming instance should be a segmented non-infected image by performing a Bernoulli trial as given below.

$$\text{Prob}(Z_i = 1) = \rho_i, \text{ where } \rho_i = \frac{\min(\alpha, f[p_i])}{\beta} \quad (18)$$

From the above equation (18), based on the Bernoulli trial ' $Z_i \in [0,1]$ ' indicates that a new segmented non-infected image should be added to update classifier upon ' $Z_i = 1$ ', with ' α ' and ' β ' representing the ratio of segmented infected image to total received instances respectively. Finally, the optimal margin classification error for kernel ' K_1, K_2, \dots, K_n ' with respect to collection of training samples ' $S = S_{a,b}^{Ninf}$ ', where ' $p_i \in S_{a,b}^{Ninf}$ ', is formulated as given below

$$\text{Min}[K\theta + \sum_{i=1}^m l(q_i(p_i))] \quad (19)$$

$$l(q_i(a_i)) = \max[(0,1) - q_i f(p_i)] \quad (20)$$

The pseudo code representation of Bernoulli Online Multiple Kernel Learning Classifier is given below.

Input: input vector ' $p = p_1, p_2, \dots, p_n$ ', weights ' $W = W_1, W_2, \dots, W_n$ ',
Output: robust plant leaf disease detection
1: Initialize Threshold ' Th ', kernel functions ' $K = (K_1, K_2, \dots, K_n)$ ', ' β '
2: Begin

- 3: **For** each input vector ' $p = p_1, p_2, \dots, p_n$ '
- 4: Obtain obtaining the optimal collections of ' m ' kernels using (16) and (17)
- 5: For each new segmented non-infected image
- 6: Perform Bernoulli trial using (18)
- 7: Perform classification with optimal margin classification error using (19) and (20)
- 8: **Return**(plant leaf disease detection)
- 9: **End for**
- 10: **End for**
- 11: **End**

Algorithm 3 Bernoulli Online Multiple Kernel Learning Classifier

As given in the above Bernoulli Online Multiple Kernel Learning Classification algorithm, the objective remains in predicting the leaf disease with minimum classification error. To achieve this objective, for each segmented infected image, optimal collections of ' m ' kernels are evaluated based on Multiple Kernel Learning. Next, with the purpose of reducing the classification error, instead of applying conventional Online Multiple Kernel Learning Classifier, in our work, a Bernoulli trial is applied that not only minimizes the computational overhead incurred by considering only the required segmented infected image and therefore reducing the classification error.

4. Experimental settings

Experimental evaluation of the proposed Covariance Kalman Geometric Graph-based Bernoulli Classifier (CKGG-BC) and existing methods AlexNet and GoogleNet CNN [1], INC-VGGN [2] are implemented using Python using plant village dataset <https://www.kaggle.com/emmarex/plantdisease/discussion> [20]. This dataset is used for Plant leaf Disease Detection which comprises a variety of plant's RGB plant leaf images like, pepper, potato, tomato, with different sizes. Different unique numbers of both healthy leaf images and disease affected images are obtained as input from the dataset and experimental evaluation is carried out accordingly in terms of computational overhead, accuracy, true positive rate and classification error respectively.

5. Discussion

6.1 Performance analysis of computational overhead

The first significant metric for leaf disease prediction is computational overhead. Lesser, the computational overhead, higher is the change of predicting the leaf disease and hence greater is the probability of safeguarding the plant. The computational overhead in our work is estimated as given below.

$$CO = \sum_{i=1}^n pl_i * MEM[LDP] \tag{21}$$

From the above equation (21), the computational overhead ' CO ' is measured based on the plant leaf image involved in simulation ' pl_i ' and the memory consumed in leaf disease prediction ' $MEM[LDP]$ '. It is measured in terms of kilobytes (KB). The performance of the proposed CKGG-BC method was compared with that of a previous plant leaf detection method AlexNet and GoogleNet CNN [1], INC-VGGN [2] in terms of computational overhead. The comparison is presented in table 1 which indicates that the proposed CKGG-BC outperformed the state-of-the-art methods.

Table 1 Comparative analysis of computational overhead

Number of images	Computational overhead (KB)		
	CKGG-BC	AlexNet and GoogleNet CNN	INC-VGGN
20	400	500	600
40	800	1200	2000
60	1200	2400	2800
80	2400	3600	4000

100	3000	4000	5000
120	3600	6000	7200
140	4200	7000	8400
160	4800	8000	9600
180	7200	9000	10800
200	8000	10000	12000

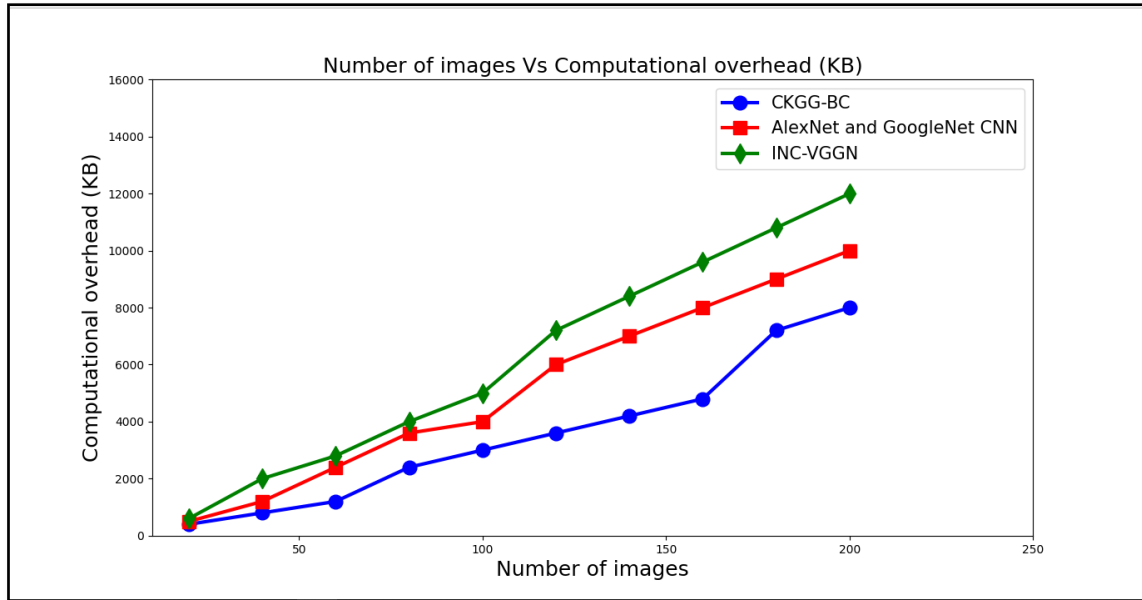


Figure 5 Graphical representation of computational overhead

Figure 5 represents the comparison graph of computational overhead for the three methods which includes the proposed Covariance Kalman Geometric Graph-based Bernoulli Classifier (CKGG-BC) method and the existing AlexNet and GoogleNet CNN [1], INC-VGGN [2] methods. When using our proposed CKGG-BC, ‘20’ plant leaf including both healthy and affected images acquired as input, the computational overhead was observed to be ‘400KB’, ‘500KB’ using [1] and ‘600KB’ using [2] respectively. Though increasing the number of plant leaf images causes an increase in the computational overhead, but with the simulation analysis results provided as above, a sharp decrease is said to be observed using CKGG-BC upon comparison with [1] and [2]. The minimization of computational overhead in CKGG-BC is due to the application of Covariance Kalman Filtered Preprocessing model. With this model, an integration of filtering and principal component analysis was taken into consideration during the preprocessing that in turn improved in the enhancement of image. Only with this enhanced image, further analysis was carried out, therefore minimizing the computational overhead using CKGG-BC by 32% compared to [1] and 45% compared to [2] respectively.

Performance analysis of accuracy

The second metric of significance for plant leaf disease prediction is the accuracy rate. Higher the accuracy, efficiency of the overall method is said to be. In other words, accuracy is measured as given below.

$$Acc = \sum_{i=1}^n \frac{Pl_{CP}}{Pl_i} * 100(22)$$

From the above equation (22), the accuracy rate ‘Acc’ is measured based on the plant leaf disease correctly predicted ‘Pl_{CP}’ to the overall plant leaf ‘Pl_i’ involved in the simulation. The accuracy is measured in terms of percentage (%). The performance of accuracy of the proposed CKGG-BC method was compared with that of previous state-of-the-art methods AlexNet and GoogleNet CNN [1], INC-VGGN [2]. The comparison is presented

in table 2 which indicates that the proposed CKGG-BC outperformed the state-of-the-art methods in terms of accuracy.

Table 2 Comparative analysis of accuracy

Number of images	Accuracy (%)		
	CKGG-BC	AlexNet and GoogleNet CNN	INC-VGGN
20	85	80	75
40	84.35	83.65	82.15
60	84	83	82
80	83.75	81.45	80.15
100	83.55	81	79.35
120	83.25	80	78
140	83	79.55	76
160	82.85	78	75.15
180	82.65	77.35	73
200	82	75	72.15

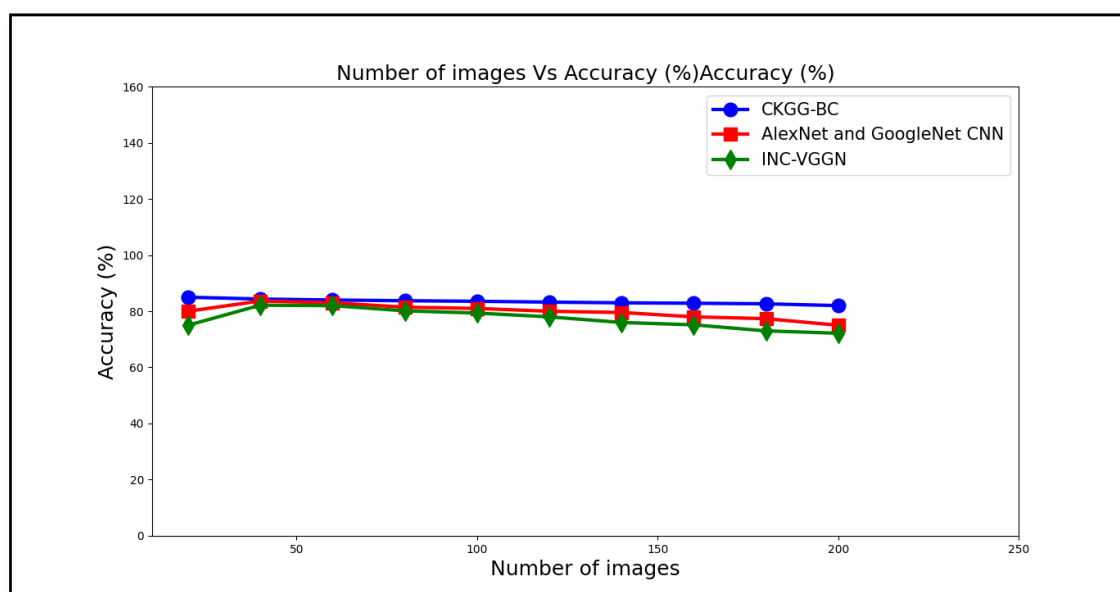


Figure 6 Graphical representation of accuracy

Figure 6 given above presents the plant leaf disease prediction accuracies for plant leaf images involving both health and infected images. In Figure 6, one can see that prediction accuracy can be improved if images are transformed using CKGG-BC method. Compared to the state-of-the-art methods AlexNet and GoogleNet CNN [1], INC-VGGN [2] tested in this experiment, CKGG-BC method performs better and improves accuracy to a greater extent than the other state-of-the-art methods, [1] and [2]. Though a smart accuracy deviation was said to be observed, simulation analysis ‘20’ numbers of healthy and infected plant leaf image provided as input, ‘85%’ of accuracy was found using CKGG-BC method, ‘80%’ using [1] and ‘75%’ using [2] respectively. The reason behind the accuracy improvement using CKGG-BC method was due to the application of Geometric Graph-based Segmented Co-occurrence Feature Extraction model. With this model, initially geometric features were utilized for segmenting the infected leaf areas and then Co-occurrence Feature Extraction was performed for extracting geometric properties of infected leaf areas. This in turn improved the accuracy rate using CKGG-BC method by 5% compared to [1] and 8% compared to [2] respectively.

Performance analysis of true positive rate and classification error

Finally, the true positive rate and classification error is analyzed in this section. True Positive rate refers to the ratio of positives (healthy plant leaf as identified with healthy and infected plant leaf identified as infected) that are correctly identified. It is estimated as given below

$$TPR = \frac{TP}{PL_{inf}} * 100(23)$$

From the above equation (23), the true positive rate 'TPR' is measured based on the number of true positives 'TP' and the total number of infected plant leaf 'PL_{inf}' considered for simulation. It is measured in terms of percentage. Finally, the classification error is measured as given below.

$$CE = \frac{S_{IC}}{S} * 100(24)$$

From the above equation (24), the classification error 'CE', is measured based on the samples incorrectly classified 'S_{IC}' to the total number of samples considered 'S' for simulation. It is expressed in terms of percentage (%). Table 3 presents the training performance of theCKGG-BC, AlexNet and GoogleNet CNN [1], INC-VGGN [2] with true positive rate and classification error details.

Table 3 Comparative analysis of true positive rate and classification error

Methods	True positive rate (%)	Classification error (%)
CKGG-BC	83.25	3.45
AlexNet and GoogleNet CNN	78.15	6
INC-VGGN	73.25	8.15

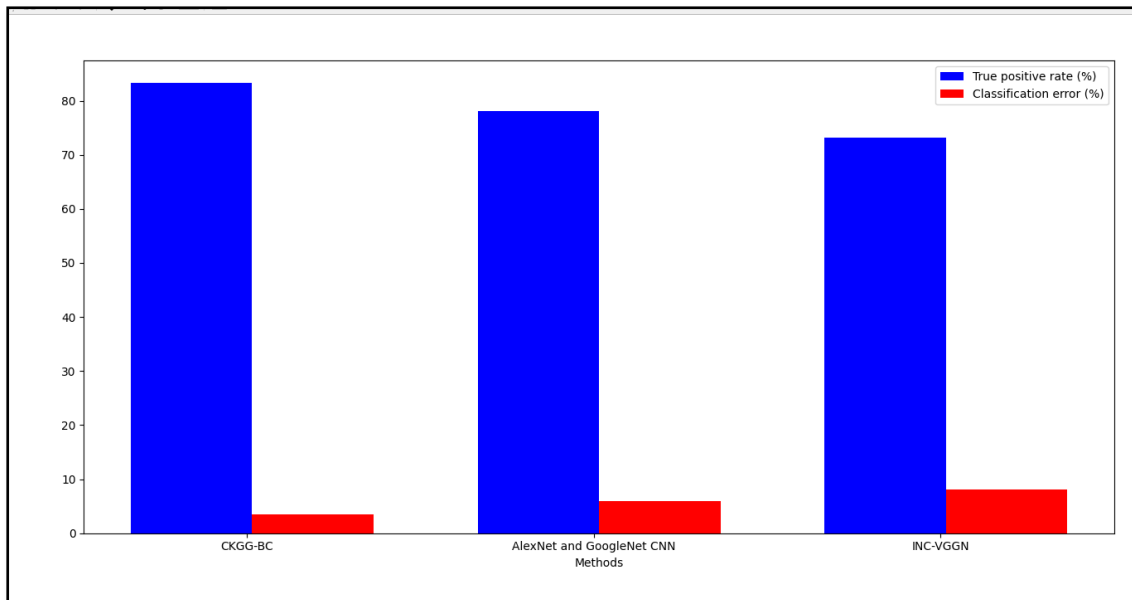


Figure 7 Graphical representations of true positive rate and classification error

Finally, figure 7 given above show the graphical representations of the true positive rate and classification error for three different methods, CKGG-BC, AlexNet and GoogleNet CNN [1], INC-VGGN [2] respectively. First, as far as true positive rate is concerned, higher amount healthy plant leaf were identified as healthy and infected plant leaf were identified as infected using the proposed CKGG-BC method upon comparison with [1] and [2]. The reason behind the improvement was due to the application of Geometric Graph-based Segmented Co-occurrence Feature Extraction algorithm. By applying this algorithm, first, the geometric properties of the infected leaf areas were significantly extracted via accurate segmentation of the infected leaf areas. Next, the Geometric Pixel were utilized and applied to the preprocessed leaf image to identify the segmented infected leaf areas. Finally, relevant feature were extracted for leaf disease prediction using Graph-based Segmented Co-occurrence Feature Extraction, therefore improving the true positive rate using CKGG-BC by 7% compared to [1] and [2].

Next, the classification error using CKGG-BC method was found to be comparatively lesser than [1] and [2]. The reason behind the reduction in classification error was due to the application of Bernoulli Online Multiple Kernel Learning Classification algorithm. By applying this algorithm, optimal collections of kernels were estimated on the basis of multiple kernel learning for only the segmented infected image. Next, instead of using the traditional Online Multiple Kernel Learning Classifier, in our work, a Bernoulli trial is applied that reduces the classification error using CKGG-BC method by 43% compared to [1] and 26% compared to [2] respectively.

6. Conclusion

In this paper, we proposed a novel Covariance Kalman Geometric Graph-based Bernoulli Classifier (CKGG-BC) method for plant leaf disease prediction. In pre-processing, to enhance the image and improve the quality of image for further analysis, principal components of filtered images are utilized and based on the principal components, advance filtering is performed. A Geometric Graph-based Segmented Co-occurrence Feature Extraction method is used for the segmentation of diseased portion and extracting the geometric properties accordingly. Finally, Bernoulli Online Multiple Kernel Learning Classifier is applied for timely and precise leaf disease prediction with minimum error even in the presence of noise. According to our experimental results, CKGG-BC outperforms state-of-the-art methods in terms of classification performance. Based on Geometric Graph-based Segmented Co-occurrence Feature Extraction and Bernoulli Online Multiple Kernel Learning Classifier, CKGG-BC can not only enhance true positive rate and accuracy, but also reduce the computational overhead and classification error, making it very practical in the field of agriculture applications.

References

- Sachin B. Jadhav, Vishwanath R. Udipi, Sanjay B. Patil, "Identification of plant diseases using convolutional neural networks", *International Journal of Information Technology*, Springer, Feb 2020 [AlexNet and GoogleNet CNN]
- Junde Chena, Jinxiu Chena, Defu Zhanga, Yuandong Sunb, Y.A. Nanehkarana, "Using deep transfer learning for image-based plant disease identification", *Computers and Electronics in Agriculture*, Elsevier, Mar 2020 [INC-VGGN] – Inception-Visual Geometry Group Network
- Mingqiang Ji, Yu Yang, Yang Zheng, Qibing Zhu, Min Huang, YaGuo, "In-field automatic detection of maize tassels using computer vision", *Information Processing in Agriculture*, Elsevier, Mar 2020
- Vijai Singh, "Sunflower leaf diseases detection using image segmentation based on particle swarm optimization", *Artificial Intelligence in Agriculture*, Elsevier, Jul 2019
- Kamlesh Golhani, Siva K. Balasundrama, Ganesan Vadamalai, Biswajeet Pradhan, "A review of neural networks in plant disease detection using hyperspectral data", *Information Processing in Agriculture*, Elsevier, May 2018
- Lawrence C. Ngugi, Moataz Abelwaha, Mohammed Abo-Zahha, "Recent advances in image processing techniques for automated leaf pest and disease recognition – A review", *Information Processing in Agriculture*, Elsevier, Apr 2020
- M. Nagaraju, Priyanka Chawla, "Systematic review of deep learning techniques in plant disease detection", *International Journal of System Assurance Engineering and Management*, Springer, May 2020
- Miaomiao Ji, Lei Zhang, Qiufeng Wu, "Automatic grape leaf diseases identification via United Model based on multiple convolutional neural networks", *Information Processing in Agriculture*, Elsevier, Oct 2019
- Monalisa Mishra, Prasenjit Choudhury, Bibudhendu Pati, "Modified ride-NN optimizer for the IoT based plant disease detection", *Journal of Ambient Intelligence and Humanized Computing*, Springer Nature, Apr 2020
- Sharada P. Mohanty, David P. Hughes, Marcel Salathé, "Using Deep Learning for Image-Based Plant Disease Detection", *Frontiers in Plant Science*, Sep 2016
- Peng Jiangi, Yuehan Chen, Bin Liu, Dongjian He, Chunquan Liang, "Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks", *IEEE Access*, IEEE. Translations and content mining, Jul 2019
- Saiqa Khan, Meera Narvekar, "Novel fusion of color balancing and superpixel based approach for detection of tomato plant diseases in natural complex environment", *Journal of King Saud University – Computer and Information Sciences*, Elsevier, Jul 2020
- Xihai Zhang, Yue Qiao, Fanfeng Meng, Chengguo Fan, Mingming Zhang, "Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks", *IEEE Access*, May 2018
- Yan Guo, Jin Zhang, Chengxin Yin, Xiaonan Hu, Yu Zou, Zhipeng Xue, Wei Wang, "Plant Disease Identification Based on Deep Learning Algorithm in Smart Farming", *Discrete Dynamics in Nature and Society*, Hindawi, Aug 2020
- Parul Sharma, Yash Paul Singh Berwal, Wiqas Ghai, "Performance analysis of deep learning CNN models for disease detection in plants using image segmentation", *Information Processing in Agriculture*, Elsevier, Nov 2019

- Hilman F. Pardede, EndangSuryawati, Vicky Zilvan, Ade Ramdan, R. Budiarianto S. Kusumo, Ana Heryana, R. Sandra Yuwana, DikdikKrisnandi, AgusSubekti, FaniFauziah, Vitria P. Rahadi, "Plant diseases detection with low resolution data using nested skip connections", *Journal of Big Data*, Springer, Jan 2020
- EsmaelHamuda, Brian Mc Ginley, Martin Glavin, Edward Jones, "Improved image processing-based crop detection using Kalman filtering and the Hungarian algorithm", *Computers and Electronics in Agriculture*, Elsevier, Feb 2018
- Anusha Rao and S.B. Kulkarni, "A Hybrid Approach for Plant Leaf Disease Detection and Classification Using Digital Image Processing Methods", *International Journal of Electrical Engineering & Education*, Aug 2020
- Feng Qin, Dongxia Liu, Bingda Sun, Liu Ruan, Zhanhong Ma, Haiguang Wang, "Identification of Alfalfa Leaf Diseases Using Image Recognition Technology", *PLOS ONE* | DOI:10.1371/journal.pone.0168274 December 15, 2016
<https://www.kaggle.com/emmarex/plantdisease/discussion>