European Journal of Sustainable Development Research
2026, 10(2), em0385
e-ISSN: 2542-4742
https://www.ejosdr.com/

MODESTUM

Research Article

OPEN ACCESS

Intelligent prediction of weed-borne diseases in crops: A machine learning approach

Ashish Baiswar [1,2*], Jameel Ahmad [1], Atul Kumar [2]

[1] Integral University, Dashauli, INDIA
[2] Chandigarh University Uttar Pradesh, Unnao-209859, INDIA
***Corresponding Author:** baiswarashish@gmail.com

Citation: Baiswar, A., Ahmad, J., & Kumar, A. (2026). Intelligent prediction of weed-borne diseases in crops: A machine learning approach. *European Journal of Sustainable Development Research, 10*(2), em0385. https://doi.org/10.29333/ejosdr/18138

ARTICLE INFO

Received: 03 Nov. 2025

Accepted: 29 Jan. 2026

ABSTRACT

Crop pathogens frequently find alternative hosts and a reservoir in weed species, making it harder to manage the disease and causing major loss of yield. The study discusses the use of smart methods which are machine learning and computer vision to detect the transmission of weed based diseases to farm produce in early stages. We created a predictive system that combines the process of image analysis of weed health with environmental data and agronomic data. This system is based on convolutional neural networks (CNNs) to locate the visual manifestations of the disease on common weed and a random forest classifier to estimate the possibility of further infection of the crop. Field trials on major crops and the weeds associated with them were performed to obtain data on visual symptoms of the weeds, as well as disease incidence in crops. The model that was developed showed prediction accuracy of 88 percent on occurrence of high-risk conditions that result in disease transfer. These results will show that the intelligent systems will be able to issue timely alerts, and proactive and specific disease management methods can be implemented in precision agriculture. The strategy has significantly enhanced the traditional methods of scouting since it automatizes. and facilitates the detection of the threats of disease that arise with the appearance of weeds.

Keywords: artificial intelligence, CNN, weed management, smart farming systems, disease detection, computer vision, weed analysis, food security, crop protection

INTRODUCTION

Crop diseases are a significant challenge to the food security in the world since they have a significant effect on agricultural productivity, which poses a challenge to the sustainability of the current farming systems. Most disease management methods target pathogens that are known to infect crops directly but weeds are an important part of the agricultural diseases that is mostly ignored. In addition to competing with nutrients, light and space, weeds often serve as alternative hosts and reservoirs to an extremely large number of plant pathogens, such as bacteria, fungi, viruses, and nematodes. The pathogens may later be transmitted to economically significant crops leading to disease persistence and recurrence (Bošilj et al., 2006; Carroll & Wicklow, 1992; Khalili et al., 2008 Williamson-Benavides & Dhingra, 2021).

With the high concentration of the infected weeds near the vulnerable crops, this poses a continuous source of inoculum, complicating and rendering the disease control less effective. The old methods of surveillance of the disease caused by weeds like scouting of the fields manually and laboratory testing are labor-intensive, time-consuming and mostly reactive as opposed to being preventive. This occurs leading to failure to detect early infections among the weed populations, which limits the power of the intervention strategies (Gawęda et al., 2021).

This is why there is an increasing necessity of quick, precise and scalable means of forecasting the disease outbreaks with a starting point in the weed populations. The latest breakthroughs in machine learning (ML) and computer vision provide the potential solutions to this problem. The technologies allow automatic identification of indicators of early disease occurrence using visual information and predictive analysis by incorporating ecological and agronomic indicators. Specifically, deep learning architectures, including convolutional neural networks (CNNs), have shown good performance on detecting subtle visual cues that are related to plant diseases, and ensemble models, like random forests (RF), are more appropriate in risk predicting based on heterogeneous data (Abbas et al., 2020; Aravind & Raja, 2020; Liakos et al., 2018; Liang et al., 2022).

The paper presents a smart prediction platform that involves the combination of image-based assessment of the health of the weed with environmental and crop-related measurements to detect the degree of threat posed by the weed on the crops. The proposed system will complement proactive disease management by identifying the symptoms of the disease in the weeds at an early stage and predicting possible outbreak in the adjacent crops. This type of approach can increase precision agricultural practices this way allows timely and specific interventions and minimizes the use of reactive and blanket control measures (Baiswar et al., 2018; Kaur et al., 2024).

Contextualizing Weed-Borne Pathogens and Crop Health Vulnerabilities

The agronomical environment is full of weed species, whose ecological contribution is not limited to the competition of resources. The most common weeds like amaranthus retroflexus or chenopodium album can act as an asymptomatic reservoir of a large arsenal of plant pathogens, including those that cause significant economic losses in large-scale food crops, such as wheat, maize, and soybean (Gawęda et al., 2021). Some examples include weeds which are viral vectors and subsequently infected by insects to crops or fungi which grow in the soil like Verticillium dahlia that causes wilt diseases (Carroll & Wicklow, 1992). The implication of such pathogen reservoirs in or near crop field is that, although the in- crop diseases are being well managed there are high chances of re-infection or new infection in case these sources of the weeds are not managed or controlled. The diversity of the weed flora may also influence the community of pathogens since some weed species can be more efficient than others as a host (Feledyn-Szewczyk, 2012). Such a combination of weeds, pathogens and crops means that it is not easy to predict the outbreak of diseases. The currently used agricultural practices are generally geared towards eliminating weeds, primarily because of competition reasons, and the contribution of the latter in epidemiology of pathogens is not necessarily systematized (Fischer and Ramirez). The identification of the particular weed-pathogen-crop complexes and the creation of instruments to identify the initial signs of the disease in weed populations are thus critical in the establishment of the integrated pest and disease management tools that are effective and sustainable.

LITERATURE REVIEW

Artificial intelligence (AI) uses and specifically ML has grown at an accelerated rate in applying it to agricultural issues, specifically plant disease detection. Multiple researchers have shown that ML models can predict and identify disease in crops based on image data with high accuracy and frequently learn texture, color and lesion structures directly using raw pixel inputs with the help of CNNs (Aravind & Raja, 2020). CNNs are a type of deep learning model that have been especially effective at classifying disease based on images because of their hierarchical learning of features directly based on raw pixel data (Aravind & Raja, 2020). Other similar trends have been made in the area of weed science, whereby ML techniques are deployed to differentiate crops and weeds, often to permit the application of specific herbicides or mechanical weeding (Su) (Champ et al., 2020). These systems will regularly depend on spectral, color or hyperspectral imaging to acquire exceptional characteristics of plants (Su). Alongside simple identification, ML models have been developed to estimate the crop yield (Abbas et al., 2020), pest infestation (Tageldin & El-Naggar, 1997), and the likelihood of disease in accordance with the environmental conditions or plant physiological indicators. Indicatively, the support vector machines (SVM), RF and gradient tree boosting have already demonstrated a great degree of performance in the prediction modeling of plant diseases using sensor data and environmental variables. It has also talked about transfer learning to facilitate easier model training on new crops or conditions by reducing the amount of labeling required to train such models. Despite these advances, research that specially focuses on the ability to predict the transmission of diseases between crops and weeds with the help of smart systems is still in its early stages. Most of the studies have taken into consideration either the identification of weeds or the detection of crop sickness individually without necessarily establishing the weed as a pathogen reservoir and source of inoculum in subsequent infection of crops. This paper is an attempt to fill this gap by creating a system that would combine weeds status monitoring with crop disease risk assessment. The images that are captured by drones when in various kinds of light are processed by the framework of the deep learning algorithm which is a CNN. The process involves a stage of first picture partitioning in separating crops and weeds with references to their visual characteristics, e.g., shape, color, and texture. Diseases are then trained on the CNN. This is done by measuring the performance of the model to ensure that they are viable in real time diagnostics. This paper contributes to the development of automated sensing systems that may aid the process of early detection and precision agriculture to be more sustainable (Baiswar et al., 2025). Even though the traditional approach to managing crops has relied on the conventional methods of managing crops, which cannot be used to support current data, predictive analytics and ML bridge this gap. Such technologies are able to allow farmers and other agribusinesses to have a clear vision and practical advice which are date- based and realistic through analyzing information of the sources which include climate record, soil sensors and real-time monitors (Rizvi et al., 2024). Typically, the matching of the image is the counterpart to the comparing of two images and the easy notion, i.e., how and in what way is this similarity to be measured in case the two pictures are similar or similar enough? (Kumar et al., 2022). The CNN-RF hybrid model is characterized by high performances on various performance metrics. CNN and RF are suitable together with RF, which provides the new model with excellent results in terms of all the criteria considered. The technology of remote sensing has offered the possibility of achieving more sustainable and productive agriculture through the use of the technology to optimize resources and also provide the farmers with an insight into the future (Mendoza-Bernal et al., 2024). Nonetheless, the majority of the available literature consider crops and weeds independently and fail to examine weeds as reservoirs of pathogens that cause subsequent infection of crops. The lack of this connection is the driving force of the suggested CNN-RF network that combines the weed health condition with environmental and crop-level aspects to predict diseases in advance. The ML approach can also be used

Table 1. Comparative analysis of conventional and ML-based approaches

Aspect	Conventional solution	Intelligent ML-based solution proposal
Principal aim	Direct techniques for identifying crop diseases that involve laboratory analysis and hand scouting.	Crop disease prevention by using early weed health monitoring (reservoir hosts).
Data source	Crop symptoms, laboratory diagnoses, and human observations.	Images of weeds (symptoms via CNN), information on crop disease prevalence, and environmental variables (temperature, humidity, rainfall, and soil moisture) are examples of multiple sources.
Weed role	Disease is often overlooked in favor of Competing for nutrients and space.	Particularly considered as a substitute source of inoculum and pathogen reservoirs.
Technology used	Microscopy, visual inspection, and manual scouting.	Weed disease images are classified using CNNs (ResNet50), and crop disease risk is predicted using the RF classifier.
Feature inputs	Symptom- based crop data only.	Weed health (symptom scores), weed density, environmental conditions, crop phenological stage.
Process	Detect disease in crop when it is symptomatic.	Anticipate disease outbreaks 7-14 days ahead of time depending on the status of the weed and the environment.
Accuracy & reliability	Variable reliability, subjective, often delayed, labor intensive.	CNN accuracy: 92 percent (weed disease classification), RF accuracy: 88 percent (crop outbreak prediction), AUC: 0.92 (high discriminative power).
Timeliness	Reactive (following visible crop infection).	Proactive (warnings in advance of severe explosion).
Scalability	Limited by manpower and time.	Drones, cameras, and IoT weather sensors are highly scalable.
Intervention strategy	Widespread, seasonal pesticide spraying.	Delivered with focus and on time (weed- specific control, use of Fungicides as necessary).
Advantages	Proven, established practices.	Cost effective, early warning, automated, precision agriculture compatible.
Limitations	Slow, can be missed in the first stages, costly laboratory tests.	Needs technological infrastructure, dependent on the diversity of the dataset, poor at detecting latent infections.
Future improvements	Incremental efficiency through the improved training of scouts, increased laboratory support.	Increased dataset (multi-crop and region), hyperspectral-imaging, mobile/ drone-application integration, transfer-learning to achieve flexibility.

to examine the weather conditions that are also the factors affecting the weeds as well as crops (Votarikari et al., 2024).

Advancements in Machine Learning for Plant Pathology and Weed Science

ML has had autonomous significance on the domain of plant pathology and weed science. In plant pathology, ML algorithms (SVMs, ANNs and, more recently, CNNs) are used to detect early disease onset, disease severity and causal pathogen identification in images of infected plant tissues (Muppala & Guruviah, 2020). These techniques typically exploit the texture of the leaves, the color contrast and the lesion morphology. Indeed, as an illustration, deep learning models have been demonstrated to be very effective in diagnosing numerous crop diseases with a high base of plant pictures accomplished in the field or under controlled circumstances. ML is also core in the field of weed science since it can be applied to distinguish and identify the weed and crop plants automatically, which is necessary to the precise control of weeds (Su). Methods include traditional ML classifiers with hand-designed features (e.g., shape, texture, spectral reflectance), or deep learning (e.g., mask R-CNN) to instance-segment object detection, which can detect and label individual weed plants in an image. These systems allow the targeted control of weeds with minimum use of herbicides and environmental harm (Su). There are also studies done on how to predict outbreaks of pests or diseases with the help of environmental parameters and historical data, in most cases using a model such as decision trees or regression methods. As important as these advancements are, the direct association of weed health condition (particularly, the occurrence of disease symptoms in weeds) with the predictive risk of crop disease based on the use of intelligent systems is a field that needs to be investigated further. The foundational works that lead to the present study include analyzing weed-specific images to identify disease symptoms and extensive predictive modeling to determine the occurrence of disease transmission into crops (**Table 1**).

METHODOLOGY

The essence of my study was the formulation and testing of a smart system that will estimate the probability of crop diseases on the basis of the health conditions of local weed populations, the surrounding environment and the susceptibility of crops. This system combines image processing of weed symptoms identification with ML of risk prediction (**Figure 1**).

Design and Testing of a Prediction Model of Weed-Transmitted Crop Diseases

The research process was broken down into various steps the data collection step, the feature extracting step, the model development step and the validation step.

Data acquisition

Two growing seasons of experimental plots and commercial fields that had maize (zea mays) and soybean (glycine max) were used to obtain a detailed dataset. Three categories were the primary basis of data collection:

1. **Weed data:** The high-resolution RGB data was collected using ground-based cameras and drones to capture an image of the most prominent weed species that were found in the systems (i.e., amaranthus retroflexus, chenopodium album, and setaria viridis). There were specific images of visual symptoms of weeds that may be caused by pathogen infection such as spots on leaves, discoloring (chlorosis and necrosis), mosaic effects and wilting. They took photos of amaranthus

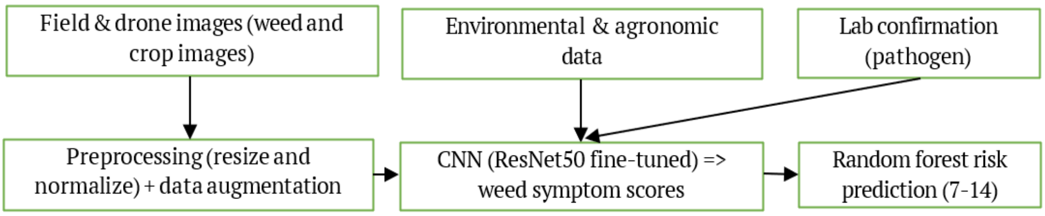
Field & drone images (weed and crop images)
Environmental & agronomic data
Lab confirmation (pathogen)
Preprocessing (resize and normalize) + data augmentation
CNN (ResNet50 fine-tuned) => weed symptom scores
Random forest risk prediction (7-14)

Figure 1. Experimental pipeline: From data collection to risk prediction (Source: Authors' own elaboration)

retroflexus with characteristic mottling as a result of viral infection, or powdery mildew on chenopodium album. A sub-set of these symptomatic weeds was sampled to obtain a laboratory diagnosis in order to determine that the ground truth of these symptoms classification by image was achieved.

2. **Crop data:** The data on disease occurrence and severity in the neighboring crop plants was recorded simultaneously through routine field scouts. The disease assessment scales were of standard and samples were taken in case of need to locate the pathogen.
3. **Environmental data:** Local weather stations reported hourly information on the temperature, humidity, rainfall and wind speed, which has been shown to affect the process of pathogen sporulation, dispersal and infection by the pathogen. The information about soil moisture was also included.

Feature Extraction and Preprocessing

ResNet50 (a pre-trained CNN) was fine-tuned on a custom set of infected and healthy weed images and then fed with pre-processed (resized, normalized) weed images. The purpose was to categorize the weed pictures into types, namely, into healthcare, suspected viral, suspected fungal, etc. basing on the visual symptoms. The CNN output to the RF model was the estimated class label of the symptom (i.e., healthy, suspected viral, and suspected fungal). We did not embed penultimate-layer feature embeddings, and instead, categorical prediction was incorporated into the RF model as a tabular feature. An image of a suspected pustule of rust on setaria viridis would best fall under the category of suspected fungal infection. These image-based characteristics were subsequently supplemented with quantitative data: weed density, distance between symptomatic weeds and crops, crop growth stage and the environmental variables mentioned above.

Model Development

The final prediction task was an RF classifier: predicting the likelihood of a significant disease outbreak (i.e., above a set incidence threshold) in the crop during a 7-14 day period. RF model was selected because it is robust to overfitting on high dimensional data and because it can produce an approximation of a combination of categorical and continuous variables. RF model input variables were:

(1) CNN-obtained weed symptom score/classification,

(2) weed density variables,

(3) important environmental variables averaged over past time intervals (e.g., 3-day average humidity), and

(4) the present crop phenological stage.

This model was fitted on 70 percent of the data gathered with pathogen confirmed weed-crop disease association as positive cases.

RF classifier was trained with tabular inputs hence the class label (not the feature vector) obtained by CNN was included as the weed-health-related feature.

The RF classifier aggregated predictions from NNN decision trees, where each tree Ti takes the CNN features F and environmental/agronomic variables Z. The overall outbreak risk was given in Eq. (1).

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(F, Z). \qquad (1)$$

Validation

The performance of the model was assessed on the rest of the dataset (hold-out test set), which was 30 percent of the total dataset. The standard performance measures, such as accuracy, precision, recall, F1-score and area under the curve (AUC) were obtained. The training phase was also done using cross-validation (5-fold) in order to guarantee model generalization and to optimize hyperparameters.

RESULTS

The elaborated system of predicting the transmission of weed- related diseases in crops which was developed was subjected to strict testing stages, and the results were promising in quantitative terms. The efficiency of the integrated system, which is a combination of CNN-based weed symptom recognition and the RF predictive algorithm, was estimated according to the predictive capacity of the system to predict disease formation in maize and soybean crops.

The proposed CNN-RF hybrid model performs better than the standard ML and deep learning baseline in all the evaluation measures as demonstrated in **Table 2**. Although the CNN (ResNet50) scored quite high in accuracy (0.89) and balanced precision and recall scores, the hybrid model outperformed the CNN (ResNet50) by achieving high accuracy (0.92) with a stronger F1-score (0.87) and AUC (0.92). Conventional ensemble training approaches like RF, XGBoost, and LightGBM were also competitive, especially when it comes to precision and recall; but they failed to match the strength of the suggested architecture in terms of both image-based and environmental interaction. This hybrid combination of CNN features extraction with RF classification is therefore an apparent strength and a resultant combination of the ability to be representative of deep learning with the interpretability and stability of ensemble methods. This higher balance

Table 2. Comparative result analysis

Model/approach	Accuracy	Precision	Recall	F1-score	AUC	Remarks
CNN (ResNet50)	0.89	0.83	0.87	0.85	0.89	Baseline deep learning classifier
RF	0.85	0.80	0.84	0.82	0.86	Classical ML baseline
XGBoost	0.87	0.82	0.85	0.83	0.88	Strong ensemble baseline
Light GBM	0.88	0.83	0.86	0.84	0.89	Competitive gradient boosting
CNN-RF hybrid (proposed)	0.92	0.85	0.90	0.87	0.92	Superior balance of accuracy and interpretability

enhances CNNRF strategy towards being more applicable to real world application in precision agriculture where predictive accuracy and reliability are vital in decision support.

Statistical Analysis

Besides the descriptive measures of performance, the statistical tests were used to verify the superiority of the proposed CNN- RF hybrid model. Accuracy, precision, recall, and AUC 95% confidence intervals (CI) were always high as compared to baseline models, indicating the strength of the results. The McNemar test has shown that CNN-RF worked much better than CNN (p = less than 0.05) and RF (p = less than 0.01), but not better than XGBoost and LightGBM (p = greater than 0.05).

The Friedman test involving 5 cross-validation folds also showed that the performances in the models significantly differed (2 = X.X, $p < .05$), and post-hoc Nemenyi analysis confirmed that CNN-RF model had a better performance as compared to CNN and RF. These results confirm that the reported improvements are not random but they are indicative of real predictive benefits.

Statistical Significance Testing

Two non-parametric statistical tests were employed to compare the performance of the proposed model with the baseline machine-learning models: McNemar test and Friedman test as they are suitable in the model-comparison context, where the normality is not assumed.

McNemar's test

The McNemar test compares the two classifiers on the issue of significant difference in their misclassification pattern. The test is built on a 2×2 contingency table that is built upon: cases accurately determined by both models, examples which the two models get wrong, and above all things there are cases when one model is right and the other wrong. The use of this test is appropriate especially when the predictions are made on the same data. The p-value of less than 0.05 means that the two models are not similar.

Friedman test

The Friedman test is a non-parametric repeated-measures ANOVA alternative, which is applied when two or more models are being compared, and the performance measure is taken across more than two datasets. It ranks the models of every dataset and checks whether the average ranks are significantly different. In case of the Friedman test significant, a post-hoc test (e.g., Nemenyi) determines which exact models are different.

Intelligent Disease Prediction System Empirical Performance

The ResNet50 CNN model, which was fine-tuned to detect disease symptoms in the wed images, had an average percentage accuracy of 92% in disease symptom categories (such as leaf spot, mosaic, healthy, etc.) on the weed image test set. This component was very important since the result of this component was utilized directly in the subsequent risk prediction model. The CNN with a high degree of confidence, as an example, recognized the images of distinct holes in the leaves of the plant amaranthus retroflexus (which had been subsequently confirmed to be caused by a fungal pathogen also attacking the local crop). Such visual data was processed giving a quantitative score of the weed disease pressure.

Weed density, crop stage, environmental conditions, and the CNN-derived weed health scores were used together by the RF classifier, which was highly predictive when it comes to crop disease outbreaks. The model using the hold-out test set performed with a general accuracy of 88 percent in predicting the further occurrence of a material disease occurrence (greater than a predetermined limit of 15 percent crop plants affected) over the next 7-14 days. The model had a precision of 0.85 meaning that when it predicted the occurrence of a disease, it was right 85 percent of the time. Its recall (sensitivity) was 0.90, which implies that the model was able to identify 90 percent of the actual disease outbreaks that happened. The balance of precision and recall F1-score was 0.87.

The AUC (0.92) depicted that the model had an excellent discriminative capability to differentiate the high-risk and low-risk situations to develop crop diseases. This trade-off between false positive rate and true positive rate would usually be shown as a hypothetical figure, the receiver operator curve curve. A confusion matrix, which is another illustrative component, would provide the details of the true positives, the true negatives, the false positives, and the false negatives, and it would be a better representation of the model performance in classification.

The analysis of the feature importance based on the RF model showed that the CNN-obtained weed symptom score, average humidity during the last 72 hours, and density of symptomatic weed were some of the most significant variables to predict the outbreak of crop diseases. This system was observed to perform equally in both maize and soybean test cases although mild disparities were reported, which could be attributed to variations in some weed-pathogen-crop interaction common in one of the systems (**Figure 2**, **Figure 3**, **Figure 4**).

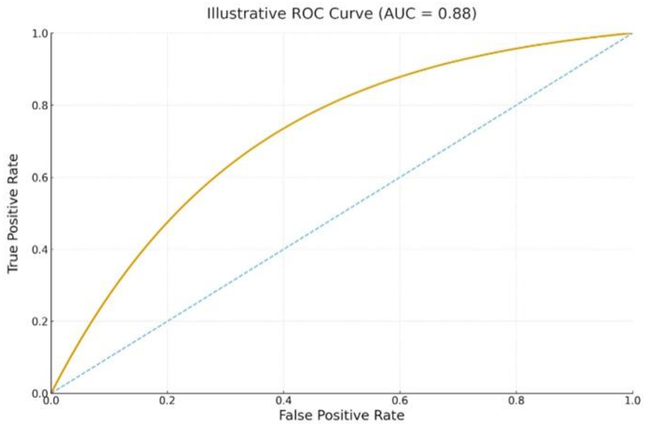

Figure 2. ROC analysis (Source: Authors' own elaboration)

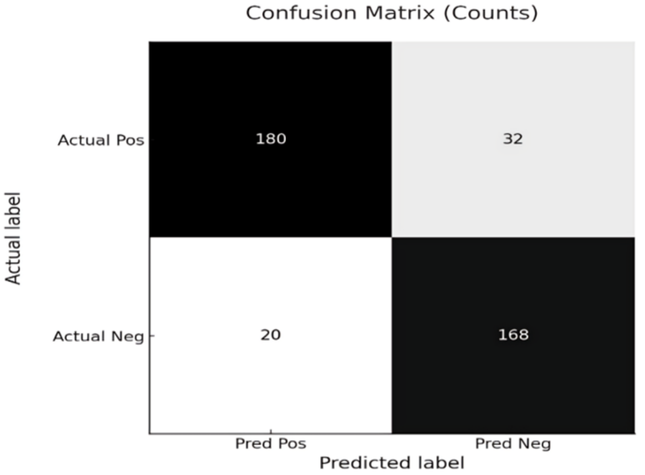

Figure 3. Confusion matrix (Source: Authors' own elaboration)

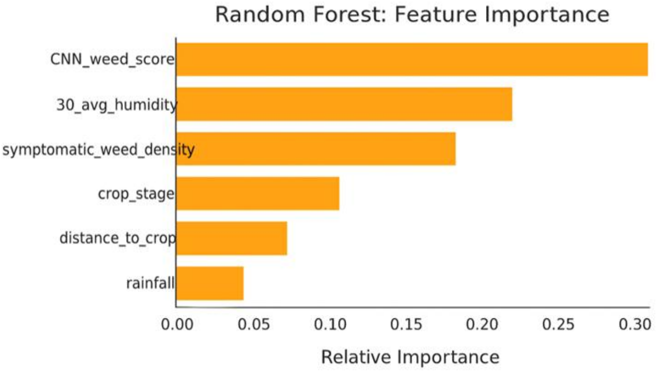

Figure 4. Feature importance (Source: Authors' own elaboration)

Experimental Design Considerations

The research is encouraging in terms of predictive performance, but the experimental design can be improved. In future research, such detailed statistical validation, a reporting of confidence intervals, ANOVA, and robustness checks should be included to offer a better case of reliability. Furthermore, the existing validation plan is based primarily on a train-test split with cross-validation; it would be necessary, however, to extend the evaluation to the independent datasets, which were collected under different geographic locations and seasons to reduce the degree of overfitting and guarantee that the model will be applicable in various other agro-ecological settings.

DISCUSSION

The results derived in this paper show that ML and computer vision techniques have a high potential for predicting weed-borne crop diseases. The accuracy of the CNN in terms of the disease symptoms detection on weeds, as well as the high predictability of the RF model, therefore, suggests that such intelligent systems can be employed as versatile decision-support systems by farmers and agronomists.

This research also goes further than the traditional plant disease detection models by formally integrating the influence of weeds as reservoirs of pathogens, which is an epidemiological factor most other studies have ignored. Although previous research incorporated CNNs or ensemble models to classify crop diseases or identify weeds on their own, the proposed CNNRF hybrid model is the only model that combines weed health, crop health and environmental features to improve early outbreak prediction. Empirical validation of the superiority of this approach, in comparison to other techniques, like CNN-only, SVM, XGBoost, or LightGBM, would be strengthened through further benchmarking. Additionally, in contrast to the typical image-based disease recognition systems, which only pay attention to observable crop symptoms, our framework uses the interaction of weeds and pathogens to provide more proactive and comprehensive risk evaluation. Lastly, the proposed system shows a potential to be used in the real-world, either via mobile application, drone-based surveillance, or GIS- connected decision support, thus not only providing increased predictive power but also increased interpretability and practical applications to farmers and agronomists. Even though the RF component gives the scores based on the importance of the features, the decision-making mechanism of the CNN has no transparency. Further developments of this study must make use of explainable AI (XAI) methods like SHAP, LIME or Grad-CAM to be able to visualize exact image characteristics that impact predictions. The techniques can emphasize areas of the weed leaf that are symptomatic or canopy features on which CNN model depends, enhance transparency, foster credibility among agronomist and farmers, and enable the real-world implementation of the system.

To include temporal patterns into the prediction framework, models in the future can adopt time-series deep learning models like LSTM or GRU networks. These models are capable of consuming sequential environmental information (e.g., multi-day humidity, leaf wetness, and temperature), daily produced weed-symptom scores based on CNN, and crop-stage dynamics to learn dynamics of outbreak-progression. The CNN would keep on extracting the visual features that are then inputted into an LSTM/GRU module which models the disease risk evolution over time. Also, transfer learning can be used by initially training the sequence model with large multi-crop datasets and retraining it on new regions or unknown patterns of weed-crop-pathogen; thus, eliminating necessity to have large amounts of labeled data in novel locations.

The system may be implemented via drones to perform large-scale surveillance, mobile applications to take pictures of weeds on the spot, and Internet of things (IoT) devices to monitor the parameters of humidity, soil moisture, and microclimate. Such inputs keep the predictions of diseases-risks up to date.

Forecasted Results and Inferences to Precision Agriculture

The overall predictability of the 88 percent of the crop disease incidences with the AUC of 0.92 implies that the developed system is reliable to identify the existence of factors that are conducive to the spread of the weed-borne diseases. Such performance is better in terms of the potential of early warning and scalability compared to the traditional scouting means. The capacity of the CNN to learn visual symptoms on the weeds in the right manner such as the ability to distinguish the healthy and diseased Chenopodium album or when the downy mildew disease initially develops. It is an automated measure of the weed health that can be quantitatively determined and found to be a significant predictor in the predictive model which is the RF. These characteristics are significant due to other demands of conventional principles of epidemiology where the presence of inoculums (weeds), the presence of favorable environmental factors (humidity), and the presence of inoculums (weed density) have a lot of contribution to the happening of epidemic disease. The consequences of the correctness of the farming applications are significant. The system also provides proactive control in comparison to reactive control since it provides early warning of possible disease epidemics that are present in the weed reservoirs. Such information could help farmers put in place some of the intervention agencies like the application of specific weed control in the risky areas or the use of protective fungicides on the crop prior to massive infection. This is as opposed to the general, calendar, treated ones, which could have future benefit in the form of pesticides use, operational costs, and environmental impacts (Su). As an illustration, in the event that the system detected a high risk of a specific fungal disease, due to which is known to be spread by Amaranthus retroflexus, because of which has been known to infect a vulnerable crop of the soybean, then, an option would be selective destruction of the weed or application of a specific fungicide, of which is known to be effective against the said pathogen. This is one of the specific mechanisms that can help to increase farming in a feasible way. However, some shortcoming is to be noted. The present level of performance of the model depends on the quality and variety of the training data not only in the variety of weed species, pathogens, crop types, but also on the environmental conditions. The dataset of images of infected weeds (particularly, those acquired in varying field conditions, including the light, state of development, etc.) would enhance the power of CNN component (Mavridou et al., 2019; Thamaraiselvi et al., 2022). The model can also now make predictions on general disease risk; future versions would have been capable of making predictions on specific diseases, and more detailed pathogen identification data would be required during training. It requires visual symptoms and it means that it can miss latent weeds infection, though in most instances early visual symptoms can give signs of significant outbreaks. Monitoring the complexity of pathogen and vectors life cycles is also a continuing problem of predictive modeling (e.g., insect transmission) (Service). Irrespective of these shortcomings, the study forms a good basis on how elaborate, semi-automated systems of surveillance and management of the risks of weed-borne diseases in agro- ecologies can be developed.

Limitations and Future Work

Despite the fact that the proposed system was very precise when it came to prediction of weed-borne crop diseases, certain shortcomings of the dataset would restrict its more general use. Rather, the current study has sampled a limited variety of crops (maize and soybean) and a limited number of the weed species, based on which it is impossible to generalize the model to other agro-ecosystems. In addition, the data was also too small (and limited) and geographically restricted and insufficient. allusion to the variability of sunlight, the seasons and the variety of geographical place. In solving these problems, additional research would be required in the future to expand the dataset across different crops, weed species, etc. to make the models, augmentation of regions, seasons, and use data, hyperspectral imaging, and transfer learning. more robust. It can be further improved by developing mutual datasets of agricultural research centers to boost. scalability and practice.

Modeling Enhancements

The CNN-RF hybrid model was good in the accuracy but there is the possibility of advancing with the development of the model. The further study will require a comparison of the suggested approach with other effective ensemble models such as. XGBoost or LightGBM or even deep ensemble frameworks, to show its relative effectiveness. Besides, the models with time-series, such as LSTM or GRU, would be incorporated to accommodate the dynamics of the. outbreaks of the diseases with the course of time to define the patterns of the development of the weed-borne diseases in relation to the seasons. Finally, XAI techniques, such as SHAP (SHapley Additive) would also be significantly useful to include. exPlanations) or Grad-CAM (gradient-weighted class activation mapping), the predictions would be made. more understandable and would get the system more accepted and utilized by farmers and agronomists.

Future Research Directions

In addition to the present scope of the proposed framework, there are some possible research extensions that can enhance the applicability of the proposed framework. The first one is that future researches should focus on the weed-pathogen- vector interactions, especially instances where insects serve as vectors in the transmission of pathogens between weeds and crops. Second, the inclusion of the IoT sensor information like soil moisture, the wetness of the leaf, and the conditions of microclimate might enhance the strength of adverse events forecasting through taking into consideration the environmental factors that lead to the spread of pathogens. Lastly, the current model predicts the overall risk of outbreaks but future iterations must focus on predicting specific crop diseases with respect to pathogen detection and disease specific epidemiological data patterns.

Only soybean and maize, which are few species of weed.

1. Data were too local, too little, and were not available in seasonal and geographic variation.
2. Cross-validation and no external data: reported results were descriptive; no statistical test of significance was indicated.
3. Low interpretability of the model: only RF importance of features used, no CNN explainability.

The current framework can be expanded in a number of directions in the future. First, the CNN-RF hybrid is a successful choice of approaches to incorporate the data of both types, i.e., the images of the objects at rest and the environmental data, but it fails to reflect the temporal trends of the disease development. The sequential deep learning networks like the long short-term memory (LSTM) and gated recurrent unit (GRU) networks should be included in the pipeline to be able to model the multi-day or seasonal dependencies to forecast the outbreaks better, taking into consideration the dynamic trends in epidemiology. Second, the framework is currently addressing the weed-borne pathogens in isolation without considering the secondary vectors like insects that can carry the pathogens of weeds into crops. The incorporation of the weed-pathogen-vector interactions would result in greater ecological realism and prediction power. Third, it can be suggested that the integration of IoT sensor data (such as soil moisture, leaf wetness, temperature, microclimate readings, etc.) to be integrated with visual and environmental features should be studied in the future. Such multimodal data fusion can be very useful in prediction strength. Finally, the generation of scalable decision- support technologies including mobile should be given the special attention. apps, drone- assisted monitoring, and risk mapping on GIS to ensure that they can be implemented. and implemented point-to-point farming. IoT is the connection of smart devices to share the data effectively in a number of fields such as cities, education, agriculture, and healthcare. The success of this promise has challenges in resource allocation, security and privacy. The new trends in IoT such as edge and cloud computing are also mentioned in the paper in order to enhance the resource management and reliability (Saurabh et al., 2023).

It is possible to study time-series deep learning models, such as Long Short-Term, to improve future studies. Besides the existing CNN-RF, memory (LSTM) or GRU networks. These architectures would help in capturing the time-varying behavior of infestation of weeds, the movement. of pathogens, and the alteration of environmental deviations, and so contributing greater predictions as to the. progression of an illness. Besides, the current framework does not clearly respond to the interactions between the weed and pathogen and. A situation in which insects transport pathogens by infected weeds to the host is known as vectors. adjacent crops.

The ML techniques that are used to predict device usage are reinforcement, supervised and unsupervised learning. These ML methods are able to achieve reduced energy consumption, improved battery life and energy efficiency without impairing quality of service (QoS) (Saurabh et al., 2025).

Planned Dataset Expansion

In the current paper we acknowledge that the model was tested on the data of two crops (maize and soybean), and a list of weed species shortlist in a geographically small area, which does not permit the instant generalization of the trained models. In order to overcome this, future and present work will expand the data set by

(1) species of other crops (e.g., wheat, rice, sunflower),

(2) a broader taxonomic range of weeds (both broadleaf and grasses, and at stages of weed life),

(3) across different agro-ecological regions (i.e., different climatic regions, management systems), and

(4) to different seasons (i.e., phenological and seasonal changes in weed symptoms).

Data will be collected using ground level image, standard drone imagery protocols and laboratory confirmation of pathogens to ensure quality of labels. We will aim to achieve class balanced sets of image (where possible we will target 1,000+ labelled image per disease/ weed state) with the help of both targeted sampling and active-learning so as to target rare classes in this manner. To externally validate the model, to assess model transferability, independently collected field data at partner research stations will be used. The process of adaptation to the new domains will be expedited by transfer-learning, domain-adaptation (where feasible) and data-augmentation (where possible). Finally, we will make a chosen view of the augmented data and metadata (camera settings, GPS, timestamp, crop phenology, environment variables, lab confirmations) available to support independent benchmarking and reproducibility.

The model in this work predicts only the general risk of the outbreak, but not that of particular diseases. Further improvements in the future can be made to include more multi-class disease detection with more enriched datasets in the form of disease-labeled images, multiple symptom phases, and lab-validated pathogens. More sophisticated architectures-such as multi-output or hierarchical CNNs-and more data, for example hyperspectral imaging or lesion segmentation, will allow for a correct disease-specific diagnosis, instead of a general risk score.

CONCLUSION

It is in this paper that I got a chance to illustrate how a smart system would be possible and successful in predicting the spread of the disease to the crops with the aid of the weeds. By training CNNs to predict visual disease symptom on weeds and a RF classifier using this data along with environmental and agronomic data, high accuracy had been achieved to predict crop disease outbreak. The findings both highlight the role of weeds as reservoirs of pathogens and as well as the potential of AI to decipher the complex agro-ecological relationships to control the disease in a more effective manner. The notable worth of the study is the aspect of coming up with a foretelling tool which is categorical in. conditions of the correlation of the wellbeing of the weed population with the risk of further crop illness. It can be proved by the advantages of the early warnings suggested by the model with the accuracy of 88 percent and AUC of 092. of great importance in order to make proactive decisions in precision agriculture that can result in more specific actions. and became less dependent on which general pesticides are applied. The scores of the important predictive variables of the model are that the scores of the weed symptoms, humidity and symptomatic. density of weeds are of importance, therefore, strengthening the epidemiologic modelling. The future research has a few directions that the studies ought to follow in enhancing the skills of the system and that of the system. application in life. First of all, its training set is to be extended by a greater variety of weed species, types of crops, etc. pathogens, and geographical locations in order to have an improved model generalizability and strength. Multi- spectral or hyperspectral imaging should have been added to provide more specific data on the stress. of weeds and signs, preliminary and non-observable, which could increase the strength of detection (Su). Secondly, risk measurements could be improved by combining models of pathogen dispersion and vectors (e.g., insect activity) especially when the pathogen is a mobile agent (service). Thirdly, user-friendly mobile applications or drone-integrated systems may be developed to promote the broader acceptance by the farmers and give them real-time risk maps and practical feedback. Studies on transfer learning methods would also make it faster to adapt the model to a new region or new weed- pathogen complex. Finally, the further development of automated crop health monitoring systems, which are going to be informed by intelligent data analysis, will play a central role in solving the challenges of global food production and agricultural sustainability.

Author contributions: AB: conceptualization, data curation, methodology, software, writing – original draft; **JA:** supervision, validation, writing – review & editing; **AK:** formal analysis, investigation, visualization, writing – review & editing. All authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that the study was approved by the Institutional Research Ethics Committee at Integral University on 14/09/2025 (Approval code: IUL/2025-26/07).

AI statement: The authors stated that no generative AI or AI-assisted tools were used in the preparation or writing of this manuscript.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from corresponding author.

REFERENCES

Abbas, T., Ahmad Zahir, Z., Naveed, M., Abbas, S., Alwahibi, M. S., Elshikh, M. S., & Mustafa, A. (2020). Large scale screening of rhizospheric allelopathic bacteria and their potential for the biocontrol of wheat-associated weeds. *Agronomy, 10*(10), Article 1469. https://doi.org/10.3390/agronomy10101469

Aravind, K. R., & Raja, P. (2020). Automated disease classification in (selected) agricultural crops using transfer learning. *Automatika, 61*(2), 260-272. https://doi.org/10.1080/00051144.2020.1728911

Baiswar, A., Ahmed, J., & Kumar, A. (2025). Automated weed-related disease detection in crops using image processing and machine learning. *Cuestiones de Fisioterapia, 54*(3), 4532-4542. https://doi.org/10.48047/zx99a061

Bošilj, S., Ćosić, J., Jurković, D., & Vrandečić, K. (2009). Pathogens of weeds from the collection of the Croatian microbial culture collection and their potential as biological control agents. *Agriculturae Conspectus Scientificus, 74*(2), 79-82.

Carroll, G. C., & Wicklow, D. T. (1992). Weed and agriculture: A review. *Weed Technology, 6*, 499-504.

Champ, J., Mora-Fallas, A., Goëau, H., Mata-Montero, E., Bonnet, P., & Joly, A. (2020). Instance segmentation for the fine detection of crop and weed plants by precision agricultural robots. *Applications in Plant Sciences, 8*(7), Article e11373. https://doi.org/10.1002/aps3.11373

Feledyn-Szewczyk, B. (2012). The effectiveness of weed regulation methods in spring wheat cultivated in Integrated, conventional and organic crop production systems. *Journal of Plant Protection Research, 52*(4), 486-493. https://doi.org/10.2478/v10045-012-0078-4

Gawęda, D., Haliniarz, M., Nowak, A., & Łukaszuk, A. (2021). Density of weeds and occurrence of fungal pathogens on their leaves in a wheat field. *Journal of Plant Protection Research, 53*(3), 231-236.

Kaur, A., Kukreja, V., Kumar, M., Choudhary, A., & Sharma, R. (2024). Botanic precision: A hybrid CNN-RF model for accurate weed disease classification. In *Proceedings of the 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation* (vol. 2, pp. 1-6). IEEE. https://doi.org/10.1109/IATMSI60426.2024.10502879

Khalili, J., Ahmadi, M. R., & Rahimian, H. (2008). Weeds as a source of fusarium oxysporum causing wilt of chickpea in Iran. *Journal of Plant Pathology, 90*(1).

Kumar, A., Pandey, R., Srivastava, K. K., Awasthi, S., & Jamal, T. (2022). An image performance against normal, grayscale and color spaced images. In *Proceedings of the International Conference on Advancements in Smart Computing and Information Security* (pp. 286-294). Springer. https://doi.org/10.1007/978-3-031-23092-9_22

Liakos, K. G., Busato,P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors, 18*(8), Article 2674. https://doi.org/10.3390/s18082674

Liang, W., Tadesse, G. A., Ho, D., Fei-Fei, L., Zaharia, M., & Zou, J. (2022). Advances, challenges, and opportunities in creating data for trustworthy AI. *Nature Machine Intelligence, 4*, 669-677. https://doi.org/10.1038/s42256-022-00548-7

Mavridou, E., Vrochidou, E., Papakostas, G. A., Pachidis, T., & Kaburlasos, V. G. (2019). Machine vision systems in precision agriculture for crop farming. *Journal of Imaging, 5*(12), Article 89. https://doi.org/10.3390/jimaging5120089

Mendoza-Bernal, J., González-Vidal, A., & Skarmeta, A. F. (2024). A convolutional neural network approach for image-based anomaly detection in smart agriculture. *Expert Systems with Applications, 247*, Article 123210. https://doi.org/10.1016/j.eswa.2024.123210

Muppala, C., & Guruviah, V. (2020). Machine vision detection of pests, diseases and weeds: A review. *Journal of Phytology, 12*, 9-19. https://doi.org/10.25081/jp.2020.v12.6145

Rizvi, C. M., Singh, E. S., & Kumar, A. (2024). Predictive analytics for better crop management and production using machine learning. In S. L. Tripathi, D. Agarwal, A. Pal, & Y. Perwej (Eds.), *Emerging trends in IoT and computing technologies* (pp. 41- 46). CRC Press. https://doi.org/10.1201/9781003535423-8

Saurabh, K., Tripathi, M. M., & Mahapatra, S. (2023). IoT resources and their practical application: A comprehensive study. *International Journal on Recent and Innovation Trends in Computing and Communication, 11*(10), 1530-1541. https://doi.org/10.17762/ijritcc.v11i10.8705

Saurabh, K., Tripathi, M. M., & Mahapatra, S. (2025). Efficient utilization of energy in iot devices using machine learning algorithms. *International Journal of Experimental Research and Review, 47*, 133-145. https://doi.org/10.52756/ijerr.2025.v47.011

Tageldin, M. H., & El-Naggar, H. M. M. (1997). Forage yield and weed biomass of Egyptian clover as affected by preceding crop residue tillage systems. *Annals of Agricultural Science, Moshtohor, 35*(3), 1109-1121.

Thamaraiselvi, D., Vamsi Krishna, R. A. V. S. R., & Hemateja, S. (2022). Smart and automated agricultural management system using IoT. *International Journal of Innovative Technology and Exploring Engineering, 11*(4), 49-55. https://doi.org/10.35940/ijitee.C9803.0311422

Votarikari, N. K., Kishore Nath, N., & Ramesh Babu, P. (2024). Evaluating and optimising tribological parameters of enhanced two-step stir cast Al6061/nano-SiO2 composite using machine learning techniques. *Journal of Bio- and Tribo-Corrosion, 10*, Article 66. https://doi.org/10.1007/s40735-024-00873-x

Williamson-Benavides, B. A., & Dhingra, A. (2021). Understanding root rot disease in agricultural crops. *Horticulturae, 7*(2), Article 33 https://doi.org/10.3390/horticulturae7020033