



Potato Leaf Disease Classification Using Optimized Machine Learning Models and Feature Selection Techniques

Marwa Radwan¹ · Amel Ali Alhussan² · Abdelhameed Ibrahim³ ·
Sayed M. Tawfeek^{4,5}



Received: 20 May 2024 / Accepted: 25 June 2024
© The Author(s) 2024

Abstract

The diseases that particularly affect potato leaves are early blight and the late blight, and they are dangerous as they reduce yield and quality of the potatoes. In this paper, different machine learning (ML) models for predicting these diseases are analysed based on a detailed database of more than 4000 records of weather conditions. Some of the critical factors that have been investigated to determine correlations with disease prevalence include temperature, humidity, wind speed, and atmospheric pressure. These types of data relationships were comprehensively identified through sophisticated means of analysis such as *K*-means clustering, PCA, and copula analysis. To achieve this, several machine learning models were used in the study: logistic regression, gradient boosting, multilayer perceptron (MLP), and support vector machine (SVM), as well as *K*-nearest neighbor (KNN) models both with and without feature selection. Feature selection methods such as the binary Greylag Goose Optimization (bGGO) were applied to improve the predictive performance of the models by identifying feature sets pertinent to the models. Results demonstrated that the MLP model, with feature selection, achieved an accuracy of 98.3%, underscoring the critical role of feature selection in improving model performance. These findings highlight the importance of optimized ML models in proactive agricultural disease management, aiming to minimize crop loss and promote sustainable farming practices.

Keywords Early blight · Feature selection · Late blight · Machine learning · Neural network · Potato leaf diseases · Predictive modelling · Sustainable agriculture · Weather parameters

Introduction

Potato is the largest non-cereal food crop worldwide, and it is widely consumed in the world, after rice, wheat, and maize. First domesticated in the Andean Mountain regions of South America, the cultivation of potatoes has, of course, been

Extended author information available on the last page of the article

spread far and worldwide to flourish under a wide range of climatic conditions. These make potatoes strong and versatile; these are the underpinning of high nutritional value and great importance for securing food safety and economic stability among varied parts of the globe. This dependence, in turn, makes the crop vulnerable to diseases whose yield and quality effects may be very severe, hence imposing high economic costs (Dolničar 2021; Singh et al. 2021). Out of these diseases, early blight and late blight are particularly destructive. Early blight results from the fungus *Alternaria solani*. Typically, it is characterized by tiny, dark-brown lesions on leaves that interfere with photosynthesis and hasten plant senescence. On the other hand, late blight, initiated by the pathogen *Phytophthora infestans*, is notorious for having caused the Irish Potato Famine in the 1840s. This disease appears as wet rot that spreads fast; it can destroy whole fields under conducive environmental conditions (Gold et al. 2020a, b).

The management of this group of diseases depends on traditional methods: routine visual surveying followed by necessary applications of chemicals. Such types of interventions are labor-intensive and may cause environmental degradation. There are efforts made in the last years to integrate artificial intelligence (AI) technology into agricultural practice as a new approach to managing crop diseases. Other AI and machine learning models are used increasingly in applications such as predicting potential disease outbreaks, optimizing the treatment plans of diseases, and further decreasing the damage to crops caused by reductions in chemical usage (Kang et al. 2023). AI is using these models through a combination of datasets like satellite images and aerial drone data—all the way to ground-level sensors that detect and analyse patterns suggestive of impending diseases in the area. Even after this type of development, the concrete linkage of weather-related conditions with the prevalence of potato diseases is still little explored. Prime conditions that have a key influence on weather variables, such as temperature, humidity, wind speed, and atmospheric pressure, have a major impact on the life cycles and spread of the pathogens of early and late blight (Gold et al. 2020a, b).

This research focuses on a complete dataset that can encompass the details of the weather parameters and records of potato leaf disease. The data and analysis received are being used in AI with the help of audio-visual techniques to create a model for disease outbreaks based on changing weather conditions. Establishing such a predictive instrument would be accommodating in refining the agricultural processes since a rapid set of procedures can be initiated to prevent crop deterioration and minimize the application of chemical intervention (Gao et al. 2021). This is why optimizing these AI models is highly significant in improving the models' predictive precision, effectiveness, and dependability. This optimization process has various technical considerations, including fine-tuning weather parameters, improving the algorithms of the artificially intelligent systems, and, perhaps, the best one—testing the models against the actual manifestation of diseases. They should also be grounded with specific practice aspects of these models in genuine agricultural environments. It must be integrated with existing systems used in managing agriculture and consider the geographical and climatic conditions in the locations. The staff users do not necessarily have to possess professional knowledge of AI in agriculture (Yang et al. 2021).

The shift brought by the integration of AI into the domain of agricultural disease management turns toward more sustainable and resilient farming practices. The models help predict the possibility of disease occurrence in relation to weather conditions and, thereby, empower farmers to take proactive and informed measures in safeguarding farm crops, in return ensuring high crop productivity, low environmental impacts, and enhanced economic returns. The research will, therefore, substantially contribute to the growing field of agricultural science because it will lend clarity to the relationship that exists between certain weather conditions and the manifestation of potato leaf diseases, therefore making better operationalized agrarian decisions (Arshaghi et al. 2023).

This research will employ current and trending machine learning (ML) methodologies commonly used to study the correlation between diseases and weather in potatoes. These compare selected models, namely, the decision trees, support vector machines, and the neural networks for predicting disease outbreak events. It includes data preprocessing, where normalization of the weather variables and encoding of categorical data are done to protect the variation in model accuracy. Therefore, optimizing all these models includes investigating their hyperparameters (Fenu and Mallocci 2020). This can be achieved through hyperparameter tuning techniques such as grid search and cross-validation for configuration seeking, which offers better performance. In the analysis phase, various significant factors about each disease and the most critical weather parameters that influence these diseases are established by feature importance analysis. This has been done by means such as the Gini importance in the decision trees or the weights in linear models to mirror the climatic factors in the development of early and late blight. The specific analyses that these levels of detailed data would enhance include a better understanding of disease mechanisms and improving predictive models to home in on the most relevant predictors for prediction and boost the efficiency and efficacy of the predictions (Meno et al. 2021).

Concerning the further development of the models used in this research, their application can be utilized to develop agri-environmental decision support systems. Barring these, it would presumably be through accessible interactive interfaces where the farmers and agriculture managers could input information about the current weather into the systems, providing nearly instantaneous and real-time responses regarding the likelihood of disease outbreaks. Such tools are meant to deliver timely and accurate treatment, for instance, applying fungicides to crops, which may be provided only if and wherever they are needed most, thus saving costs and minimizing the burden on the environment (King et al. 2020). In this research, further tests on the scalability of the models were to be explored in terms of their applicability in other environments to enhance their performance in various regions and climates. For this to be achieved, one would have to optimize weather information and diseases in a way that makes the tool fitting and suitable for use across the various global settings. The adaptation elements increase the effectiveness and value of the models in this research. As such, it becomes a value addition for international agriculture groups and stakeholders (Waaswa et al. 2022; Tang et al. 2024).

The main output of this work is to introduce an AI-based tool for predicting diseases in potato leaves with high accuracies, using real or forecasted weather

data. This can enable more precise and proactive disease management in potato cultivation, which can cumulatively translate into huge improvements in crop yield and pesticide use. Other general implications are as follows: improving food security, minimizing environmental degradation from chemicals used in farming, and more sustainable farming practices across the world. This current research can be considered a step forward in the integration of the current AI technology into this field of research in agricultural disease management. The proposed methodology in this work consists of several stages to achieve the required output. Pre-processing and analysis are the first stages. The *K*-means clustering and the principal component analysis methods are employed first to preprocess the dataset. The dataset for this research holds 4020 records with robust weather information, including temperature, humidity, and wind speed (Yeasmin 2023). Data analysis is done based on copula analysis. Copula analysis is a statistical technique employed to investigate the relationships between a collection of variables. The next stage is the feature selection, including applying the binary Greylag Goose Optimization (bGGO) (El-kenawy et al. 2024) and the binary Waterwheel Plant Algorithm (bWWPA) (Alhussan et al. 2023). This stage helps improve predictive accuracy by isolating the most relevant features. The final stage is the classification. The ML models, including logistic regression, gradient boosting, multilayer perceptron (MLP), support vector machine, and *K*-nearest neighbors models, are employed for potato leaf disease classification. The models are applied with and without the feature selection process to show the importance of feature selection.

The contributions made by this paper encompass a range of significant advancements and insights into the field of agricultural disease management through the application of machine learning techniques. These contributions include the following:

1. The study provides a detailed investigation into potato leaf diseases, specifically early blight and late blight, highlighting their impact on crop yield and quality.
2. Utilizes a comprehensive dataset of over 4000 weather records, including parameters such as temperature, humidity, wind speed, and atmospheric pressure, to predict disease outbreaks.
3. Employs advanced data preprocessing techniques like *K*-means clustering and principal component analysis (PCA) to uncover significant data relationships and improve model training.
4. Uses copula analysis for a deeper exploration of the relationships between various weather parameters and disease outbreaks.
5. Implements and compares various machine learning models, including logistic regression, gradient boosting, multilayer perceptron (MLP), support vector machine (SVM), and *K*-nearest neighbors (KNN).
6. Applies feature selection algorithms such as binary Greylag Goose Optimization (bGGO) and binary Waterwheel Plant Algorithm (bWWPA) to enhance predictive accuracy by isolating the most relevant features.
7. Evaluates the performance of machine learning models both with and without feature selection, demonstrating the significant improvement in accuracy when feature selection is applied.

8. Reports that the MLP model, with feature selection, achieved an accuracy of 98.3%, highlighting the effectiveness of feature optimization in disease prediction.
9. Emphasizes the importance of optimized machine learning models in proactive agricultural disease management, aiming to minimize crop loss and promote sustainable farming practices.

Related Work

The integration of artificial intelligence (AI) and machine learning (ML) in agriculture is a transformative shift from traditional empirical practices. This section reviews the existing literature and highlights key advancements in this field, organizing the discussion into subsections that address various aspects of AI and ML applications in managing agricultural diseases.

Revolutionary Use of AI and ML in Agriculture

The use of technology, especially AI and ML, in managing agricultural diseases is largely revolutionary compared to traditional practices—empirical and manual (Sharma et al. 2021). The history of plant diseases will show that the prognosis and control depended on empirical approaches, where the records of outbreaks and the weather data were used for forecasting. That helps, but it is an imprecise approach that does not provide flexibility in accommodating the variability and unpredictability of the conditions (Viana et al. 2021). In agriculture, the induction of AI is bringing its own new set of ultra-modern tools that can even analyse a complex set of data and be able to predict, for example, when and where diseases might attack (Hamrani et al. 2020). These would, therefore, be using data inputs from satellite images showing large-scale environmental changes to fine detail and high-resolution images taken from drones, all coupled with real-time data from the ground-based sensing of micro-climatic conditions (Shin et al. 2020). In this case, machine learning algorithms, of course, are much more complicated than simple neural networks, and they process the data to find the patterns and anomalies within it. And these could very well be signals of the onset of a developing disease (Cravero et al. 2022).

Key Climatic Factors in Disease Prediction

Part of this has included the development of AI models for processing inputs related to several key climatic factors, particularly those required in the context of managing diseases such as early blight and late blight. When considering the development of diseases like late blight, temperature and humidity are two important factors (Garske et al. 2021). Speed and direction of the wind can also be major considerations for the spread of fungal spores and, hence, should be characteristics that need to be considered

in predictive models. With these and other related variables in mind, the AI systems can give timely projections about disease outbreaks, allowing actions beforehand that can avert massive destruction of crops (Qazi et al. 2022). Their optimization is key, needing to make sure that the performance and reliability of the models should excel. This goes far beyond the choice of relevant features that lead to higher precision in predictions. Fine-tuned algorithms can then also be adjusted to specific conditions within another farming region (Wang et al. 2020; Tang et al. 2021; Zhang et al. 2022). Techniques like the analysis of feature importance make it easier to understand which of the model's variables influence predictions most significantly. For example, the analysis of the feature importance might point out that humidity and temperature levels at some thresholds predict the blight outbreak very well and guide the model parameter adjustments to focus even more precisely on such ranges (Zhang et al. 2020; Wang et al. 2021).

Development of User-Friendly AI Tools

Besides, the very pivotal development for the accessibility of high-tech solutions by the agri-community will be the development of more user-friendly interfaces for these AI tools. The interfaces typically have a dashboard showing real-time data and forecast data with advice on the action to take in return to the forecasted conditions. Its design should be thoroughly user-centric for adoption and its optimum utilization since it should make sure that non-technical farmers can even use high-tech tools to protect their crops (Bhat and Huang 2021). As AI technology expands its tentacles, so does its applicability in agriculture. Current studies are pushing the envelope, with researchers applying AI to genetic data to predict how different potato varieties may react when exposed to certain diseases after being grown in different weather patterns (Khan et al. 2021). This integration of genetic and environmental data promises a revolution in how we approach disease prevention and management in agriculture. It will raise the hope of further increasing crop yields and sustainability by ensuring disease-free, healthier plants (Shaikh et al. 2022).

Broader Impacts of AI on Agricultural Operations

In addition to the direct impact of AI on agriculture in the form of disease prediction, the effect of AI on agricultural operations is felt in multiple other aspects. AI systems are now becoming more dominant in agricultural activities, and this heralds the advent of accurate farming. This practice is marked by an intensive manner of natural resource utilization, which reduces waste and increases output through the accurate usage of water, fertilizers, and chemicals. Precise forecasting of disease epidemics facilitates the implementation of pre-emptive action that is only necessary when the situation demands it, thus lowering the environmental footprint of agricultural activities (Zhang et al. 2021). Such a transition is spurred on by the continually growing body of research that focuses on the intricate relationships between plants, pests, diseases, and the surrounding environment. AI models analyse and process a vast array of data from different sources to detect patterns that cannot be easily identified via simple observational methods. To illustrate, the relationship of minor climate

changes with disease outbreaks can improve the breeding of drought-resistant crop varieties and effective crop management strategies (Subeesh and Mehta 2021).

Socio-economic Benefits of AI in Agriculture

The impact of the inventions on the community is also prominent. Using AI technologies in agriculture combined with the rise of predictable crop yields and a drop in losses due to pests and diseases, AI can achieve financial stability in the world of farmers, especially those in vulnerable regions where climate change variability creates numerous impediments. Furthermore, accurate and timely predictions of AI can contribute to better planning and resource use, which is essential to building a more sustainable economy in the agricultural sector (Nawaz et al. 2022). The tech-driven nature of this development is achieved through a synergy between agronomists, data scientists, and farmers. Each group provides an important viewpoint and expertise; thus, the approaches devised consider scientifically sound and applied possibilities. Agronomists research the biological and ecological aspects of farming, data scientists will continue to improve prediction models, and farmers will continue to provide pragmatic feedback on the application of these models in real situations (Linaza et al. 2021).

Future Directions and Innovations

The current trend with AI indicates that its integration in agriculture will continue to grow in the future thanks to further research, which will lead to more elaborate uses. Such tasks include the employment of AI for automated pest identification and control, real-time soil health monitoring, and the development of optimal single-crop rotation methods that enhance soil fertility and crop yields. Every one of these promised advancements in AI technology will greatly improve the efficiency, sustainability, and output of agricultural systems across the world (Gupta et al. 2020).

This research, therefore, adds to the body of knowledge in agricultural science by elaborating that weather conditions play a critical role in the prevalence of potato leaf diseases and how AI models can be optimized for effective prediction; in other words, such models have higher power for real field use. It provides practical tools that can be used to increase the resilience of potato crops against these climatic challenges, thereby supporting sustainable agricultural development and assuring food security in the era of changing global climate to facilitate prediction accuracy.

Proposed Methodology

The proposed methodology aims to harness machine learning techniques to forecast potato leaf disease outbreaks based on various weather conditions. This involves careful data collection, preprocessing, feature selection, and the application of different machine-learning models. The methodology ensures that critical information is identified and prioritized for predictive accuracy, leading to proactive agricultural management.

Dataset

The dataset for this research holds 4020 records with powerful weather information such as temperature, humidity, wind speed, wind direction, visibility, and atmospheric pressure (Yeasmin 2023). The interactions between these environmental variables determine breeding areas and the way diseases spread. Two columns, ‘Disease name’ and ‘Due to a number of diseases’, describe types of potato leaf diseases; additional information is added, early blight and late blight, necessary for the machine learning model training. Early blight, due to *Alternaria solani*, is most favoured in warm and humid conditions, producing dark brown spots on the leaves that are then unable to complete photosynthesis. In contrast, late blight, caused by *Phytophthora infestans*, is more likely in chilly and humid regions and can ruin a whole field at the speed of light under favourable conditions. The table with weather and disease descriptions extends the ability to build models that give forecasts of the disease based on changing climate.

The correlation matrix of Fig. 1 demonstrates a variety of feature relationships. The highest correlations (near 1) between several weather parameters enable us to determine which conditions are contributing to diseases the most, allowing for the most accurate prediction of outbreaks. For instance, these variables, such as temperature and humidity, may shoot to the sky with a strong positive correlation, leading to the occurrence of both early blight and late blight.

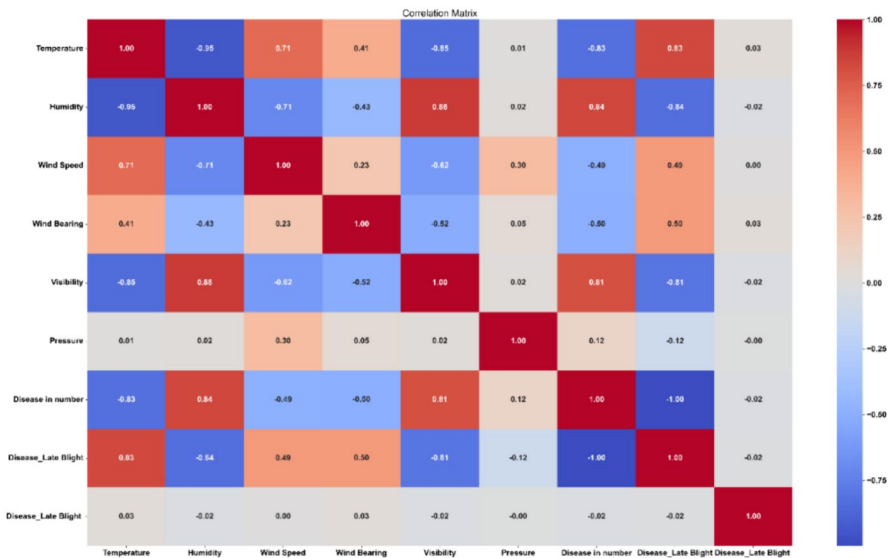


Fig. 1 Original dataset features’ correlation matrix

Data Preprocessing

Data preprocessing appears to be of utmost importance in data cleaning, forming consistency, and developing an appropriate dataset that can be subjected to analysis. In this research, categorical variable normalization and encoding are conducted to normalize the data. PCA is an elementary dimensionality reduction technique of the dataset that is applied to its smaller set of uncorrelated variables (principal components), which makes data processing and visualization easier (Aditya Shastry and Sanjay 2021). Figure 2 shows the way the PCA method decreases the space of dimensions of the dataset and reflects the largest unique characteristics of the features accordingly. Every factor stands for the specific features which jointly determine the variation. PCA explains why features behave similarly, which is why these groups may be correlated with an appearance or lack of diseases.

Clustering with *K*-means goes even further to organize the points based on similarity, which reveals the underlying patterns that can help predict the progression of disease. These preprocessing steps make sure machine learning models have a training set of high-quality data upon which to learn, thus reducing such biases and improving predictive performance (Javidan et al. 2023). In Fig. 3, data points are classified into clusters with similar values based on *K*-means clustering. Clustering reminds the inherent associations within the data, unveiling combinations of weather characteristics grouped. These clusters can then be connected to disease labels so that it would be easier to understand which weather patterns correlate better with potato diseases.

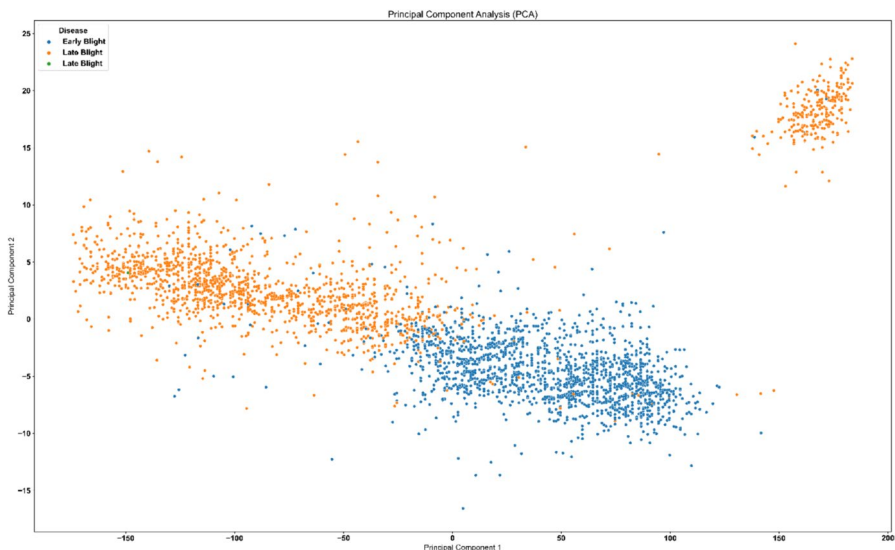


Fig. 2 Principal component analysis (PCA) of tested dataset

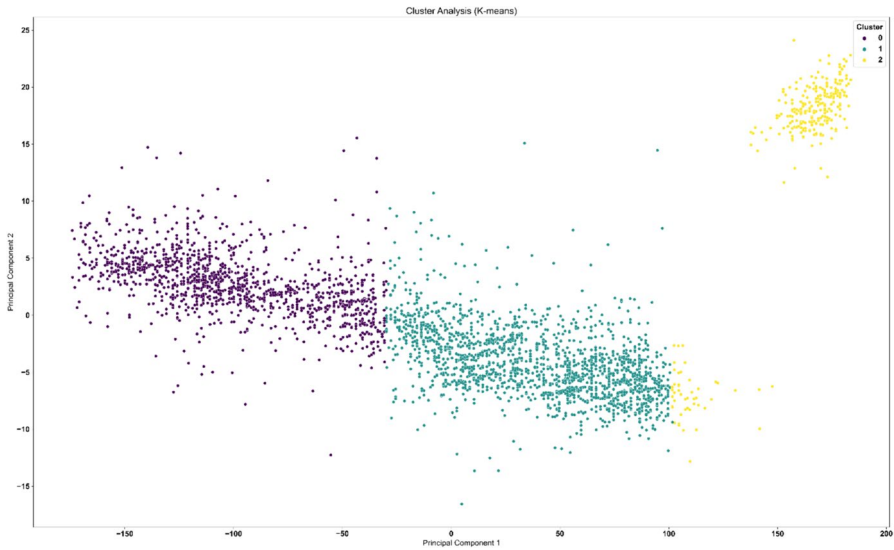


Fig. 3 Cluster analysis (K -means) of tested dataset

Copula Analysis

Copula analysis is a statistical technique employed to investigate the relationships between a collection of variables. It enables the exploration of relationships between variables that would otherwise be concealed. Through the copula function, the model generates synthetic datasets that can be used to simulate future scenarios. This kind of synthetic data, in turn, is useful in training machine learning models by creating several scenarios that not only allow machines to identify specific patterns as well as correlations between the weather and disease occurrences but also increase the machines' power to preserve such patterns for the future. The synthetic dataset has a correlation matrix between the variables, which is useful for identifying important weather-disease interactions and helping further develop predictive models (Albulescu et al. 2020; Das et al. 2022). Through copula synthesis in Fig. 4, a synthetic dataset is obtained, which reflects similar statistical features as the original data. The synthetic dataset is the correlation matrix, which is designed to test machine learning models in different scenarios and validate their robustness. The above table describes how weather processes that are linked or interacted are affected during disease outbreaks.

Figure 5 indicates how different weather attributes vary with the weather, meaning each of them individually has different behaviours. By recognizing the extent and distribution of these features, models can be improved in their ability to detect less evident but very important symptoms of disease outbursts, improving their forecasting.



Fig. 4 Synthetic dataset features’ correlation matrix

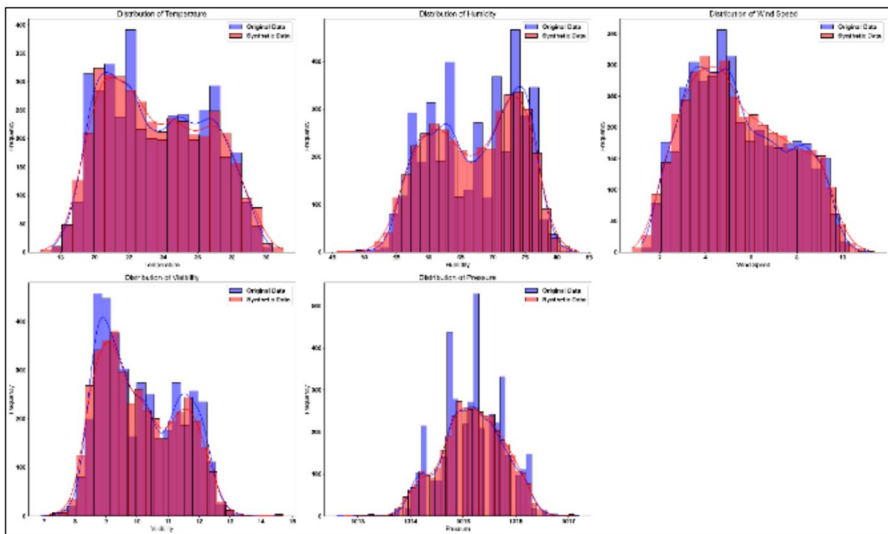


Fig. 5 Distribution of dataset features

Feature Selection

Feature selection acts as the key determinant for finding the most crucial variables predictive against diseases in hand. In this research, binary feature selection will be

a transformation of the features into binary format (0 or 1) based on the predefined thresholds. In other words, this approach helps to weed out unimportant or duplicate properties from the models and allows them to concentrate on the most significant parameters (Suruliandi et al. 2021). By paring down the feature set, models can result in better execution of the task given and higher accuracy, as they no longer must process irrelevant data. For example, because of the process, temperature and humidity may come out at the top of the list, and the models will rely heavily on them when forecasting the outbreak of the diseases (Dhal and Azad 2022).

The most recent binary optimization algorithms employed in this stage are the binary Greylag Goose Optimization (bGGO) (El-kenawy et al. 2024) and the binary Waterwheel Plant Algorithm (bWWPA) (Alhussan et al. 2023). Other state-of-the-art binary algorithms, including binary Grey Wolf Optimizer (bGWO), binary Particle Swarm Optimization (bPSO), binary Whale Optimization Algorithm (bWOA), binary Biogeography-Based Optimization (bBBO), binary Multi-Verse Optimizer (bMVO), binary Stochastic Bayesian Optimization (bSBO), binary Genetic Algorithm (bGA), are also tested. This stage helps improve predictive accuracy by isolating the most relevant features.

Table 1 shows the criteria for evaluating feature selection results, using several key metrics. The metrics include best fitness, worst fitness, average error, average fitness, average fitness size, and standard deviation (Ali et al. 2024).

Machine Learning Models

The research employs a diverse array of machine learning models (Sharma et al. 2021; Benos et al. 2021; Saleem et al. 2021; Ayoub Shaikh et al. 2022), each bringing unique strengths to the task of disease prediction:

1. Logistic regression: Effective in binary classification problems, logistic regression identifies the likelihood of specific diseases based on the correlation between features.
2. Neural network (MLP): MLP is a deep learning model capable of capturing complex data patterns. It is ideal for datasets with non-linear relationships.

Table 1 Criteria for evaluating feature selection results

Metric	Formula
Best fitness	$\min_{i=1}^M S_i^*$
Worst fitness	$\max_{i=1}^M S_i^*$
Average error	$\frac{1}{M} \sum_{j=1}^M \frac{1}{N} \sum_{i=1}^N \text{mse}(\hat{V}_i - V_i)$
Average fitness	$\frac{1}{M} \sum_{i=1}^M S_i^*$
Average fitness size	$\frac{1}{M} \sum_{i=1}^M \text{size}(S_i^*)$
Standard deviation	$\sqrt{\frac{1}{M-1} \sum_{i=1}^M (S_i^* - \text{mean})^2}$

3. Random forest: An ensemble method that builds multiple decision trees and merges them to improve predictive accuracy, particularly beneficial for datasets with varied attributes.
4. Support vector machine: It uses hyperplanes to separate disease and non-disease instances efficiently, making them ideal for classification problems.
5. *K*-nearest neighbors (KNN): KNN classifies data points by measuring their distance from known instances, making it highly effective for recognizing patterns in proximity.
6. Naive Bayes: This probabilistic model calculates conditional probabilities based on past data, excelling at predicting categorical outcomes.
7. Decision tree: A hierarchical structure that maps decision rules, providing clear paths for disease identification based on different weather variables.
8. Gradient boosting: This ensemble method sequentially builds models to minimize predictive errors, resulting in high accuracy for disease prediction.
9. SVM (rbf kernel): A variant of the support vector machine, this model employs linear or non-linear kernels for effective classification.

This research aims to compare those models regarding their accuracy, sensitivity, and specificity to select the best for a potato’s disease forecasting purposes using the given dataset.

Table 2 shows the classification model evaluation criteria, detailing various metrics used to assess the performance of machine learning models.

These evaluation criteria are essential for comprehensively assessing the performance of classification models, ensuring that the models not only make accurate predictions but also effectively identify and distinguish between positive and negative instances (Zaki et al. 2023).

Experimental Results

The experimental outcome section outlines an in-depth analysis of the machine learning models’ ability to predict potato leaf diseases using the weather dataset. Feature selection results will be grouped into two categories: the models

Table 2 Classification model evaluation criteria

Evaluation criteria	Value
Accuracy	$\left(\frac{\text{Number of correct predictions}}{\text{Total predictions}} \right)$
Sensitivity (recall)	$\left(\frac{\text{True positives}}{\text{True positives} + \text{false negatives}} \right)$
Specificity	$\left(\frac{\text{True negatives}}{\text{True negatives} + \text{false positives}} \right)$
Positive predictive value	$\left(\frac{\text{True positives}}{\text{True positives} + \text{false positives}} \right)$
Negative predictive value	$\left(\frac{\text{True negatives}}{\text{True negatives} + \text{false negatives}} \right)$
F1 score	$\left(2 \times \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}} \right)$

evaluated with and without the implementation of feature selection. Such grouping allows us to compare the accuracy, sensitivity, and specificity of the models with and without applying the feature selection, which provides evidence of how feature selection improves the performance of the model by eliminating the noise that might be present in the data.

Machine Learning Models Without Feature Selection

A detailed comparison of various machine learning models' classification results for the tested dataset without feature selection is shown in Table 3. The highest accuracy (0.9489) in this case is for logistic regression, representing its ability to classify the data accurately. MLP follows, indicating that deep learning techniques can identify patterns of disease outbreaks. Despite that, random forest and support vector machine models also provide high performances with accuracies over 0.93. This implies that any of these models can model intricate associations between meteorological variables and disease incidences. The sensitivity (true positive rate) and specificity (true negative rate) scores are also uniformly high in all comparisons, being essential for false-positive and false-negative reduction. Starting results provide a satisfactory level for continuing optimization.

Figure 6 depicts the amount of exactness before the feature selection using a line graph. Logistic regression is number one in terms of accuracy, but the rest of the strategies, such as neural network and random forest, are not too far behind. This depiction illustrates that classical classification techniques, such as logistic regression, can represent incredible power and ensemble methods, and deep neural

Table 3 Performance metrics results of machine learning models without feature selection

Models	Accuracy	Sensitivity (TRP)	Specificity (TNP)	<i>P</i> -value (PPV)	<i>P</i> -value (NPV)	FScore
Logistic regression	0.9489	0.9452	0.9429	0.9303	0.9473	0.9377
Neural network (MLP)	0.9429	0.9429	0.9351	0.9323	0.9453	0.9376
Random forest	0.9324	0.9406	0.9351	0.9321	0.9432	0.9364
Support vector machine	0.9356	0.9384	0.9329	0.9299	0.9410	0.9341
<i>K</i> -nearest neighbors	0.9345	0.9384	0.9307	0.9278	0.9409	0.9330
Naive Bayes	0.9245	0.9384	0.9307	0.9278	0.9409	0.9330
Decision tree	0.9136	0.9384	0.9286	0.9257	0.9408	0.9320
Gradient boosting	0.9101	0.9338	0.9329	0.9295	0.9370	0.9317
SVM (rbf kernel)	0.8951	0.9184	0.9164	0.9136	0.9107	0.9109

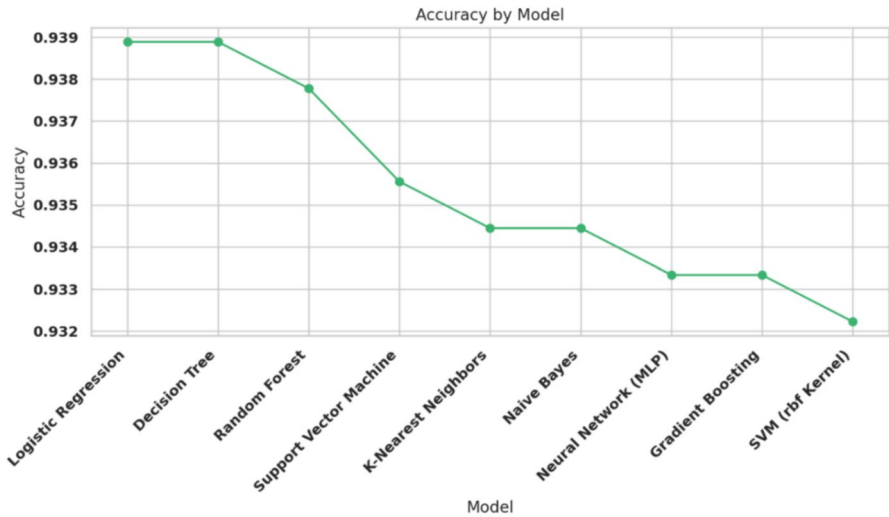


Fig. 6 Accuracy by model for machine learning models without feature selection

networks can offer alternative robust methods that can cope with complex data. Examples like SVM (rbf kernel) and gradient boosting will also approximate accuracy but will also remain alongside the competition. The symbolic story illustrates the contrast of the models, which is a basis of measurement after the sets of features are selected.

Figure 7 shows pair plots of various metrics for models before feature selection. For each scatter plot, prediction metrics, including accuracy, sensitivity, and specificity, are compared and are also used to reveal correlations and trade-offs. Firstly, models with good accuracy are likely also to have good sensitivity and specificity, as in logistic regression and MLP, for instance. The plots also show how random forest and SVM (rbf kernel) can still have equivalent predictive power via different performance metrics. This analysis helps understand in what sense each model outperforms or underperforms being utilized as the base for the model’s personalized feature selection.

Feature Selection Results

A detailed inspection of the feature selection result and comparison of different algorithms is presented in Table 4, which are bGGO, bWWPA, and bGWO. The data results indicate that bGGO attains the lowest average mistake (0.350) and even keeps a small standard deviation. The high accuracy shows that bGGO can carry out feature selection effectively, thereby enhancing the model’s performance. Other algorithms, such as bPSO and bWOA, also have good outputs, but their bigger errors and deviations indicate this optimization process is not systematic and consistent. The analysis shows the crucial role played by the feature selection method in avoiding the nuisance of unnecessary data and instead achieving maximum predictive accuracy.

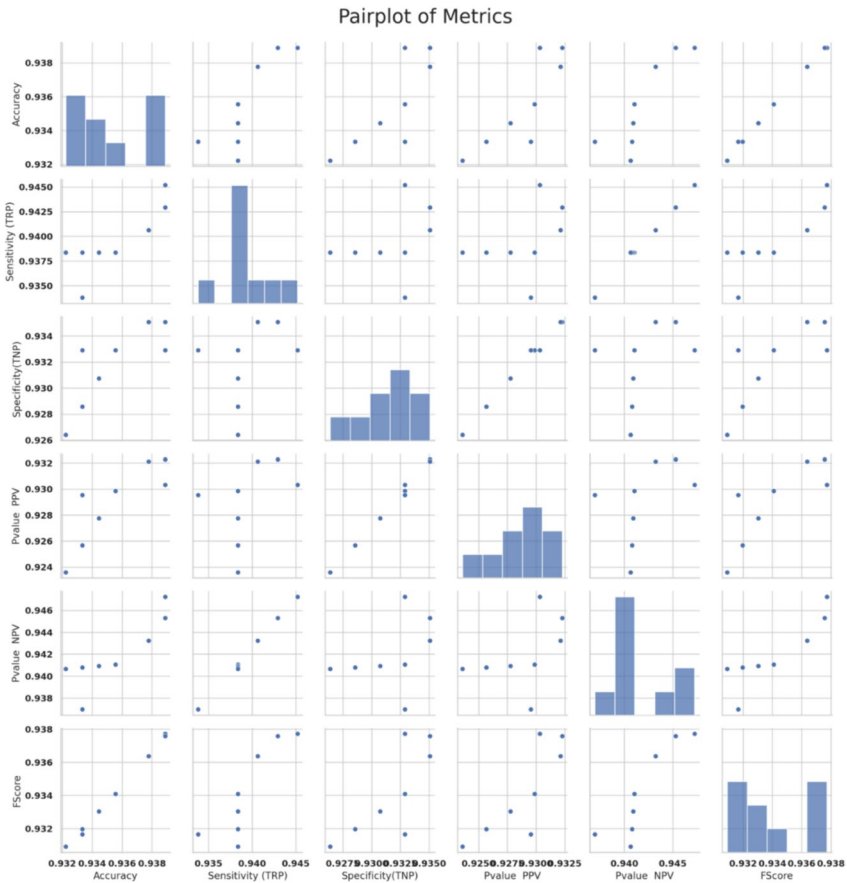


Fig. 7 Pair plot of metrics for machine learning models without feature selection

Table 4 Feature selection results

	bGGO	bWWPA	bGWO	bPSO	bWOA	bBBO	bMVO	bSBO	bGA
Average error	0.350	0.375	0.437	0.431	0.431	0.399	0.408	0.439	0.411
Average select size	0.333	0.533	0.666	0.533	0.696	0.697	0.629	0.703	0.475
Average fitness	0.443	0.460	0.468	0.458	0.466	0.464	0.488	0.498	0.471
Best fitness	0.345	0.380	0.421	0.438	0.430	0.453	0.413	0.441	0.374
Worst fitness	0.444	0.447	0.531	0.506	0.506	0.540	0.531	0.520	0.489
Standard deviation fitness	0.266	0.270	0.289	0.270	0.272	0.315	0.320	0.331	0.272

Figure 8 confirms that among the feature selection methods that are being used, methods such as bGGO and bWWPA are the most accurate, as they display a consistently low error rate. Such high accuracy points out that these features are important in detecting diseases or health conditions. The consistent grouping

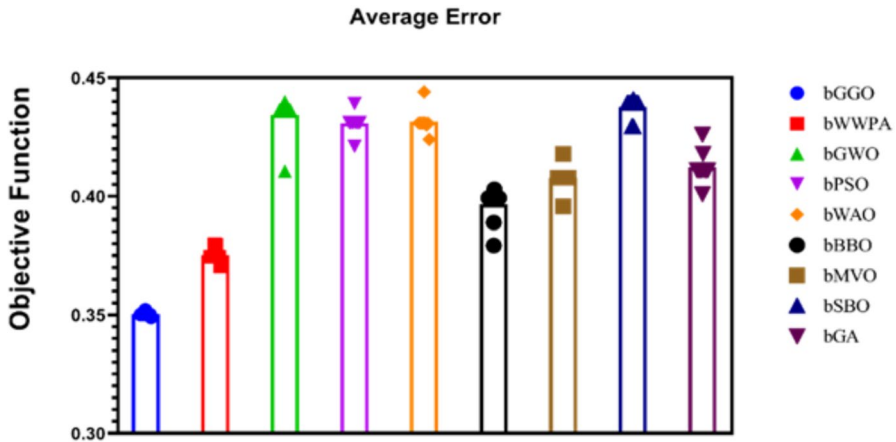


Fig. 8 Average error plot of feature selection results

around the low error rate levels provides evidence of their reliability; it can be deduced that these methods, as feature selection for the agricultural datasets, are very suitable. It is this knowledge that will help eliminate mistakes made during the process of selecting reliable methods to minimize predictive error.

Figure 9 presents the histogram of average error plots, emphasizing the occurrence distribution of errors produced by different approaches to feature selection. The close gatherings around the common value for methods like bGGO and bWWPA accomplish the purpose of high-precision results. Differently, the wide range of other methods shows that some points are effective, and others prove to be not so useful. This histogram delineates a clear view of how well each method of reducing errors works in practice, with the data scientists afterward being able to make informed decisions about approaches that can accurately reduce errors.

The statistics analysed for feature selection methods are presented in Table 5. The performance metrics include mean, median, and standard deviation, and the robustness of methods such as bGGO and bPSO are disclosed. As with the sampling time, reducing the variance and mean deviation among the values at hand shows the effective way the method operates. Looking into this, however, we can see that the differences in readings at higher levels and fluctuations in other methods could mean

Fig. 9 Histogram of average error plot of feature selection results

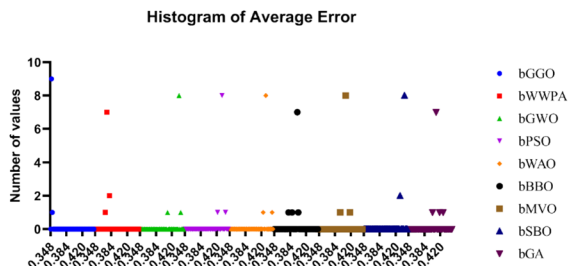


Table 5 Statistical analysis of feature selection results

	bGGO	bWWPA	bGWO	bPSO	bWAO	bBBO	bMVO	bSBO	bGA
Number of values	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000
Minimum	0.349	0.371	0.411	0.421	0.424	0.379	0.396	0.430	0.401
25% percentile	0.350	0.375	0.437	0.431	0.431	0.397	0.408	0.437	0.411
Median	0.350	0.375	0.437	0.431	0.431	0.399	0.408	0.440	0.411
75% percentile	0.350	0.376	0.437	0.431	0.431	0.399	0.408	0.440	0.413
Maximum	0.352	0.380	0.440	0.439	0.444	0.403	0.418	0.441	0.426
Range	0.003	0.009	0.029	0.018	0.020	0.024	0.022	0.011	0.025
Mean	0.350	0.375	0.434	0.431	0.432	0.397	0.408	0.438	0.412
Std. deviation	0.001	0.003	0.008	0.004	0.005	0.007	0.005	0.004	0.006
Std. error of the mean	0.000	0.001	0.003	0.001	0.002	0.002	0.002	0.001	0.002
Sum	3.503	3.752	4.344	4.309	4.315	3.966	4.077	4.376	4.122

variability, resulting in less accurate outcomes. These comprehensive statistical studies confirm the significance of using stable robot feature selection methods.

Table 6 shows the analysis of the Wilcoxon signed-rank test, where the significant values are compared for feature selection methods according to actual medians and theoretical means. It was shown that the *P*-values confirm statistical significance and high ranks of bGGO and bWWPA methods validate their superiority over the theoretical baseline. A conclusion is drawn that, given the positive results through all the tests, these procedures are seen to generate effective predictive models.

Machine Learning Model Results with Feature Selection

Table 7 evaluates the after-feature selection performance of machine learning algorithms. MLP is gallant, consists of precision, and achieves an accuracy of 0.983, which judges the effectiveness of the refining above the preselection ones. KNN

Table 6 Wilcoxon signed-rank test for feature selection results

	bGGO	bWWPA	bGWO	bPSO	bWAO	bBBO	bMVO	bSBO	bGA
Sum of signed ranks (<i>W</i>)	55	55	55	55	55	55	55	55	55
Sum of positive ranks	55	55	55	55	55	55	55	55	55
Sum of negative ranks	0	0	0	0	0	0	0	0	0
<i>P</i> -value (two tailed)	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact	Exact
<i>P</i> -value summary	**	**	**	**	**	**	**	**	**
Significant ($\alpha=0.05$)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
How big is the discrepancy?									
Discrepancy	0.3502	0.3746	0.4367	0.4311	0.4310	0.3993	0.4079	0.4395	0.4110

Table 7 Performance metrics results of machine learning models with feature selection

Models	Accuracy	Sensitivity (TRP)	Specificity (TNP)	<i>P</i> -value (PPV)	<i>P</i> -value (NPV)	FScore
Neural network (MLP)	0.983	0.980	0.977	0.965	0.982	0.972
<i>K</i> -nearest neighbors	0.967	0.977	0.969	0.967	0.980	0.972
Random forest	0.967	0.975	0.969	0.966	0.978	0.971
Decision tree	0.970	0.973	0.967	0.964	0.975	0.968
Logistic regression	0.969	0.973	0.965	0.962	0.975	0.967
Naive Bayes	0.959	0.973	0.965	0.962	0.975	0.967
Support vector machine	0.948	0.973	0.963	0.960	0.975	0.966
Gradient boosting	0.944	0.968	0.967	0.964	0.971	0.966
SVM (rbf kernel)	0.929	0.953	0.951	0.948	0.945	0.945

and random forest show quite good improvements, too—with accuracies of 0.967 and above. This highlights the fact that feature selection serves the purpose of noise removal, and the resulting models focus on including the most important predictive parameters. Sensitivity and specificity scales have also been upgraded, which demonstrates that these decisions with improved classes can more precisely recognize diseases from the no-disease cases.

Figure 10 presents the bar chart that gives the results of different models after the feature selection. Thinking about MLP, which outperforms the other two algorithms,

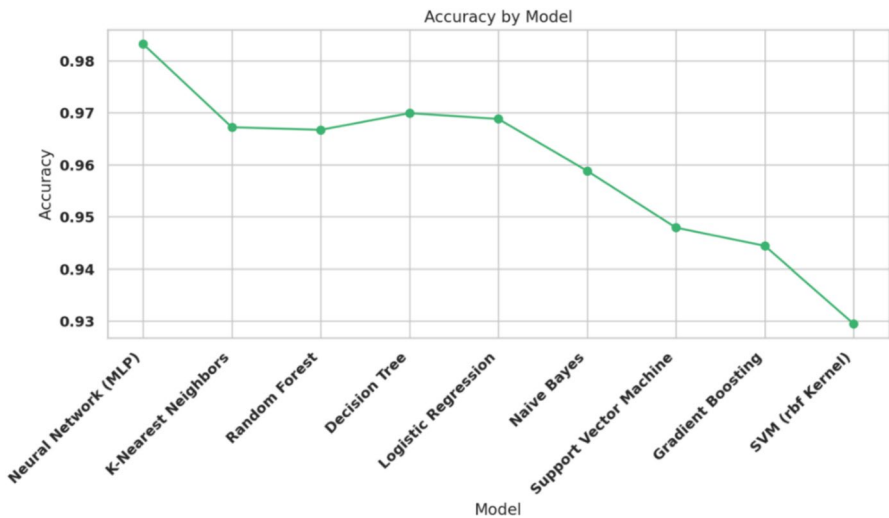


Fig. 10 Accuracy by model for machine learning models with feature selection

we can say for sure that *K*-nearest neighbors and random forest also have good accuracy rates. An improvement in precision for all models symbolizes that feature selection can be considered a significant factor that boosts accuracy prediction for machine learning models. The fact that this diagram can be directly compared with the initial results after selection smokes out the importance of working on the consistencies of the key elements.

Figure 11 contains the pair plot of performance metrics—both before and after the feature selection. Such comparison impresses as error-tolerance, detecting more details, and identifying details after the new features. To illustrate, MLP and random forest now continuously post good scores across all metrics. By representing the one-after-another quality of the feature selection performance of all models, we can be sure that using the best covariance will help multiple models extend their capability of accurately predicting to the next level.

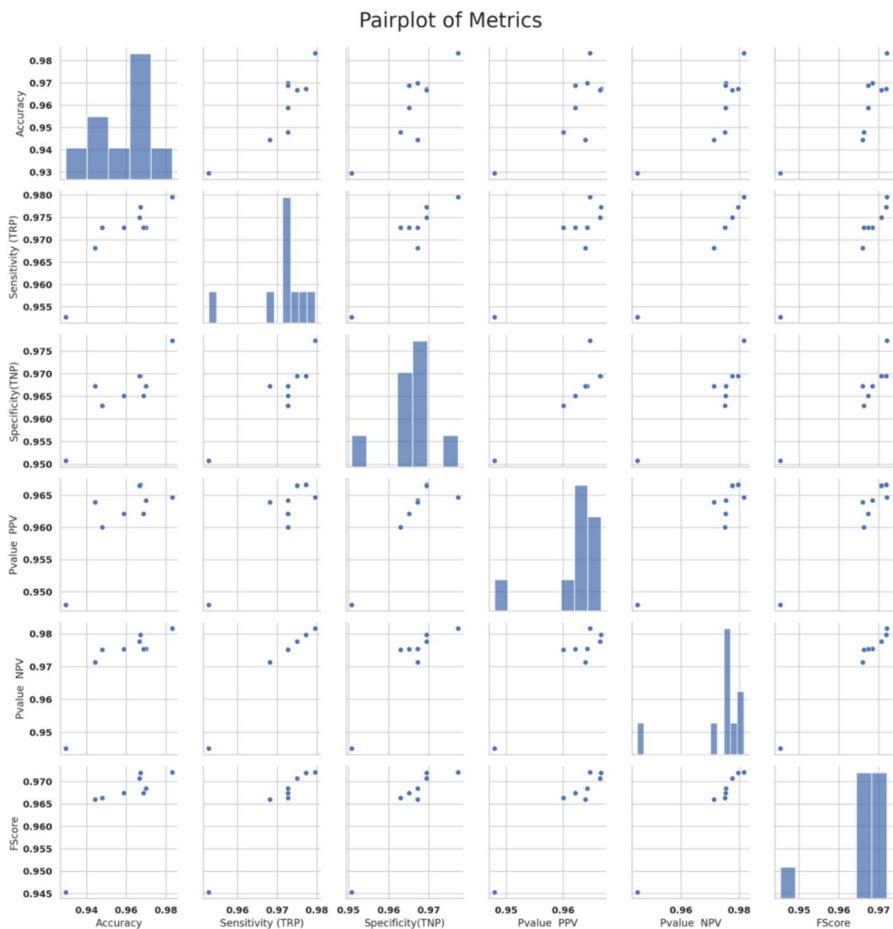


Fig. 11 Pair plot of metrics for machine learning models with feature selection

Limitations of the Study

Despite the promising outcomes of this study, several limitations must be addressed. The dataset used has geographic and temporal biases, as it primarily includes data from specific regions and periods, potentially limiting the model's effectiveness in other areas. There may also be data imbalance issues, affecting the model's ability to accurately predict less prevalent diseases. Additionally, while the feature selection techniques like bGGO and bWWPA have shown to improve model accuracy, their effectiveness may vary with different datasets or types of features. The models, tailored for potato leaf diseases, might not perform as well for other crops or diseases without significant adjustments. Practical implementation on a larger scale poses challenges, such as extensive data collection and ensuring user-friendly tools for farmers. The study's models need external validation using independent datasets to confirm robustness and generalizability, and their long-term performance requires continuous monitoring. Addressing these limitations involves expanding the dataset, exploring additional techniques, and focusing on practical applications to enhance the model's applicability and impact on sustainable agriculture.

Conclusion and Future Direction

To summarize, this work focused on how we can use all these weather parameters to classify potato leaf diseases using different machine-learning techniques. To determine the predictive accuracy of each model, it was tested against the training set data, and the results revealed that the models constructed using the logistic regression (LR) and the neural network (MLP) algorithms had the highest level of accuracy, which was above (94%) before applying the feature selection method. It is, however, correct to say that all the models have benefited from feature selection techniques, which have led to the observed improvements in predictive performance evaluation. The specific algorithms, for example, bGGO and bWWPA, demonstrated additional favourable performance, measuring the mistake and defining important characteristics at a higher level. Finally, MLP boasted the highest level of accuracy in its results (98.3%) and was the most enhanced among all the networks for the best data reduction, which enhanced its predictive performance. They want to indicate how it is imperative that machine learning models are optimized in fighting agricultural diseases. In short, using state-of-the-art methods in predictive models, farmers can be assured that the resulting information has high accuracy and can be used for strategic planning regarding prevention and losses.

In future work, expanding the dataset to extend the study to other crops and diseases is essential. Therefore, the models become primarily universal in the sense that they are developed for prediction. Thus, it is essentially vital to apply more advanced feature selection methods, and other parameters should be tuned to reduce the prediction error and to raise the accuracy of the model. Developing simple and intuitive management tools addressed to farmers and agricultural managers will

increase their prevalence and use in as many cases as possible at the non-adverse and environmentally friendly agriculture stage. In conclusion, this study offers a firm ground for the collaboration between artificial intelligence and agriculture, as the practices will be made efficient and sustainable in disease control.

Acknowledgements Princess Nourah bint Abdulrahman University Researchers Supporting Project number PNURSP2024R 308, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contribution All authors have contributed equally.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Data Availability Available upon request.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aditya Shastry K, Sanjay HA (2021) A modified genetic algorithm and weighted principal component analysis based feature selection and extraction strategy in agriculture. *Knowl-Based Syst* 232:107460. <https://doi.org/10.1016/j.knosys.2021.107460>
- Albulescu CT, Tiwari AK, Ji Q (2020) Copula-based local dependence among energy, agriculture and metal commodities markets. *Energy* 202:117762. <https://doi.org/10.1016/j.energy.2020.117762>
- Alhussan AA, Abdelhamid AA, El-Kenawy El-S M, Ibrahim A, Eid MM, Khafaga DS, Em AA (2023) A binary waterwheel plant optimization algorithm for feature selection. *IEEE Access* 11:94227–94251. <https://doi.org/10.1109/ACCESS.2023.3312022>
- Ali MZ, Abdullah A, Zaki AM, Rizk FH, Eid MM, El-Kenway EM (2024) Advances and challenges in feature selection methods: a comprehensive review. *J Artif Intell Metaheuristics* 7(1):67–77. <https://doi.org/10.54216/JAIM.070105>
- Arshaghi A, Ashourian M, Ghabeli L (2023) Potato diseases detection and classification using deep learning methods. *Multimed Tools Appl* 82(4):5725–5742. <https://doi.org/10.1007/s11042-022-13390-1>
- Ayoub Shaikh T, Rasool T, Rasheed Lone F (2022) Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Comput Electron Agric* 198:107119. <https://doi.org/10.1016/j.compag.2022.107119>

- Benos L, Tagarakis AC, Dolias G, Berruto R, Kateris D, Bochtis D (2021) Machine learning in agriculture: a comprehensive updated review. *Sensors* 21(11):3758. <https://doi.org/10.3390/s21113758>
- Bhat SA, Huang N-F (2021) Big data and AI revolution in precision agriculture: survey and challenges. *IEEE Access* 9:110209–110222. <https://doi.org/10.1109/ACCESS.2021.3102227>
- Cravero A, Pardo S, Sepúlveda S, Muñoz L (2022) Challenges to use machine learning in agricultural big data: a systematic literature review. *Agronomy* 12(3):748. <https://doi.org/10.3390/agronomy12030748>
- Das S, Das J, Umamahesh NV (2022) Copula-based drought risk analysis on rainfed agriculture under stationary and non-stationary settings. *Hydrol Sci J* 67(11):1683–1701. <https://doi.org/10.1080/02626667.2022.2079416>
- Dhal P, Azad C (2022) A comprehensive survey on feature selection in the various fields of machine learning. *Appl Intell* 52(4):4543–4581. <https://doi.org/10.1007/s10489-021-02550-9>
- Dolničar P (2021) Importance of potato as a crop and practical approaches to potato breeding. In: Dobnik D, Gruden K, Ramšak Ž, Coll A (eds) *Solanum tuberosum: methods and protocols*. Springer, US, New York, NY, pp 3–20
- El-kenawy El-SM, Khodadadi N, Mirjalili S, Abdelhamid AA, Eid MM, Ibrahim A (2024) Grey-lag Goose Optimization: nature-inspired optimization algorithm. *Expert Syst Appl* 238(Part E):122147. <https://doi.org/10.1016/j.eswa.2023.122147>
- Fenu G, Mallocci FM (2020) Artificial intelligence technique in crop disease forecasting: a case study on potato late blight prediction. In: Czarnowski I, Howlett RJ, Jain LC (eds) *Intelligent decision technologies*. Springer, Singapore, pp 79–89
- Gao J, Westergaard JC, Sundmark EHR, Bagge M, Liljeroth E, Alexandersson E (2021) Automatic late blight lesion recognition and severity quantification based on field imagery of diverse potato genotypes by deep learning. *Knowl-Based Syst* 214:106723. <https://doi.org/10.1016/j.knosys.2020.106723>
- Garske B, Bau A, Ekardt F (2021) Digitalization and AI in European agriculture: a strategy for achieving climate and biodiversity targets? *Sustainability* 13(9):4652. <https://doi.org/10.3390/su13094652>
- Gold KM, Townsend PA, Chlus A, Herrmann I, Couture JJ, Larson ER, Gevens AJ (2020) Hyperspectral measurements enable pre-symptomatic detection and differentiation of contrasting physiological effects of late blight and early blight in potato. *Remote Sensing* 12(2):286. <https://doi.org/10.3390/rs12020286>
- Gold KM, Townsend PA, Herrmann I, Gevens AJ (2020) Investigating potato late blight physiological differences across potato cultivars with spectroscopy and machine learning. *Plant Sci* 295:110316. <https://doi.org/10.1016/j.plantsci.2019.110316>
- Gupta N, Khosravy M, Patel N, Dey N, Gupta S, Darbari H, Crespo RG (2020) Economic data analytic AI technique on IoT edge devices for health monitoring of agriculture machines. *Appl Intell* 50(11):3990–4016. <https://doi.org/10.1007/s10489-020-01744-x>
- Hamrani A, Akbarzadeh A, Madramootoo CA (2020) Machine learning for predicting greenhouse gas emissions from agricultural soils. *Sci Total Environ* 741:140338. <https://doi.org/10.1016/j.scitotenv.2020.140338>
- Javidan SM, Banakar A, Vakilian KA, Ampatzidis Y (2023) Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning. *Smart Agric Technol* 3:100081. <https://doi.org/10.1016/j.atech.2022.100081>
- Kang F, Li J, Wang C, Wang F (2023) A lightweight neural network-based method for identifying early-blight and late-blight leaves of potato. *Appl Sci* 13(3):1487. <https://doi.org/10.3390/app13031487>
- Khan N, Ray RL, Sargani GR, Ihtisham M, Khayyam M, Ismail S (2021) Current progress and future prospects of agriculture technology: gateway to sustainable agriculture. *Sustainability* 13(9):4883. <https://doi.org/10.3390/su13094883>
- King BA, Stark JC, Neibling H (2020) Potato irrigation management. In: Stark JC, Thornton M, Nolte P (eds) *Potato production systems*. Springer International Publishing, Cham, pp 417–446
- Linaza MT, Posada J, Bund J, Eisert P, Quartulli M, Döllner J, Pagani A, Olaizola IG, Barriguinha A, Moysiadis T, Lucat L (2021) Data-driven artificial intelligence applications for sustainable precision agriculture. *Agronomy* 11(6):1227. <https://doi.org/10.3390/agronomy11061227>
- Meno L, Escuredo O, Rodríguez-Flores MS, Seijo MC (2021) Looking for a sustainable potato crop. Field assessment of early blight management. *Agric For Meteorol* 308–309:108617. <https://doi.org/10.1016/j.agrformet.2021.108617>

- Nawaz SA, Li J, Bhatti UA, Shoukat MU, Ahmad RM (2022) AI-based object detection latest trends in remote sensing, multimedia and agriculture applications. *Front Plant Sci* 13. <https://doi.org/10.3389/fpls.2022.1041514>
- Qazi S, Khawaja BA, Farooq QU (2022) IoT-equipped and AI-enabled next generation smart agriculture: a critical review, current challenges and future trends. *IEEE Access* 10:21219–21235. <https://doi.org/10.1109/ACCESS.2022.3152544>
- Saleem MH, Potgieter J, Arif KM (2021) Automation in agriculture by machine and deep learning techniques: a review of recent developments. *Precision Agric* 22(6):2053–2091. <https://doi.org/10.1007/s11119-021-09806-x>
- Shaikh FK, Memon MA, Mahoto NA, Zeadally S, Nebhen J (2022) Artificial intelligence best practices in smart agriculture. *IEEE Micro* 42(1):17–24. <https://doi.org/10.1109/MM.2021.3121279>
- Sharma A, Jain A, Gupta P, Chowdary V (2021) Machine learning applications for precision agriculture: a comprehensive review. *IEEE Access* 9:4843–4873. <https://doi.org/10.1109/ACCESS.2020.3048415>
- Shin J-Y, Kim KR, Ha J-C (2020) Seasonal forecasting of daily mean air temperatures using a coupled global climate model and machine learning algorithm for field-scale agricultural management. *Agric For Meteorol* 281:107858. <https://doi.org/10.1016/j.agrformet.2019.107858>
- Singh R, Kaur S, Aggarwal P (2021) Exploration of potato starches from non-commercial cultivars in ready to cook instant non cereal, non glutinous pudding mix. *LWT* 150:111966. <https://doi.org/10.1016/j.lwt.2021.111966>
- Subeesh A, Mehta CR (2021) Automation and digitization of agriculture using artificial intelligence and internet of things. *Artif Intell Agric* 5:278–291. <https://doi.org/10.1016/j.aiaa.2021.11.004>
- Suruliandi A, Mariammal G, Raja SP (2021) Crop prediction based on soil and environmental characteristics using feature selection techniques. *Math Comput Model Dyn Syst* 27(1):117–140. <https://doi.org/10.1080/13873954.2021.1882505>
- Tang J, Xiao D, Wang J, Fang Q, Zhang J, Bai H (2021) Optimizing water and nitrogen managements for potato production in the agro-pastoral ecotone in North China. *Agric Water Manag* 253:106945. <https://doi.org/10.1016/j.agwat.2021.106945>
- Tang J, Bai H, Zhang S, Xiao D, Tianzhu Z, Li Liu D, Wang B, Feng P (2024) Adaptations of potato production to future climate change by optimizing planting date, irrigation and fertilizer in the Agro-Pastoral Ecotone of China. *Clim Risk Manag* 44:100604. <https://doi.org/10.1016/j.crm.2024.100604>
- Viana CM, Santos M, Freire D, Abrantes P, Rocha J (2021) Evaluation of the factors explaining the use of agricultural land: a machine learning and model-agnostic approach. *Ecol Ind* 131:108200. <https://doi.org/10.1016/j.ecolind.2021.108200>
- Waaswa A, Oywaya Nkurumwa A, Mwangi Kibe A, Ngeno Kipkemoi J (2022) Climate-smart agriculture and potato production in Kenya: review of the determinants of practice. *Climate Dev* 14(1):75–90. <https://doi.org/10.1080/17565529.2021.1885336>
- Wang C, Zang H, Liu J, Shi X, Li S, Chen F, Chu Q (2020) Optimum nitrogen rate to maintain sustainable potato production and improve nitrogen use efficiency at a regional scale in China. A Meta-Analysis. *Agron Sustain Dev* 40(5):37. <https://doi.org/10.1007/s13593-020-00640-5>
- Wang Y-P, Pan Z-C, Yang L-N, Burdon JJ, Friberg H, Sui Q, Zhan J (2021) Optimizing plant disease management in agricultural ecosystems through rational in-crop diversification. *Front Plant Sci* 12. <https://doi.org/10.3389/fpls.2021.767209>
- Yang H, Li F, Hu Y, Yu K (2021) Hyperspectral indices optimization algorithms for estimating canopy nitrogen concentration in potato (*Solanum tuberosum* L.). *Int J Appl Earth Obs Geoinformation* 102:102416. <https://doi.org/10.1016/j.jag.2021.102416>
- Yeasmin T (2023) Potato leaf disease based on weather details. <https://www.kaggle.com/datasets/tamim1530/potato-leaf-disease-based-on-weather-details>. Accessed 5 May 2024
- Zaki AM, Abdelhamid AA, Ibrahim A, Eid MM, El-Kenawy E-SM (2023) Enhancing K-nearest neighbors algorithm in wireless sensor networks through stochastic fractal search and particle swarm optimization. *J Cybersecurity Inf Manag* 13(1):76–84. <https://doi.org/10.54216/JCIM.130108>
- Zhang Y, Qu H, Yang X, Wang M, Qin N, Zou Y (2020) Cropping system optimization for drought prevention and disaster reduction with a risk assessment model in Sichuan Province. *Glob Ecol Conserv* 23:e01095. <https://doi.org/10.1016/j.gecco.2020.e01095>
- Zhang P, Guo Z, Ullah S, Melagraki G, Afantitis A, Lynch I (2021) Nanotechnology and artificial intelligence to enable sustainable and precision agriculture. *Nat Plants* 7(7):864–876. <https://doi.org/10.1038/s41477-021-00946-6>

Zhang S, Fan J, Zhang F, Wang H, Yang L, Sun X, Cheng M, Cheng H, Li Z (2022) Optimizing irrigation amount and potassium rate to simultaneously improve tuber yield, water productivity and plant potassium accumulation of drip-fertigated potato in northwest China. *Agric Water Manag* 264:107493. <https://doi.org/10.1016/j.agwat.2022.107493>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

**Marwa Radwan¹ · Amel Ali Alhussan² · Abdelhameed Ibrahim³ ·
Sayed M. Tawfeek^{4,5}**

✉ Marwa Radwan
marwa.radwan@deltauniv.edu.eg

¹ Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 11152, Egypt

² Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, 11671 Riyadh, Saudi Arabia

³ Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt

⁴ Delta Higher Institute for Engineering and Technology, Mansoura 35511, Egypt

⁵ MEU Research Unit, Middle East University, Amman 11831, Jordan