International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V71I2P242

Volume 71 Issue 2, 408-420, February 2023

Original Article

Hybrid Features based Classification of Insect and Leaf Disease of Soybean Plants using Random Forest Classifier

Chandrakant Mahobiya[1], Sailesh Iyer[2], Savita Kolhe[3]

[1,2] Department of Computer Science & Engineering, School of Engineering and Applied Sciences, Rai University, Ahmedabad, Gujarat, India
[3] ICAR, Indian Institute of Soybean and Research, Indore, MP, India

[1]Corresponding Author : ckmahobiya26486@gmail.com

Received: 23 September 2022 Revised: 07 February 2023 Accepted: 18 February 2023 Published: 25 February 2023

Abstract - *With the growth in the global population, agricultural productivity must expand. Since insects (pests) and crop diseases are among the difficulties farmers encounter, they can cause significant agricultural loss. It is critical to creating solutions that reduce losses in order to boost production. Some of these technologies are environmentally friendly, such as those designed for the automated and early diagnosis of diseases using image processing techniques in conjunction with deep learning computational algorithms. In India, more than 40 species of insect pests of this crop were registered, one of the most relevant being Blue beetle adult/ mrl Larvae/ mrl G.gemma Larvae/mrl, .acuta, Heliothis, Grey weevil adult/ mrl, Stem fly incidence % Plant inf./ mrl Girdle, beetle. The primary goal of this research is to explore the accuracy and efficiency of computational approaches used in the problem of soybean leaf disease detection and insect classification, which are implemented with the utilization of hybrid features. Soybean insect and leaf diseases automatic classification and the prediction model are presented with hybrid features created by extracting deep features from a convolutional neural network (CNN) and texture features (Acquired from Gabor Wavelet and Harris Corner Method). The proposed hybrid features are then classified by a random forest classifier. MATLAB-based simulations exhibit the performance for insects and disease detection and classification.*

Keywords - *Convolutional Neural Network, Deep learning, Gabor wavelet, Harris corner, Random forest classifier, Soybean.*

1. Introduction

Smart agriculture, also known as precision agriculture, can help address these challenges by providing farmers with real-time information about their crops and soil. By using sensors, drones, and other technologies, farmers can monitor crop growth and soil conditions, detect pest and disease outbreaks, and make informed decisions about when to apply fertilizers and pesticides. Smart agriculture can also help farmers make more sustainable decisions. For example, by using precision irrigation systems, farmers can reduce water waste and conserve resources. Using sensors to monitor soil health and nutrient levels can avoid over-fertilization and reduce the risk of soil degradation and water pollution. In addition, smart agriculture can help improve food safety by providing farmers with the tools they need to detect and respond to foodborne illness outbreaks quickly. For example, by using sensors to monitor the temperature and humidity of food storage facilities, farmers can ensure that their food remains at safe temperatures and reduce the risk of spoilage and contamination. Overall, smart agriculture has the potential to revolutionize the way we produce food and address some of the biggest challenges facing the world's food system. By providing farmers with the information and tools they need to make informed decisions, smart agriculture can help ensure that we have enough food to feed a growing global population and preserve natural resources for future generations.

Indeed, the authors of [1] and [2] are correct in highlighting the importance of smart agriculture in addressing these critical issues in agriculture. In particular, the use of technology and data-driven approaches can greatly improve decision-making and help farmers increase productivity while reducing their impact on the environment and ensuring food safety. Indeed, while pesticides can effectively control pests and diseases, they can also negatively impact the environment and human health. Pesticides can enter the food chain and contaminate the food we eat, leading to potential health risks. They can also harm beneficial insects such as bees and butterflies, which play a critical role in pollination and maintaining biodiversity [3].

In addition, pesticides can contribute to soil degradation, water pollution, and air pollution, further harming the environment and compromising the health of both people and wildlife. The use of pesticides can also lead to the development of pesticide-resistant pests, making it more difficult and expensive to control future outbreaks [2].
With the advancement of technology in the field, vast data and information collected on field conditions can reduce the usage of pesticides in the agricultural environment; this integration is known as precision agriculture. Precision farming benefits both the producer and the environment by allowing farmers to apply pesticides at the optimal location and time. However, when done manually and without the assistance of smart agricultural technology, the effort of checking and assessing the status of plants in a culture becomes difficult, resulting in additional labour and adding to uncertainty in decision-making [4].

Early disease diagnosis is crucial for farmers, as it can help prevent crop losses and reduce the need for pesticides. By using machine learning algorithms to analyze crop images, researchers can identify disease symptoms, such as discoloration or wilting, and provide farmers with the information they need to make informed decisions about when and how to treat their crops [5].

In precision agriculture, there is still room for improvement in classifying insect and leaf diseases in soybean plants. One of the main research gaps in this area is the lack of large, high-quality datasets for training and evaluating machine learning models. In order to accurately diagnose insect and leaf diseases in soybean plants, researchers need access to a large number of images of healthy and diseased crops, along with corresponding annotations and labels. However, collecting such data can be challenging, requiring significant resources and expertise. Another research gap is the limited ability of current machine learning models to handle variability in the images of crops effectively. For example, images of crops can vary widely in terms of lighting conditions, angle, and resolution, which can affect the performance of machine learning algorithms. As a result, researchers need to develop more robust models that can handle this variability and accurately classify insect and leaf diseases in soybean plants.

There is a need for better methods for evaluating the performance of machine learning models for diagnosing insect and leaf diseases in soybean plants. It includes developing better metrics for assessing model accuracy and developing new methods for evaluating the robustness of models in the face of variability in image data.

While machine learning techniques, such as Convolutional Neural Networks (CNNs), have advanced the ability to diagnose plant diseases accurately, there are still some challenges that need to be addressed. By using these techniques, researchers can develop computer vision systems that can accurately detect signs of disease in crops [6] [7] [8]. Current technologies applied to precision agriculture are capable of helping to identify problems in farming, such as soybean leaf disease, a problem that is the motivation of this research. Using these technologies can lead to overcoming many challenges in agriculture, especially about pathologies that can affect plantations. In this context, the classification of patterns contained in images of plant foliage has proved to be a very useful alternative, in addition to being cheap, in the automatic detection and recognition of the main diseases and pests that affect a considerable range of agricultural products [9] [10] [11] [12] [13].

Images, such as those shown in Figure 1, can be used in the construction of a data classifier [14], performed as follows: (i) initially, a classification model (classifier) is induced from a set of training data (labelled data), in which each object (e.g., the image of a leaf) is labelled according to the class to which it belongs (e.g., “healthy_plant”, “sick_plant”); (ii) subsequently, the obtained classifier can then be used to infer the class of new unlabelled (and unobserved during training) objects. Algorithms that generate these classifiers can be implemented computationally to automatically recognise different pathological agents that attack the most diverse cultures, such as soybeans. The main problems faced by soybean producers are proportional to the areas planted and their exports, mainly due to production losses resulting from diseases caused by fungi, bacteria and viruses, as well as those caused by environmental factors and misuse of chemical products [15].

1.1. Insects of Soybean

In India, more than 40 species of insect pests of soybean crop were registered, one of the most relevant being Blue beetle adult/ mrl Larvae/ mrl G.gemma Larvae/mrl, .acuta, Heliothis, Grey weevil adult/ mrl, Stem fly incidence % Plant inf./ mrl Girdle, beetle.

Fig. 1 Example of different patterns found in soybean leaves: (a), (b) and (c) refer, respectively, to a healthy plant, a plant affected by Red Root Rot and a plant affected by powdery mildew

1.2. Soybean Crop Diseases

Soybean is subject to a large number of pathologies of economic importance that affect mainly its leaves, with their frequency and intensity varying according to the producing region [16]. According to the authors of [17], among the main diseases that affect the crop, we can mention Asian rust (Phakopsora pachyrhizi Syd. & P. Syd.), target spot (Corynespora cassiicola Berk. & M.A. Curtis), anthracnose [Colletotrichum dematium var. truncata (Schwein.) Arx], septoria or brown spot (Septoria glycines Hemmi), brown eye spot (Cercospora sojina Hara), downy mildew [Peronospora manshurica (Naumov) Syd.], powdery mildew [Erysiphe diffuse (Cooke & Peck) U. Braun & S Takam.], white mold [Sclerotinia sclerotiorum (Lib.) of Bary], bacterial blight (Pseudomonas syringae pv. glycinea) and bacterial pustule (Xanthomonas axonopodis pv. glycines).

Despite being a culture studied and cultivated intensively, soybean still suffers from some obstacles in phytosanitary management. Despite Asian rust being the main disease of soybean, the complex of diseases at the end of the cycle has been causing concern to farmers due to the higher incidence and severity that has been observed in the most recently released varieties in India. The pathogens involved in seedling disease complex (SDC) settle in the early stages of development. Due to the long latency period, the symptoms will only be visible in the later stages of the culture. SDC diseases such as brown eye spot, brown spot and target spot reduce photosynthetic efficiency, impairing grain filling and reducing productivity [18].

Knowledge of the region (history of diseases), the characteristics of the cultivar planted and monitoring of the crop (assessment of climatic conditions and stage of the crop) are essential for decision making, that is, knowing which product to apply and when mainly in order to avoid unnecessary applications if there are no ideal conditions to start the epidemiology of diseases of economic importance [19]. According to the authors of [20], grain losses have been reduced in recent years thanks to the efficient control performed with fungicides. Among the various diseases that affect cultivars, for the purposes of this article, those associated with Red Root Rot (RRR) or sudden death syndrome and powdery mildew stand out. In Indian lands, the RRR is caused by one of three fungi, namely: Fusarium brasiliense, F. tucumaniae and F. crassistipitatum. As its name implies, it starts at the plant's root with a simple reddish spot located a few centimetres below ground level.

RRR can cause a loss of 20 to 80% of production depending on some key factors, such as the stage of development of the culture at the time of infection and the way of cultivating this legume. As the infection progresses, the spot –significantly small – expands, encircling the root, changing its color from purplish-red to reddish-brown, then turning black. This whole process causes the leaves to acquire a precocious yellowish color.

The following cultivation conditions are considered optimal for the proliferation of patches in RRR: poorly compacted soils with inefficient drainage systems and temperatures between 22 and 24◦C [21] [22]. Powdery mildew, in turn, is characterized by showing a whitish color in its initial phase in its foliage, gradually covering the entire leaf surface. Over time, this color changes to grayish-brown, giving the plant a dirty appearance [42]. The most common condition for the proliferation of disease is centered on mild temperatures (18 to 24◦C) at the beginning of flowering.

This paper is focused on texture and deep features extraction of soybean leaf dataset images. It is implemented in the image processing toolbox of MATLAB 2020a, aiming to evaluate the impacts of the final classification results for soybean insect classification and soybean leaf disease classification using a random forest classifier. This study will be able to guide the development of the automatic diagnosis of diseases that may be affecting crops, especially soybean. The novelty of a hybrid feature-based classification approach for insect and leaf disease detection in soybean plants using a random forest classifier lies in its combination of multiple features derived from both visible light and near-infrared spectral imaging.

This approach integrates texture, color, shape, and spectral features to improve the detection and classification accuracy of soybean plant diseases caused by insects and pathogens. Using a random forest classifier enables efficient and effective classification of the features to accurately identify and distinguish between different types of insect and leaf diseases affecting soybean plants. Overall, this hybrid feature-based approach has the potential to significantly improve the detection and management of soybean plant diseases, thereby enhancing crop yield and quality. Section two provides a literature review in the field of plant leaf disease detection. The proposed methodology of this research is described in section three. Section four presents the results achieved with this research, and finally, section five describes the authors' conclusions regarding the future direction of this study.

2. Literature Review

According to the authors of [24], the causes of a significant reduction in the quality and quantity of world agricultural production stem from the diseases occurring in these cultivars' plants. Among the procedures most used in identifying any pest or disease in the plantation, the traditional method of observation stands out, which consists - as its name suggests - of visual analysis of the disease, mainly with regard to the color changes of the foliage of the plants. It is a laborious, inefficient and difficult procedure for plantations of large proportions, with relatively low accuracy and which, above all, requires a qualified and well-trained professional to perform such a function [25].

An alternative (or solution) to this problem is the automatic detection and recognition of the main diseases and pests that affect agricultural production, based, for example, on Machine Learning (ML) [17] [26] [27]. In this context, the features extracted from images of leaves/foliage (attributes) can provide significant clues for identifying and treating diseases in their various stages. ML algorithms can act together, recognizing and providing a diagnosis [25].

Each leaf carries substantial information about the plant of which it is a constituent, and, as a result, any problem or anomaly can be revealed by certain characterizations in them [28]. In [29], the authors propose five groups of descriptive parameters for the automatic analysis of leaves, namely: diameter; physiological length; physiological width; area and perimeter of the sheet. In [43], textures, colors, shapes and combinations of these characteristics are analyzed.
Regarding the detection of diseases by systems that automatically inspect plant leaves, such as the one being developed in the present research, the following procedure can be used [30,31]:

- Acquisition of images of the leaves through a digital camera or utilize a benchmark dataset;
- Pre-processing of the images obtained (noise removal and other adjustments).
- Image segmentation (e.g., with the removal of the “background”);
- Extraction of features/attributes (e.g., using image descriptors);
- Classification using a machine learning algorithm.

An algorithm that uses image processing techniques to detect diseases evidenced by leaf spots is implemented in [32]. In this work, a set of leaf images was created. These images, conditioned to the traditional color system (RGB), were transformed into YCbCr, HIS and CIE-LAB color spaces after being submitted to a filter to remove spots. The component "A" (which describes the variation from green to red color) was then extracted from the CIE-LAB; the "H" component (which describes a pure color and is usually related to the wavelength of light) of the HSI; and the "Cr"(color) component of the YCbCr color space, seeking to detect disease spots on the leaves. The segmented images of the disease spot, obtained by all three methods, were compared to find the best for disease detection.

In [29], the authors apply Neural Networks with image processing techniques to obtain a plant classification system by analysing their foliage. Schematically, the entire procedure is performed as follows: i) digital capture of the leaf image; ii) image processing; iii) feature extraction; iv) analysis of the main components extracted; v) network training; vi) test of the trained network; vii) comparison of the results obtained.

Detecting diseases in plants through analysing images obtained from their leaves was also one of the objectives exposed in [33]. The methodology is similar to that performed in (WU et al., 2007), that is: i) image acquisition; ii) pre-processing; iii) feature extraction; iv) classification and diagnosis. Neural Networks are also used as a classification tool. Likewise, in [34], the adopted methodology involves the following steps: i) image acquisition; ii) pre-processing; iii) image segmentation using the K-Means grouper; iv) extraction of characteristics through the Gray Level Co-occurrence Matrix (GLCM); v) classification with Support Vector Machines (SVM).

After a general analysis of articles selected for the development of the study, the choice of technique and culture to be used in this study was made. After choosing soybean as the crop to be studied and CNNs as the standard technique for the application of the study, a separate survey was carried out containing two studies that also used techniques to detect diseases in soybean leaves.

The two studies that used the soybean crop were studies presented by the authors of [35] and [36]. Both studies chose to detect soybean leaf diseases, using an image bank to carry out the experiments. The authors of [35] used a third-party image bank, which is called PlantVillage, while the authors of [36] made use of their own image bank, using a digital camera to acquire images in a real scenario. Regarding the number of images, the authors of [35] used 4775 images for training and testing the algorithms, while the authors of [36] used 65,760 images for training, testing and validating the algorithms created. This shows the variety in the number of images used by the studies.

Regarding the techniques used for detecting diseases, the authors of [35] used the MATLAB programming language as a standard technique, together with the k-means algorithm. The authors of [36], on the other hand, opted for the use of deep convolutional neural networks as a standard technique for detecting diseases in soybean leaves. The results obtained by the studies are different; while the authors of [35] reached an accuracy equal to 85.65% of accuracy, the authors of [36] obtained a result equal to 94.13% of accuracy. With the results obtained by both studies, it is correct to say that the authors of [36] achieved better results with the training of its algorithms when compared to the authors of [35].

There are many variables between the two studies, such as the number of images, type of image used, techniques used to detect diseases, pre-processing and division of the set of images for training and tests. Taking into account that the authors of [36] used their own images and taken from a real scenario, the positive result becomes even more important, as well as the creation of a group of images with a high number of images, which favours the training of the algorithms created.

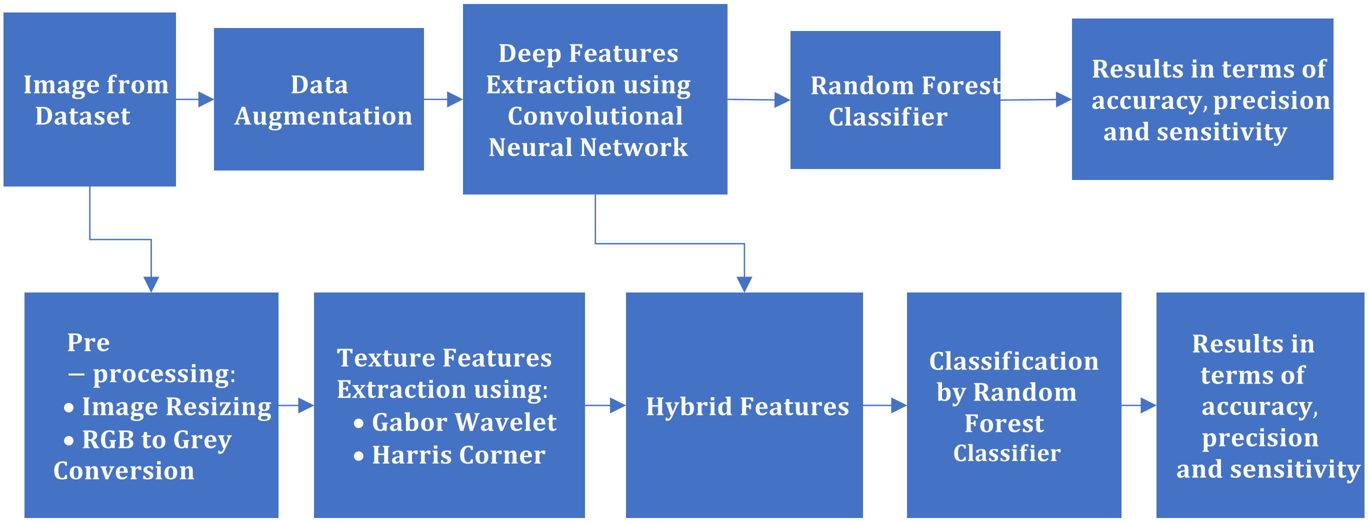

Fig. 2 A hybrid features method based on deep learning for identifying and classifying insects and leaf diseases in soybean plants

Automatic and early detection of diseases has evolved over the years and will one day increasingly help farmers' day-to-day lives. Below we point out why they are not yet a reality observed in farmers' toolkits. In this literature review, we show a variety of studies in the area adopting different ways of detecting plant diseases.

In terms of a repeatable research protocol, we found that the lack of detail makes it difficult to carry out an adequate scientific method. We did not find a study with a follow-up of the techniques responding to the proposed approaches. Likewise, another reason why these technologies are not expressly reported as being adopted in practice is also that most studies have difficulties that make it difficult or prevent the replication of their work by other researchers. This is important to obtain feasibility for adoption.

Another negative point that makes implementing an ideal scientific method difficult is the unavailability of the image database used by the studies. In order to provide an unbiased empirical knowledge base, data sets must be shared across the scientific community. However, the lack of studies considering real scenarios makes its adoption in practice very difficult, embedded with low reliability.

To overcome these issues, future studies in the area should adopt methods of comparison with other similar proposals for disease detection. Thus, this review of the literature provides great value for future research, as it: provides the literature in the area with methods of comparison; enhances state-of-the art with a map for future analysis of related work; it serves as a guide for selecting the studies that presented the best performance in the detection of the disease, being also useful for decision making. The procedures and techniques adopted for this work are explained below.

3. Proposed Methodology

This section presents the methodology and research protocol used in the proposed approach. The main focus of this study is to develop a CNN-based approach for soybean leaf disease detection and classification utilizing a random forest classifier. Along with the use of the CNN, the research varies in terms of data augmentation techniques and feature extraction methods, as well as the algorithms used during the training of the CNNs.

This work proposes a new hybrid architecture to make the best use of deep learning, Gabor wavelet, and Harris corner features as a set of hybrid features that enable the classifier to achieve the best optimal accuracy. Soybean insect and leaf diseases automatic classification and prediction model with hybrid features created by extracting deep features from a convolutional neural network (CNN) and texture features.

The proposed hybrid features are classified by a random forest classifier. The suggested method's block diagram is shown in Figure 2, and the technique is described in the following subsections.

3.1. Dataset

Weekly Blue beetle adult/mrl larval populations G.gemma larvae/mrl Grey weevil adult/mrl, Larvae/mrl, acuta, Heliothis, Stem Fly Incidence Percentage Plant Inf./mrl In order to determine the impact that climatic factors had on the incidence of this insect on soybeans, the girdle beetle incidence data gathered from the ICAR, Indian Institute of Soybean Research Indore under Crop Pest Surveillance from 2009 to 2018 was reviewed. Larvae were seen throughout the soybean growing season, with the greatest abundance between the 1st and 3rd weeks of August.

3.2. Data Augmentation

This implies that the model must have been able to understand the key characteristics of a data set during training. To do this, it is essential that:

- Data Space: The learning data space includes all potential outcomes and the widest variety of examples pertinent to the setting where the model will be utilized.
- Features Space: The training data's feature space includes the whole range of potential outcomes or the most representations of each feature that might be used.

Thus, to satisfy the first criterion, we need to collect the most diverse set of training images pertinent to the usage environment and the objective of the proposed model.

To meet the second condition, we must use data augmentation techniques for the training images we have accessible, the most well-liked of which are affine transformations (vertical and/or horizontal flip, rotation). Additionally, there are non-affine modifications, including, for instance, wrap (perspective), variations in brightness and contrast, scaling, random crop (an arbitrary portion of an image), cut-out (squares random blacks), or jitter (random noise).

3.3. Pre-Processing

By downsizing the input image to 300×450 pixels, the pre-processing is accomplished. This image will be transformed from RGB to Grey format to achieve texture and deep feature extraction.

3.4. Texture Features Extraction

The texture, which consists of a collection of visual statistical primitives structured in accordance with certain placement guidelines, enables the solution of the issue presented when the colour distributions are closely spaced. For the extraction of texture features in this research, Gabor wavelet and Harris corner techniques are used:

3.4.1. Gabor Wavelet

Texture feature descriptors are extracted by convolving the image with a Gabor Wavelets filter bank. Following this procedure, each image pixel is associated with a feature vector.

A Gaussian function modulated by a sine wave produces a 2D Gabor filter in the spatial domain. Equation (1) describes this filter's mathematical expression:

$$g(x,y) = e^{-\left[\frac{(x-x_0)^2}{2\sigma_x^2}+\frac{(y-y_0)^2}{2\sigma_y^2}\right]} e^{-ik(x-x_0)} \tag{1}$$

The spatial frequency of a wave in the complex plane with the wave normal along the x-axis is given by the expression $k = \frac{2\pi}{\lambda}$. And (x_0, y_0) symbolizes the centre of the Gaussian wave, σ_x and σ_y are the variances of the Gaussian along the x and y axes, respectively, and λ symbolizes the wavelength. It employs Gabor Wavelets filters, which are comparable in function to themselves. If the function $g(x,y)$ is considered as the Gabor Wavelets matrix, then a bank of filters similar to themselves can be created by scaling and rotating the function $g(x,y)$ through the equations (2), (3) and (4):

$$g_{mm} = g(x', y') \tag{2}$$

$$x' = \alpha^{-m}(x\cos\Theta_n + y\sin\Theta_n) \tag{3}$$

$$y' = \alpha^{-m}(-x\sin\Theta_n + y\cos\Theta_n) \tag{4}$$

Gabor Wavelets filters are used in a design with seven scales and five distinct orientations. The filter settings are chosen so that they will overlap by 50% at their highest magnitudes in the frequency spectrum. Equations (5), (6), and (7) provide the expressions needed to keep these conditions in place.:

$$\alpha = \left[\frac{U_h}{U_l}\right]^{\frac{1}{M-1}} \tag{5}$$

$$\sigma_x = \frac{(\alpha+1)\sqrt{2\ln(2)}}{2\pi(\alpha-1)U_h} \tag{6}$$

$$\sigma_y = \frac{\sqrt{2\ln(2)-\left(\frac{\ln(2)}{\pi\sigma_x U_h}\right)^2}}{2\pi\tan\left(\frac{\pi}{2N}\right)\left\{U_h - 2\ln\left(\frac{1}{2\pi^2\sigma_x^2 U_h}\right)\right\}} \tag{7}$$

U_h and U_l stand for the highest and lowest frequencies of interest, respectively, where α is a filter scale factor. N is the number of orientations, while M is the number of scales.
The filter response's statistical characteristics are used to produce a useful texture description. This is done by calculating the image's mean and non-normalized standard deviation.

The goal is to split the image into groups of overlapping, mesh-cantered rectangular blocks. A texture vector is generated for each block and assigned to the relevant point in the model. The mesh's resolution is the same as the image's resolution in pixels.

The convolution of a coral image with the filter mask is analogous to the mean value across a tiny block. The Gaussian mask is utilized to enhance the outcomes of the smoothing process. The following formulae in equations (8) and (9) provide the texture features:

$$\mu_{mn}(x,y) = c_{mn}(x,y) * gs_{mn}(x,y) \tag{8}$$

$$\sigma_{mn}(x,y) = \sqrt{\{c_{mn}(x,y) - \mu_{mn}(x,y)\}^2 * gs_{mn}(x,y)} \quad (9)$$

Where $c_{mn}(x,y)$ is the response to channel mn, corresponding to scale m and orientation n, while $gs_{mn}(x,y)$ is given by the expression in equation (10):

$$gs_{mn}(x,y) = \exp\left[-\frac{x^2}{2\rho_x^2} - \frac{y^2}{2\rho_y^2}\right] \quad (10)$$

3.4.2. Harris Corner Detection Method

The corners in the input image I_I must be found via the Harris corner detector. In order to produce a gradient image, I_I is first filtered via a Gaussian Mask. The Harris corner procedure is then enforced. Figure 3 shows a generalized flow diagram for Harris corner detection.

Harris corner detection algorithm is explained as follows:

1. Determine the horizontal and vertical gradients that are represented by the following:

$$\mathrm{M} = \begin{pmatrix} P & R \\ R & Q \end{pmatrix} = \begin{pmatrix} I_x^2 & I_{xy} \\ I_{xy} & I_y^2 \end{pmatrix} \quad (11)$$

2. Calculate the image's x and y derivatives:

$$I_x = G_\sigma^x * I, \quad I_y = G_\sigma^y * I \quad (12)$$

3. Calculate the derivative product at each pixel:

$$I_x^2 = I_x * I_x, \quad I_y^2 = I_y * I_y, \quad I_{xy} = I_x * I_y \quad (13)$$

4. Filter the image with a Gaussian filter:

$$w_{\mu,\mathrm{v}} = \exp\left(\frac{-1}{2}\,(\mu^2 + \mathrm{v}^2)\,/\,\delta^2\right) \quad (14)$$

Where, $w_{\mu,\mathrm{v}}$ symbolizes the Gaussian window.

5. Determine the pixel's r value.:

$$R_r = \left\{I_x^2 + I_y^2 - \left(I_x I_y\right)^2\right\} - k\left\{I_x^2 + I_y^2\right\}^2 \quad (15)$$

6. Local extreme points are chosen.
7. Determine the threshold and specify the corner point.

3.4.3. Deep Features Extraction by Convolutional Neural Network

Deep learning (DL) is a Machine learning (ML) technique that teaches computers to perform tasks that are natural to humans, such as learning from examples, so that they can solve problems such as image and speech recognition. This technique is increasingly being applied to the biological sciences [37].

Each of the neurons in a typical Neural Network (NN) produces a string of real-value activations. Neurons are tiny, linked processors. While weighted connections between previously activated neurons excite the remaining neurons, sensors that monitor the environment activate the input neurons [37].

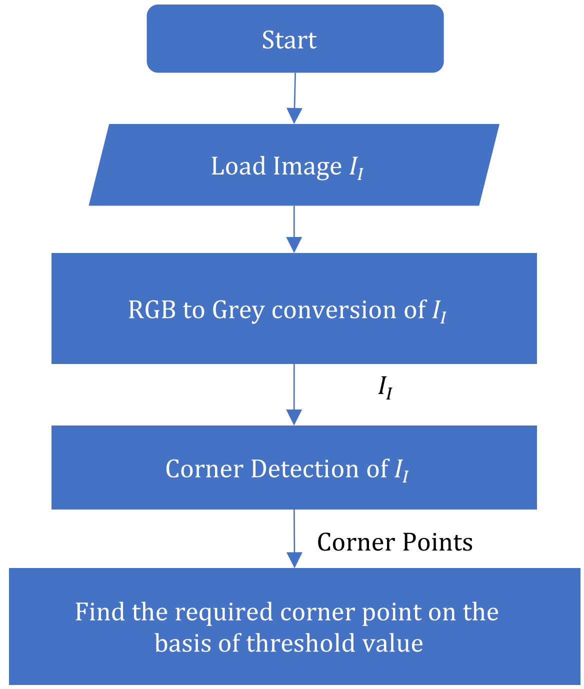

Fig. 3 Harris Corner detection flowchart

This study applied a deep CNN model to extract feature vectors of soybean leaf images. As shown in Figure 4 in the basic architecture of CNN, features were extracted from soybean leaf input images with successive convolution and pooling layers in CNN. At this stage, the established model and CNN are mentioned.

A convolution set (convolution-pooling) is made that separates and defines various features of the soybean leaf image, and this process is called feature extraction.

Convolutional Layer

The convolution process, which is the foundation of convolutional neural networks, tries to apply a filter matrix to the input and use the results for the subsequent layer. Small filters, including 2×2, 3×3, and 5×5, are applied to the whole soybean leaf image in this layer. As a result, a new image is created by deleting more identifying elements from the original. During the convolutional neural network's learning phase, the weights of the filter matrix used for the convolution operation are chosen. The convolution process is used after the predetermined amount has moved the filter matrix. The outcome from this layer, if not the last layer, is provided as input. The output image is represented by this layer if it is the final one. The windows of the same size (w) in the image are multiplied and summed to determine the filter coefficients (f), as shown in equation (16). Consequently, a fresh image built upon recognizable high-level elements is produced [38].

$$w(x,y) * f(x,y) = \sum_{i=-m}^{m} \sum_{j=-n}^{n} w(i,j) f(x+i, y+j) \quad (16)$$

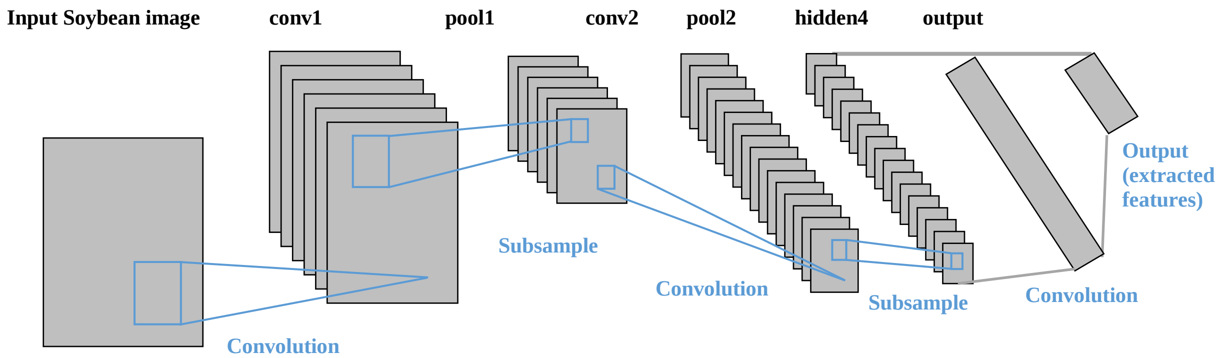

Fig. 4 Architecture for deep convolutional neural network

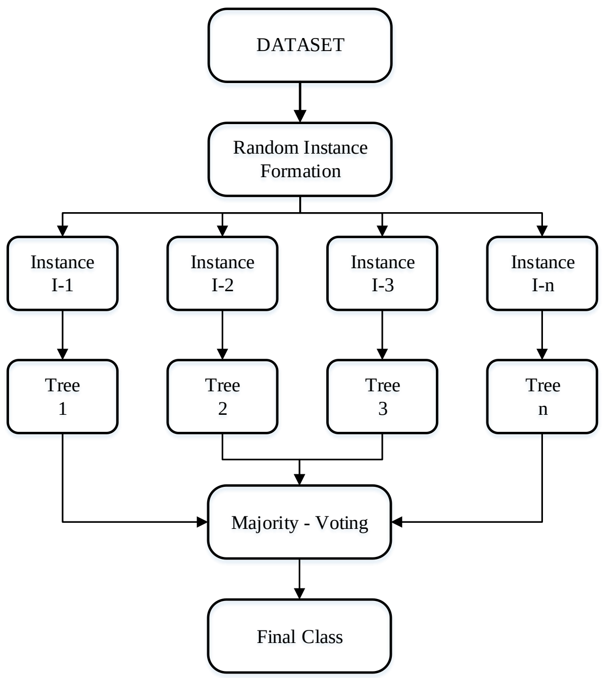

Fig. 5 Generalized block diagram for random forest classifier [39]

Pooling Layer

By lowering network settings, this layer helps to lighten the computational strain. When using maximum pooling or average pooling, the output pixel retains the highest value out of all the remaining pixel values in the filter window and the average of all the remaining pixel values in the filter window. The image's aspect ratio is lowered at the conclusion of the pooling procedure [38].

3.5. Classification by using Random Forest Classifier

The random forest methodology enhances the tree Bagging method by introducing a de-correlation requirement between them. The goal of this strategy is to diminish correlation without significantly increasing variance. The concept is to select a subset of variables at random that will be examined at each level of the tree's selection of the best node.

Consider the training set $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$, where a is the number of characteristics of X samples. Consider S_t to be a bootstrap comprising m instances acquired by resampling with S replaced. Let $\{h_1, \dots h_t\}$ be a collection of T decision trees. S_t is used to construct each tree h_t. The partitioning attribute is picked for each node of the tree by taking into account a number $f(f < a)$ of randomly selected characteristics (among the attributes a). The random forest classifier uses a uniformly weighted majority vote of classifiers in that collection to categorise a new instance x. This idea is shown by the algorithm.

Algorithm

$Input: lS = \{(x_1, ly_1), \dots, (x_m, ly_m)\}, lthe\ ltraining\ lset.$

$Input: lT, lthe\ lnumber\ lof\ ldecision\ ltrees\ lin\ lthe$
$random\ lforest.\ l\ l$

$lFor\ lt = 1, \dots, T\ ldo\ l\ l$

1. $Generate\ la\ lBootstrap\ lsample\ lS_t\ lof\ lsize\ lm$
   $from\ lS\ \ l l\ l$
2. $Create\ la\ ldecision\ ltree\ lh_t\ lfrom\ lS_t\ lby$
   $lrecursively\ lrepeating\ lfor\ leach\ lnode\ lof\ lthe$
   $tree, lthe\ lfollowing\ lsteps: l\ l$
   a. $Randomly\ lselect\ lf\ lattributes\ lamong\ la\ lattributes.$
   b. $Choose\ lthe\ lpartitioning\ lattribute\ lamong\ lf\ l$
   c. $Partition\ lthe\ lnode\ linto\ ltwo\ lchild\ lnodes\ l$

$l\ l\ lEnd\ lfor\ l$

$Output: lH, lthe\ lrandom\ lforest\ lclassifier\ l$

4. Simulation and Results

4.1. Evaluation Parameters

Table 1. Evaluation parameters

TP (True Positive)	"Indicated the Soybean with the disease that were classified as correctly classified."
TN (True Negative)	"Indicated the Soybean with a disease that was classified as not classified correctly."
FP (False Positive)	"Indicated the Soybean with the disease that were classified as incorrectly classified."
FN (False Negative)	"Indicated the Soybean with a disease that was classified as not classified incorrectly."

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (17)$$

$$Precision = \frac{TP}{TP+FP} \quad (18)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (19)$$

$$Specificity = \frac{TN}{TN+FN} \quad (20)$$

$$Error\ Rate = \frac{FP+FN}{TP+TN+FP+FN} \quad (21)$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP+TN} \quad (22)$$

$$F-Score = \frac{2TP}{2TP+FP+FN} \quad (23)$$

$$Matthews\ Correlation\ Coefficient\ (MCC) = \frac{(TP\times TN)-(FP\times FN)}{\sqrt{(TP+FN)(TP+FP)(TN+FN)(TN+FP)}} \quad (24)$$

$$Kappa\ Statistics = \frac{2(TP\times TN - FN\times FP)}{(TP+FP)\times(FP+TN)+(TN+FN)\times(FN+TN)} \quad (25)$$

4.2. Simulation Results

4.2.1. Soybean Insect Classification Result

Table 2. Tabular representation of confusion matrix for a soybean insect classification

	Predict Class1	Predict Class2	Predict Class3	Predict Class4	Predict Class5
Actual Class 1	1	0	0	0	0
Actual Class 2	0	4	0	0	0
Actual Class 3	0	0	4	0	0
Actual Class 4	0	0	0	1	0
Actual Class 5	2	0	0	0	2

Table 3. Multi-Class confusion matrix output using CNN-based soybean insect classification

	True Positive	False Positive	False Negative	True Negative
Actual Class 1	1	2	0	11
Actual Class 2	4	0	0	10
Actual Class 3	4	0	0	10
Actual Class 4	1	0	0	13
Actual Class 5	2	0	2	10

Table 4. Calculations for multi-class confusion matrix for a soybean insect classification

Calculations for Actual Class 1: *Here, TP=1, TN=11, FP=2, FN=0* $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{1+11}{1+11+2+0} = 85.71\%$ $Precision = \frac{TP}{TP+FP} = \frac{1}{1+2} = 33.34\%$ $Sensitivity = \frac{TP}{TP+FN} = \frac{1}{1+0} = 100\%$ $Specificity = \frac{TN}{TN+FN} = \frac{11}{11+0} = 100\%$ $F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times1}{2\times1+2+0} = 50\%$	***Calculations for Actual Class 2:*** *Here, TP=4, TN=10, FP=0, FN=0* $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{4+10}{4+10+0+0} = 100\%$ $Precision = \frac{TP}{TP+FP} = \frac{4}{4+0} = 100\%$ $Sensitivity = \frac{TP}{TP+FN} = \frac{4}{4+0} = 100\%$ $Specificity = \frac{TN}{TN+FN} = \frac{10}{10+0} = 100\%$ $F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times4}{2\times4+0+0} = 100\%$
Calculations for Actual Class 3: *Here, TP=4, TN=10, FP=0, FN=0* $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{4+10}{4+10+0+0} = 100\%$ $Precision = \frac{TP}{TP+FP} = \frac{4}{4+0} = 100\%$ $Sensitivity = \frac{TP}{TP+FN} = \frac{4}{4+0} = 100\%$ $Specificity = \frac{TN}{TN+FN} = \frac{10}{10+0} = 100\%$ $F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times4}{2\times4+0+0} = 100\%$	***Calculations for Actual Class 4:*** *Here, TP=1, TN=13, FP=0, FN=0* $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{1+13}{1+13+0+0} = 100\%$ $Precision = \frac{TP}{TP+FP} = \frac{1}{1+0} = 100\%$ $Sensitivity = \frac{TP}{TP+FN} = \frac{1}{1+0} = 100\%$ $Specificity = \frac{TN}{TN+FN} = \frac{13}{13+0} = 100\%$ $F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times1}{2\times1+0+0} = 100\%$
Calculations for Actual Class 5: *Here, TP=2, TN=10, FP=0, FN=2* $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{2+10}{2+10+0+2} = 85.71\%$ $Precision = \frac{TP}{TP+FP} = \frac{2}{2+0} = 100\%$ $Sensitivity = \frac{TP}{TP+FN} = \frac{2}{2+2} = 50\%$ $Specificity = \frac{TN}{TN+FN} = \frac{10}{10+2} = 83.34\%$ $F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times2}{2\times2+0+2} = 66.67\%$	

Table 5. Overall result for a soybean insect classification

Accuracy	94.28%
Error	5.72%
Sensitivity	90%
Specificity	96.66%
Precision	86.66%
False Positive Rate	3.34%
F-score	83.33%
Matthews Correlation Coefficient	0.8353
Kappa	0.5536

4.2.2. Soybean Leaf Diseases Detection Result

Table 6. Tabular representation of confusion matrix for soybean leaf diseases detection

	Predict Class1	Predict Class2	Predict Class3	Predict Class4
Actual Class 1	9	0	0	0
Actual Class 2	1	6	0	0
Actual Class 3	0	0	10	0
Actual Class 4	0	0	0	9

Table 7. Multi-Class confusion matrix output using CNN-based soybean leaf diseases detection

	True Positive	False Positive	False Negative	True Negative
Actual Class 1	9	1	0	25
Actual Class 2	6	0	1	28
Actual Class 3	10	0	0	25
Actual Class 4	9	0	0	26

Table 8. Calculations for multi-class confusion matrix for soybean leaf diseases detection

Calculations for Actual Class 1: *Here, TP=9, TN=25, FP=1, FN=0* $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{9+25}{9+25+1+0} =$ 97.14% $Precision = \frac{TP}{TP+FP} = \frac{9}{9+1} =$ 90% $Sensitivity = \frac{TP}{TP+FN} = \frac{9}{9+0} = 100\%$ $Specificity = \frac{TN}{TN+FN} = \frac{25}{25+0} = 100\%$ $F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 9}{2\times 9+1+0} = 94.74\%$	***Calculations for Actual Class 2:*** *Here, TP=6, TN=28, FP=0, FN=1* $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{6+28}{6+28+0+1} =$ 97.14% $Precision = \frac{TP}{TP+FP} = \frac{6}{6+0} =$ 100% $Sensitivity = \frac{TP}{TP+FN} = \frac{6}{6+1} = 85.71\%$ $Specificity = \frac{TN}{TN+FN} = \frac{28}{28+1} = 96.55\%$ $F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 6}{2\times 6+0+1} = 92.31\%$
Calculations for Actual Class 3: *Here, TP=10, TN=25, FP=0, FN=0* $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{10+25}{10+25+0+0} =$ 100% $Precision = \frac{TP}{TP+FP} = \frac{10}{10+0} = 100\%$ $Sensitivity = \frac{TP}{TP+FN} = \frac{10}{10+0} = 100\%$ $Specificity = \frac{TN}{TN+FN} = \frac{25}{25+0} = 100\%$ $F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 10}{2\times 10+0+0} = 100\%$	***Calculations for Actual Class 4:*** *Here, TP=9, TN=26, FP=0, FN=0* $Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{9+26}{9+26+0+0} =$ 100% $Precision = \frac{TP}{TP+FP} = \frac{9}{9+0} =$ 100% $Sensitivity = \frac{TP}{TP+FN} = \frac{9}{9+0} = 100\%$ $Specificity = \frac{TN}{TN+FN} = \frac{26}{26+0} = 100\%$ $F-Score = \frac{2TP}{2TP+FP+FN} = \frac{2\times 9}{2\times 9+0+0} = 100\%$

Table 9. Overall result for soybean leaf diseases detection

Accuracy	98.57%
Error	1.43%
Sensitivity	96.43%
Specificity	99.14%
Precision	97.5%
False Positive Rate	0.86%
F-score	96.76%
Matthews Correlation Coefficient	0.9600
Kappa	0.9238

Table 10. Comparison with previous research works

Method	Accuracy
Deep Learning based Soybean Leaf Diseases detection [15]	93.71%
Soybean Leaf Diseases detection using SVM [35]	90.00%
Classification of soybean insects using deep learning [40]	93.82%
Disease detection in plant leaf using CNN and Bayesian optimized SVM [41]	92.2%
Hybrid features based proposed method of soybean insects classification	94.28%
Hybrid features based proposed method of soybean leaf disease detection	98.57%

The abovementioned results in the table suggest that machine learning methods, particularly deep learning and hybrid feature-based methods, can effectively detect and classify soybean leaf diseases and insects.

The deep learning-based method for detecting soybean leaf disease [15] and the method for classifying soybean insects using deep learning [40] perform well with an accuracy of 93.71% and 93.82%, respectively. The SVM-based method for soybean leaf disease detection [35] has a lower accuracy of 90.00%, which suggests that using a single classification algorithm may not be as effective as using a hybrid approach. Furthermore, the method for disease detection in plant leaves using CNN and Bayesian optimized SVM [41] has an accuracy of 92.2%, which indicates that the use of Bayesian optimization can further improve the accuracy of the SVM-based classification.

The proposed hybrid features-based method for soybean insect classification also performs well, with an accuracy of 94.28%. This method uses a similar approach of combining multiple features, including visible light and near-infrared spectral imaging, to improve the classification accuracy of soybean insects. The high accuracy of this method suggests that using a combination of features can improve the accuracy of insect classification.

The highest accuracy of 98.57% is achieved by the proposed hybrid feature-based method for soybean leaf disease detection. This method combines multiple features to improve the detection and classification accuracy of soybean leaf diseases. The high accuracy of this method could be due to the use of a variety of features that capture different aspects of the disease. These features are then used to train a random forest classifier model to identify and classify different soybean leaf disease types accurately.

5. Conclusion

The results of validation techniques showed a minimum significant difference between observed and predicted values. The overall accuracy of the random forest classifier in soybean insect classification and leaf disease detection is 94.28 percent and 98.57 percent, respectively. The work presented in this paper clearly outperforms the previous research works in terms of classification accuracy. In this sense, it is intended, in future works, to investigate more deeply the role of the other feature extraction techniques in applications such as the one discussed in this article, also involving optimization in the classifier, seeking even better results with less computational time. It is also intended to expand the database, including more samples, including for more soybean diseases.

References

[1] Andreas Kamilaris, and Francesc X. Prenafeta-Boldú, “Deep Learning in Agriculture: A Survey,” *Computers and electronics in agriculture*, vol. 147, pp. 70-90, 2018. *Crossref,* https://doi.org/10.1016/j.compag.2018.02.016

[2] Justine Boulent et al., “Convolutional Neural Networks for the Automatic Identification of Plant Diseases,” *Frontiers in plant science*, vol. 10, 2019. *Crossref,* https://doi.org/10.3389/fpls.2019.00941

[3] Hiago de O. Gomes et al., “A Socio-environmental Perspective on Pesticide use and Food Production,” *Ecotoxicology and Environmental Safety*, vol. 197, p.110627, 2020. *Crossref,* https://doi.org/10.1016/j.ecoenv.2020.110627

[4] Antonio Monteiro, Sergio Santos, and Pedro Gonçalves, “Precision Agriculture for Crop and Livestock Farming—Brief Review,” *Animals*, vol. 11, no. 8, p. 2345, 2021. *Crossref,* https://doi.org/10.3390/ani11082345

[5] R.S. Latha et al., “Automatic Detection of Tea Leaf Diseases using Deep Convolution Neural Network,” *2021 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1-6, 2021. *Crossref,* https://doi.org/10.1109/ICCCI50826.2021.9402225

[6] Sachin B. Jadhav, Vishwanath R. Udupi, and Sanjay B. Patil, “Identification of Plant Diseases using Convolutional Neural Networks,” *International Journal of Information Technology*, vol. 13, pp. 2461-2470, 2021. *Crossref,* https://doi.org/10.1007/s41870-020-00437-5

[7] Zhou Libo, Huang Tian, and Guan Chunyun, “Wireless Multimedia Sensor Network for Rape Disease Detections,” *EURASIP Journal on Wireless Communications and Networking*, 2019. *Crossref,* https://doi.org/10.1186/s13638-019-1468-3

[8] Uday Pratap Singh et al., “Multilayer Convolution Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease,” *IEEE Access*, vol. 7, pp. 43721-43729, 2019. *Crossref,* https://doi.org/10.1109/ACCESS.2019.2907383

[9] Siddharth Singh Chouhan, Uday Pratap Singh, and Sanjeev Jain, “Applications of Computer Vision in Plant Pathology: A Survey,” *Archives of Computational Methods in Engineering*, vol. 27, pp. 611-632, 2020. *Crossref,* https://doi.org/10.1007/s11831-019-09324-0

[10] Luiz F.S. Coletta et al., “Novelty Detection in UAV Images to Identify Emerging Threats in Eucalyptus Crops,” *Computers and Electronics in Agriculture*, vol. 196, p. 106901. 2022. *Crossref,* https://doi.org/10.1016/j.compag.2022.106901

[11] Marwan Adnan Jasim, and Jamal Mustafa Al-Tuwaijari, “Plant Leaf Diseases Detection and Classification Using Image Processing and Deep Learning Techniques,” *2020 International Conference on Computer Science and Software Engineering (CSASE)*, pp. 259-265, 2020. *Crossref,* https://doi.org/10.1109/CSASE48920.2020.9142097

[12] Sumit Kumar, Veerendra Chaudhary, and Supriya Khaitan Chandra, “Plant Disease Detection using CNN,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 12, pp. 2106-2112, 2021.

[13] Lili Li, Shujuan Zhang, and Bin Wang, “Plant Disease Detection and Classification by Deep Learning—A Review,” *IEEE Access*, vol. 9, pp. 56683-56698, 2021. *Crossref,* https://doi.org/10.1109/ACCESS.2021.3069646

[14] S. Vieira, W.H.L. Pinaya, and A. Mechelli, *Introduction to Machine Learning, Machine learning*, Academic Press, 2020.

[15] Qiufeng Wu, Keke Zhang, and Jun Meng, "Identification of Soybean Leaf Diseases via Deep Learning," *Journal of The Institution of Engineers (India): Series A*, vol. 100, pp. 659-666, 2019. *Crossref*, https://doi.org/10.1007/s40030-019-00390-y

[16] Laercio Zambolim et al., "How to Cope with the Vulnerability of Site Specific Fungicides on the Control of Asian Soybean Rust," *International Journal of Research in Agronomy*, vol. 4, no. 1, pp.14-25, 2021. *Crossref*, http://dx.doi.org/10.33545/2618060X.2021.v4.i1a.44

[17] U. Shruthi, V. Nagaveni, and B.K. Raghavendra, "A Review on Machine Learning Classification Techniques for Plant Disease Detection," *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pp. 281-284, 2019. *Crossref*, https://doi.org/10.1109/ICACCS.2019.8728415

[18] Heng-An Lin, Maria B. Villamil, and Santiago X. Mideros, "Characterization of Septoria Brown Spot Disease Development and Yield Effects on Soybean in Illinois," *Canadian Journal of Plant Pathology*, vol. 43, no. 1, pp. 62-72, 2021. *Crossref*, https://doi.org/10.1080/07060661.2020.1755366

[19] M.D.C.Ferreira, F. Griesang, and A.F. Barreto, "Effects of Nozzle Types, Adjuvants and Spray Volumes on the Spray Application Quality and Control of Asian Soybean Rust," *Aspects of Applied Biology*, vol. 137, pp. 167-174, 2018.

[20] Gustavo C. Beruski et al., "Performance and Profitability of Rain-based Thresholds for Timing Fungicide Applications in Soybean Rust Control," *Plant Disease*, vol. 104, no. 10, pp. 2704-2712, 2020. *Crossref*, https://doi.org/10.1094/PDIS-01-20-0210-RE

[21] Vipendra Parmar et al., "Relative Efficacy of Bio Pesticides in Managment of Dry Root Rot and Collar Rot in Soyabean and Chickpea," *International Journal of Chemical Studies*, vol. 6, no. 2, pp. 83-86, 2018.

[22] Munmi Borah, "Identification of Soybean Diseases in Assam," *International Journal of Recent Scientific Research*, vol. 10, pp .34154-34159, 2019.

[23] D. Sheema et al., "The Detection and Identification of Pest-FAW Infestation in Maize Crops Using Iot-Based Deep-Learning Algorithm," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 9, no. 12, pp. 180-188, 2022. *Crossref*, https://doi.org/10.14445/23488379/IJEEE-V9I12P116

[24] Eftekhar Hossain, Md. Farhad Hossain, and Mohammad Anisur Rahaman, "A Color and Texture Based Approach for the Detection and Classification of Plant Leaf Disease Using KNN Classifier," *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1-6, 2019. *Crossref*, https://doi.org/10.1109/ECACE.2019.8679247

[25] Y.M. Oo, and N.C. Htun, "Plant Leaf Disease Detection and Classification using Image Processing," *International Journal of Research and Engineering*, vol. 5, no. 9, pp. 516-523, 2018.

[26] K.B. Prakash, and G.R. Kanagachidambaresan, "Pattern Recognition and Machine Learning," *Programming with TensorFlow* (pp. 105-144). Springer, Cham.

[27] Shima Ramesh et al., "Plant Disease Detection using Machine Learning," *2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C)*, pp. 41-45, 2018. *Crossref*, https://doi.org/10.1109/ICDI3C.2018.00017

[28] Silky Sachar, and Anuj Kumar, "Survey of Feature Extraction and Classification Techniques to Identify Plant Through Leaves," *Expert Systems with Applications*, vol. 167, p. 114181, 2021. *Crossref*, https://doi.org/10.1016/j.eswa.2020.114181

[29] Adnan Mohsin Abdulazeez et al., "Leaf Identification Based on Shape, Color, Texture and Vines Using Probabilistic Neural Network," *Computación Y Sistemas*, vol. 25, no. 3, pp. 617-631, 2021. *Crossref*, https://doi.org/10.13053/cys-25-3-3470

[30] Gargi Sharma, and Gourav Shrivastava, "Crop Disease Prediction using Deep Learning Techniques - A Review," *SSRG International Journal of Computer Science and Engineering*, vol. 9, no. 4, pp. 23-28, 2022. *Crossref*, https://doi.org/10.14445/23488387/IJCSE-V9I4P104

[31] R. Manavalan, "Automatic Identification of Diseases in Grains Crops Through Computational Approaches: A Review," *Computers and Electronics in Agriculture*, vol. 178, p. 105802, 2020. *Crossref*, https://doi.org/10.1016/j.compag.2020.105802

[32] Abirami Devaraj et al., "Identification of Plant Disease using Image Processing Technique," *2019 International Conference on Communication and Signal Processing (ICCSP)*, pp. 749-753, 2019. *Crossref*, https://doi.org/10.1109/ICCSP.2019.8698056

[33] Sushil R. Kamlapurkar, "Detection of Plant Leaf Disease using Image Processing Approach," *International Journal of Scientific and Research Publications*, vol. 6, no. 2, pp. 73-76, 2018.

[34] P. Perumal, "Guava Leaf Disease Classification using Support Vector Machine," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 7, pp. 1177-1183, 2021. *Crossref*, https://doi.org/10.17762/turcomat.v12i7.2728

[35] Sukhvir Kaur, Shreelekha Pandey, and Shivani Goel, "Semi-automatic Leaf Disease Detection and Classification System for Soybean Culture," *IET Image Processing*, vol. 12, no. 6, pp. 1038-1048, 2018. *Crossref*, https://doi.org/10.1049/iet-ipr.2017.0822

[36] Sambuddha Ghosal et al., "An Explainable Deep Machine Vision Framework for Plant Stress Phenotyping," *Proceedings of the National Academy of Sciences*, vol. 115, no. 18, pp. 4613-4618, 2018. *Crossref*, https://doi.org/10.1073/pnas.1716999115

[37] Everton Castelao Tetila et al., "Automatic Recognition of Soybean Leaf Diseases using UAV Images and Deep Convolutional Neural Networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 5, pp. 903-907, 2020. *Crossref*, https://doi.org/10.1109/LGRS.2019.2932385

[38] Wei Lu et al., "Soybean Yield Preharvest Prediction Based on Bean Pods and Leaves Image Recognition Using Deep Learning Neural Network Combined with GRNN," *Frontiers in Plant Science,* vol. 12, 2021. *Crossref,* https://doi.org/10.3389/fpls.2021.791256

[39] V. Brindha Devi et al., "An Efficient and Robust Random Forest Algorithm for Crop Disease Detection," *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, pp. 1-4, 2022. *Crossref,* https://doi.org/10.1109/IC3IOT53935.2022.9767937

[40] Everton Castelao Tetila et al., "Detection and Classification of Soybean Pests Using Deep Learning with UAV Images," *Computers and Electronics in Agriculture*, vol. 179, p. 105836, 2020. *Crossref,* https://doi.org/10.1016/j.compag.2020.105836

[41] Ashutosh Kumar Singh et al., "Hybrid Feature-based Disease Detection in Plant Leaf using Convolutional Neural Network, Bayesian Optimized SVM, and Random Forest Classifier," *Journal of Food Quality*, 2022. *Crossref,* https://doi.org/10.1155/2022/2845320

[42] Lidong Dong et al., "Genetic Basis and Adaptation Trajectory of Soybean from its Temperate Origin to Tropics," *Nature Communications,* vol. 12, 2021. *Crossref,* https://doi.org/10.1038/s41467-021-25800-3

[43] Mohammad Keivani et al., "Automated Analysis of Leaf Shape, Texture, and Color Features for Plant Classification," *Traitement du Signal*, vol. 37, no. 1, pp. 17-28, 2020. *Crossref,* http://dx.doi.org/10.18280/ts.370103