Environ Monit Assess (2025) 197:366
https://doi.org/10.1007/s10661-025-13744-w

RESEARCH

Leveraging ML to predict climate change impact on rice crop disease in Eastern India

Satiprasad Sahoo · Chiranjit Singha · Ajit Govind · Mamta Sharma

Received: 29 October 2024 / Accepted: 4 February 2025 / Published online: 8 March 2025

Abstract Rice crop disease is critical in precision agriculture due to various influencing components and unstable environments. The current study uses machine learning (ML) models to predict rice crop disease in Eastern India based on biophysical factors for current and future scenarios. The nine biophysical parameters are precipitation (Pr), maximum temperature (T_{max}), minimum temperature (T_{min}), soil texture (ST), available water capacity (AWC), normalized difference vegetation index (NDVI), soil-adjusted vegetation index (SAVI), normalized difference chlorophyll index (NDCI), and normalized difference moisture index (NDMI) by Random forest (RF), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGB), Artificial Neural Net (ANN), and Support vector Machine (SVM). The multicollinearity test Boruta feature selection techniques that assessed interdependency and prioritized the factors impacting crop disease. However, climatic change scenarios were created using the most recent Climate Coupled Model Intercomparison Project Phase 6 (CMIP6) Shared Socioeconomic Pathways (SSP) 2–4.5 and SSP5-8.5 datasets. The rice crop disease validation was accomplished using 1105 field-based farmer observation recordings. According to the current findings, Purba Bardhaman district experienced a 96.72% spread of *rice brown spot* disease due to weather conditions. In contrast, *rice blast* diseases are prevalent in the north-western region of Birbhum district, affecting 72.38% of rice plants due to high temperatures, water deficits, and low soil moisture. *Rice tungro* disease affects 63.45% of the rice plants in Bankura district due to nitrogen and zinc deficiencies. It was discovered that the link between NDMI and NDVI is robust and positive, with values ranging from 0.8 to 1. According to SHAP analysis, Pr, T_{min}, and T_{max} are the top three climatic variables impacting all types of disease cases. The study's findings could have a substantial impact on precision crop protection and meeting the United Nations Sustainable Development Goals.

Keywords Food security · Rice disease · ML · Remote sensing

S. Sahoo (✉) · A. Govind
International Center for Agricultural Research in the Dry Areas (ICARDA), 2 Port Said, Victoria Sq, Ismail El-Shaaer Building, Maadi, Cairo 11728, Egypt
e-mail: satispss@gmail.com

A. Govind
e-mail: A.Govind@cgiar.org

S. Sahoo
Prajukti Research Private Limited, Baruipur 743610, West Bengal, India

C. Singha
Department of Agricultural Engineering, Institute of Agriculture, Visva-Bharati (A Central University), Sriniketan, Birbhum, West Bengal 731236, India
e-mail: singha.chiranjit@gmail.com

M. Sharma
International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, Telangana 502324, India
e-mail: mamta.sharma@icrisat.org

Introduction

The global population is expected to exceed nine billion by 2050, necessitating a 50% increase in food production above current levels (Tripathi et al., 2019). Maintaining long-term food production without expanding the area under cultivation is a huge difficulty. Plant breeding is critical for increasing production and resistance to both current and emerging biotic and abiotic stressors. Cereals are the most important staple crops for global food security, accounting for approximately 42.5% of total food calories. Rice and wheat account for over half of cereal consumption, with projections of 703 and 503 million tons in 2017/2018, respectively.

Rice, one of the world's most important food crops, makes a significant contribution to India's food and nutritional security. Rice production has evolved throughout the decades, from simple cultivation procedures to complex cultivation to boost output. Rice yields have increased significantly, particularly since the 1960s, owing to the advent of high-yielding semi-dwarf cultivars, which demand additional inputs such as chemical fertilizers, water, and other resources (Shivappa et al., 2021). Disease damage to rice can significantly diminish production. Bacteria, viruses, and fungus primarily cause them. Planting a disease-resistant variety is the easiest and, in many cases, the least expensive way to manage disease. Rice is grown in practically every state in India. Rice plant diseases pose a major threat to crops' quality and quantity. Various diseases, such as foot rot and sheath blight, can affect rice plants, leading to a decline in agricultural productivity (Prismantoro et al., 2024). Paddy diseases can devastate rice production and farmers' livelihoods. It is one of India's most significant food crops, accounting for one-quarter of the total planted area (Sahoo et al., 2024). Moreover, half of the world's population relies on rice as their primary food source. India is second after China in world rice output. The total rice production during 2022–2023 is 125 million tons and 567 million tonnes by 2030 (Mohidem et al., 2022). The overall area under rice cultivation in 2022–2023 is 45.5 million hectares, with an average production of 4.1 tonnes/ha (USDA, 2022). Paddy is most farmed in India during the Kharif season. Tropical and subtropical hot and humid temperatures are ideal for its growth. Understanding the causes, symptoms, and treatment options for paddy diseases is critical for preventing their spread and reducing their effect. Changing climatic circumstances facilitate the spread of diseases to new areas and intensify their impact (Mwangi et al., 2023). In addition to the previously extensively spread rice diseases such as blast, tungro, sheath blight, fake, and bacterial leaf blight, new diseases such as false smut of rice and wheat blast are becoming increasingly dangerous (Azizi & Lau, 2022). To reduce the impact of these diseases, environmentally friendly and cost-effective preventative and control strategies are required. The development and use of disease-resistant cultivars is the most successful, cost-effective, and environmentally responsible method of managing these risks (Li et al., 2023). Thus, the present goal is to employ machine learning (ML) algorithms to predict Kharif rice crop disease in eastern India while taking into consideration biophysical and meteorological factors. ML models offer several advantages over traditional methods. It excels at capturing complex, nonlinear relationships among variables, which are challenging for conventional approaches, i.e., multi-criteria decision-making analysis (Sarker, 2021). ML models provide higher predictive accuracy by leveraging advanced algorithms and large datasets (Mumuni & Mumuni, 2024). These are scalable and automated, enabling efficient processing of vast amounts of data, and can integrate diverse data types (e.g., spatial, temporal, and categorical) for more comprehensive analyses. Unlike static traditional methods, ML models adapt to new data, improving their performance over time (Yaqoob et al., 2023). Additionally, ML models demonstrate robustness against noisy or missing data and are effective in simulating future scenarios, such as climate change impacts, based on historical and synthetic datasets (Singha et al., 2024).

A few researchers employed ML algorithms to predict rice crop disease during changing conditions in the environment. According to Jackulin and Murugavalli (2022), a comprehensive assessment of the numerous strategies used in plant disease control has been conducted using ML and deep learning techniques. Yan et al. (2022) developed an artificial inoculation technique for the artificial induction of bakanae disease that uses mung bean medium to accelerate the proliferation of Fusarium fujikuroi spores. This approach will allow for the rapid assessment of bakanae disease resistance, which will be beneficial for rice breeding. Aggarwal et al.

(2022) investigated rice disease identification, seedling health, and grain quality utilizing sophisticated artificial intelligence and ML approaches, as well as better agribusiness, to broaden the notion of rice. Xinyue et al. (2023) presented their findings on the features, categorization, and molecular mechanisms of necrotic lesion formation. It also overlooked the molecular regulatory pathway of genes involved in rice disease resistance, summarized the relationship between resistance and rice yield using newly developed gene editing, and discussed the use of molecular design technology to better reproduce disease prevention and high-yield varieties. Singh et al. (2023) suggested a custom convolutional neural network (CNN) architecture for identifying and categorizing rice plant diseases by lowering the number of network parameters and including 1400 on-field healthy rice leaf image datasets to identify disease-free plants. Lu et al. (2023) proposed an enhanced rice disease identification approach that combines a CNN with a bidirectional gated recurrent unit (BiGRU). This study also identified four types of rice disease and offered a reliable approach to disease detection. Nayak et al. (2023) began processing smartphone photographs of rice plant sections into numerous categories, as well as real-time validation images, to detect rice disease and nutrient shortages. Different image segmentation algorithms were used to isolate the affected areas. Zheng et al. (2023) examined current and projected trends in remote sensing for rice crop monitoring. This paper goes over the mechanics and applications of numerous data sources for monitoring rice disease and pests, as well as a summary of current monitoring approaches such as statistical discriminant type, ML, and deep learning algorithms. It also includes a framework for monitoring unknown diseases and pests, as well as a discussion of obstacles and prospects in rice disease and pest monitoring using remote sensing. Kamarudin et al. (2024) analyzed whole-genome resequencing data for blast resistance with kernel elongation features in the Mahsuri Mutant, mutant line, and parental line in Malaysia.

Most prior research has focused on utilizing computer vision technology for classifying various rice diseases (Deng et al., 2020, 2021; Mahadevan et al., 2024; Udayananda et al., 2022). However, these studies have largely overlooked the integration of geo-environmental factors and the impact of climate variability on different spatial locations. To address this research gap, the study aims to spatially predict Kharif rice crop diseases using ML methods that incorporate the effects of climate change, facilitating the development of both current and future scenarios. The peculiarity of this study is that it is the first-time rice crop disease prediction has been considered considering climate change. However, nine biophysical and climatic parameters, including precipitation (Pr), maximum temperature (T_{max}), minimum temperature (T_{min}), available water capacity (AWC), soil texture (ST), normalized difference vegetation index (NDVI), normalized difference moisture index (NDMI), soil adjusted vegetation index (SAVI), and normalized difference chlorophyll index (NDCI), were used to predict rice crop diseases. The Model for Interdisciplinary Research on Climate 6 (MIROC6) SSP2-4.5 and SSP5-8.5 data from the CMIP6 climate model were used to estimate climate change scenarios. The multicollinearity test and the Boruta algorithm were used to identify the influencing biophysical and climatic components for rice crop disease prediction.

Study area

The study area in West Bengal, covering 37,836.5 km^2, includes eight districts: Murshidabad, Birbhum, Purba Bardhaman, Paschim Bardhaman, Bankura, Purulia, Jhargram, and Paschim Medinipur (Fig. 1). The district, located in the heart of West Bengal, has a humid climate. The Birbhum district, also known as "The Land of Red Soil," has undulating topography and strong reddish clay soil in its western half, "Radh," and rich alluvial soil in its eastern half, "Bagri." Its rivers include Ajay, Bakreshwar, Brahmani, Bansloi, Dwarka, Hinglo, Kopai, and Mayurakshi.

Materials and methods

Crop disease sampling

During the survey conducted in October and November 2023, data on Kharif rice crop diseases were collected from 1105 agricultural fields. In the study, a total of seven types of rice diseases were identified, including rice brown spot, rice blast, rice tungro, bakanae, stem rot, and leaf blight. However, only three major diseases—rice brown spot, rice blast, and rice tungro—were chosen for

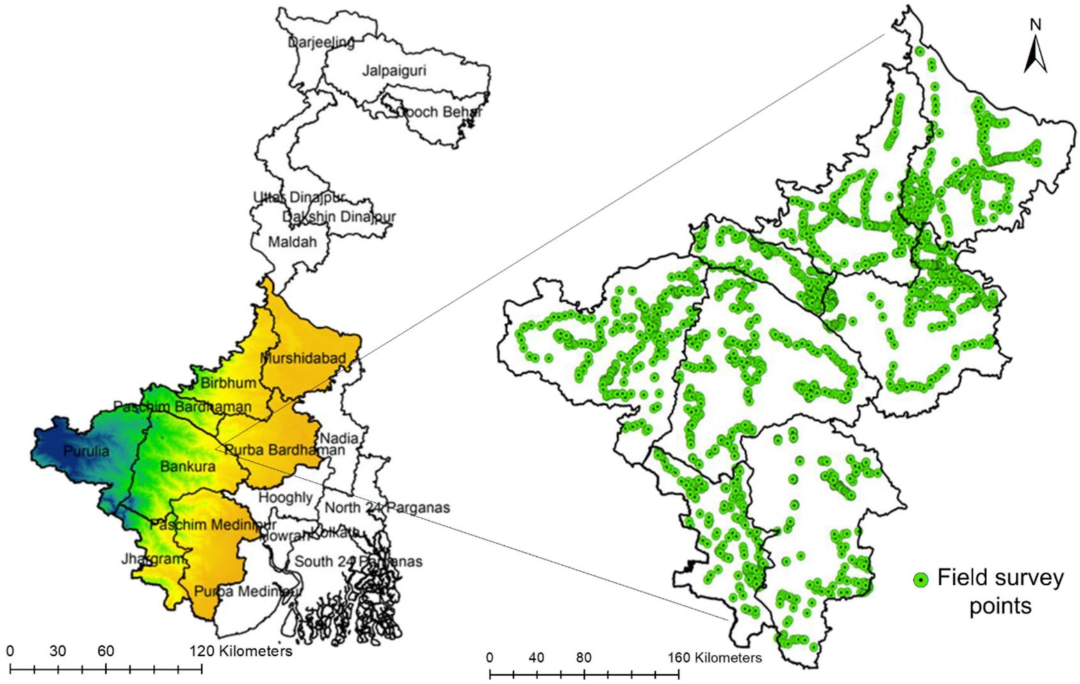


Fig. 1 Location of the study research

analysis due to the availability of data required for ML model evaluation. These were geo-tagged using a handheld GPS device to map disease areas in Eastern India. The field survey also recorded various Kharif rice varieties being cultivated, including Swarna, Ratna, Lalat, Varam, IR36, IR41, MTU7029, Badsha Bhog, Dhanaraj, Chaitali, Maharaj, Minikit, Gutka, and Gobindo Bhog. A random sampling analysis was conducted using a handheld GPS (Garmin Ltd., Olathe, KS, USA) to determine the coordinates at each sampling location. Random sampling techniques were employed to select farms or plots for data collection, minimizing the risk of systematic biases that could distort the results. During the survey, detailed information on rice yield was collected, covering aspects such as farming methods, fertilizer use, irrigation plans, soil conditions, the impact of soil moisture on climatic hazards, market trends, and more, incorporating the insights and experiences of local farmers for a comprehensive analysis. The study involved visiting various locations in the area to gather samples and document the experiences and observations of local farmers regarding recent changes in their crop rotation. This crop disease database was used as the dependent variable for the crop disease classification analysis. The crop diseases classification analysis identified three types of diseases: brown spot (labeled as 1), rice blast (labeled as 2), and rice tungro (labeled as 3). Previous study indicated that using 70% of the data for training and 30% for testing in ML classification analyses provided reliable model performance, despite potential inconsistencies (Al-Sheriadeh & Daqdouq, 2024; Chowdhury et al., 2024). For the ML tenfold cross-validation analysis, the disease samples of each label were randomly divided in a 70:30 ratio, with brown spot comprising approximately 260 training instances and 112 testing instances out of a total of 372, rice blast including around 320 training instances and 138 testing instances out of 458, and rice tungro consisting of 192 training instances, and 83 testing instances out of 275.

Crop diseases contributing factors

In this study, 9 crop disease conditioning factors (CDCFs) were chosen, encompassing climatological, soil, and vegetation indices elements. These selections were based on a review of existing literature, data availability, and the specific characteristics of the study area under examination (Lu et al., 2023; Nayak et al., 2023; Singh et al., 2023). The climatological CDCFs included Pr, T_{max}, and T_{min}. The soil-related CDCFs include ST and AWC. The NDVI, SAVI, NDCI, and NDMI were the selected vegetation indices CDCFs (Table 1). The future prediction of rice crop diseases is generated through the coupled model intercomparison project-6 (CMIP-6) future projections of Pr, T_{max}, and T_{min} data. We downloaded the climatic parameters to a daily scale from Coupled Model Intercomparison Project-6 CMIP-6 Global Circulation Models (GCMs) through the Google Earth Engine cloud (GEE). Finally, each of these CDCFs was converted into a raster format with a 30 m × 30 m spatial resolution using the bilinear resampling technique in R software v4.3.2 (Fig 2).

In this study, the average annual Pr, T_{max}, and T_{min} were calculated based on meteorological data spanning 40 years (1990–2030) sourced from Terra:Climate database (spatial resolution of 4638.3 m). All the climatological dataset was mapped using the Inverse Distance Weighted (IDW) interpolation algorithm in ArcGIS software v10.7. In the study area, the precipitation (mm) value ranges between 967.04 and 1351.92 mm. However, the T_{max} and T_{min} (°C) values varied from 30.11 to 32.39 (°C) and 19.10 to 22.13 (°C) respectively.

This research utilized Sentinel-2 optical satellite imagery to generate spectral indicators for rice crop disease mapping analysis. Sentinel-2 is a European multispectral satellite developed through the European Space Agency (ESA). All the vegetation indices are derived from the mosaicking cloud-free Sentinel 2 images (10-m spatial resolution) during the peak growing season of kharif rice (October–November 2023). Accordingly, 157 images were filter preprocessed (cloud mask, radiometric, geometric correction, median composite, and layer stacking) through the GEE cloud in the peak rice growing season. Additionally, four vegetation indices, including the NDVI, NDMI, SAVI, and NDCI were computed with the following Table 2. In this study, the NDVI values range between − 0.273 and 0.682. The NDMI value varies from − 0.387 to 0.439. The SAVI value ranges between − 1.103 and 0.092. While the NDCI value varies between − 0.021 and 0.310 (Table 3).

The ST vector layer was obtained from the National Bureau of Soil Survey & Land Use Planning (NBSS & LUP, India) and imported into ArcGIS for ST layer mapping. Subsequently, the ST maps were classified into eight major categories, including clay, silt clay loam, clay loam, loam, sandy clay loam, silt clay, silt loam, and sandy loam. The GEE extracted the AWC factor from Soil Grids 250 m ISRIC World Soil Information data (15–30-m depth). In the study area, the AWC value ranges from 5 to 15.63 (Fig. 2).

Future projections of T_{max}, T_{min}, and precipitation data, from the CMIP-6 were utilized to forecast the area affected by crop diseases in the study

Table 2 Vegetation indices formulas employed in the research

Index	Formula	Source
NDVI	$\frac{(NIR-RED)}{(NIR+RED)}$	Rouse et al., 1973
NDMI	$\frac{(NIR-SWIR1)}{(NIR+SWIR1)}$	
SAVI	$\frac{(NIR-RED)}{(NIR+RED+0.5)} \times (1+0.5)$	Huete, 1988
NDCI	$\frac{(Red\ Edge\ 1-RED)}{(Red\ Edge\ 1+RED)}$	Mishra et al., 2012

Table 1 Description of data sources crop disease susceptibility

Parameters	Description	Source
Precipitation (mm), T_{max} (°C), T_{min} (°C)	IDAHO_EPSCOR/TERRACLIMATE, 4638.3 m (1958–2023)	Abatzoglou et al., 2018
Soil texture	NBSS & LUP, India, RF 1:10,000	URL: http://14.139.123.73/bhoomigeoportal
Available water capacity (AWC) (volumetric fraction) with FC = pF 2.3	ISRIC — World Soil Information, SoilGrids250m, (15–30 cm depth)	URL: https://data.isric.org
NDVI, NDMI, SAVI, NDCI	COPERNICUS/S2_SR, 10 m (2023)	URL:https://scihub.copernicus.eu/
Future precipitation (mm), T_{max} (°C), T_{min} (°C)	NASA/GDDP-CMIP6, 27,830 m, (1990–2030), (ssp245, ssp585)	(URL: https://registry.opendata.aws/nex-gddp-cmip6/)

Table 3 Details statistics of the crop diseases' contributing factors

Parameters	Min	Max	Mean	Standard Deviation
Precipitation (mm)	967.04	1351.92	1090.0	5.92
T_{max} (°C)	30.11	32.39	31.81	2.58
T_{min} (°C)	19.10	22.13	21.06	3.93
Available water capacity (AWC)	5	15.63	9.88	0.67
Normalized Difference Vegetation Index (NDVI)	−0.273	0.682	0.40	0.09
Normalized Difference Moisture Index (NDMI)	−0.387	0.439	0.15	0.09
Soil Adjusted Vegetation Index (SAVI)	−1.103	0.092	−0.75	0.12
Normalized Difference Chlorophyll Index (NDCI)	−0.021	0.31	0.15	0.03

area. Specifically, we utilized two combined Shared Socioeconomic Pathway (SSP) scenarios, namely SSP-245 and SSP-585. To ensure consistency in the projected climatic data, we averaged the datasets derived from the MIROC6 GCMs spanning 40 years from 1990 to 2030.

Methodology

The overall methodology deployed in this study is as follows (Fig. 3). Primarily, crop disease inventory data and nine CDCFs were applied to generate training, validation, and testing datasets for the ML model system, after performing Pearson's correlation and multicollinearity analysis. The feature selection of the CDCFs is selected through the Boruta analysis. Next, crop disease susceptibility maps were developed for the study area based on the spatial intra-association between CDCFs and crop disease inventory data, and the ML model's performance was assessed using certain performance indicators. In addition, the future crop disease susceptibility maps expanded through the CMIP6 datasets. Subsequently, an evaluation was carried out to ascertain the trained model's capacity to forecast crop disease susceptibility in the study area. Ultimately, the model results and the significance of CDCFs were analyzed using the most effective model, employing the explainable artificial intelligence (XAI) technique known as SHapley Additive exPlanations (SHAP).

Multicollinearity test

An analysis for multicollinearity was performed using the Variance Inflation Factor (VIF) technique to evaluate the interdependencies among the CDCFs. Multicollinearity presents a challenge in accurately estimating model outputs, potentially misrepresenting the importance of variables within statistical models due to high inter-correlations (Chang et al., 2019). Factors exhibiting a VIF greater than 5 or a Tolerance (TOL) less than 0.1 are affected by multicollinearity and are recommended for exclusion. The formula for independent variables was represented as x = {×1, ×2,…, xn}, with R2j signifying the coefficient of determination that quantifies the linear correlation of the ith independent variable with the others. The formula for calculating VIF is as described (Eq. (1)).

$$VIF = \frac{1}{1 - R_i^2} = \frac{1}{\text{Tolerance}} \quad (1)$$

Boruta feature selection

The Boruta algorithm was implemented to prioritize the identified factors influencing crop disease. Utilizing the combined dataset, the algorithm, integrated within the R statistical package, extends the principles of the Random Forest classifier. By introducing additional randomness into the system and aggregating outcomes from a collection of randomized samples, Boruta mitigates the influence of incidental variations and correlations. This augmented randomness offers a more discernible perspective on the genuine significance of each property. Notably, the Boruta algorithm has previously demonstrated efficacy in predicting apple yield prediction in India (Singha et al., 2023).

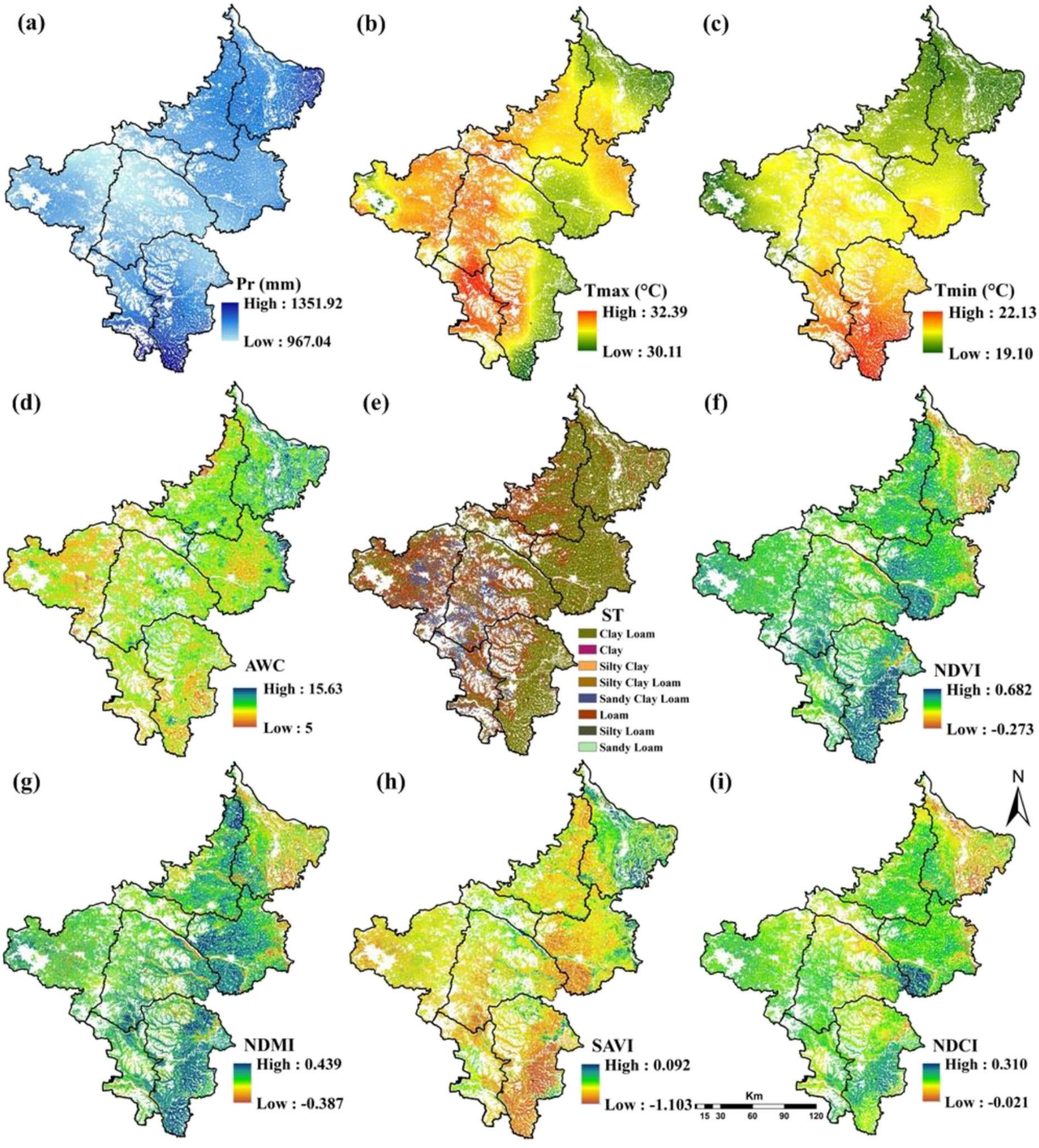


Fig. 2 Crop diseases contributing factors

ML Methods

Random Forest (RF)

Random Forest (RF) stands as a prominent algorithm within ML for both regression and classification tasks, introduced by Breiman in 2001. This method leverages bagging to diminish variance, particularly effective for models like decision trees that tend to have high variance but low bias. Essentially, RF enhances the bagging approach by incorporating numerous decision trees. Each tree within the RF

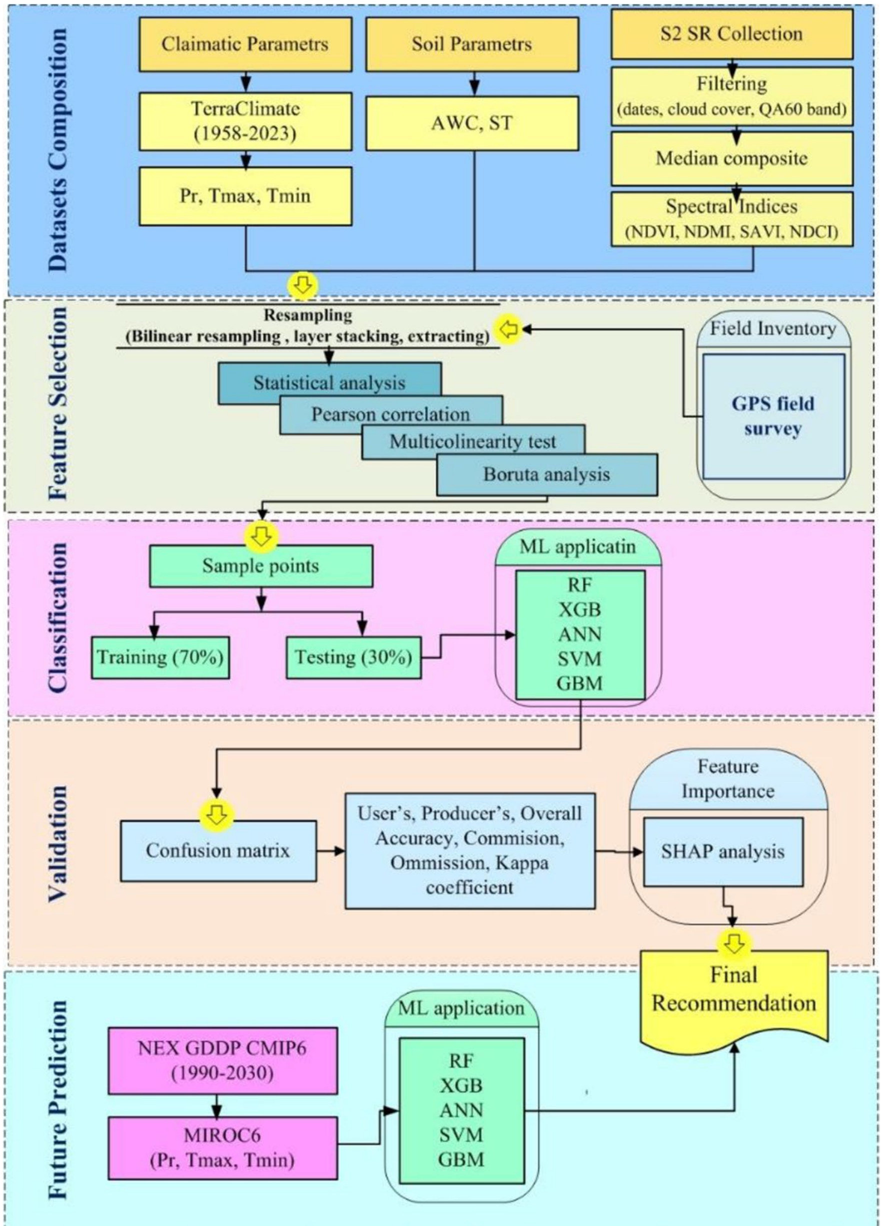


Fig. 3 Flowchart of this study's disease susceptibility modeling process

model, which acts as its foundational element, undergoes training with a distinct subset of the data. For classification objectives, the prediction is determined by the consensus among the trees and the collaborative nature of RF in decision-making processes. In the current study, accuracy was used to select the optimal model using the largest value and mtry is 5, the number of trees 500, the number of variables tried at each split is 5, and additional sampling using up-sampling respectively.

Gradient Boosting Machine (GBM)

GBM stands as a powerful supervised ML method rooted in decision tree ensembles, initially presented by Jerome H. Friedman in 2001. This technique aims to bolster the predictive accuracy of basic classifiers by iteratively building decision trees. Each tree in the sequence is crafted to correct the mistakes of its predecessor, effectively learning from refined residual data. A notable feature of GBM is its application of differentiable loss functions during the boosting process, significantly improving its robustness against outlier data. Here a gradient-boosted model with multinomial loss function is employed for disease classification in the study area. The present study was performed with the final values of the model where shrinkage was 0.1, "n.minobsinnode" was 10, n. Trees were 100, interaction. Depth was 3, and the number of iterations was 100 respectively.

Extreme Gradient Boosting (XGB)

XGB stands as a high-performance ensemble, scalable tree-boosting system, engineered for efficiency, and rapid processing (Chen & Guestrin, 2016). Diverging from methods that aggregate independent trees, XGB sequentially constructs decision trees. Each tree in the sequence is built based on the prediction residuals from its predecessor, thereby homing in on instances with greater prediction error or uncertainty. This iterative refinement ensures that the algorithm adaptively focuses on challenging samples. The culmination of these models, through successive addition, yields the ultimate prediction result. In this study, there are many adaptable parameters in the XGB algorithm, including "gamma" was 0.1, min_child_weight was 1, nrounds was 50, max_depth was 3, eta was 0.4,colsample_bytree was 0.8, min_child_weight was 1, and subsample was 1 respectively.

Artificial Neural net (ANN)

ANN are advanced computational models that mimic the human brain's structure and function, using a backpropagation technique for adjusting weights to learn from training data (Hecht-Nielsen, 1992). Structured with three primary layers—input, hidden, and output—ANNs process data through neurons in each layer, utilizing forward and backpropagation algorithms to make and refine predictions. Each layer is comprised of neurons, each endowed with multiple weights that influence data processing. These are powerful tools for solving complex classification and regression challenges but face limitations, particularly when training data distribution is uneven, which can affect their ability to generalize to new data effectively. In this study the ANN model performed with a hidden layer of 9–5-3 network with 68 weights, the model was sized with 5, and softmax modeling decay was 0.1 respectively.

Support vector Machine (SVM)

The principle of SVM involves detecting an ideal separating hyperplane in the dataset that maximally distinguishes between two categories, ensuring the widest possible margin between the hyperplane and the nearest data points from each category, known as support vectors due to their pivotal role in defining the hyperplane (Hearst et al., 1998). SVM employs kernel functions (i.e., radial, linear) to project low-dimensional input data into a higher-dimensional space, thus transforming non-linearly separable datasets into linearly separable ones within this new space. Despite its advantages, SVM faces challenges, including stringent demands on data selection and the necessity for data normalization. Moreover, the application of SVMs to vast datasets can lead to significantly increased requirements for computational time and memory usage. In this study for SVM adaptable parameters are included, namely, kernel function was radial basis, "sigma" was 0.1076459, number of support vectors was 606, objective function Value are −314.5845 , −285.6889, −258.128, and C parameters was 1.

Model validation

In this study, all ML tenfold cross-validation analysis is evaluated with various statistical metrics like kappa coefficient, overall accuracy, producers' accuracy, users' accuracy, omission, and commission were employed to assess the performance of the models (Eqs. (2)–(5)) (Bai & Feng, 2018). These metrics are commonly utilized in disease susceptibility modeling research for evaluation purposes.

$$\text{Kappa coefficient} = \frac{P_{\text{obs}} - P_{\text{exp}}}{1 - P_{\text{exp}}} \quad (2)$$

$$UA(k) = \frac{S(kk)}{\sum_{p=1}^{9} S(kp)} \quad (3)$$

$$PA(k) = \frac{S(kk)}{\sum_{p=1}^{9} S(pk)} \quad (4)$$

$$OA = \frac{\sum_{p=1}^{9} S(pp)}{\sum_{k=1}^{9} \sum_{p=1}^{9} S(kp)} \quad (5)$$

where $UA(k)$ and $PA(k)$ represent the user and producer accuracies of disease type k in the disease samples, respectively; OA denotes the overall accuracy of the disease samples; $S(kk)$ represents the location of appropriately labeled k; and $S(kp)$ is the location of k that is imperfectly labeled into the disease type p.

Model explainability and feature importance

The concept of making ML models understandable falls within the realm of XAI, which underscores the significance of various factors in influencing the model's predictions, particularly in disease susceptibility evaluations. This research adopts a specific local XAI technique, known as the SHAP method, to delve into the key determinants affecting prediction accuracy and their respective impacts at an individual prediction level. Unlike other methodologies, SHAP not only identifies the importance of each feature but also discerns the direction of their impact—whether they contribute positively or negatively to the predicted outcomes. The computation of SHAP values involves determining the average incremental effect of each feature by considering all possible combinations and permutations of the features involved with Eq. 6.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} [v(S \cup \{i\} - v(S))] \quad (6)$$

where ϕi denotes the influence of factor i,N is the group of all factors, n is the number of factors in N, S is the subgroup of N comprising factor i, and v N is the reference value. The model results for each sample point are controlled by summing the SHAP scores for all the factors related to that sample point. Consequently, the instructive model is matched as follows:

$$g(z') = \phi_0 + \Sigma_{i=1} M \phi_i z' \quad (7)$$

Where $z\prime \in \{0, 1\}^M$ is the number of features.

A SHAP force plot is a visualization that illustrates the importance of each feature to the model's prediction for an individual sample. It indicates the base value estimated by the regression model and highlights the contribution of features to the model's prediction for that specific sample, integrating these effects at the end of the summation. The *shap.force_plot()* function is used to determine the direction (positive or negative) of the relationship between each predictive feature and the target feature. This type of plot effectively shows how features influence the model's prediction for a given sample, making it a valuable tool for understanding model predictions and their alignment with individual observations.

Perturbation sensitivity analysis

The perturbation approach is widely utilized for conducting sensitivity analysis in multiclass classification tasks (Franceschini et al., 2018). This method is extended to identify the most influential features for a specific class and measure the impact of feature perturbations on the model's performance. It involves systematically altering one feature at a time while keeping all other features constant, leveraging the predict_proba function. The multiclass perturbation sensitivity was calculated using the equation provided below (Eq. (8)). The sensitivity of the feature P_i to a specific class C can be expressed as (Scardi & Harding, 1999):

$$S_C(P_i) = \frac{1}{N} \sum_{j=1}^{N} \left| X_C^{\text{original}}(P_j) - P_C^{\text{perturbed}}(P_j) \right| \quad (8)$$

where $X_C^{\text{original}}(P_j)$ is predicted probability of class C for sample j before perturbation, $P_C^{\text{perturbed}}(P_j)$ is the predicted probability of class C for sample j after perturbing P_i

Results

Correlation

The correlation between AWC, NDCI, NDMI, and T_{max} is weakly negative, with values ranging from − 0.1 to − 0.2. The link between AWC and NDVI is weakly negative, with values ranging from − 0.2 to − 0.4. In contrast, the relationship between AWC and Pr is weakly positive, with values ranging from 0.2 to 0.4. Two individual variables, AWC and SAVI, have a weakly negative relationship to values ranging from 0 to − 0.1. The connection between AWC and T_{min} is moderately negative, with a value of − 0.4. The moderately favorable link between NDCI and NDMI ranges from 0.4 to 0.6, whereas the positive relationship between NDCI and NDVI ranges from 0.6 to 0.8. The link between NDCI and Pr is weakly positive, with values ranging from 0 to 0.2. The relationship between NDCI and SAVI is weakly positive, whereas NDCI and ST have no relationship (value 0). The relationship between NDCI, T_{max}, and T_{min} is also weakly positive. The relationship between NDCI and T_{max} is 0 to 0.1, while NDCI and T_{min} are 0.1 to 0.2. The relationship between NDMI and NDVI is strong and favorable, with values ranging from 0.8 to 1. The negative correlation between NDMI and Pr is minimal, with values ranging from 0 to – 0.1. A slight negative association exists between NDMI, SAVI, and ST, with measured values ranging from – 0.2 to – 0.4. The data shows a weak negative relationship between NDMI and T_{max}, with values ranging from – 0.1 to – 0.2, and a weak positive relationship between NDMI and T_{min} (values ranging from 0.2 to 0.4).

There is no correlation between NDVI and Pr. The relationship between NDVI with SAVI and ST ranges from – 0.1 to – 0.2, whereas NDVI with T_{max} ranges from 0 to – 0.1. However, NDVI with T_{min} has a moderately favorable relationship (0.4 to 0.5). The association between Pr and SAVI and ST is weakly negative, with values ranging from – 0.2 to – 0.3 and – 0.3 to – 0.4, respectively. However, Pr and T_{min} have a slightly positive relationship, with values ranging from 0 to 0.1. It was observed that SAVI and ST have a strong positive association when the value is 1. SAVI has a modest positive correlation with T_{max} (0.4–0.5) and a very poor positive correlation with T_{min} (0–0.2). ST and T_{max} have a moderately positive association (0.4–0.5), while ST and T_{min} have a slightly positive relationship (0–0.2). T_{max} and T_{min} show a marginally positive connection with values under 0.2 to 0.3. Rapidly rising temperatures have been linked to less precipitation. As a result, soil has become gritty and dry, resulting in decreased moisture, limited vegetation coverage, and lower chlorophyll-a content in plants. As a result of the ideal conditions, various plant viruses emerged, causing rice diseases to spread over the region.

Multicollinearity test

Multicollinearity testing is one type of statistical analysis to avoid model overfitting through VIF. In the present research, this testing method has been applied for many parameters direct and indirect impact of rice disease prediction purposes. As a result, eight parameters finally selected the best-fit ML methods before tuning for crop disease identification purposes (Narmilan et al., 2022). It was observed that the highest VIF values of 4.620 and 3.517 from NDMI and NDVI compared to other parameters and the minimum VIF value shown in AWC. It was also observed that no significant collinearity issues were coming during analysis periods. The details of VIF results for predicting rice diseases are shown in Table 4. The tolerance values ranged from 0.216 (NDMI) to 0.843 (AWC).

Boruta feature selection

The Boruta algorithm is a wrapper for the random forest classification technique in the R package, offering a fast, parameter-free, and numerical feature relevance assessment. Using ML approaches, it was used for feature selection to predict kharif rice crop disease. It was also discovered that all selected parameters complied with modeling requirements. Boruta completed 203 iterations in 1.457303 min, confirming 9 attributes as important: AWC, NDCI, NDMI, NDVI, Pr, and four others. No attributes were found to be unimportant. The highest mean, median, min, and max were recorded in precipitation, maximum, and minimum temperature (Table 5). As a result, it was determined that three factors are significant for detecting kharif rice crop disease. The minimal impact on crop diseases was observed with the lowest mean importance value, specifically AWC (3.106). In this study, all nine factors under consideration were deemed more important than the shadow factor, as confirmed in crop disease prediction analysis. This means that duplicates of the original attributes were generated by randomly mixing the features, referred to as Shadow Attributes.

Analysis of rice disease predictions

Diseases are the most significant biotic limitations to rice yields, reducing them by 20–100% depending on severity. Major diseases including blast, brown spot, bacterial blight, sheath blight, and tungro continue to cause increasing damage, as well as new minor diseases. Thus, the current study focuses on employing ML approaches (RF, GBM, SVM, XGB, and ANN) to forecast rice disease using biophysical and meteorological characteristics (Fig. 4). However, three rice diseases, namely brown spot, rice blast, and tungro, have been found through large farmer surveys. The fungus Helminthosporium Oryzae causes brown spot-on rice. It is a disease that is transmitted by seeds. This disease can occur in rice crops at any time, from seedling to grain production stage. This disease is more widespread in soils with low potash levels. This disease targets plant leaves and grains, however, the symptoms are more obvious on the leaves. The specks form a golden circle with a yellowish-brown center. Magnaporthe Grisea is a fungal infection that causes rice blasts. Rice fever and rotting neck are other names for rice blast disease. This disease was initially discovered in India in 1918. This disease is widespread in around 80 rice-growing countries. Crop loss might reach 70–80% in the event of a major infestation of this disease. The disease affects the plant's leaves, nodes, and neck, with symptoms being more noticeable on the leaves. Tungro disease is caused by two viruses: Rice Tungro Bacilliform Virus (RTBV) and Rice Tungro Spherical Virus (RTSV). Leafhoppers spread this disease in paddy fields. The majority of diseases are spread by green leafhoppers. Tungro disease impacts crops at all phases of growth, however, it is most common during vegetative growth. The disease-affected plants' growth is stunted, and tillering is diminished.

Rice brown spot diseases were detected in Purba Bardhaman district, northern Paschim Bardhaman, south-eastern Paschim Medinipur, and western Purulia district using various techniques (RF, GBM, SVM, XGB, ANN) in 2023. Purba Bardhaman, also known as West Bengal's rice bowl, is a significant rice farming district. However, the rice brown spot fungus heavily afflicts the district's rice plants. Bipolaris Oryzae is responsible for rice brown spot, a fungal disease. The Purba Bardhaman district had a 96.72% rice brown spot disease spread due to weather circumstances. Diseases may develop in areas with high relative humidity (86–100%) and temperatures ranging from 16 to 36 °C. In the Purba Bardhaman district, farmers have limited access to nitrogen-based fertilizers, despite their importance in soil fertility and availability in organic forms. The western section of the Purulia district has 56.86% rice brown spot disease due to water stress, high temperatures, and nutrient-deficient soil. In the Paschim Medinipur district's south-east region, rice brown spot disease affected 44.79% of crops due to N, P, or K deficiencies, poor soil drainage, temperatures ranging from 20 to – 30 °C for 2 months, and gloomy weather.

Table 4 VIF results for rice disease prediction

Variables	Tolerance	VIF
AWC	0.843	1.186
NDCI	0.380	2.629
NDMI	0.216	4.620
NDVI	0.284	3.517
Pr	0.806	1.240
ST	0.749	1.335
Tmax	0.685	1.460
Tmin	0.691	1.448

Magnaporthe Oryzae causes rice blast, a devastating fungal disease of rice. Rice blast diseases are prevalent in the north-western region of Birbhum district, affecting 72.38% of rice plants. The disease thrives in agroclimatic conditions such as high rainfall, 20–24 °C temperature, water deficit, and low soil moisture. Rice blast infections have been identified in the southern regions of Jhargram and Paschim Medinipur District, as well as the northern part of Murshidabad District, using the ANN approach. Rice blast disease was observed at 51.02% in Paschim Medinipur district's north-western region, attributed to dew development on leaves and temperature changes between day and night. Rice tungro diseases cause significant yield loss in Bankura district, Murshidabad district (excluding ANN techniques), and western Purulia district. Rice tungro disease is caused by RTBV and RTSV, a combination of viruses. Bankura district's rice plant suffers from rice tungro disease at 63.45% due to nitrogen and zinc deficits and rising groundwater levels. Murshidabad district has 53.17% rice tungro disease due to water scarcity and insect infestation, whereas Purulia district's northern region has 35.64% due to high water poverty, dry weather, plant nutrient insufficiency, and rat damage (Table 6).

The RF model outperformed other ML models regarding crop disease detection. It was stated that several ML models, including RF, GBM, ANN, XGB, and SVM, are used to identify three types of rice diseases: brown spot, rice blast, and tungro disease. The RF model had the highest overall accuracy (0.70) and kappa value (0.53). The GBM model's overall accuracy rate is 0.683, with a kappa value of 0.52. The ANN model's overall accuracy is 0.680, with a kappa value rate of 0.51. The XGB model has an overall accuracy of 0.67 and a kappa of 0.50, whereas

Table 5 Boruta results for rice disease sensitivity analysis

Variable	meanImp	medianImp	minImp	maxImp	normHits	Decision
AWC	3.106	3.118	−0.298	6.126	0.611	Confirmed
NDCI	3.905	3.916	0.231	6.673	0.798	Confirmed
NDMI	7.702	7.599	4.223	10.806	0.995	Confirmed
NDVI	8.153	8.091	4.637	11.686	0.990	Confirmed
Pr	17.189	17.135	14.349	20.663	1.000	Confirmed
SAVI	10.962	10.913	8.497	14.172	1.000	Confirmed
ST	10.931	11.010	8.859	12.973	1.000	Confirmed
Tmax	18.556	18.517	15.846	22.331	1.000	Confirmed
Tmin	20.311	20.318	16.491	23.571	1.000	Confirmed

the SVM model has an overall accuracy of 0.66 and a kappa of 0.49. However, the RF model outperforms conventional algorithms for rice disease detection and classification in the research area. Thus, the RF model results investigate the many viewpoints for breakdown patterns and aspects relevant to three forms of rice crop disease. Break-down (BD) plots offer a potential solution by showcasing "variable attributions." These plots break down the model's prediction into contributions from various explanatory variables. This technique is implemented in the Explain Prediction R package (Robnik-Sikonja & Kononenko, 2008). However, the breakdown profile shows how the contribution of distinct explanatory variables affects the mean (Fig. 5). This profile depicts the variable contribution in a clear graphical format for the three-rice crop disease evaluation objectives. The green bars represent positive changes, while the red bars indicate negative changes in the mean predictions, showing the contributions attributed to the explanatory variables. The breakdown profile for brown spot disease shows significant positive contributions from T_{min} (+0.09), T_{max} (+0.009), SAVI (+0.035), ST (+0 0.033), NDVI (+0.017), NDCI (+0.122), and AWC (+0.03), while Pr (−0.008) and NDMI (−0.032) contribute significantly negative effects. For rice blast disease, all factors have significantly negative effects, with T_{min} showing the greatest negative impact at −0.096. Similarly, for tungro disease, the factors Pr (+0.074), T_{min} (+0.006), T_{max} (+0.023), and NDMI (+0.047) show positive contributions, while SAVI (−0.029), ST (−0.032), NDVI (−0.011), NDCI (−0.061), and AWC (−0.063) exhibit negative contributions. In this profile, the highest predicted probability was observed for rice brown spot at 0.698. Figure 6 depicts calculating the score for all input features in a model by RF model to determine the significance of each feature in the decision-making process for detecting rice brown spot, rice blast, and rice tungro rice crop disease. The top five factors, i.e., Pr, T_{min}, T_{max}, NDCI, and NDMI—are the most important for all three types of disease instances. Conversely, NDVI, AWC, ST, and SAVI are the least important for all disease instances in the study area. Finally, a district-by-district farmer field survey including field images was conducted to validate actual rice crop production conditions (Fig. 7).

Furthermore, it was shown that rice brown spot disease is most prevalent in Purba Bardhaman, Purulia, and Paschim Medinipur. It was also noted that the highest regions covered by rice blast and tungro disease in Murshidabad, Birbhum, Paschim Medinipur, and Jhragram and Murshidabad, Bankura, and Purulia validated by field survey photographs.

Analysis of future rice disease predictions

Analysis of future SSP2-4.5 predictions

It was discovered that many sorts of rice diseases were specified using SSP2-4.5 in 2030 as prospects (Fig. 8). Rice brown spot was identified in Purba Bardhaman and Paschim Bardhaman districts using RF, GBM, SVM, and XGB techniques, but not ANN. This is due to high rainfall intensity, relative humidity, wind speed, and increased sodium ions in the soil. In this study area, it was observed that certain percentage of randomly distributed rice brown spot infections in Purulia, Bankura district excluding ANN technique caused in these districts had faced very high temperatures and excessive heat that is blowing in these areas and lacks of essential nutrients in this

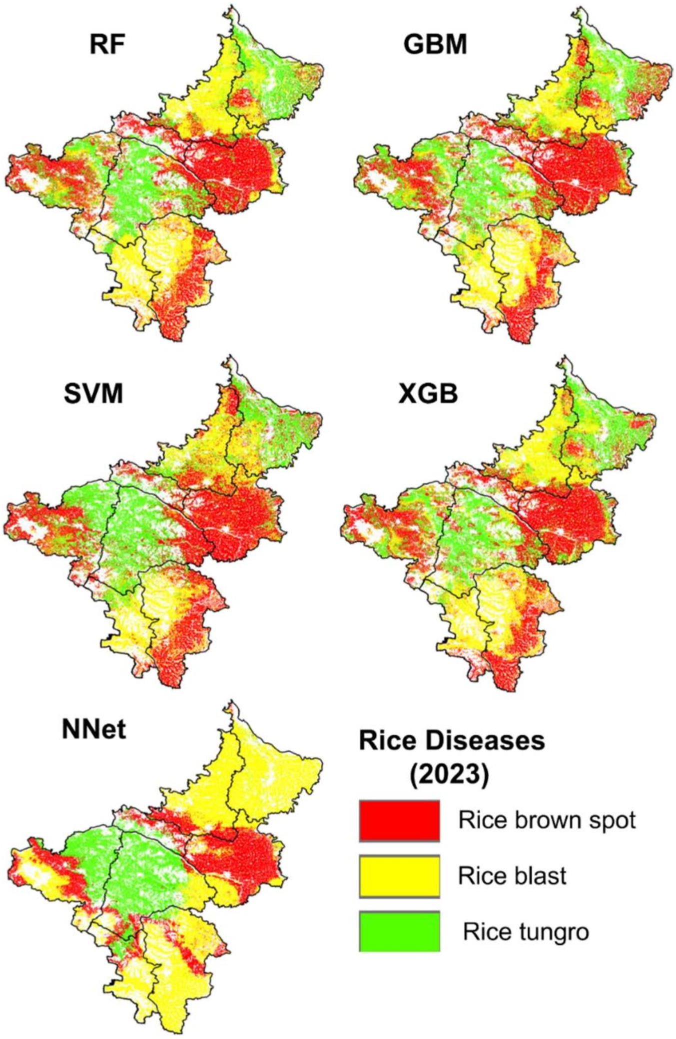


Fig. 4 Rice disease mapping by five ML techniques of 2023

district's soil like nitrogen, phosphorus, and lime, etc. and experienced water stress problem, on the other way, due to abundant rainfall, humidity and salinization is a burning problem in the Paschim Medinipur district that occurs rice brown spot diseases.

The most severe rice blast disease has been detected in the majority of Birbhum district using RF, GBM, SVM, ANN and XGB approaches, owing to the district's high rainfall intensity from June to October, increased wind speed and humidity, and the progressive impact of climate change. Groundwater in Birbhum is overexploited, and groundwater levels are fast falling. It was discovered that soil salinity occurs in Murshidabad, hence this disease has been observed in several areas of the Murshidabad district. Nutrient loss is the main

Table 6 Percentage changes of rice disease mapping by the RF model

Districts name	Rice brown spot	Rice blast	Rice tungro
Murshidabad	16.09	**30.72**	**53.17**
Birbhum	15.58	**72.38**	12.03
Paschim Bardhaman	14.74	0.85	3.29
Purba Bardhaman	**96.72**	14.19	9.99
Bankura	24.11	12.43	**63.45**
Purulia	**56.86**	19.22	**35.64**
Paschim Medinipur	**44.79**	**51.02**	1.89
Jhargram	2.60	**32.10**	2.37

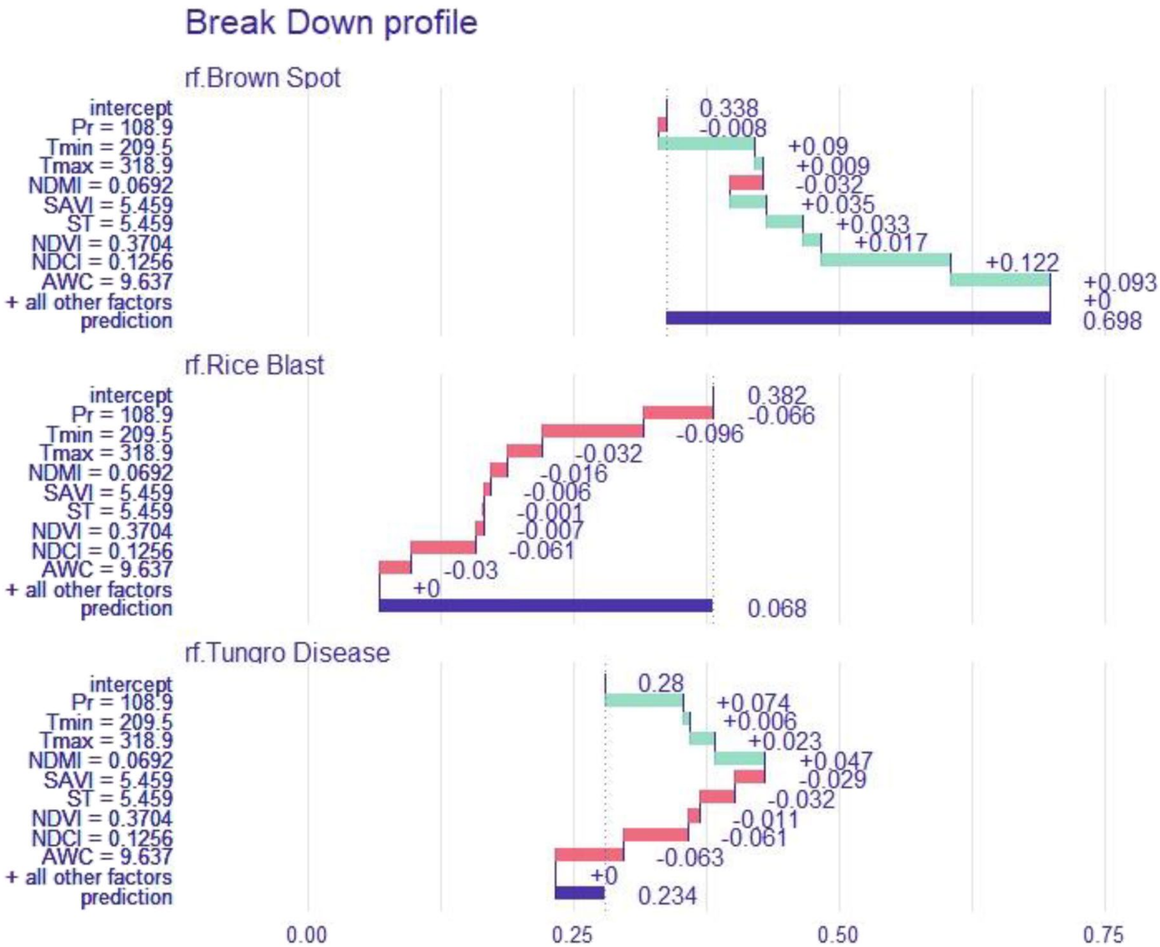


Fig. 5 Break down the profile of the three rice disease for the best model of RF

cause of heavy rainfall in the middle of Jhargram, northern and southern Paschim Medinipur districts, with maximum temperatures reaching 40 °C and heatwave conditions in only a few areas of Purulia district, except for the SVM technique. As a result, rice blast disease continues to spread alarmingly. However, in the ANN technique, rice blast disease is viewed in the north-western region, such as the western part of Birbhum, Paschim Bardhaman, and Purulia district, as well as the south-eastern region, such as the extreme southern portion of Purba Bardhaman, the eastern part of Bankura, and the northern part of Paschim Medinipur. Rice tungro disease has increased in Bankura, Purulia, and Murshidabad districts using RF, GBM, and XGB techniques due to hardpan and heavy metal soil types. These soils restrict water infiltration, leading to waterlogging and poor drainage, causing water stress in plants and deficiencies in essential elements for plant growth. Additionally, water shortage variability is severe in the Murshi dabad district. Rice tungro disease was found in several districts, including Purba Bardhaman, which was excluded from the SVM approach due to infrastructure concerns. However, in the ANN technique, the disease remained in the northern to south-western regions. As a result, this region encompassed the majority of Murshidabad district, the eastern part of Birbhum, the entire Purba Bardhaman, the eastern part of Paschim Bardhaman, the majority of Bankura district, the southern part of Purulia district, and the northern half of Jhargram and Paschim Medinipur districts.

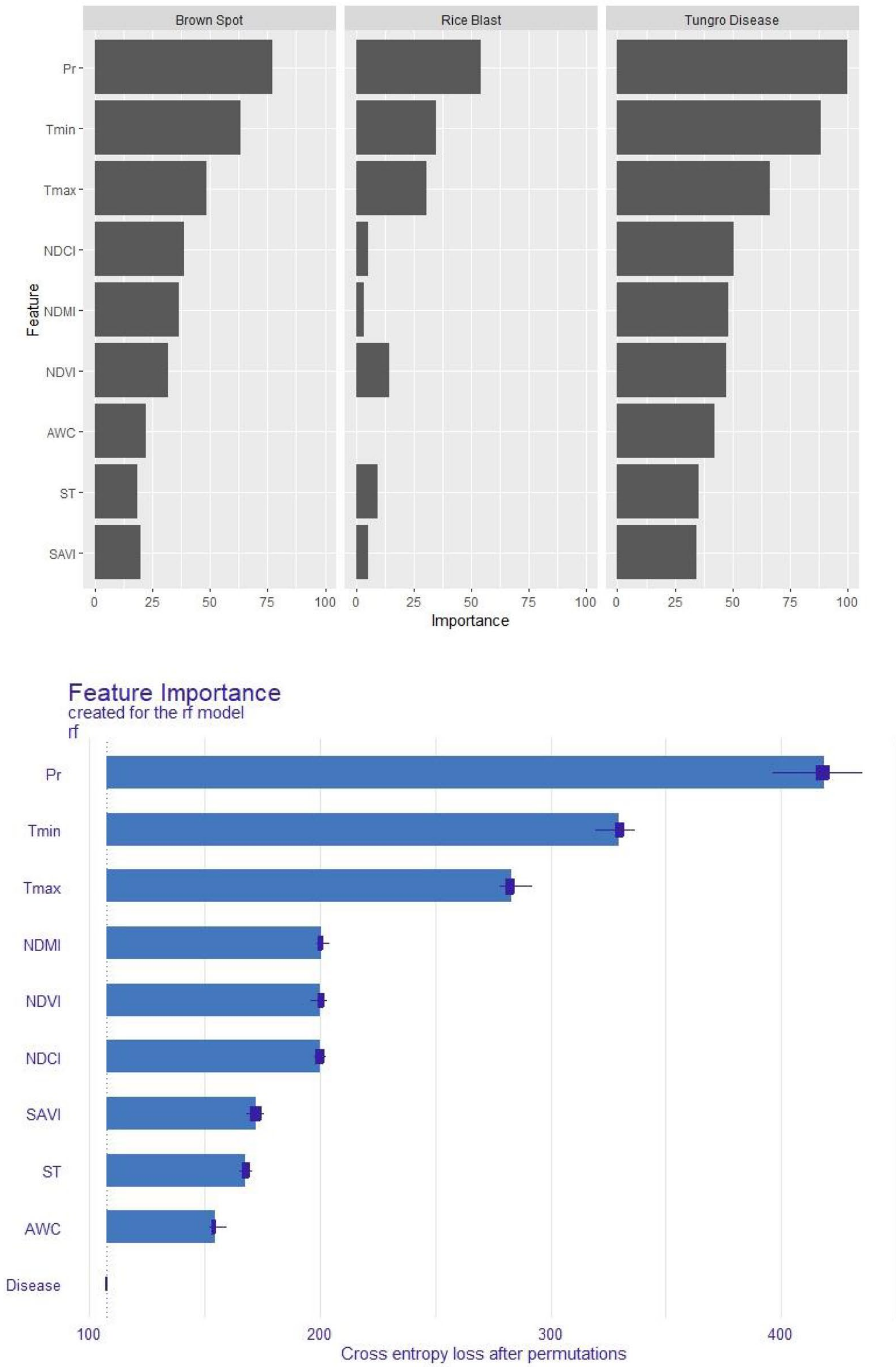


Fig. 6 Biophysical influencing features important of rice crop disease for RF model

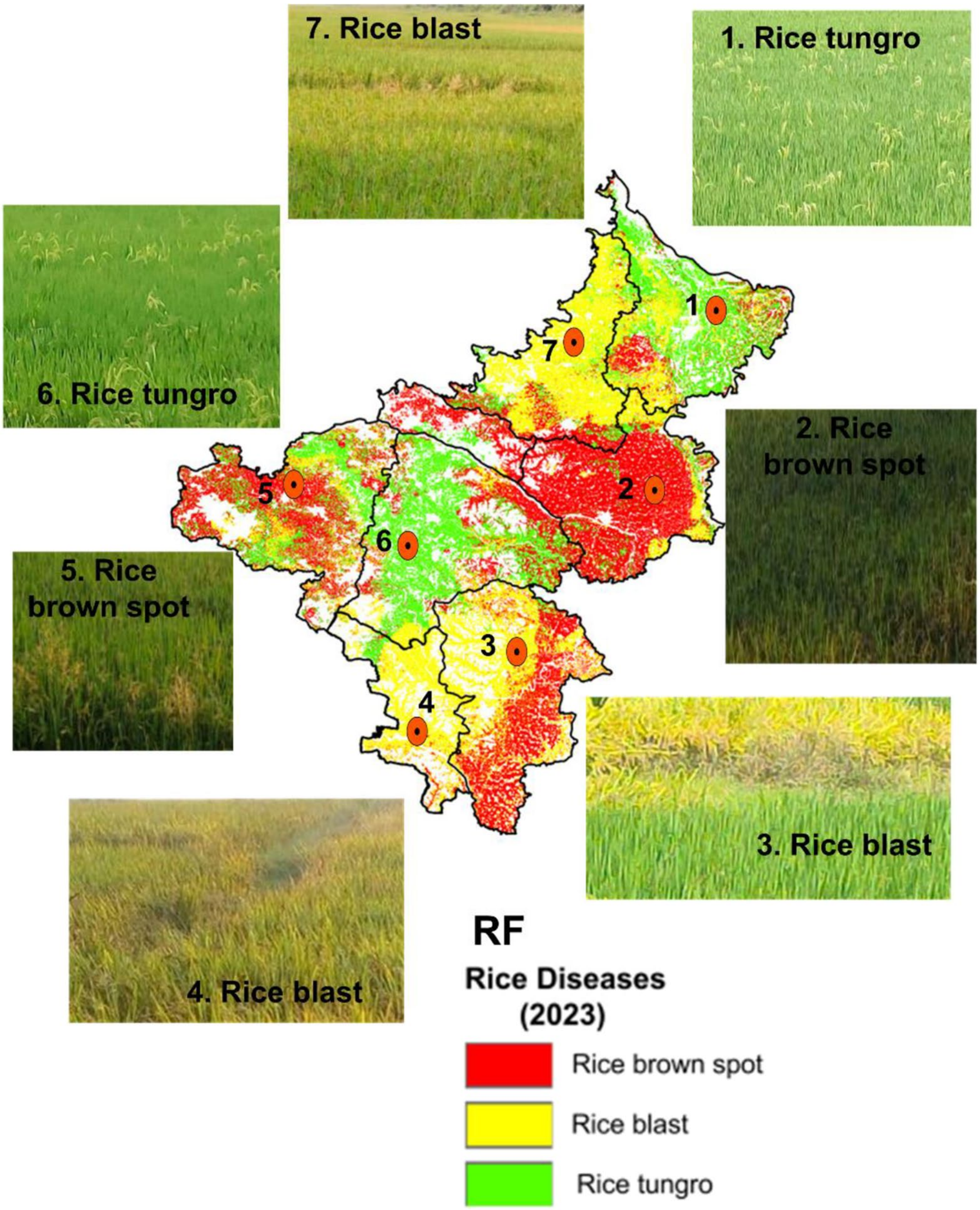


Fig. 7 Validation of the best ML model of RF with field observation photographs

Analysis of future SSP5-8.5 predictions

The study region discovered various rice diseases in the 2030 edition of SSP5-8.5. The majority of rice brown spot disease was detected in the Purba Bardhaman district as a result of anthropogenic climate change, which had an impact on rice harvests. Rice brown spot disease was caused by increased rainfall intensity, a lack of nitrogen-based fertilizers, and other factors. The present study spread rice brown spot disease to Purulia, Paschim Bardhaman, and Paschim Medinipur districts. In the Purulia district has warmer temperatures, heat waves are expected to become more frequent, affected moisture stress can produce this rice

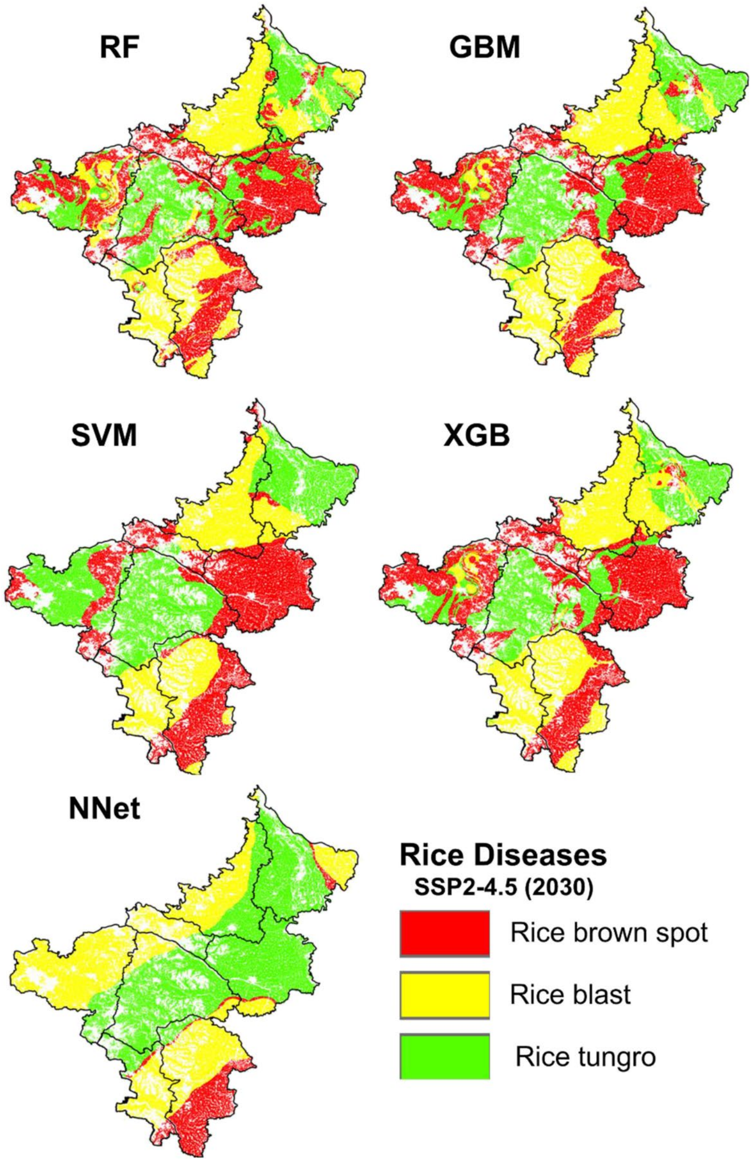
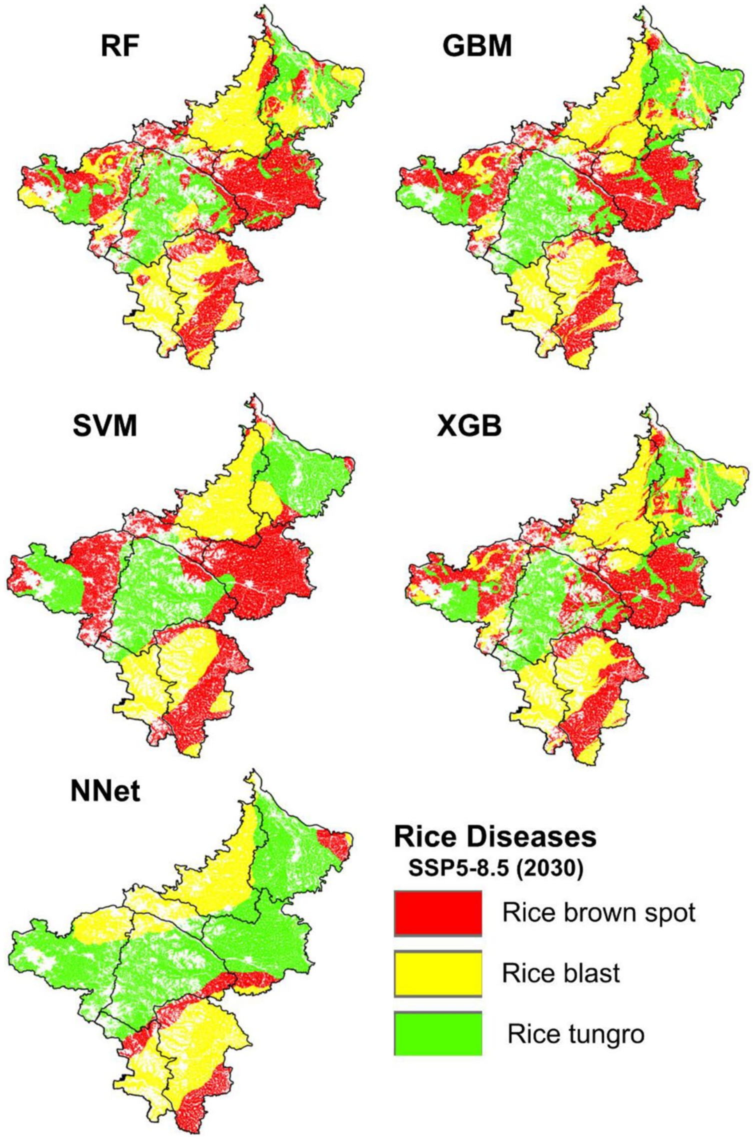


Fig. 8 Future rice disease mapping by five ML techniques of 2030 under SSP2-4.5 of the MIROC6 model

brown spot disease and in Paschim Bardhaman and Paschim Medinipur districts have noticed that high rainfall intensity with increasing wind speed and the exposure of high humidity, lack of nutrient in the soil and suspected to others is that to be responsible for these diseases. Rice brown spot infections were found to be widely distributed in Murshidabad, Bankura district, as well as a small area of Birbhum, Jhargram district, utilizing RF, GBM, SVM, and XGB approaches, but not ANN. Rice brown spot disease has been detected in a small area in southern Purba Bardhaman, the border area of Bankura and Paschim Medinipur district, south-eastern Paschim Medinipur, and northern Jhargram using the ANN approach (Fig. 9).

Rice blast disease affected all the districts in the research area, but Birbhum was the most vulnerable, and it spread throughout the district by 2030 utilizing the RF, GBM, SVM, ANN and XGB approaches. Birbhum district has arid conditions with high temperatures, strong wind speed, increasing relative humidity, and rainfall intensity, as well as a scarcity of water for irrigation and poor soil water content, which can contribute to rice blast disease. Rice blast disease is prevalent in Jhargram and Paschim Medinipur

Fig. 9 Future rice disease mapping by five ML techniques of 2030 under SSP5-8.5 of the MIROC6 model



districts, caused by saline-sodic soils, water shortage, wind velocity, and high relative humidity. It was also showed that scattered distributed in Murshidabad, Purulia, Bankura, the northern part of Paschim Bardhaman, and the north-western part of Purba Bardhaman district except SVM technique but in ANN technique, rice blast disease notably that north-western regions like a maximum portion of Birbhum, Paschim Bardhaman, north-western part of Bankura and northern part of Purulia districts and south-eastern region like the extreme southern part of Purba Bardhaman, extreme eastern part of Bankura, and maximum portion of Paschim Medinipur, Jhargram districts.

Rice tungro is one of the most common viral infections affecting rice. Climate change, nutrient loss from heat waves, decreased zinc and nitrogen concentrations in rice plants, and hardpan and heavy metal soil types all contribute to the prevalence of rice tungro disease in the Bankura district. The disease spread across Murshidabad and Purulia districts. In Purulia district, the disease was caused by groundwater depletion, rising temperatures, and heat waves. In Murshidabad district, rice tungro disease was linked to extreme weather, pest infestations, and soil salinity. This disease is also found in a scattered area in Paschim Bardhaman district using RF, GBM, SVM, ANN and XGB approaches, as well as Purba Bardhaman district, except for the SVM methodology. Rice tungro disease has been found to spread from the north-east to the south-west, affecting the majority of Murshidabad, Purba Bardhaman, Bankura, and Purulia districts, as well as a small portion of the north-east and south-west parts of Birbhum, the eastern part of Paschim Bardhaman, and the northern part of Jhargram district, according to the ANN technique.

SHAP analysis

SHAP values represent each feature's contribution to the target's expected value, which aids in identifying the most important features for prediction. The green bars represent positive impact, while the red bars indicate negative impact in the mean predictions, showing the contributions attributed to the explanatory variables. In the context of brown spot disease, T_{min}, Pr, T_{max}, SAVI, and NDVI are identified as the top five influential variables. In contrast, NDMI and ST show moderate influence, while NDCI and AWC contribute less significantly. The Shapley results indicate that all input predictors for rice blast disease contribute negatively. In this analysis, T_{min} and Pr emerge as the most critical factors in predicting rice blast. Conversely, NDMI, NDCI, and AWC are of lesser importance. The analysis reveals that T_{min}, Pr, T_{max}, SAVI, and NDVI factors make notable negative contributions to Tungro disease prediction, while NDMI and ST show moderate importance. NDCI and AWC, however, rank lowest with minimal contributions. T_{min}, T_{max}, and Pr were determined to be the most useful metrics for detecting rice crop illnesses using ML approaches. Figure 10 depicts the detailed shapely data for the best RF model.

SHAP summary plot

Figure 11a displays the SHAP value feature summary plots positively and negatively, where each point represents the SHAP value for a specific feature across individual samples. The color gradient, ranging from purple to blue, indicates the feature values from high to low. The horizontal axis represents the SHAP values, while the vertical axis lists the features, arranged in descending order of their importance. Climatic factors play a significant role in influencing the spatial variability of rice crop diseases (Ansari et al., 2021). Figure 11a highlights that Pr, T_{min}, and T_{max} are key features with substantial impacts on all three rice diseases, although their effects vary across different districts. Increasing temperatures and shifting rainfall patterns can result in conditions of flooding and drought, which are highly susceptible to pest and disease outbreaks, ultimately impacting crop productivity (Ansari et al., 2023). ST, and SAVI make moderate contributions to brown spot and rice blast, whereas their influence is minimal in cases of tungro diseases. All the vegetation indices such as NDMI, NDCI, and NDVI demonstrated a positive influence on the classification analysis of rice crop diseases. These satellite-derived indices have shown significant potential in assessing crop health, monitoring phenology, and identifying crop stress conditions at the field scale.

SHAP force plot

SHAP force plots illustrate model explanations based on test samples for brown spot, rice blast, and tungro diseases, as shown in Fig. 11b. For brown spot, the Pr value of 1079.12 and the value of 20.917 contributed most positively to the model's output, whereas SAVI, NDMI, NDCI, ST, AWC, T_{max}, and NDVI had negative impacts. In the case of Rice Blast, climatic factors such as T_{min} (20.867), T_{max} (31.84), and Pr (1113.98) provided the most significant positive contributions, while soil features and vegetation indices negatively influenced the model. For tungro, the force plots indicated a close positive prediction with a value of 0.23. Features like SAVI (4.0) and Pr (1047.64) contributed positively, whereas T_{min}, NDCI, NDMI, and NDVI were identified as the most negative influencers.

Feature sensitivity analysis

Figure 12 illustrates the sensitivity of each feature for individual disease classes using the "perturbation" sensitivity analysis method. AWC and SAVI input features exhibited negative sensitivity for brown spot and rice blast but showed positive sensitivity for tungro disease. Rice blast disease showed negative sensitivity, reaching up to − 0.0015, while NDCI, ST, and T_{min} demonstrated positive sensitivity in the classification analysis for brown spot and tungro diseases. Vegetation indices such as NDVI and NDMI

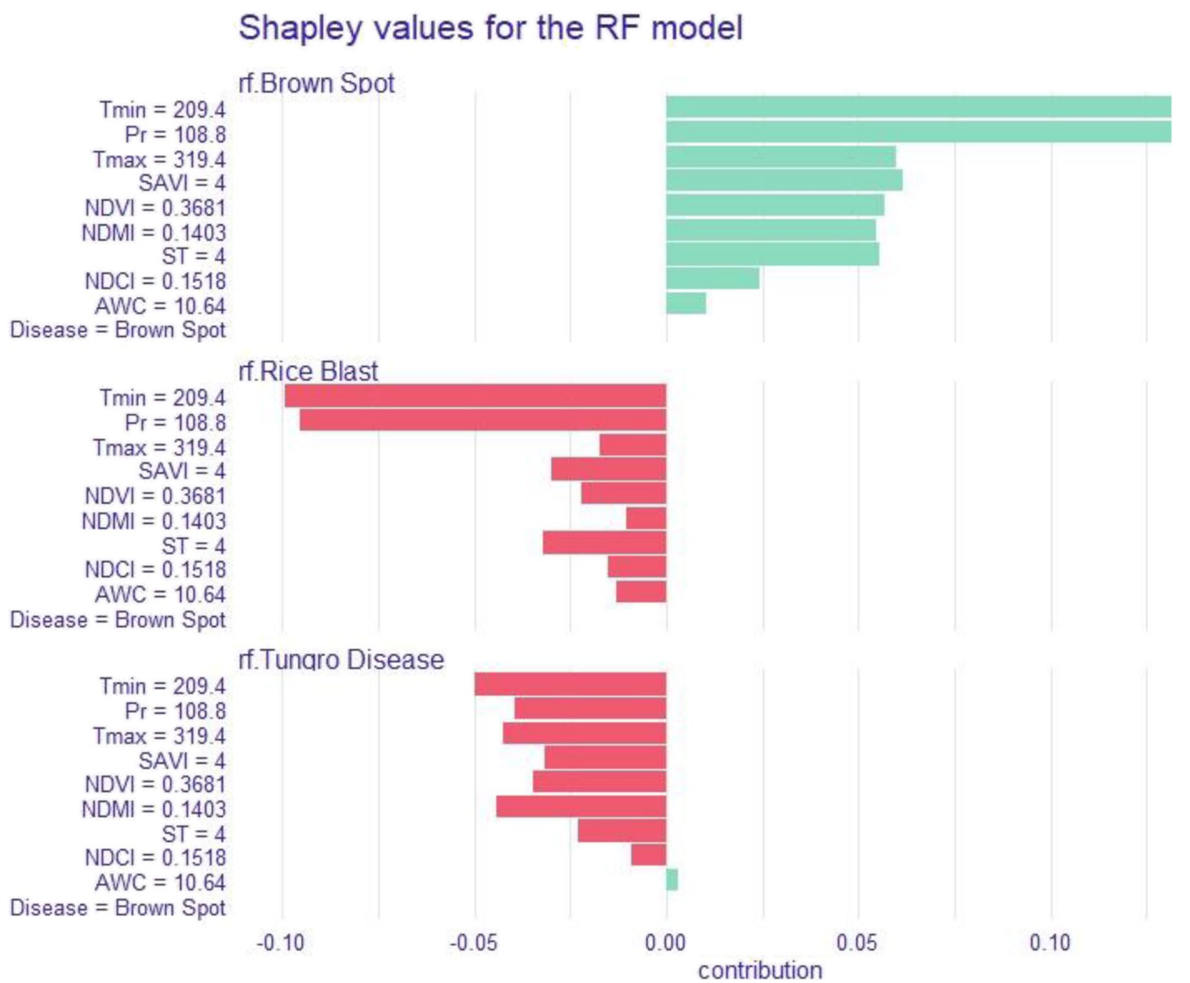


Fig. 10 SHAP summary of influencing factors for rice disease mapping

displayed a positive average sensitivity of up to 0.015 for brown spot disease, while exhibiting negative sensitivity of up to − 0.011 for rice blast and tungro diseases. The Pr feature exhibited positive sensitivity for rice blast, while demonstrating negative sensitivity of up to − 0.0033 for brown spot and tungro diseases.

The T_{max} feature exhibited an average positive sensitivity of 0.006 for brown spot, whereas rice blast and tungro diseases showed negative average sensitivity.

Discussions

The current study focuses on utilizing ML models to identify rice disease in existing and future scenarios while taking biophysical aspects into account. It is critical for the food and nutritional security of India. Rice disease has been validated using farmers' field survey data from eight districts in West Bengal by several enumerators. To accomplish the United Nations Sustainable Development Goals, existing research relies entirely on primary data. Xinyue et al. (2023) introduced their research on necrotic lesion formation's characteristics, classification, and molecular mechanism. They reviewed the molecular regulatory pathway of genes involved in rice disease resistance and summarized the relationship between resistance and yield using newly developed gene editing. They also discussed a rational and accurate breeding strategy that uses molecular design technology to breed disease-resistant rice varieties. This study examined the relationship between rice disease

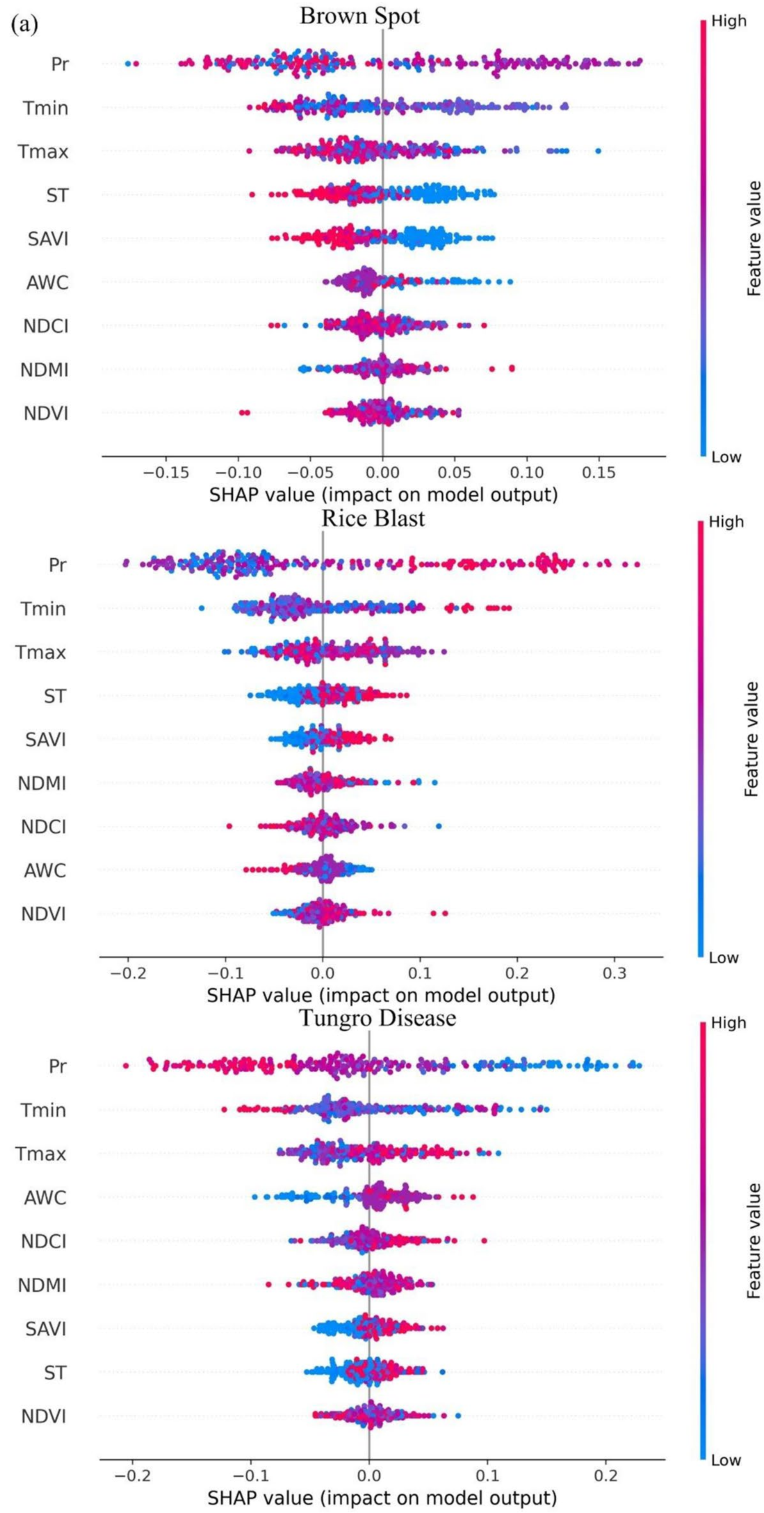
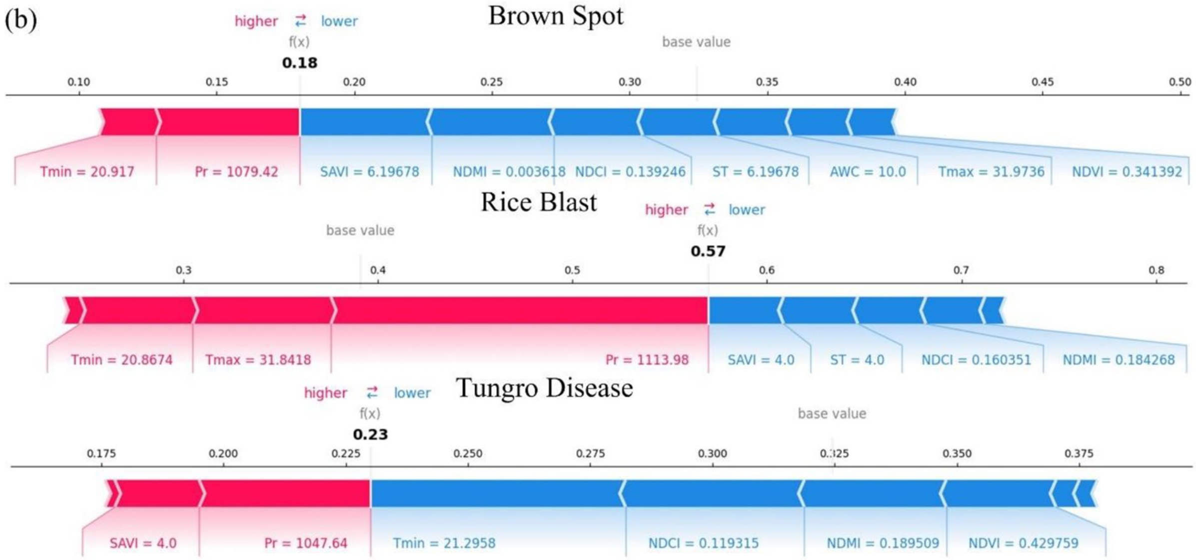


Fig. 11 **a** Local SHAP summary for rice disease classification and **b** SHAP force plot for rice disease classification of different samples

Fig. 11 (continued)



dispersion and favorable growing conditions. In this study, we summarized the criteria (AWC, NDCI, NDMI, NDVI, Pr, SAVI, ST, T_{max}, and T_{min}) that contribute to rice disease development for sustainable agriculture production. The SHAP analysis revealed that the three primary climatic factors—Pr, T_{min}, and T_{max}—are crucial for all instances of the three disease types. A similar study indicated that climatic factors play a significant role in the development of rice diseases (Mousumi et al., 2023). Singh et al. (2023) developed a Custom CNN design to identify and classify rice plant disease by decreasing network parameters. The present study used RF, GBM, SVM, XGB, and ANN algorithms to detect rice diseases and predict the future using SSP2-4.5 and SSP5-8.5 climate data for 2030. Kamarudin et al. (2024) analyzed whole genome re-sequencing data for blast resistance and kernel elongation features in the Mahsuri Mutant rice variety and its parental line. The researchers collected data in Malaysia for validation purposes. The present research uses ML models to analyze the prevalence of rice brown spot, rice blast, and rice tungro disease in eight West Bengal, India districts. Rice brown spot and rice blast disease were found to have the highest area coverage in eight districts. Lu et al. (2023) identified four types of rice diseases: rice blast, sheath blight, brown spot, and leaf blight, and developed an effective method for disease detection. We discovered three forms of rice diseases: rice brown spot, rice blast, and rice tungro, which were diagnosed using various methodologies in current and future scenarios in Murshidabad, Birbhum, Purba, and Paschim Bardhaman, Purulia, Bankura, Paschim Medinipur, and Jhargram.

Jackulin and Murugavalli (2022) researched identifying and classifying plant leaf diseases. Nayak et al. (2023) investigated on-field detection of complications such as crop disease incidences, macro-nutrient deficits, and diseases associated with nutrient deficiencies. Our study focused on rice diseases such as brown spot, blast, and tungro, which can be caused by high relative humidity, high temperature, water stress, nutrient-deficient soil, low soil moisture, cloudy weather, and inadequate input management. Yan et al. (2022) analyzed a simple and cost-effective system for artificially inducing bakanae disease, which is crucial for breeding resistant rice varieties. The system focuses on rapid propagation of Fusarium fujikuroi pathogenic spores, efficient inoculation of rice seeds with spore solution, and growth conditions of rice seedlings after inoculation. Our study examines the current distribution of rice brown spot, rice blast, and rice tungro disease in 2023 and predicts their future spread in the region by 2030. Aggarwal et al. (2022) analyzed articles from 2012 to 2019 using various ML methods, including Integrated Pest Management, SVM, CNN, and k-nearest neighbor (kNN), to identify and detect rice diseases. The present study used RF, GBM, SVM, XGB, and ANN technologies

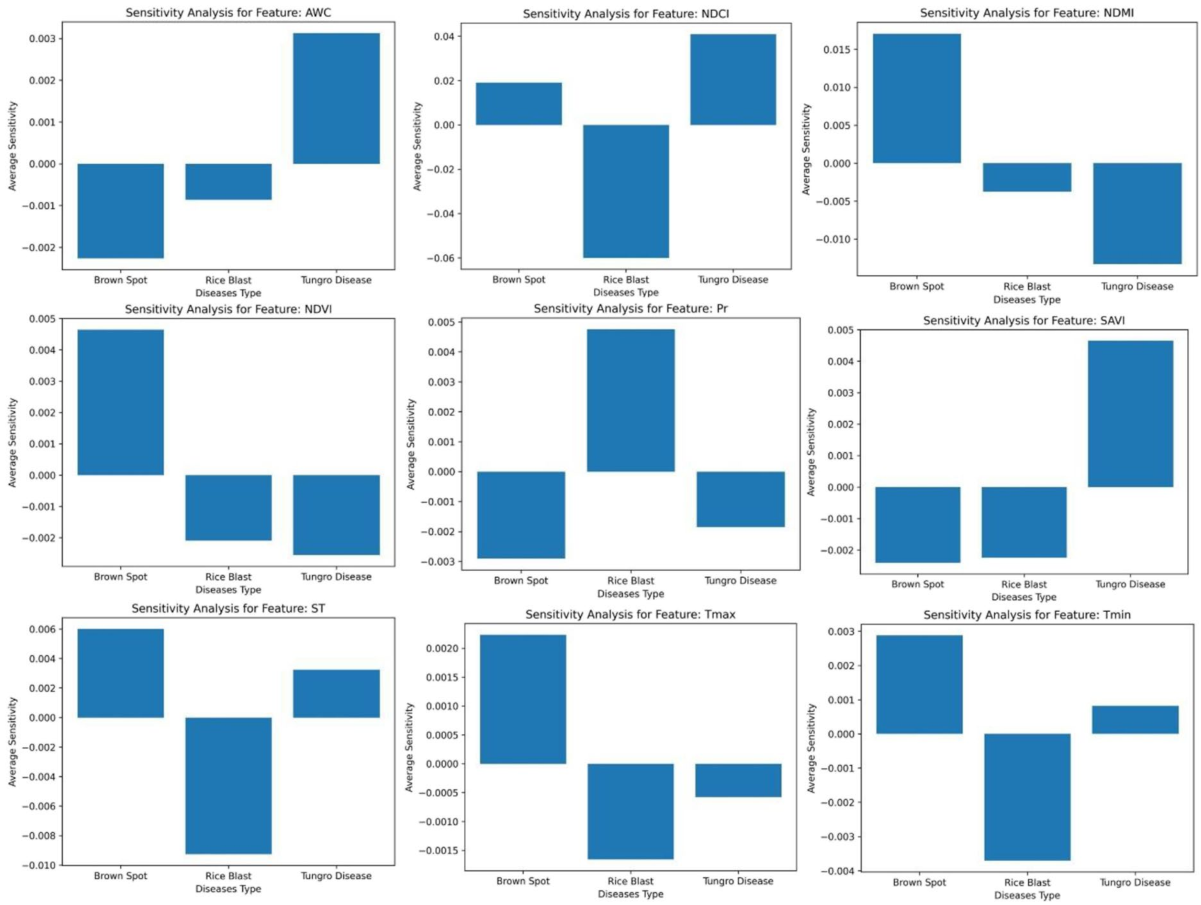


Fig. 12 Perturbation sensitivity analysis for rice disease classification

to model rice disease distribution in 2023 and estimate future development patterns of agriculture activities. Zheng et al. (2023) discuss remote sensing monitoring of rice diseases and pests, covering several data sources such as hyperspectral, multispectral, thermal infrared, fluorescence, and multi-source data fusion. Furthermore, historical and current studies are comprehensively compared in terms of materials, methods, outcomes, and limits to improve the agricultural system and solve global food security challenges (Fig. 13). Figure 14 showed a weak negative correlation (-0.1 to -0.2) between AWC, NDCI, NDMI, and Tmax.

Adaptation strategies and policy recommendations

Climate change significantly impacts sustainable agriculture, affecting crop development, growth, and paddy production. Adaptability is crucial to mitigate risks and maintain crop output, while solutions like high temperatures can help overcome challenges, especially disease-related issues. Rice's water-use efficiency is low, and water-saving irrigation technology could shift land from anaerobic to aerobic, impacting sustainability, weed, insect, disease ecology, and nutrient and soil organic matter dynamics. Rice growers in the research area primarily use conventional paddy farming, focusing on field preparation and water conservation through bund creation and leveling, and building channels for excess water drainage. Farmers grow multiple paddy crops on the same field using a monocropping system but opt for zero tillage due to time constraints. Farmers often avoid crop insurance on rainfed and irrigated land due to a lack of understanding and information about the insurance procedure and claiming processes, which

Fig. 13 Representation of sustainable development goals correlated with our present research work

can be complicated due to time constraints and unexpected expenses. Paddy growers often use medium-level adaptation solutions to climate change, largely due to a lack of simple strategies like direct seeding or varied cropping systems. The study identifies major barriers to agricultural adaptation strategies in the area, including a lack of climate change knowledge, specific institutions, and a better institutional framework.

Diseases management strategies

Rice brown spot diseases

This disease is more prevalent in soils with low levels of potash. In the Purba Bardhaman and Purulia districts, farmers face limited access to nitrogen-based fertilizers, despite their crucial role in maintaining soil fertility and their availability in organic forms. In the south-east region of Paschim Medinipur district, rice brown spot disease impacted 44.79% of crops, primarily due to deficiencies in nitrogen, potassium, or phosphorus, as well as poor soil drainage. To monitor and manage the spread of rice brown spot in this region, several strategies were implemented, including the use of disease-free seeds and the planting of resistant rice varieties such as PY 4, ADT 44, CO 44, CORH 1, Cauvery, TPS 4, Dhanu, and Bhavani. Silicon, combined with calcium silicate slag, was applied before planting, and regular irrigation scheduling was maintained in areas experiencing water stress. For chemical control methods, Mancozeb (2.0 g/L) was sprayed two to three times at intervals of 11 to 16 days. Additionally, seed treatment with Ceresan/Agrosan at 2.5 g/kg was performed to prevent seedling blight during the early stages of growth.

Rice blast diseases

Rice blast disease is widely prevalent in the north-western region of Birbhum district, impacting 72.38% of rice plants. Additionally, rice blast infections have

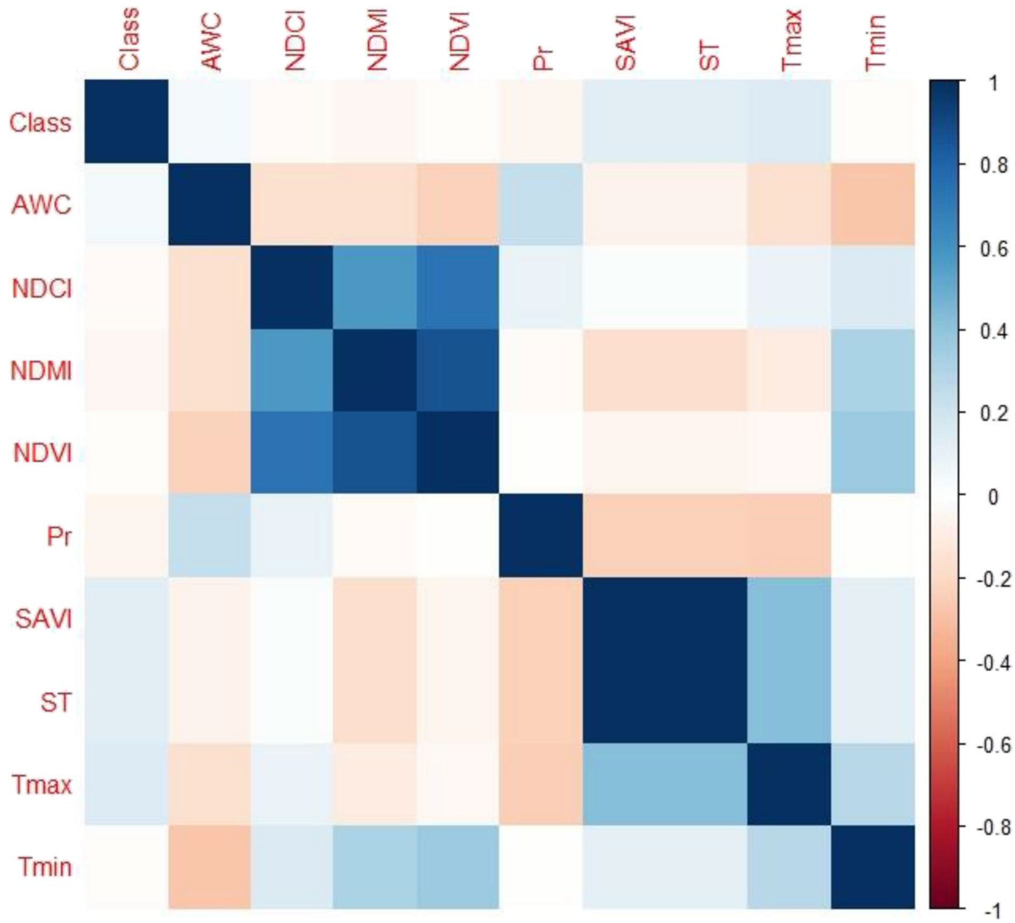


Fig. 14 Correlation analysis of influencing factors for rice disease mapping

been reported in the southern areas of Jhargram and Paschim Medinipur districts, as well as in the northern portion of Murshidabad district. The north-western part of Paschim Medinipur district is particularly vulnerable, likely due to dew formation on rice leaves and fluctuations in temperature between day and night. To manage rice blast disease in this region, it is recommended to maintain appropriate doses of nitrogen fertilizer and regularly remove weed hosts. Disease management can also involve the use of tolerant rice cultivars, such as CO 50, CO 47, TPS 3, White Ponni, CORH, BPT 5204, Swarnamukhi, Swathi, Palghuna, Prabhat, IR-36, IR-64, Jaya ADT 37, ADT 36, ADT 16, ADT 20, ADT 39, ADT 19, and ADT 44, to limit disease spread. Preventive strategies include delayed planting and minimum tillage, the proper management of straw and stubble. For chemical control, spraying Carbendazim and Edifenphos at a concentration of 1 g/L during the flowering stage is effective (Source: http://www.agritech.tnau.ac.in).

Tungro diseases

Rice tungro disease has caused considerable yield losses in Bankura district, Murshidabad district, and western Purulia district. In Murshidabad district, 53.17% of rice tungro disease cases are attributed to water scarcity, while in the northern region of Purulia district, 35.64% of cases are linked to water shortages, dry weather, insufficient plant nutrients, and rat damage. Management strategies include the use of light traps to control the rice tungro population. Cultural methods such as regular maintenance of fallow lands to eliminate weed hosts updated the planting date, and the promotion of effective crop rotation, inter cropping, were also adopted. Additionally, oilseed crops and resistant rice varieties—IR 50, IR 36, ADT 37, Co 45, Ponmani, Co 48, Vikramarya, Surekha, White Ponni, and Bharani—were implemented as part of disease management strategies. The recommended application rates for nitrogen, zinc, and organic fertilizers, along with neem cake at 12.5 kg/20, were applied effectively, and optimal use of groundwater was ensured. In this region, rice tungro disease can be controlled by using a recommended dose of 2% urea combined with Mancozeb at 2.5 gm/liter, along with treatments such as Carbofuran, Phorate, and spraying combinations of Urea + Mancozeb. Additionally, foliar spraying of Multi-K and other chemical methods were utilized to manage the disease.

Local stakeholders can adopt bio-control agents like *Bacillus*, *Pseudomonas*, and *Trichoderma* as effective methods to control brown spot disease (Mondal et al., 2024). According to Carvalho et al. (2010), consistent and efficient nutrient management is an effective strategy to lower the risk of brown spot disease in the affected region. Surendhar et al. (2022) noted that silicon improves the immune responses of host plants to various biotic stressors while also reducing the incidence of brown spot disease. Additionally, they recommended the use of biofungicides, specifically Hexaconazole (0.2%) and Propiconazole (0.1%), which have shown high levels of effectiveness in managing this disease. To combat rice blast disease, Skamnioti and Gurr (2009) proposed a collaborative approach in which farmers prune infected areas and maintain regular field cleaning to prevent the disease's spread. Furthermore, Saha et al. (2008) highlighted that farmer in Purulia, Bankura, and Birbhum implemented plant protection strategies, including applying Edifenphos fungicide at a concentration of 2 ml per liter of water, as a targeted approach to manage the disease.

Recommendations and a participatory approach are suggested to help local farmers and the community manage rice crop diseases in the study area. Weather information services should be accurate, timely, and location-specific to manage disease outbreaks effectively (Rengalakshmi et al., 2018). Rice farmers need training on using weather forecasts, identifying diseases and pests, managing diseases, applying pesticides, and understanding weather conditions for rice diseases. Weather information should also be provided in the local language with clear interpretations (Mousumi et al., 2023). Farmers need to be trained on utilizing weather information services for making agricultural decisions. While most farmers own mobile phones and smartphones, many are unfamiliar with accessing these services. ICT specialists should collaborate with the meteorology department, University teacher, academician, agricultural extension agents, Krishi Vigyan Kendra (KVK) managers, and rice disease experts to create customized Weather and Climate Information Services for climate-resilient rice disease management. NGOs and private organizations could support rice farmers by distributing timely, location-specific weather information, addressing the shortage of extension agents and limited funding. Farmers in the study region are implementing adaptive measures due to knowledge, experience, resource availability, and lack of institutional framework. Specialized interventions, technology demonstrations, training, research, and institutional frameworks are being implemented. Education, outreach, and extension activities can enhance human capital, decision-making capability, and collective adaptability. Strengthening grassroots extension bodies is crucial. Paddy producers adopt monocropping, promoting integrated farming systems for long-term income and productivity. Crop insurance helps farmers weather climate change effects, promoting policy initiatives.

Limitation and future scope

The Survey limits its capacity to cover and accurately portray information. Because each district has a set budget and sample size, the survey may miss important details in places that are not covered by the available resources. Some areas, particularly those with large forests and conservative tribal nations, are difficult to reach, creating challenges and potentially painting an incorrect picture of precision agricultural techniques. Limited road access also makes it difficult for surveyors to reach specific field plots, reducing overall coverage. Many farmers may be ignorant of the reason for the poll survey or interview, which may impact their responses since they are anxious about its potential political implications. Farmers' lack of awareness, training, and education is critical for understanding rice disease patterns. A lack of knowledge about suitable crops and growing approaches exacerbates the underutilization of these significant agricultural resources. However, while accuracy is a useful indicator for assessing a model's overall performance, it can be misleading in some circumstances, especially when the dataset is skewed, or the misclassification costs are unequal. A key limitation of this study is the analysis of only three rice diseases due to the unavailability of extensive field survey data. Variability in rice planting durations complicates disease classification, while small landholdings and coarse-resolution data further contribute to uncertainty. Additionally, inconsistencies in irrigation, pesticide, insecticide, and fertilizer application rates among farmers make it challenging to accurately determine the causes and timing of disease outbreaks.

Future research should expand field surveys to include a broader range of rice diseases while considering the effects of climate change. It will include recommendations for machines that use sophisticated computer vision-reliant deep learning and artificial intelligence methodologies for rice disease identification. These strategies produce abnormally fruitful results for the detection of diseases involving images of leaves, harvest fields, or seeds. The future effort will focus on a more in-depth examination of the accuracy of agribusiness to broaden the understanding of rice, which is one of the world's most important crops. It is recommended to systematically store a database focused on specific crop diseases associated with particular crops. Additionally, it is important to compile a synthetic data-driven and user-driven system that considers various treatment conditions for the respective crops. These conditions include tailored service practices such as planting dates, irrigation schedules, real time weather forecasting apps, soil health status, fertilizer application rates, seed rates, levels of farm mechanization, tillage practices, crop residue management, farmer socioeconomic conditions, and the use of farmyard manure.

Conclusions

This study employs five machine learning models to predict Kharif rice crop diseases in both current and future geospatial scenarios in Eastern India. Various vegetation indices were analyzed to assess crop health, though early disease symptoms may not be immediately detectable. Future climate projections for disease detection were generated using CMIP6 global climate models under SSP2-4.5 and SSP5-8.5 scenarios. In the region high humidity, water scarcity, nutrient-deficient soil, and poor management contribute to the prevalence of diseases like brown spot, blast, and tungro. This study examined three rice diseases for 2023 and 2030 using RF, GBM, SVM, XGB, and ANN models. Among these, the RF model performed best, achieving a maximum test accuracy of 0.70. The findings reveal that rice brown spot disease has spread by 96.72% in Purba Bardhaman. Rice blast disease affects 72.38% of rice plants in northwestern Birbhum, driven by high temperatures, water deficits, and low soil moisture. Meanwhile, due to nitrogen and zinc deficiencies, rice tungro disease impacts 63.45% of rice crops in Bankura. SHAP analysis identified Pr, T_{min}, and T_{max} as the key climatic factors influencing all three rice diseases. The perturbation sensitivity analysis showed that AWC and SAVI negatively affected brown spot and rice blast but positively influenced tungro disease. NDVI, NDMI, and T_{max} exhibited varying sensitivities, with T_{max} positively impacting brown spot but negatively affecting rice blast and tungro. This approach enhances crop production optimization, aiding in doubling farmer income and supporting decision-makers in monitoring crop disease disasters. It also contributes to achieving the United Nations Sustainable Development Goals.

Acknowledgements We acknowledge the project "Integration of Digital Augmentation for Sustainable Agroecosystem in Western Lateritic Zone under National Hydrology Project, West Bengal" under which this work is mapped. The author also conveys special thanks to the International Centre for Agricultural Research in the Dry Areas (ICARDA) for supporting the necessary logistics for this research work. We would like to express our heartfelt appreciation to all enumerators, farmers, professionals, and anyone who contributed to this research.

Authors' contributions Conceptualization, S.S., C.S., A.G.; methodology, S.S., C.S., A.G.; software, S.S., C.S., A.G; validation, S.S., C.S., A.G., M.S; formal analysis, S.S., C.S., A.G; investigation, S.S., C.S., A.G., M.S; resources, S.S., C.S., A.G., M.S; data creation, S.S., C.S.; writing—original draft preparation, S.S., C.S. A.G.; writing—review and editing, S.S., C.S., A.G., M.S; visualization, S.S.; supervision and revision, A.G., M.S., S.S; project administration, A.G.; funding – A.G.; All authors have read and agreed to the published version of the manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Ethics statement Not applicable: This manuscript does not include human or animal research.

Competing interest The authors declare no competing interests.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data, 5*, 170191.

Aggarwal, Shruti, M. Suchithra, N. Chandramouli, Macha Sarada, Amit Verma, D. Vetrithangam, Bhaskar Pant, and Biruk Ambachew Adugna. "Rice disease detection using artificial intelligence and ML techniques to improvise agro-business." Scientific Programming 2022 (2022).

Al-Sheriadeh, M. S., & Daqdouq, M. A. (2024). Robustness of ML algorithms to generate flood susceptibility maps for watersheds in Jordan. Geomatics Natural Hazards and Risk, 15(1). https://doi.org/10.1080/19475705.2024.2378991

Ansari, A., Lin, Y., & Lur, H. (2021). Evaluating and adapting climate change impacts on rice production in Indonesia: A case study of the Keduang Subwatershed. *Central Java. Environments, 8*(11), 117. https://doi.org/10.3390/environments8110117

Ansari, A., Pranesti, A., Telaumbanua, M., Alam, T., Taryono, N., Wulandari, R. A., Nugroho, B. D. A., & Supriyanta, N. (2023). Evaluating the effect of climate change on rice production in Indonesia using multimodelling approach. *Heliyon, 9*(9), e19639. https://doi.org/10.1016/j.heliyon.2023.e19639

Azizi, M. M. F., & Lau, H. Y. (2022). Advanced diagnostic approaches developed for the global menace of rice diseases: A review. *Canadian Journal of Plant Pathology, 44*(5), 627–651. https://doi.org/10.1080/07060661.2022.2053588

Bai, Y., Feng, M. (2018). Data fusion and accuracy evaluation of multi-source global land cover datasets. *Acta Geogr Sin, 73*, 2223–2235.

Breiman, L. (2001). Random forests. *Machine learning, 45*, 5–32.

Carvalho, M. P., Rodrigues, F. A., Silveira, P. R., Andrade, C. C. L., Baroni, J. C. P., Paye, H. S., et al. (2010). Rice resistance to brown spot mediated by nitrogen and potassium. *Journal of Phytopathology., 158*, 160–166.

Chang, K. T., Merghadi, A., Yunus, A. P., Pham, B. T., & Dou, J. (2019). Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based ML techniques. *Scientific Reports, 9*(1), 1–21.

Chen, T., Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).

Chowdhury, M. E., Islam, A. S., Zzaman, R. U., & Khadem, S. (2024). A ML-based approach for flash flood susceptibility mapping considering rainfall extremes in the northeast region of Bangladesh. *Advances in Space Research*. https://doi.org/10.1016/j.asr.2024.10.047

Deng, R., Jiang, Y., Tao, M., Huang, X., Bangura, K., Liu, C., et al. (2020). Deep learning-based automatic detection of productive tillers in rice. *Computers and Electronics in Agriculture, 177*, 105703. https://doi.org/10.1016/j.compag.2020.105703

Deng, R., Tao, M., Xing, H., Yang, X., Liu, C., Liao, K., & Qi, L. (2021). Automatic diagnosis of rice diseases using deep learning. Frontiers in Plant Science, 12. https://doi.org/10.3389/fpls.2021.701038

Franceschini, S., Gandola, E., Martinoli, M., Tancioni, L., Scardi, M. (2018). Cascaded neural networks improving fish species prediction accuracy: The role of the biotic information. Scientific Reports, 8(1). https://doi.org/10.1038/s41598-018-22761-4

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232.

Hearst, M. A. Dumais, S. T. Osuna, E. Platt J. and Scholkopf, B. (1998). Support vector machines, in IEEE intelligent systems and their applications, 13(4), 18–28,1998, https://doi.org/10.1109/5254.708428.

Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In *Neural Networks for Perception* (pp. 65–93). Elsevier.

Huete, A. R., 1988, "A soil-adjusted vegetation index (SAVI)," Remote Sensing of Environment, Vol 25, 295–309." (ESRI, 2018)

Jackulin, C., & Murugavalli, S. (2022). A comprehensive review on detection of plant disease using ML and deep learning approaches. *Measurement: Sensors, 24*, 100441.

Kamarudin, S. A. A., Ahmad, F., Hisham, S. N., Yusof, S. N., Hassan, A. A., Hussein, S., ... & Shamsudin, N. A. A. (2024). Whole genome resequencing data and grain quality traits of the rice cultivar Mahsuri and its blast disease resistant mutant line, Mahsuri Mutant. Data in Brief, 52, 109974.

Li, R., Chen, S., Matsumoto, H., Gouda, M., Gafforov, Y., Wang, M., & Liu, Y. (2023). Predicting rice diseases using advanced technologies at different scales: Present status and future perspectives. *Abiotech, 4*(4), 359–371.

Lu, Y., Wu, X., Liu, P., Li, H., & Liu, W. (2023). Rice disease identification method based on improved CNN-BiGRU. *Artificial Intelligence in Agriculture, 9*, 100–109.

Mahadevan, K., Punitha, A., Suresh, J. (2024). Automatic recognition of rice plant leaf diseases detection using deep neural network with improved threshold neural network. e-Prime - Advances in Electrical Engineering Electronics and Energy, 8, 100534. https://doi.org/10.1016/j.prime.2024.100534

Mishra, S., & Mishra, D. R. (2012). Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sensing of Environment, 117*, 394–406.

Mohidem, N. A., Hashim, N., Shamsudin, R., & Man, H. C. (2022). Rice for food security: Revisiting its production, diversity, rice milling process and nutrient content. *Agriculture, 12*(6), 741. https://doi.org/10.3390/agriculture12060741

Mondal, D., Akhtar, S. R., Mukhopadhyay, R. (2024). A review on important diseases of rice in West Bengal. International Journal of Research in Agronomy, 7(4S), 67–70. https://doi.org/10.33545/2618060x.2024.v7.i4sa.533

Mousumi, M. A., Paparrizos, S., Ahmed, M. Z., Kumar, U., Uddin, M. E., and Ludwig, F. (2023). Common sources and needs of weather information for rice disease forecasting and management in coastal Bangladesh. NJAS Impact in Agricultural and Life Sciences, 95(1). https://doi.org/10.1080/27685241.2023.2191794

Mumuni, A., & Mumuni, F. (2024). Automated data processing and feature engineering for deep learning and big data applications: A survey. *Journal of Information and Intelligence*. https://doi.org/10.1016/j.jiixd.2024.01.002

Mwangi, R. W., Mustafa, M., Charles, K., Wagara, I. W., & Kappel, N. (2023). Selected emerging and reemerging plant pathogens affecting the food basket: A threat to food security. *Journal of Agriculture and Food Research, 14*, 100827. https://doi.org/10.1016/j.jafr.2023.100827

Narmilan, A., Gonzalez, F., Salgadoe, A. S. A., & Powell, K. (2022). Detection of white leaf disease in sugarcane using ML techniques over UAV multispectral images. *Drones, 6*(9), 230.

Nayak, A., Chakraborty, S., & Swain, D. K. (2023). Application of smartphone-image processing and transfer learning for rice disease and nutrient deficiency detection. *Smart Agricultural Technology, 4*, 100195.

Prismantoro, D., Akbari, S. I., Permadi, N., Dey, U., Anhar, A., Miranti, M., Mispan, M. S., & Doni, F. (2024). The multifaceted roles of Trichoderma in managing rice diseases for enhanced productivity and sustainability. *Journal of Agriculture and Food Research, 18*, 101324. https://doi.org/10.1016/j.jafr.2024.101324

Rengalakshmi, R., Manjula, M., & Devaraj, M. (2018). Making climate information communication gender sensitive - lessons from Tamil Nadu. *Economic and Political Weekly/Economic & Political Weekly, 53*(17). https://www.cabdirect.org/cabdirect/abstract/20183269227

Robnik-Sikonja, M., & Kononenko, I. (2008). Explaining classifications for individual instances. *IEEE Transactions on Knowledge and Data Engineering, 20*(5), 589–600. https://doi.org/10.1109/tkde.2007.190734

Rouse, J. W. Jr., Haas, R. H., Schell, J. A., Deering, D. W. (1973). Monitoring vegetation systems in the great plains with erts. In *Proceedings of the 3rd ERTS-1 Symposium (NASA SP-351)* (pp. 309–317).

Saha, A., Dutta, A., Ghosh, S., & Mallick, G. K. (2008). Rice diseases in the red lateritic zone of West Bengal. *India, Environment and Ecology, 26*(4), 1691–1692.

Sahoo, S., Singha, C., Govind, A., & Moghimi, A. (2024). Review of climate-resilient agriculture for ensuring food security: Sustainability opportunities and challenges of India. Environmental and Sustainability Indicators, 100544. https://doi.org/10.1016/j.indic.2024.100544

Sarker, I. H. (2021). ML: Algorithms, real-world applications and research directions. *SN Computer Science, 2*(3). https://doi.org/10.1007/s42979-021-00592-x

Scardi, M., & Harding, L. W., Jr. (1999). Developing an empirical model of phytoplankton primary production: A neural network case study. *Ecological Modelling, 120*, 213–223. https://doi.org/10.1016/S0304-3800(99)00103-9

Shivappa, R., Navadagi, D. B., Baite, M. S., Yadav, M. K., Rathinam, P. S., Umapathy, K., ... & Rath, P. C. (2021). Emerging minor diseases of rice in India: Losses and management strategies. Integrative Advances in Rice Research.

Singh, S. P., Pritamdas, K., Devi, K. J., & Devi, S. D. (2023). Custom convolutional neural network for detection and classification of rice plant diseases. *Procedia Computer Science, 218*, 2026–2040.

Singha, C., Gulzar, S., Swain, K. C., & Pradhan, D. (2023). Apple yield prediction mapping using ML techniques through the Google Earth Engine cloud in Kashmir Valley. *India. Journal of Applied Remote Sensing, 17*(1), 014505.

Singha, C., Sahoo, S., Tinh, N. D., Ditthakit, P., Lu, Q., El-Magd, S. A., & Swain, K. C. (2024). Climate-resilient strategies for sustainable groundwater management in Mahanadi River basin of Eastern India. *Acta Geophysica*. https://doi.org/10.1007/s11600-024-01466-5

Skamnioti, P., & Gurr, S. J. (2009). Against the grain: Safeguarding rice from rice blast disease. *Trends in biotechnology, 27*(3), 141–150.

Surendhar, M., Anbuselvam, Y., & Ivin, J. (2022). Status of rice brown spot (Helminthosporium oryz) management in India: A review. *Agricultural Reviews, 43*(2), 217–222. https://doi.org/10.18805/ag.R-2111

Tripathi, A. D., Mishra, R., Maurya, K. K., Singh, R. B., & Wilson, D. W. (2019). Estimates for world population and global food availability for global health. In The role of functional food security in global health (pp. 3–24). Academic Press.

Udayananda, G. K. V. L., Shyalika, C., & Kumara, P. P. N. V. (2022). Rice plant disease diagnosing using ML techniques: A comprehensive review. *SN Applied Sciences, 4*(11). https://doi.org/10.1007/s42452-022-05194-7

USDA, (2022). Foreign agricultural service. *India Rice: Production down due to a decline in planted area*, 1–10. https://ipad.fas.usda.gov. Accessed Oct 12 2024.

Xinyue, H., Yuping, W. A. N. G., Qian, Q. I. A. N., & Deyong, R. E. N. (2023). Molecular mechanism of rice necrotic lesion for optimized yield and disease resistance. Rice Science.

Yan, Y. X., Zhang, X. Y., Tan, Y. Y., Huang, J. Z., Jankuloski, L., & Shu, Q. Y. (2022). Establishment of an artificial inoculation system for the efficient induction of rice bakanae disease. *Crop Design, 1*(2), 100016.

Yaqoob, A., Aziz, R. M., & Verma, N. K. (2023). Applications and techniques of ML in cancer classification: A systematic review. *Human-Centric Intelligent Systems, 3*(4), 588–615. https://doi.org/10.1007/s44230-023-00041-3

Zheng, Q., Huang, W., Xia, Q., Dong, Y., Ye, H., Jiang, H., ... & Huang, S. (2023). Remote sensing monitoring of rice diseases and pests from different data sources: A review. Agronomy, ***13*****(7), 1851.**