Cureus Journal of Computer Science

Open Access Original Article

Crop Health Detection Using Image Processing and Machine Learning for Better Yield Production

Received 02/07/2025
Review began 02/25/2025
Review ended 03/21/2025
Published 03/25/2025

DOI:
https://doi.org/10.7759/s44389-025-03237-9

Neha Sharma [1] , Pradeep Kumar [2]

1. Computer Science and Engineering, Graphic Era Hill University, Haldwani, IND 2. Computer Science and Engineering, Shri Ram Murti Smarak College of Engineering Technology and Research, Bareilly, IND

Corresponding authors: Neha Sharma, neha2504sharma@gmail.com, Pradeep Kumar, pradeep2002@gmail.com

Abstract
Agriculture plays a vital role in every nation, as a healthy population relies on robust yields to ensure food security. With the continuous growth of the population, the excessive use of pesticides and fertilizers has become prevalent, which can negatively impact crop health. In this paper, we introduce a solution that utilizes images of crops, employing image processing and machine learning techniques to classify them as healthy or unhealthy. Various feature detection and extraction methods are available, but we specifically compare Oriented FAST and Rotated BRIEF and scale-invariant feature transform in this work. Both techniques can effectively extract features from the images, and we can use the matcher function from OpenCV to determine if the extracted features correspond to those of a trained image. If there is a match, it indicates an unhealthy crop, while a lack of matching suggests the crop is healthy. Additionally, machine learning classifiers can be employed to enhance training on these extracted features, leading to improved results and predictions.

Categories: Image Processing and Analysis, Data Analysis, Machine Learning (ML)
Keywords: image processing, opencv, sift, orb, feature extraction

Introduction
Healthy agriculture is the backbone of any country. Therefore, solution to diagnose early crop disease is very efficient. Here, in this paper, image processing and machine learning are used to classify healthy and unhealthy classes on the basis of a dataset. The global agricultural industry faces significant challenges in ensuring healthy crop production due to the prevalence of plant diseases. These diseases not only reduce crop yield but also affect the quality of produce, ultimately impacting food security and economic stability. Traditional methods for detecting plant diseases typically rely on manual inspection, which is time-consuming, labor-intensive, and prone to human error. As a result, there is an increasing need for automated and efficient disease detection systems that can accurately classify crops as healthy or unhealthy.

Recent advancements in image processing techniques have shown great promise in addressing this need. By analyzing high-resolution images of crops [1], these methods can detect early signs of diseases, enabling prompt intervention and minimizing crop loss. Among these techniques, Oriented FAST and Rotated BRIEF (ORB) and scale-invariant feature transform (SIFT) are powerful tools that can extract key features from images, enabling the classification of crops based on their visual characteristics.

The goal of this approach is to identify crop diseases by analyzing images to determine whether they are infected or not. This method employs the SIFT model, aiming for both precise and rapid classification of fruits. SIFT features are localized and derived from the object's appearance at specific points of interest, remaining consistent regardless of changes in image scale or rotation. They are also resilient against variations in lighting, noise, and slight alterations in viewpoint, according to image processing theories. SIFT offers considerable advantages, including high accuracy and ease of extraction, facilitating reliable object identification with minimal risk of errors.

This paper utilizes wheat crop images for training and testing purposes. The images display various diseases, such as fungal infections and black rust, which may impact the overall yield. By employing image processing techniques, we can identify disease characteristics, and then utilize OpenCV matcher function to align the extracted features with the original or test image, allowing us to visualize the final mapping. In the future, machine learning methods can be integrated to automate this mapping process, enabling us to assess crop health from the input images. While there are numerous methods available for feature extraction, this study focuses on the SIFT algorithm to identify key points within the images. Therefore, a combination of machine learning and image processing techniques is being developed. Various techniques have been proposed by researchers in the recent past.

This paper does not explicitly address the application of SIFT or ORB in detecting crop diseases. Rather, it emphasizes features related to color, shape, and texture, utilizing the Sobel edge detector along with an ensemble model [1].

A segmentation method has been introduced for intricate real-field backgrounds. Individual testing of standard features showed that gray-level co-occurrence matrix texture features produced the highest accuracy. However, combining features did not lead to significant improvements in model performance. On the other hand, cepstral coefficients delivered outstanding outcomes for classifying both species and diseases. The accuracy rates for first-level classifications were 94.33%, 94.11%, and 98.44% for the respective crop types. In terms of second-level disease classification, the accuracies were 97.75%, 96.66%, and 97.95%. These results confirm the effectiveness of cepstral coefficients across various conditions. Looking ahead, the focus will shift towards refining segmentation techniques to enhance crop yields [2].

The document classifies methods for detecting crop diseases, highlighting image analysis approaches such as SIFT and ORB. These techniques effectively extract features from images to identify and categorize plant diseases [3].

Plant diseases pose a significant challenge to agricultural yield and global food security. Timely identification and precise diagnosis of these diseases are essential for effective management. Recent advancements in machine learning, especially deep learning methods like convolutional neural networks (CNNs), have demonstrated substantial promise in detecting plant diseases. In this research, we introduce a system designed for real-time detection of plant diseases using image processing alongside CNNs. The system accepts an image of the affected plant, employs image processing strategies to extract key features, and then these features are processed through a CNN model trained on a comprehensive dataset of plant disease images to determine the specific disease. By enabling early detection and management of plant diseases, this system aims to help farmers mitigate crop losses and enhance agricultural productivity [4].

The article does not cover SIFT or ORB methods for detecting crop diseases. Instead, it emphasizes the application of NLP and machine learning to analyze crop [5].

Farming plays a vital role in driving the economy and advancement of our nation. However, precision agriculture is still evolving in the realm of technology-enhanced growth. Over the years, various plant diseases have inflicted suffering on millions globally, leading to an estimated annual yield loss of 14%. Utilizing computer-based disease segmentation and diagnosis from leaf images offers a more efficient alternative to traditional methods. The process involves image capture, preprocessing, and segmentation, followed by augmentation, feature extraction, and classification using models for automatic plant disease diagnosis. This framework utilizes deep learning models such as VGG-16, ResNet-50, AlexNet, DenseNet-169, and InceptionV3 to identify plant diseases from photos in the Plant Village Dataset and accurately categorize them into two classes. The experimental results indicated that ResNet-50 achieved the highest accuracy at 97.80% compared to the other deep learning models used for disease classification [6].

In this study, we explore the application of CNNs for crop disease detection systems. Our findings suggest that there is still potential for improving outcomes in this area. During our research, we observed that many authors did not deploy a unified approach that incorporates all relevant variables. By considering the impact on crop yield simultaneously, we can better estimate the yields. Additionally, adopting a neural network approach could further enhance these results [7].

The study centers on employing the SIFT model to identify diseases in fruits, highlighting its precision and reliability. It does not concentrate on the application of ORB for detecting diseases in crops [8].

Detecting diseases in crops is a crucial yet labor-intensive aspect of agricultural practices that demands significant time and skilled personnel. This study introduces a clever and effective method for identifying crop diseases through the application of computer vision and machine learning techniques. The proposed system can accurately recognize 20 different diseases across five commonly cultivated plants, achieving an accuracy rate of 93% [9].

This paper proposes the identification of 38 different disease classes, encompassing a wide variety of plant ailments. The methodology developed leverages deep learning to effectively recognize these diseases, which includes both unhealthy and healthy plants. To enhance user interaction with the proposed system, an interface has been created using the Django framework, facilitating a user-friendly experience [10].

Materials And Methods

The objective of this methodology is to leverage SIFT and ORB feature extraction techniques to accurately classify crops as either healthy or unhealthy based on image data. Both techniques are robust in handling scale variations, rotation, and partial occlusions, making them ideal for detecting early signs of disease in crops through visual inspection.

Data collection

To begin, high-quality images of crop leaves or plants are collected under controlled or real-world conditions. These images must encompass both healthy and unhealthy crops (with visible symptoms of disease) to create a balanced dataset. Datasets may include images from different angles, lighting conditions, and environments, simulating realistic field scenarios.

Image preprocessing

Before applying feature extraction methods, the images are preprocessed to enhance quality and minimize noise. The typical preprocessing steps are as follows:

Resizing: Images are resized to a fixed resolution to maintain consistency.

Grayscale Conversion: Since both ORB and SIFT focus on feature points, converting the images to grayscale simplifies the task by reducing complexity.

Noise Reduction: Techniques like Gaussian blur or median filtering may be applied to reduce noise from the images.

Feature extraction using SIFT

SIFT is a robust algorithm for detecting distinctive keypoints (features) in images, which can then be used for matching and classification. The SIFT algorithm operates in the following steps:

Scale-Space Extrema Detection: The first step involves detecting keypoints at different scales by creating a series of images with progressively blurred versions. These keypoints correspond to significant variations in pixel intensity, often marking boundaries or textures that are sensitive to diseases.

Keypoint Localization: For each detected keypoint, a more precise localization is performed by refining its position and scale using a Taylor series expansion. This process eliminates points that are not stable across different scales.

Orientation Assignment: Each keypoint is assigned an orientation based on the local image gradient directions, making the detection rotation-invariant. This step ensures that the keypoints can be matched consistently despite changes in orientation (e.g., a plant turning or being photographed from different angles).

Descriptor Creation: A 128-dimensional descriptor vector is generated for each keypoint by analyzing its local neighborhood. This descriptor encodes information about the keypoint's local texture and gradient, which is essential for distinguishing healthy from unhealthy crops.

Keypoint Matching: Once keypoints and descriptors are extracted from the input images, they can be matched with keypoints in images from other crops (either healthy or diseased). A nearest neighbor search is often used to find the best matches, which can be used to assess the likelihood of the crop being healthy or unhealthy.

Feature extraction using ORB

ORB is an efficient, fast alternative to SIFT that combines the advantages of both FAST, i.e. Features from Accelerated Segment Test and BRIEF, i.e. Binary Robust Independent Elementary Features methods. ORB is preferred for real-time applications due to its speed and lower computational overhead while still maintaining robust performance in feature extraction. The ORB algorithm proceeds as follows:

Keypoint Detection Using FAST: ORB uses the FAST corner detector to identify potential keypoints in the image. FAST is particularly good at identifying corners and edges, which are often useful features for distinguishing plant diseases.

Keypoint Orientation Assignment: To achieve rotation invariance, ORB assigns an orientation to each keypoint based on the local image gradient directions, similar to the process in SIFT.

BRIEF Descriptor Calculation: ORB then computes a binary descriptor for each keypoint by analyzing the intensity comparisons of pixel pairs in the surrounding region of the keypoint. This binary descriptor is more computationally efficient than SIFT's 128-dimensional descriptor and provides a fast way to compare keypoints.

Keypoint Matching: The descriptors are compared using Hamming distance (appropriate for binary descriptors), allowing for quick and efficient matching between keypoints from different images.

Feature matching and classification

Once the features (keypoints and descriptors) are extracted using either SIFT or ORB, the next step is to match features between healthy and unhealthy crop images. The keypoints that are matched across images allow the system to assess the presence of disease symptoms, such as leaf discoloration, spots, or irregularities in the plant's texture.

Feature Matching

A K-nearest neighbor (KNN) or FLANN (Fast Library for Approximate Nearest Neighbors) search is commonly used to match the keypoints between images. The closer the matches, the more confident the system is that the image belongs to a specific class (healthy or unhealthy).

Decision-Making

After matching, the system classifies the crop based on the frequency and type of detected keypoints. If a higher number of keypoints associated with disease symptoms are detected, the crop is classified as unhealthy, while a healthy crop will have fewer or no disease-related keypoints.

Machine learning model for final classification

To improve accuracy and robustness, the features extracted by SIFT and ORB can be used to train a machine learning classifier [7], such as support vector machine (SVM), random forest, k-NN, and CNNs) for end-to-end learning.

These models are trained using the feature descriptors (SIFT or ORB) extracted from a labeled dataset (healthy vs. unhealthy). The classifier will learn to associate specific features with healthy and unhealthy crop patterns. After training, the classifier can be used to predict the health status of new crop images.

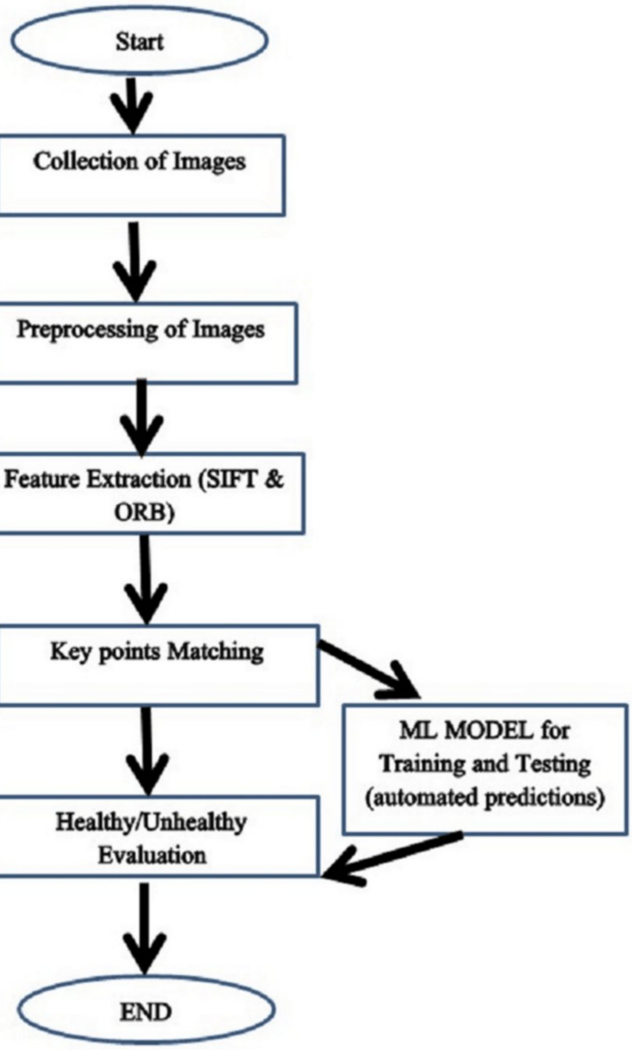

FIGURE 1: Crop disease detection methodology

Source: [9]

The methodology outlined for crop disease detection using SIFT, ORB, and machine learning model involves (Figure 1):

1. Collecting high-quality crop images.

2. Preprocessing these images (grayscale conversion, resizing, noise reduction).

3. Extracting distinctive keypoints using SIFT and ORB.

4. Matching these keypoints across images to identify disease patterns.

5. Using machine learning classifiers to predict whether a crop is healthy or unhealthy based on these features.

Results

Figure 2 illustrates the results of crop disease detection using image processing and machine learning.

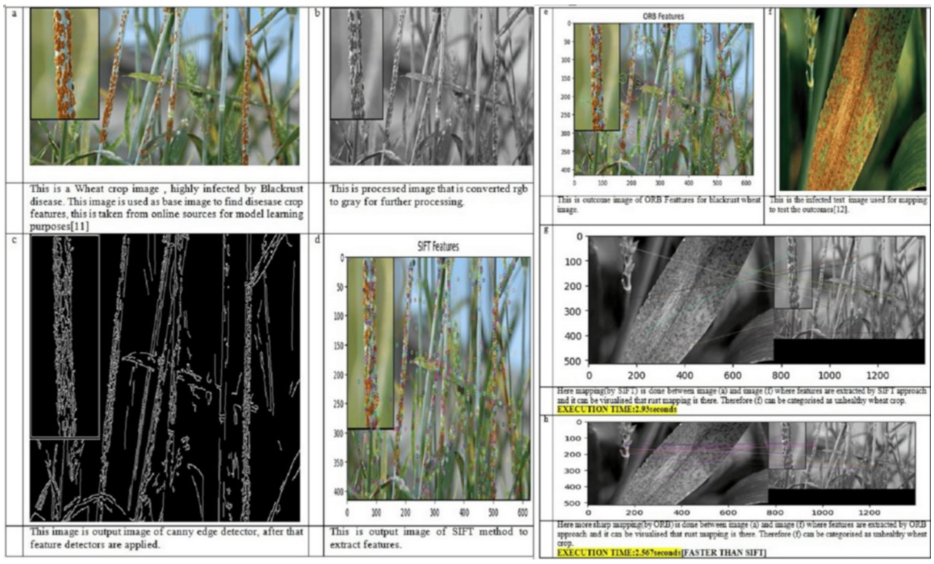

FIGURE 2: Result of crop disease detection using image processing and machine learning. (a) Original wheat image. (b) RGB to gray image. (c) Processed image. (d) Feature extraction by SIFT. (e) Feature extraction by ORB. (f) Test image. (g, h) Feature mapping for prediction

RGB, Red, Green, and Blue; SIFT, Scale-Invariant Feature Transform; ORB, Oriented FAST and Rotated BRIEF

Table *1* presents a comparison of execution times between the SIFT and ORB methods for feature extraction. The SIFT method, labeled as slow, takes 2.93 seconds, whereas the ORB method, labeled as fast, completes the task in 2.567 seconds.

Execution Time Comparison	
SIFT Method (Slow)	ORB Method (Fast)
2.93 seconds	2.567 seconds

TABLE 1: Execution time comparison of SIFT and ORB method

SIFT, Scale-Invariant Feature Transform; ORB, Oriented FAST and Rotated BRIEF

Implementation setup

For this implementation, important requisites are as follows:

1. Python library-supportable software: here Jupyter notebook is used.

2. Crop images with disease: here images of wheat with rust disease are used.

3. Open CV2: Open CV2 is used for image manipulation and to perform further task.

Both SIFT and ORB are effective for feature extraction and can be paired with matcher functions to identify or recognize infected areas on crops. Upon visualizing the results, it is clear that ORB surpasses SIFT in both execution time and the number of features extracted.

Discussion

The use of feature extraction methods like SIFT and ORB has proven to be effective for classifying diseased crops based on image analysis. It is shown that SIFT takes 2.93 seconds for execution and ORB is much faster than it, as it takes 2.567 seconds. SIFT is a well-established algorithm for detecting and describing local features in images that are invariant to scale, rotation, and affine transformations. It performs well in scenarios where the disease-related patterns in crop images may appear at varying scales or orientations. However, SIFT has computational drawbacks, such as its relatively high processing cost, which can be a limitation for real-time applications or when dealing with large datasets. Additionally, SIFT is a patented algorithm, which may limit its accessibility for certain commercial applications.

Conclusions

ORB is a more computationally efficient alternative to SIFT, designed to offer similar performance with a lower computational cost. It combines the FAST keypoint detector and the BRIEF descriptor, both of which are faster and more efficient than their counterparts in SIFT. However, ORB may not be as precise as SIFT in cases where the disease patterns are very subtle or require very high-detail extraction. SIFT is preferable when accuracy and robustness to variations in crop images are paramount, especially when dealing with diverse environmental conditions. ORB is more suitable for real-time applications and environments where computational efficiency is critical, while still providing acceptable accuracy for crop disease classification. A hybrid approach that combines both feature extraction techniques with deep learning models could also be explored to leverage the strengths of both methods. This could potentially lead to even more robust and accurate crop disease detection systems, capable of scaling to large agricultural datasets.

Different methods with hybrid approach can enhance the results as SIFT, ORB, CNN, and deep learning. Different environmental conditions (lightning, noises in image, angles of capturing, etc.) of acquiring dataset can also be considered in future to enhance the application.

Appendices

GitHub Repository link for dataset: https://github.com/68neha/dataset_crop_detection.git

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Neha Sharma, Pradeep Kumar

Acquisition, analysis, or interpretation of data: Neha Sharma, Pradeep Kumar

Drafting of the manuscript: Neha Sharma, Pradeep Kumar

Critical review of the manuscript for important intellectual content: Neha Sharma, Pradeep Kumar

Supervision: Neha Sharma, Pradeep Kumar

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Rabbi MMH, Hossain MJ, Masud MA, Rahman MA: An image recognition approach for crop disease detection in agro-field from infected plant area. Global Mainstream Journal of Innovation, Engineering & Emerging Technology. 2024, 3:11-18. 10.62304/jieet.v3i05.219
2. Bhagwat R, Dandawate Y: Towards robust crop disease detection for complex real field background images. Vietnam Journal of Science and Technology. 2024, 62:990-1004. 10.15625/2525-2518/18914
3. Batham S, Sharma AK, Rathore AS: A research paper on crop disease detection using deep learning model. International Journal for Research in Applied Science & Engineering Technology. 2023, 11:399-406. 10.22214/ijraset.2023.56002
4. Paithankar V, Awari A, Raskar A, Patil S, Jamdar N: Plant disease detection using machine learning. International Journal of Advanced Research in Science Communication and Technology. 2023, 3:267-272. 10.48175/ijarsct-9297
5. Prashanthi T, Susmitha T, Mishra RK: Crop disease detection using NLP and deep learning. Proceedings of the 2nd International Conference on Cognitive and Intelligent Computing. Kumar A, Ghinea G, Merugu S (ed):

Springer, Singapore; 2023. 197-203. 10.1007/978-981-99-2742-5_21
6. Bande P, Dewangan K: Plant disease detection using deep learning. International Journal for Research in Applied Science & Engineering Technology. 2022, 10:858-865.
7. Gaikwad D, Sonone A, Patil S, Limbole S, Jain N: Crop disease detection system using deep learning method. SAMRIDDHI A Journal of Physical Sciences Engineering and Technology. 2022, 14:291-295. 10.18090/samriddhi.v14spli02.16
8. Ayeesha, Zeela F, Vijetha: Classifying disease in fruit using machine learning. International Journal of Scientific Research in Computer Science Engineering and Information Technology. 2021, 7:309-313. 10.32628/cseit217341
9. Kulkarni P, Karwande A, Kolhe T, Kamble S, Joshi A, Wyawahare M: Plant disease detection using image processing and machine learning [PREPRINT]. arXiv:2106.10698. 10.48550/arXiv.2106.10698
10. Prathyusha G, Reddy V: Leaf based plant disease detection using deeplearning. Journal of Emerging Technologies and Innovative Research. 2021, 8:47-51.