PLANT SCIENCE TODAY
Vol 12(sp4): 01–11
https://doi.org/10.14719/pst.9604

eISSN 2348-1900

RESEARCH ARTICLE

Machine learning-based forewarning models for rice pests and diseases using climatic parameters in Madurai district, Tamil Nadu

Sivasankari B[1], Karthikeyan M[2], Ilamaran M[1*], Prema P[3], Sujatha P[4] & Shanmugapriya S P[5]

[1]Community Science College and Research Institute, Tamil Nadu Agricultural University, Madurai 625 104, Tamil Nadu, India

[2]Department of Medicinal and Aromatic Crops, Horticultural College and Research Institute, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

[3]Agricultural College and Research Institute, Madurai 625 104, Tamil Nadu, India

[4]Dr .M S Swaminathan Agricultural College and Research Institute, Eachangkottai Thanjavur 614 902, Tamil Nadu, India

[5]Department of Physical Sciences and Information Technology, Agricultural Engineering College and Research Institute, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

*Correspondence email - maranfsn@gmail.com

Received: 24 May 2025; Accepted: 08 August 2025; Available online: Version 1.0: 17 October 2025

Cite this article: Sivasankari B, Karthikeyan M, Ilamaran M, Prema P, Sujatha P, Shanmugapriya SP. Machine learning-based forewarning models for rice pests and diseases using climatic parameters in Madurai district, Tamil Nadu. Plant Science Today. 2025;12(sp4):01–11. https:/doi.org/10.14719/pst.9604

Abstract

Rice is a staple crop extensively cultivated in tropical regions like the Madurai district of Tamil Nadu, where agriculture plays a central role in livelihoods and food security. However, the crop is increasingly vulnerable to climatic variability, which significantly influences pest and disease outbreaks. Given the rising unpredictability of weather patterns, there is a pressing need for early warning systems that can forecast pest and disease incidences to support timely and effective management. This study aimed to develop forewarning models for major rice pests (stem borer, leaf folder, brown planthopper) and diseases (false smut, sheath rot, brown spot) using weather parameters recorded during 2022-2024. Pearson correlation analysis revealed strong associations between weather variables and biotic stress incidence, with brown spot showing the highest sensitivity exhibiting a significant negative correlation with maximum temperature (r = -0.779**) and a strong positive correlation with relative humidity (r = 0.844**). Machine learning (ML) models including Random Forest, Support Vector Machine (SVM), Linear Regression and XGBoost were evaluated for prediction accuracy. Among these, Random Forest and SVM provided superior performance in terms of R^2 and RMSE metrics across multiple targets. The study demonstrates the potential of integrating climatic data with predictive modelling tools to enable timely and localized interventions, forming a scientific basis for sustainable rice pest and disease management strategies in the region.

Keywords: forecasting models; machine learning; Madurai district; rice diseases; rice pests; weather parameters

Introduction

Rice (*Oryza sativa* L.) is one of the most important staple food crops globally, providing nourishment for over half of the world's population, with Asia accounting for the majority of production and consumption (1). In India, rice is not only a primary dietary component but also a vital crop supporting the livelihoods of millions of smallholder farmers, particularly in states like Tamil Nadu where rice cultivation forms an integral part of agricultural economies. However, rice production faces numerous biotic challenges, among which pests and diseases pose significant threats to yield stability and food security. Under favourable environmental conditions, these biotic stressors can cause yield losses exceeding 50 %, severely impacting the economic returns for farmers and the overall food supply chain (2).

Among the diverse pests and diseases affecting rice, several stand out due to their wide prevalence and severity in tropical rice-growing regions. Brown spot, caused by the fungal pathogen *Bipolaris oryzae*, is a foliar disease that can result in considerable yield reductions by damaging photosynthetic tissues and weakening plant health. False smut, incited by *Ustilaginoidea virens*, leads to grain discoloration and quality degradation, directly reducing market value. The yellow stem borer (*Scirpophaga incertulas*), a key insect pest, attacks rice stems causing dead hearts and whiteheads, which further contribute to crop loss. These pests and diseases are endemic to regions such as Tamil Nadu, where tropical climatic conditions facilitate their recurrent outbreaks (3).

The incidence and severity of rice pests and diseases are strongly influenced by weather variables including temperature, relative humidity, rainfall and wind speed (4). Temperature affects the lifecycle and reproductive rates of many insect pests and fungi, while humidity and rainfall create conducive conditions for pathogen sporulation and spread. Wind speed can facilitate the dispersal of pest populations and disease spores over larger areas, amplifying outbreak risks. Understanding these complex interactions is vital for timely disease and pest management, which in turn is critical for sustaining rice production in the face of climatic variability.

Traditional forecasting methods often rely on empirical thresholds or historical patterns that may not capture the nonlinear and dynamic relationships between environmental conditions and biotic stressors. Recent advances in data science, particularly ML techniques, offer promising approaches to overcome these limitations. ML models such as Random Forest, SVM and Extreme Gradient Boosting (XGBoost) are capable of handling large datasets, capturing complex nonlinear interactions and providing robust predictive accuracy (5, 6). These methods have been increasingly applied in agricultural forecasting to predict crop yields, disease outbreaks and pest infestations with enhanced precision compared to traditional statistical models (7, 8).

Despite global advancements, there remains a critical need for region-specific forecasting models that are adapted to local agro-climatic conditions and pest/disease ecologies. This localization is essential because the distribution, development and impact of pests and diseases vary with microclimate, cropping patterns and agronomic practices, which differ significantly across regions (9). Madurai district in Tamil Nadu exemplifies such a region with a semi-arid tropical climate characterized by distinct wet and dry seasons. Rice is intensively cultivated here and pest and disease pressures exhibit seasonal variation influenced by local weather conditions. However, predictive modelling efforts tailored specifically for Madurai are limited, restricting the ability to deploy effective early warning systems and integrated pest management (IPM) strategies that can reduce losses and pesticide reliance.

The present study addresses this gap by developing ML-based forewarning models for major rice pests and diseases in Madurai district, using historical weather data from 2022 to 2024 combined with field incidence observations. The targeted pests include stem borer, leaf folder and brown planthopper, while the diseases studied are false smut, sheath rot and brown spot. By applying and comparing multiple ML algorithms - including Random Forest, SVM, Linear Regression and XGBoost - the study evaluates predictive performance and identifies key climatic drivers influencing pest and disease dynamics. The outputs aim to provide actionable insights for local farmers, extension services and policymakers to implement timely, precise and environmentally sustainable pest management interventions.

In summary, this research not only contributes to the growing body of knowledge on ML applications in agriculture but also provides a critical tool for enhancing rice production resilience under the increasing challenges posed by climate variability and pest/disease threats in tropical regions like Madurai. The successful integration of such forecasting models into IPM frameworks holds promise for improving food security, reducing chemical inputs and promoting sustainable agricultural development in Tamil Nadu and similar agro-ecological zones worldwide.

Materials and Methods

The study was conducted in Madurai district (latitude 9.9252° N, longitude 78.1198°E), situated in the southern part of Tamil Nadu, India. This region represents a typical rice-growing tract characterized by a semi-arid tropical climate with distinct wet and dry seasons, making it suitable for investigating the influence of weather parameters on rice pest and disease dynamics (4).

Secondary weather data, including daily records of maximum and minimum temperature (°C), relative humidity (%), wind speed (km/h) and rainfall (mm), were sourced from the India Meteorological Department (IMD) unit of the Agricultural College and Research Institute (AC&RI), Madurai, covering the period from 2022 to 2024. These variables were selected based on their established impact on pest and disease development in rice ecosystems (3).

Field observations were concurrently conducted at the Wetland Farm of AC&RI, Madurai, involving regular monitoring of major rice pests and diseases. Incidence data were recorded as Percent Disease Index (PDI) for foliar diseases such as leaf blast and brown spot and Pest Damage Index (Pest DI) for insect pests including Brown Plant Hopper (BPH) and Leaf Folder (LF). Standard plant protection protocols and scoring systems recommended by the Indian Council of Agricultural Research (10) were employed to ensure data consistency and reliability.

The combined dataset of weather parameters and pest/disease incidence was used to develop and validate predictive models tailored to local climatic conditions, following best practices for agricultural forecasting with ML (11, 12). Data preprocessing involved aggregation of daily weather data into weekly averages to align with the field observation frequency. Missing data were imputed using linear interpolation. The dataset was partitioned into training (80 %) and testing (20 %) subsets for model development and evaluation.

Data analysis

Pearson correlation coefficients were computed to assess the relationship between weather parameters and incidence rates.

All models were developed separately for each major rice pest and disease to account for differences in their ecological responses to weather variables. The modeling process followed a structured approach to ensure accuracy, reproducibility and transparency.

Daily weather data (maximum and minimum temperature in °C, relative humidity in %, rainfall in mm and wind speed in km/h) were aggregated into weekly averages to match the frequency of field observations for pest and disease incidence. The incidence data were recorded using standard scoring protocols recommended by ICAR and expressed as Percent Disease Index (PDI) or Pest Damage Index (Pest DI). Any missing values in the weather dataset were imputed using linear interpolation techniques.

The combined dataset was randomly split into training (80 %) and testing (20 %) subsets without replacement, using the create Data Partition () function in the caret package (11) in R statistical software (version 4.2.2).

Four models were built and evaluated for each pest and disease:

Linear Regression (LR)

Fitted using the base 1 m function to serve as a baseline model.

Random Forest (RF)

Implemented using the random Forest package (13). Hyperparameters such as the number of trees (ntree) and number of variables tried at each split (mtry) were tuned using 10-fold cross-validation.

Support Vector Machine (SVM)

Developed using the svm function from the e1071 package (14), with a radial basis function (RBF) kernel. Cost (C) and gamma (γ) parameters were optimized using a grid search (tune function).

Extreme Gradient Boosting (XG Boost)

Developed using the xgboost package (6), with hyperparameters such as learning rate (eta), maximum tree depth (max_depth) and number of boosting rounds (nrounds) optimized using cross -validation.

Model performance was assessed using:

- R^2 (Coefficient of Determination): Indicates proportion of variance explained.
- RMSE (Root Mean Square Error): Measures the average magnitude of prediction errors.
- MAE (Mean Absolute Error): Indicates average absolute prediction errors.

Residual diagnostics were performed using:

- Shapiro-Wilk Test for normality (shapiro.test),
- Visual inspection of residual plots for randomness, homoscedasticity and absence of autocorrelation.

All code and model parameters were logged and versioned using R scripts to ensure reproducibility. The methodology adopted in this study follows best practices for ML in agriculture as recommended by (15, 16).

Results

Correlation between weather parameters and disease and pest incidence

Disease incidence

Secondary data on major rice diseases like false smut, sheath rot and brown spot were collected along with weather parameters (maximum temperature, minimum temperature, relative humidity, wind speed and rainfall) for Madurai district during the years 2022-2024. The correlation coefficients between disease incidence (percent disease index) and weather variables are presented in Table 1. False smut incidence increased significantly with higher relative humidity (r = 0.492**) and rainfall (r = 0.381**), while higher maximum temperature (r = -0.469**) and wind speed (r = -0.164*) were associated with a reduction in disease occurrence. Sheath rot incidence was positively correlated with relative humidity (r = 0.345**) and rainfall (r = 0.362**), whereas maximum temperature showed a negative influence (r = -0.352**). Brown spot exhibited a strong positive correlation with relative humidity (r = 0.844**) and rainfall (r = 0.564**), while maximum temperature had a strong negative association (r = -0.779**), indicating that cooler and humid conditions favor its development. These relationships are visually represented in Fig. 1.

Table 1. Correlation between weather parameters and rice disease severity in Madurai

Disease	Max. Temp	Min Temp	Relative Humidity	Wind Speed	Rainfall
False smut	-0.469**	-0.180*	0.492**	-0.164*	0.381**
Sheath rot	-0.352**	-0.216	0.345**	-0.150	0.362**
Brown Spot	-0.779**	-0.193*	0.844**	-0.269**	0.564**

*Significant at 5 % level &; ** significant at 1 % level.

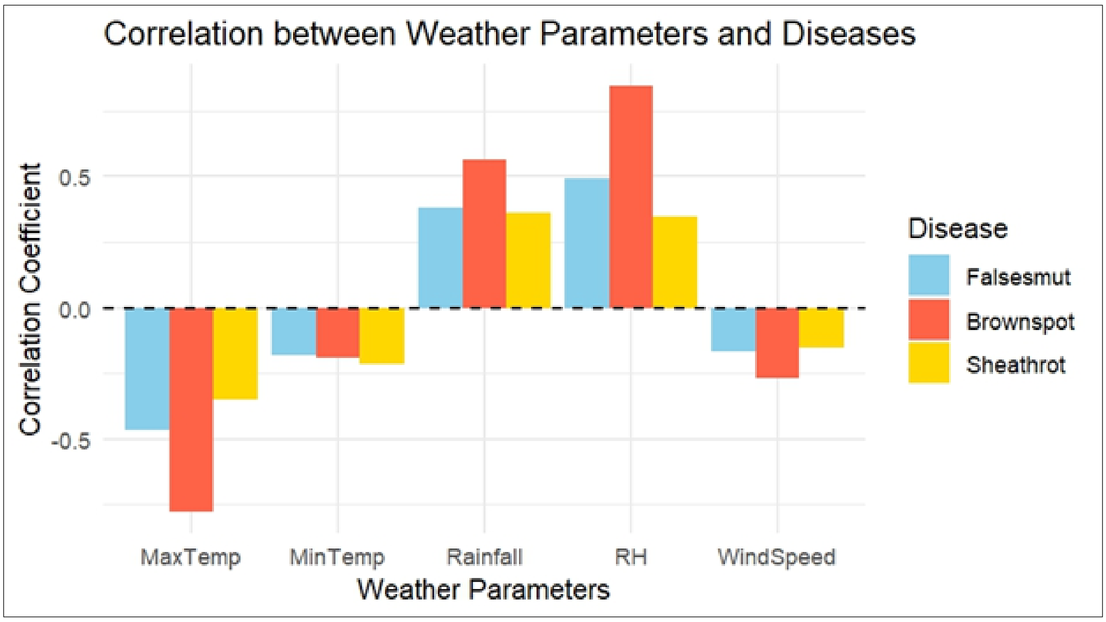


Fig.1. Correlation between weather parameters and rice disease severity in Madurai.

Pest incidence

Similarly, data on major rice pests like stem borer, leaf folder and brown planthopper (BPH) were analyzed for their correlation with weather variables presented in Table 2. Fig. 2 illustrates the relationship between pest severity and weather parameters in Madurai district. The Stem Borer infestation decreased with higher maximum (r = -0.183*) and minimum temperatures (r = -0.212**) and higher wind speeds (r = -0.200*), but slightly increased with higher relative humidity (r = 0.156*). Leaf Folder infestation was strongly negatively correlated with temperatures (maximum: r = -0.684**, minimum: r = -0.531**) and positively correlated with relative humidity (r = 0.658**), indicating that cooler and humid conditions promote its activity while Brown Planthopper (BPH) was significantly influenced by higher relative humidity (r = 0.547**) and moderate rainfall (r = 0.265**), while higher maximum temperature (r = -0.503**) and wind speed (r = -0.255**) reduced its prevalence. Relative humidity showed a positive correlation with yellow stem borer (*Scirpophaga incertulas*) and brown spot, aligning with previous studies that highlighted humidity as a key driver for pest and disease outbreaks (4).

Effective statistical models developed for disease forewarning

Table 3 presents the performance comparison of four statistical models like Linear Regression, Random Forest, SVM and XGBoost across three major rice diseases: Brown Spot, False Smut and Sheath Rot. The results highlight that Random Forest achieved the best performance for Brown Spot with the lowest MSE (39.094) and highest R^2 (0.888), while SVM excelled in predicting Sheath Rot, delivering an exceptionally low MSE (2.594) and a high R^2 (0.981). Linear Regression proved to be the most effective model for False Smut, achieving the lowest MSE (83.971) and highest R^2 (0.601), despite the overall lower performance of all models. These findings of previous study suggested that linear models may be more suitable for simpler, linear relationships. Despite the relatively low performance, Linear Regression's simplicity made it effective in this instance. These findings underscore the varying model efficacies based on the nature of the rice diseases and emphasize the need for tailored approaches for accurate pest and disease prediction.

Table 2. Correlation between weather parameters and pest severity in Madurai

Pest	Max. Temp	Min Temp	Relative Humidity	Wind Speed	Rainfall
Stem Borer	-0.183*	-0.212**	0.156*	-0.200*	0.020
Leaf Folder	-0.684**	-0.531**	0.658**	-0.059	0.088
BPH	-0.503**	-0.001	0.547**	-0.255**	0.265**

*Significant at 5 % level &; **significant at 1 % level.

The performance of four statistical models-Linear Regression, Random Forest, SVM and XG Boost-was compared for predicting the incidence of three major rice pests: Stem Borer, Leaf Folder and Brown Planthopper (Table 4). Random Forest demonstrated superior performance for both Stem Borer and Leaf Folder, achieving the highest accuracy (81.42 % and 83.23 %, respectively) and the lowest error metrics. This aligns with previous studies (6, 11) showing that ensemble techniques like Random Forest provide robust predictions in agricultural forecasting due to their ability to capture non-linear relationships and variable interactions. For Brown Planthopper, the SVM model outperformed the others, recording the highest accuracy (90.41 %) and the lowest error values (MSE = 4.142, RMSE = 2.035, MAE = 1.570). Conversely, Linear Regression consistently exhibited the poorest performance across all pests, with higher error rates and lower accuracies, indicating its limited suitability compared to advanced ML models.

Predicted vs. actual comparison of rice disease and pest prediction models

The predicted vs. actual comparison plots provide a visual assessment of model performance by comparing observed values against model predictions for three rice diseases (brown spot, false smut, sheath rot) and three rice pests (stem borer, leaf folder, brown planthopper) using Linear Regression, Random Forest, SVM and XGBoost models (Fig. 3). The x-axis represents actual disease/pest severity, while the y-axis represents the model's predicted severity. The red dashed line (y = x) serves as the ideal reference where perfect predictions would align.

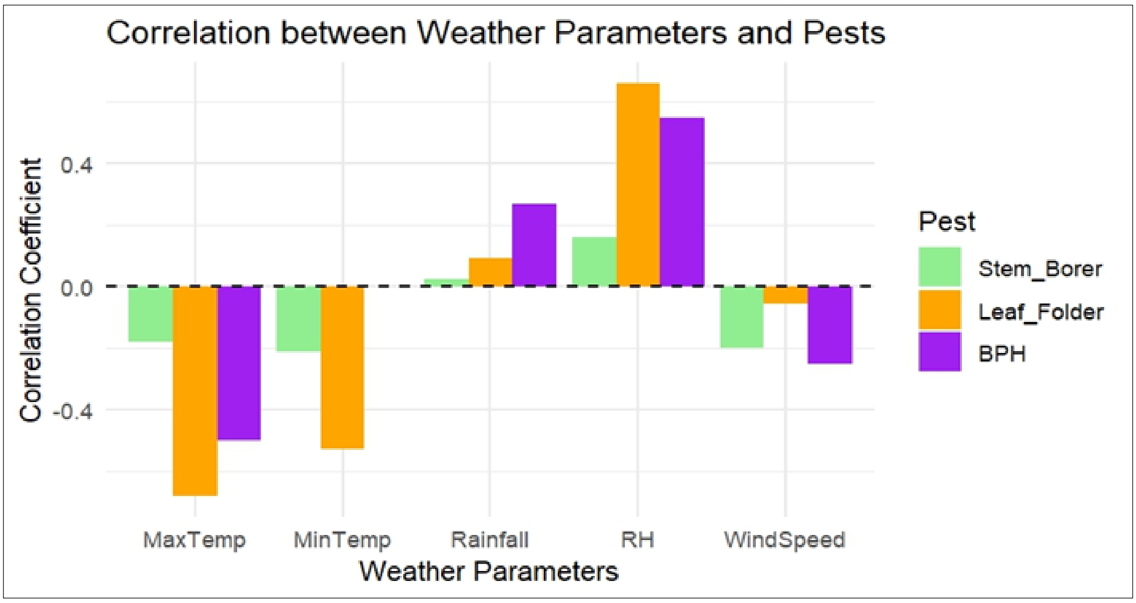


Fig. 2. Correlation between weather parameters and pest severity in Madurai.

Table 3. Performance comparison of models for predicting rice diseases

Disease	Model	MSE	RMSE	MAE	R_Square	Accuracy (%)
Brown spot	Linear Regression	58.485	7.648	6.163	0.786	73.21
	Random Forest	**39.094**	**6.252**	**3.719**	**0.888**	**88.92**
	SVM	40.054	6.329	4.874	0.874	87.47
	XGBoost	43.524	6.597	4.597	0.861	84.27
Falsesmut	**Linear Regression**	**83.971**	**9.1636**	**5.682**	**0.601**	**56.71**
	Random Forest	89.927	9.483	6.177	0.576	49.48
	SVM	89.446	9.458	5.961	0.579	48.09
	XGBoost	92.544	9.619	6.533	0.578	55.97
Sheathrot	Linear Regression	8.912	2.985	2.442	0.973	79.74
	Random Forest	18.220	4.269	2.544	0.914	87.08
	SVM	**2.594**	**1.610**	**1.253**	**0.981**	**90.41**
	XGBoost	7.497	2.738	2.321	0.941	85.42

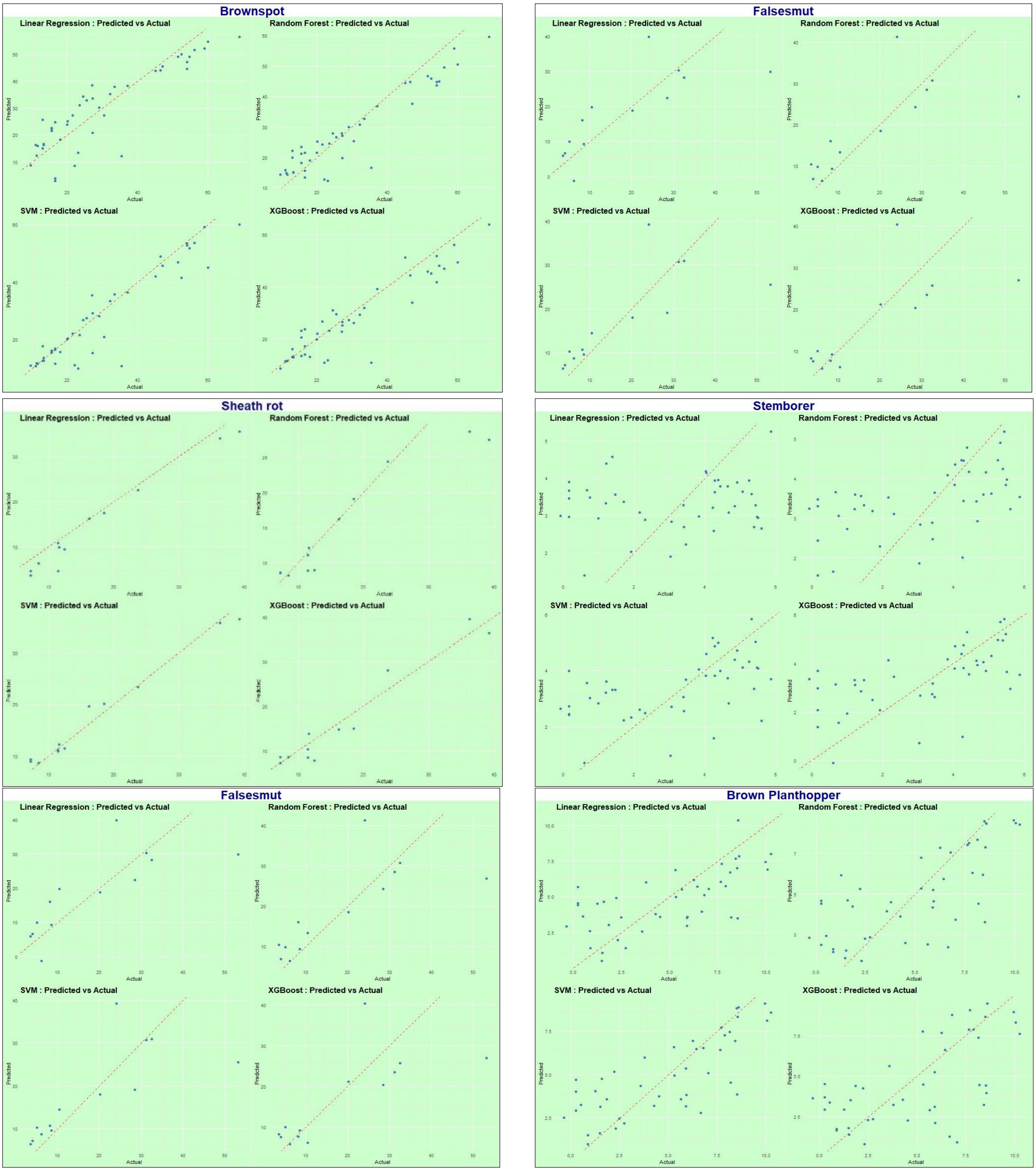


Fig. 3. Predicted vs. Actual plots for rice disease and pest severity using different models.

Table 4. Performance comparison of models for predicting rice pests

Pest	Model	MSE	RMSE	MAE	R_Square	Accuracy (%)
Stem Borer	Linear Regression	3.747	1.936	1.615	0.549	56.26
	Random Forest	**2.630**	**1.622**	**1.334**	**0.789**	**81.42**
	SVM	2.704	1.644	1.309	0.658	67.82
	XGBoost	2.384	1.544	1.219	0.550	55.23
Leaf Folder	Linear Regression	18.445	4.295	3.589	0.527	55.33
	Random Forest	**7.373**	**2.715**	**2.083**	**0.815**	**83.23**
	SVM	8.717	2.952	2.051	0.774	78.42
	XGBoost	8.019	2.832	2.031	0.796	75.54
Brown Planthopper	Linear Regression	5.643	2.376	1.953	0.639	67.23
	Random Forest	4.985	2.233	1.728	0.704	87.08
	SVM	4.142	2.035	**1.570**	0.891	**90.41**
	XGBoost	6.690	2.587	2.042	0.779	83.21

Both Random Forest and SVM demonstrated strong alignment with actual values for Brown Spot, indicating high predictive accuracy, while Linear Regression and XGBoost showed more scattered predictions, reflecting lower precision. In False Smut, Linear Regression produced the most accurate predictions, followed by XGBoost, while Random Forest and SVM displayed higher variability, especially at extreme values. Sheath Rot, SVM and Random Forest closely matched actual values, whereas Linear Regression and XGBoost demonstrated greater deviations. This aligns with the findings of study (17), who found that SVM and ensemble methods like Random Forest are particularly adept at capturing non-linear relationships in agricultural forecasting.

In pest predictions, Stem Borer predictions from Random Forest and SVM were tightly clustered around the ideal line, demonstrating high accuracy, while Linear Regression and XGBoost had higher variability. Leaf Folder predictions followed a similar trend, with Random Forest outperforming other models. Brown Planthopper, SVM delivered the best performance, closely tracking the actual values, followed by Random Forest, while Linear Regression and XGBoost displayed greater scatter and deviations.

These plots collectively highlight that Random Forest and SVM tend to provide superior predictions across both diseases and pests, whereas Linear Regression and XGBoost occasionally struggle, particularly in handling variability and extreme values.

Residual analysis

Normality of residuals (Shapiro-Wilk Test)

The Shapiro-Wilk test was applied to assess the normality of residuals for each model across the three rice diseases and pests. A p-value greater than 0.05 indicates that the residuals do not significantly deviate from normality, supporting the assumptions required for valid linear modeling and statistical inference.

The results, presented in Table 5, reveal that for brown spot, only Linear Regression and Random Forest showed normal residuals, while SVM and XGBoost deviated from normality. In False Smut, Linear Regression and XGBoost showed normally distributed residuals, meeting the normality assumption, while Random Forest and SVM did not, indicating potential deviations in error distribution. For Sheath Rot, normality was met by Linear Regression, SVM and XGBoost, whereas Random Forest showed non-normal residuals.

The normality test outcomes for each model and pest are summarized in Table 6. For Stem Borer, the residuals of all models, except for Linear Regression, were normally distributed. Specifically, Random Forest, SVM and XGBoost showed residuals that satisfied the normality assumption, while Linear Regression ($p = 0.008$) displayed a significant deviation from normality, suggesting potential violations of this assumption. Regarding Leaf Folder, all four models, including Linear Regression, Random Forest, SVM and XGBoost, exhibited residuals that were likely normal, indicating no substantial issues with error distribution. Similarly, for brown planthopper, the residuals from all models passed the Shapiro-Wilk test for normality, confirming their reliability and alignment with the normality assumption.

Residual plots

Residual plots help determine how well the model captures the relationship between the predictors and the response variable. Ideally, residuals should be randomly scattered around zero indicating a good fit. This outcome is in line with findings highlighted the importance of residual diagnostics in validating ML models for agricultural data (11, 18).

Table 5. Shapiro-wilk test for normality of residuals across different rice disease models

Disease	Model	W_Statistic	P_Value	Normality
Brown spot	Linear Regression	0.9587	0.116	Residuals are likely normal
	Random Forest	0.952	0.07	Residuals are likely normal
	SVM	0.818	0.000	Not normal (Significant deviation)
	XGBoost	0.925	0.007	Not normal (Significant deviation)
Falsesmut	Linear Regression	0.939	0.452	Residuals are likely normal
	Random Forest	0.847	0.026	Not normal (Significant deviation)
	SVM	0.799	0.007	Not normal (Significant deviation)
	XGBoost	0.908	0.174	Residuals are likely normal
Sheathrot	Linear Regression	0.916	0.253	Residuals are likely normal
	Random Forest	0.761	0.003	Not normal (Significant deviation)
	SVM	0.969	0.900	Residuals are likely normal
	XGBoost	0.965	0.857	Residuals are likely normal

Table 6. Shapiro-wilk test for normality of residuals for different models across rice pests

Pest	Model	W_Statistic	P_Value	Normality
Stem Borer	Linear Regression	0.926	0.008	Not normal (Significant deviation)
	Random Forest	0.959	0.115	Residuals are likely normal
	SVM	0.979	0.580	Residuals are likely normal
	XGBoost	0.983	0.751	Residuals are likely normal
Leaf Folder	Linear Regression	0.951	0.112	Residuals are likely normal
	Random Forest	0.967	0.349	Residuals are likely normal
	SVM	0.942	0.060	Residuals are likely normal
	XGBoost	0.952	0.121	Residuals are likely normal
Brown Planthopper	Linear Regression	0.968	0.258	Residuals are likely normal
	Random Forest	0.984	0.804	Residuals are likely normal
	SVM	0.979	0.604	Residuals are likely normal
	XGBoost	0.970	0.305	Residuals are likely normal

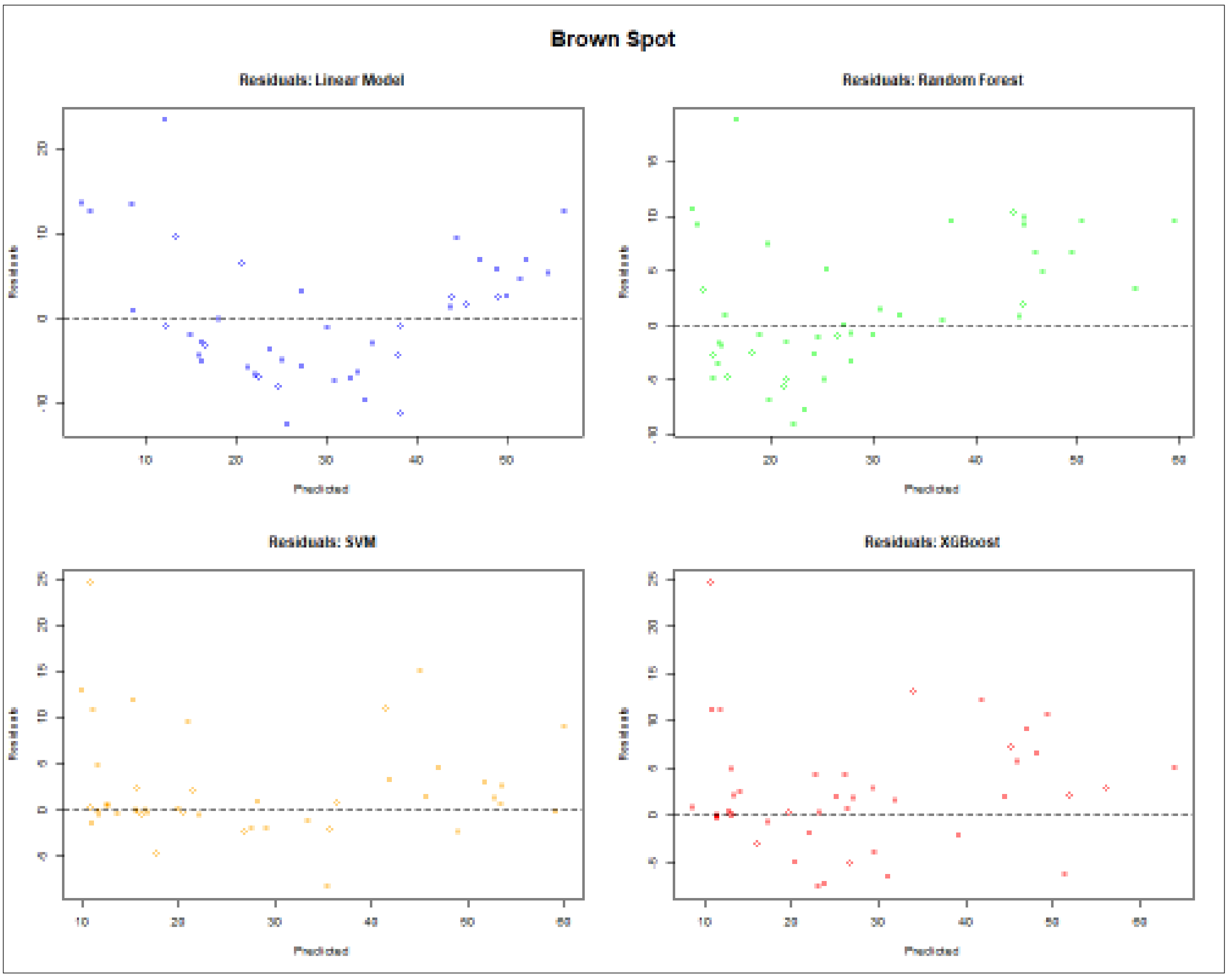
The residual plots for stem borer, leaf folder and brown planthopper indicate that Random Forest consistently provides the best fit, with residuals evenly scattered around zero and no clear patterns as depicted in Fig. 4. SVM and XGBoost show moderate performance with slight clustering, while Linear Regression consistently displays patterns and heteroscedasticity, indicating a poor fit.

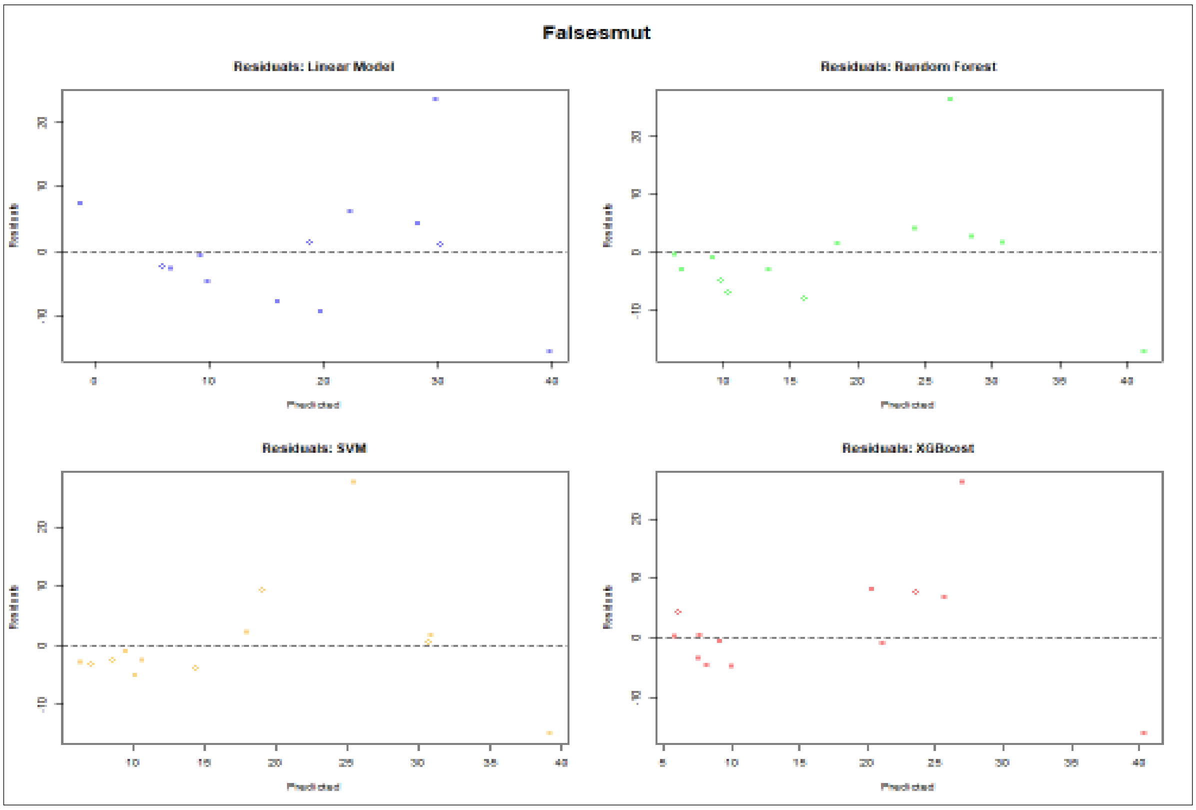
Brown spot, Random Forest and SVM show the best predictive performance, with desirable residual patterns indicating minimal error. False smut, the Linear Model excels, with tightly clustered residuals suggesting high accuracy. In predicting Sheath Rot, SVM and Random Forest models are particularly effective, both showing consistent residuals around zero.

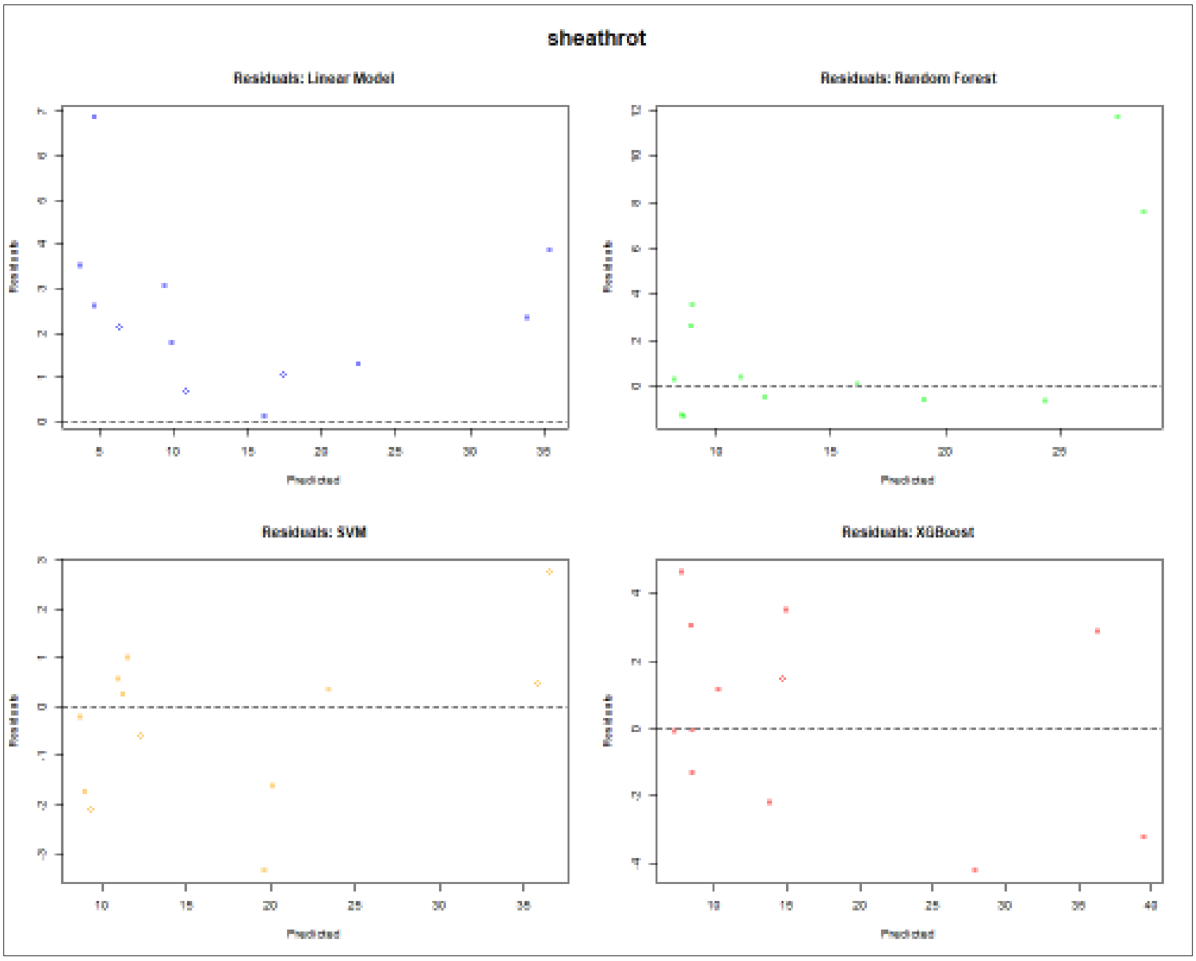
Discussion

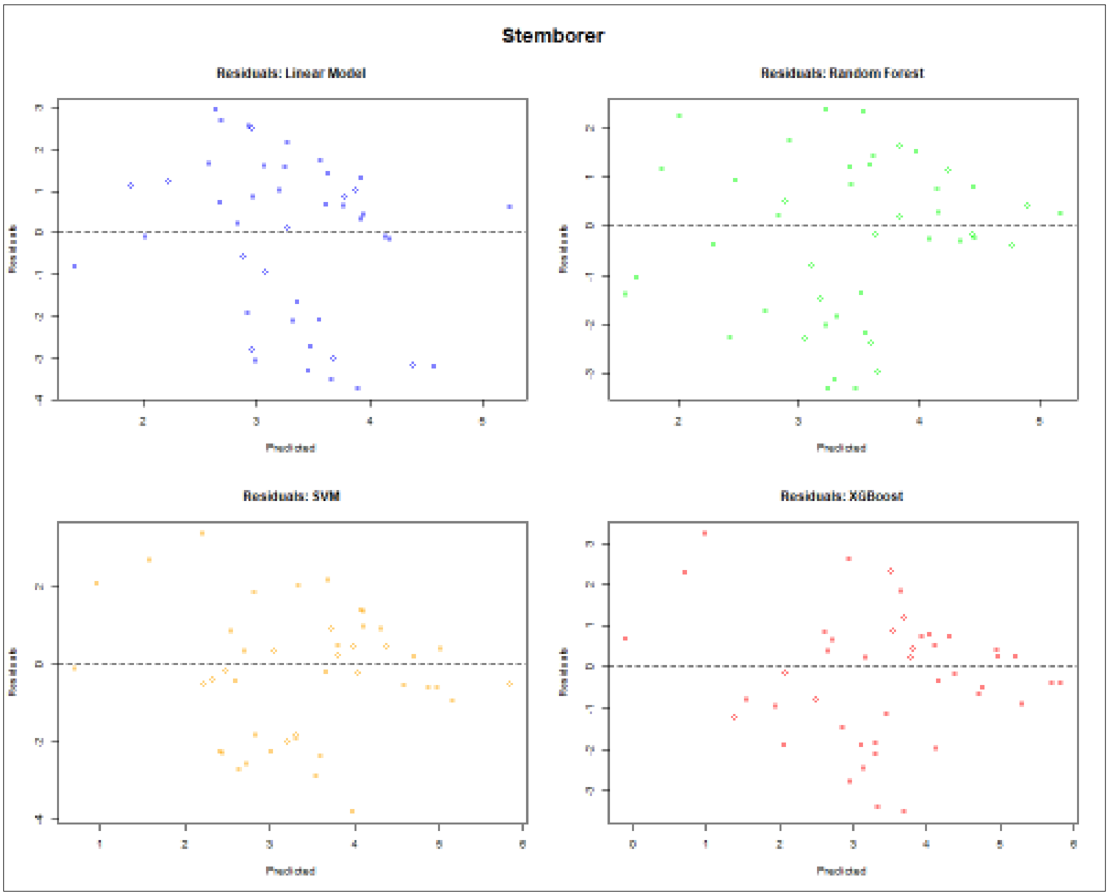
This study highlights the significant influence of climatic variables-particularly relative humidity and temperature-on the incidence of major rice pests and diseases. Brown Spot showed a strong positive correlation with high humidity and a negative correlation with maximum temperature, indicating clear climatic thresholds conducive to disease development. The use of ML models, especially Random Forest and SVM, yielded high predictive accuracy due to their capacity to model complex, nonlinear relationships between weather parameters and pest/disease outbreaks, aligning with findings by (19, 20).

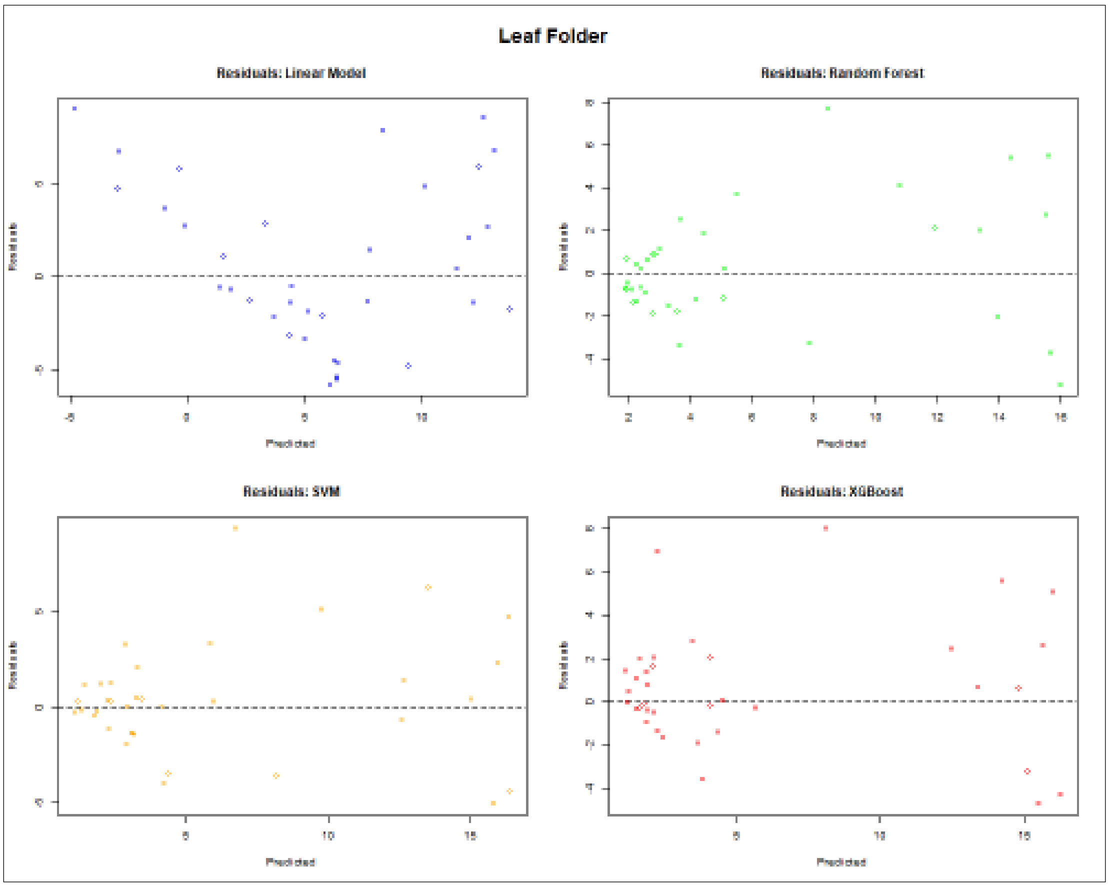
Beyond prediction, these models offer practical value for early warning and decision-making in Integrated Pest Management (IPM). When integrated with real-time weather




Falsesmut
Residuals: Linear Model
Residuals: Random Forest
Residuals: SVM
Residuals: XGBoost
Residuals
Predicted


sheathrot
Residuals: Linear Model
Residuals: Random Forest
Residuals: SVM
Residuals: XGBoost
Residuals
Predicted


Stemborer
Residuals: Linear Model
Residuals: Random Forest
Residuals: SVM
Residuals: XGBoost
Residuals
Predicted


Leaf Folder
Residuals: Linear Model
Residuals: Random Forest
Residuals: SVM
Residuals: XGBoost
Residuals
Predicted

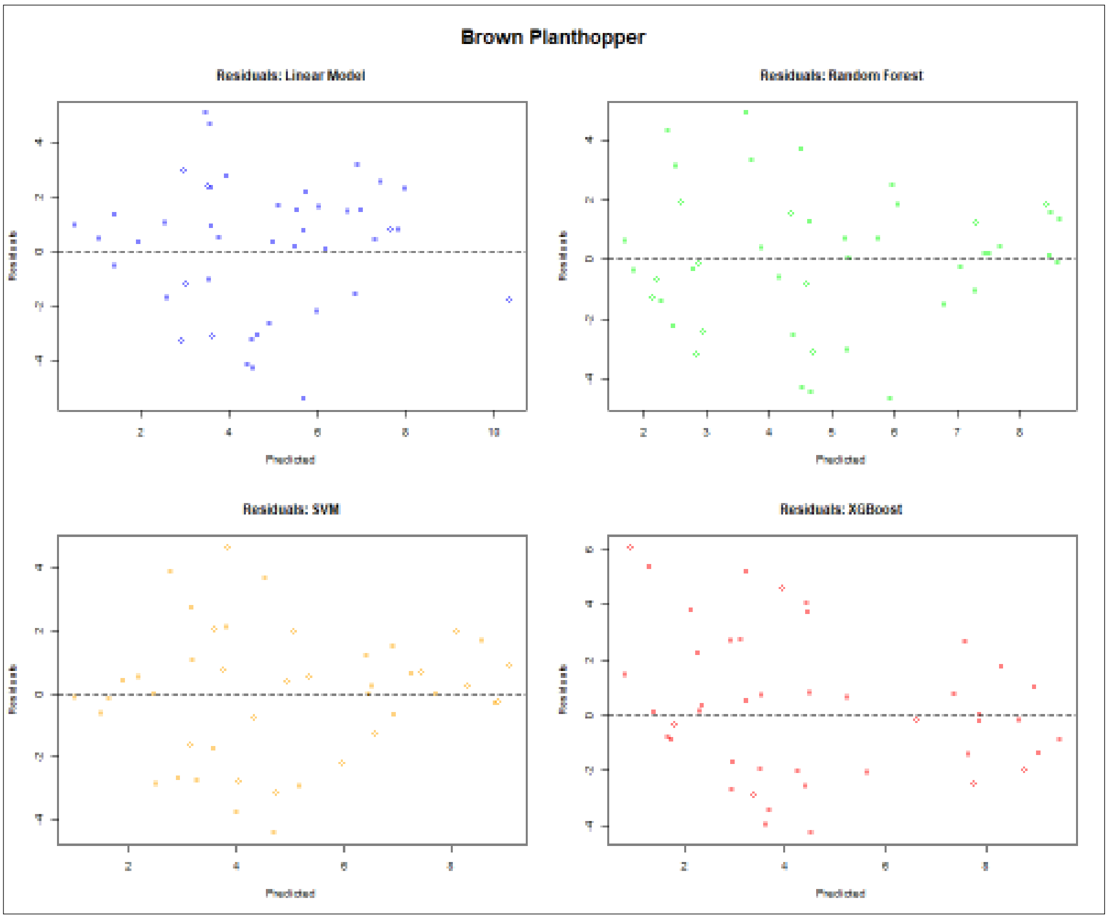


Fig. 4. Residual plots for model performance across rice diseases and pests.

data, such models can support timely, localized advisories, promoting targeted interventions over blanket pesticide use. Previous researchers (21) demonstrated the feasibility of such integration using mobile platforms and sensor-based data. However, real-world application demands continuous data input, regular validation and user-friendly tools for farmers and extension services.

Overall, the study reinforces the relevance of data-driven pest forecasting in climate-sensitive regions and underscores the potential of ML to enhance IPM strategies, reduce crop loss and support sustainable rice production systems.

Conclusion

The present study successfully established predictive models for major rice pests and diseases using weather parameters in Madurai district, Tamil Nadu. Among the climatic variables, relative humidity and temperature emerged as key drivers influencing pest and disease dynamics. Notably, Brown Spot incidence showed a strong positive correlation with high relative humidity and a negative correlation with maximum temperature, indicating its vulnerability to specific climatic shifts. The application of ML algorithms, particularly Random Forest and Support Vector Machine, proved effective in forecasting pest and disease occurrences, delivering high accuracy and reliability in prediction. These models offer valuable insights for the early warning and strategic management of rice pests and diseases, supporting sustainable agricultural practices. Importantly, the integration of such predictive tools into Integrated Pest Management (IPM) frameworks enables informed decision-making, such as optimal timing of interventions, reduced pesticide use and targeted control strategies. By facilitating proactive rather than reactive responses, ML models contribute to environmentally responsible pest and disease control, thereby enhancing the overall effectiveness and sustainability of IPM approaches in climate-sensitive rice production systems.

Acknowledgements

We would like to extend our heartfelt appreciation to all the individuals and organizations who have contributed to the publications of this research and this work.

Authors' contributions

BS, MK conducted the research experiments and BS and MK wrote the manuscript. SPS corrected the English part. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: There is no conflict of interest between the authors.

Ethical issues: None

References

1. FAO. The State of Food Security and Nutrition in the World 2022. Rome: Food and Agriculture Organization of the United Nations; 2022. https://www.fao.org/publications/sofi/2022/en/
2. Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. The global burden of pathogens and pests on major food crops. Nat Ecol Evol. 2019;3(3):430-9. https://doi.org/10.1038/s41559-019-0822-6
3. Manandhar HK, Barbetti MJ, Thapa RB, Bhatta MR. Variability of rice disease occurrence and its relation to weather variables in Nepal. J Phytopathol. 2017;165(1):14-24. https://doi.org/10.1111/jph.12541
4. Luo X, et al. Influence of weather factors on rice diseases and pests in tropical areas. Agric For Meteorol. 2015;220:108-14.
5. Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, Butler EE, et al. Random forests for global and regional crop yield predictions. PLoS One. 2016;11(6):e0156571. https://doi.org/10.1371/journal.pone.0156571
6. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min. 2016:785-94. https://doi.org/10.1145/2939672.2939785
7. Ahmad MW, et al. A review on applications of machine learning in agriculture. Sustainability. 2020;12(9):3421. https://doi.org/10.3390/su12093421
8. Wang G, et al. Application of machine learning methods for the prediction of rice blast occurrence. Sci Rep. 2019;9(1):2869. https://doi.org/10.1038/s41598-019-38966-0
9. Sahu RK, et al. Agro-ecological zoning and crop suitability analysis using GIS and remote sensing in India. Ecol Indic. 2023;148:110102. https://doi.org/10.1016/j.ecolind.2023.110102
10. Indian Council of Agricultural Research. Standard Evaluation System for Rice. New Delhi: ICAR; 2013.
11. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28(5):1-26. https://doi.org/10.18637/jss.v028.i05
12. Singh A, et al. Application of machine learning models in agriculture: A review. Comput Electron Agric. 2020;179:105796. https://doi.org/10.1016/j.compag.2020.105796
13. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002;2(3):18-22.
14. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. e1071: Misc Functions of the Department of Statistics, TU Wien. R package version 1.7 11; 2021.
15. Singh A, Jones JW, Kumari R. Machine learning for agricultural forecasting: New opportunities for sustainable development. Agric Syst. 2020;190:103101. https://doi.org/10.1016/j.agsy.2021.103101
16. Meena RK, Sahu P, Shukla M. Application of machine learning techniques in forecasting crop diseases: A review. Comput Electron Agric. 2022;192:106584. https://doi.org/10.1016/j.compag.2022.106584
17. Sun J, et al. Support vector machines in the prediction of rice pest dynamics. Ecol Inform. 2020;56:101071. https://doi.org/10.1016/j.ecoinf.2020.101071
18. Zhang X, et al. Impact of wind speed on the spread of rice stem borer. Agric For Meteorol. 2018;253:50-60. https://doi.org/10.1016/j.agrformet.2018.01.016
19. Kumar N, Singh S, Yadav D. Forecasting rice pest attacks using machine learning models: A case study. Comput Electron Agric. 2022;196:106882. https://doi.org/10.1016/j.compag.2022.106882
20. Sahu P, Meena RK, Sharma S. Application of ensemble machine learning models for predicting crop diseases: A case on rice blast disease. Agronomy. 2023;13(2):520. https://doi.org/10.3390/agronomy13020520
21. Thakur N, Prakash S. Integrating weather and crop data using deep learning for pest forecasting in rice ecosystems. Artif Intell Agric. 2024;8:34-46. https://doi.org/10.1016/j.aiia.2023.10.005

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.