frontiers in Climate

TYPE Original Research
PUBLISHED 22 June 2026
DOI 10.3389/fclim.2026.1733203

Check for updates

OPEN ACCESS

EDITED BY
Anthony Lupo,
University of Missouri, United States

REVIEWED BY
Sumit Chakravarty,
Uttar Banga Krishi Viswavidyalaya, India
Karam Alsafadi,
Xiamen University, China

*CORRESPONDENCE
A. P. Ranjini
ccpranju@gmail.com

RECEIVED 27 October 2025
REVISED 25 May 2026
ACCEPTED 27 May 2026
PUBLISHED 22 June 2026

CITATION
Ranjini AP, Machenahalli S, Patil S, Giri MS, Naika R, Rajan S, Babou C and Senthilkumar M (2026) Weather-driven epidemiology and forecasting of Myrothecium leaf spot in coffee seedlings: machine learning approach for predicting the disease.
Front. Clim. 8:1733203.
doi: 10.3389/fclim.2026.1733203

COPYRIGHT
© 2026 Ranjini, Machenahalli, Patil, Giri, Naika, Rajan, Babou and Senthilkumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Weather-driven epidemiology and forecasting of Myrothecium leaf spot in coffee seedlings: machine learning approach for predicting the disease

A. P. Ranjini[1]*, Santoshreddy Machenahalli[1], Somashekhargouda Patil[2], Madhu S. Giri[1], Raja Naika[3], Soundara Rajan[1], C. Babou[4] and M. Senthilkumar[4]

[1]Division of Plant Pathology, Central Coffee Research Institute, Chikkamagaluru, Karnataka, India, [2]Division of Plant Physiology, Central Coffee Research Institute, Chikkamagaluru, Karnataka, India, [3]Department of Applied Botany, Kuvempu University, Shivamogga, Karnataka, India, [4]Central Coffee Research Institute, Chikkamagaluru, Karnataka, India

Myrothecium leaf spot, caused by *Myrothecium roridum*, is an emerging disease threat to coffee (*Coffea arabica* L. cv. Chandragiri) nurseries, resulting up to 30% seedling loss and occasionally affecting young field transplants. Although *M. roridum* was historically considered a weak pathogen, recent increases in its prevalence suggest that climate variability and shifts in local weather patterns may be driving its resurgence. To evaluate the influence of meteorological factors on disease development, a study was conducted during 2018 and 2019, focusing on cumulative rainfall, maximum and minimum temperatures and relative humidity. Disease incidence and severity were recorded weekly and correlated with weather data. Interannual comparisons revealed that, during the year 2019 characterized by higher rainfall and humidity and lower temperatures exhibited substantially greater disease index (DI) values and spore spread potential (Sv) than 2018. The dataset consisted of 44 weekly observations (22 per year). Pearson correlation analysis revealed strong positive associations between DI and rainfall (r = 0.76) and minimum RH (r = 0.74) and between Sv and both rainfall (r = 0.72) and DI (r = 0.84). Among six forecasting models tested for one-week-ahead prediction of DI, XGBoost achieved the best performance (RMSE ≈ 5.85, R^2 ≈ 0.63). Granger causality analysis confirmed rainfall as a significant temporal predictor of DI and maximum RH as a key driver of Sv, highlighting the role of climate-sensitive weather parameters in disease epidemiology. However, the moderate predictive accuracy and absence of independent validation indicate that the models should presently be considered exploratory rather than operational forecasting tools. These findings suggest that interannual weather variability, particularly fluctuations in rainfall and humidity, may influence the risk of *Myrothecium* outbreaks in coffee nurseries. Integrating weather-based forecasting models into early warning systems may support the future development of climate-resilient disease management strategies.

KEYWORDS
climate variability, coffee seedlings, early warning systems, forecasting models, *Myrothecium* leaf spot, plant disease, plantation crops, XGBoost

1 Introduction

Coffee is a major exporting commodity and generates income to many coffee growing countries including India. The success of new planting in a perennial crop like coffee, depends primarily on planting of vigorous, pest and disease-free seedlings in the field. In such a condition, additional care has to be taken to bring up healthy seedlings. Nursery plays a major role in setting-up, extending or reviving a coffee plantation, as the healthy and vigorous seedlings are selected for planting in the main field from the nursery. This gains a long-term productivity and in turn long term income to the planter (Ranjini et al., 2017, Ranjini and Rajanaika, 2018). Leaf spot disease, caused by the pathogen *Myrothecium roridum* on coffee seedlings is of major concern in coffee nurseries and in the field conditions (Dutta et al., 2013). This disease is of economic importance, as the pathogen causes death of seedlings, leading to a loss upto 30% in coffee nurseries and cause considerable damage to leaves and berries on grown up plants under field conditions (Daivasikamani et al., 2016). The pathogen *M. roridum*was listed under the minor pathogens of coffee, but in the recent years it is widespread in all the coffee regions and gaining importance because of the variations in the weather conditions. The change in the climatic conditions might have favored the spread and multiplication of the pathogen (Sudhakar Bhat et al., 2011). Though factors like pathogen virulence and susceptibility of host are important in the disease development, environmental conditions play a major role in determining the course of epidemics of the plant diseases. Weather is one of the important abiotic factors that has a major impact on initiation, development and spread of the disease.

To address the existing knowledge gap, this study was designed around the following research questions (i) how do key weather variables influence the development of *Myrothecium* leaf spot in coffee seedlings and (ii) can short-term disease incidence be reliably predicted using statistical and machine learning models.

Based on these questions, it was hypothesized that (i) increased rainfall and relative humidity significantly enhance disease incidence and severity and (ii) machine learning models, particularly ensemble approaches such as machine learning models may better capture nonlinear disease-weather relationships than conventional statistical approaches in forecasting disease dynamics under variable climatic conditions. The formulation of the present hypotheses is grounded in established evidence that environmental factors, particularly moisture and humidity, play a critical role in the epidemiology of fungal plant diseases. Previous studies have demonstrated that increased rainfall and high relative humidity create favorable conditions for pathogen infection, sporulation and disease spread (Chakraborty, 2005; Garrett et al., 2006; Pautasso et al., 2012; Velasquez et al., 2018). In coffee systems, climate variability has been shown to significantly influence pest and disease dynamics, including major outbreaks under favorable environmental conditions (Jaramillo et al., 2011; Bunn et al., 2015; Avelino et al., 2015; Chakraborty and Newton, 2011).

Furthermore, advances in predictive analytics indicate that machine learning approaches are effective in modeling complex and nonlinear relationships between environmental variables and plant disease dynamics (Elith et al., 2008; Jeong et al., 2016; de Oliveira et al., 2020; Cohen et al., 2020). These findings provide a strong basis for formulating the hypotheses of the present study. Although several studies have examined weather-disease relationships in plantation crops, comparatively few investigations have integrated epidemiological observations with machine learning-based forecasting approaches for coffee diseases under Indian conditions. Previous studies on coffee disease prediction have largely focused on coffee rust and berry diseases using conventional statistical methods. Recent advances in deep learning and machine learning have demonstrated considerable potential for automated detection and prediction of coffee diseases (Martinez et al., 2022; Elisaye et al., 2023; Topal et al., 2024; Nawaz et al., 2024). Therefore, the present study contributes novelty by comparatively evaluating multiple time-series and machine learning frameworks for forecasting *Myrothecium* leaf spot outbreaks under weather-variable conditions.

The specific objectives of this study were (a) to quantify the relationship between weather variables and disease progression (b) to identify key climatic drivers influencing disease outbreaks and (c) to evaluate and compare multiple forecasting models for short-term prediction of disease incidence.

To achieve these objectives, a suite of six models representing both classical statistical (VAR, DLNM) and advanced data-driven approaches (XGBoost, HMM, FDA-FPCA, DTW-KNN) were employed. These models were selected to capture both linear and nonlinear relationships, temporal dependencies and complex interactions among climatic variables. While traditional models such as VAR provide interpretable linear relationships, they may fail under non-stationary conditions. In contrast, machine learning models such as XGBoost offer higher predictive flexibility but may be prone to overfitting. No structural modification of these models was required; however, parameter tuning and validation using RMSE, MAE and R^2 were conducted to ensure robustness and comparability.

2 Materials and methods

The nursery experiments were conducted at Central Coffee Research Institute (CCRI), Balehonnur, Koppa taluk, Chikkamagaluru district, Karnataka State, India. The study area is situated at 13° 22″ North Latitude and 75° 28″ East Longitude at an elevation of 848 meters above MSL. During the experiment period of years 2018 and 2019, the annual rainfall varied between 3073.6 and 3859.9 mm, received mainly from May to October months with an annual average rainfall of 3466.75 mm. The maximum temperatures ranged from 20.1 to 36.4 °C and minimum temperature from 10 to 20 °C. The relative humidity (RH) ranged from 45 to 100%. Maximum sunshine was 9 h.

2.1 Biological material used for the study

The seedlings of *Coffea arabica* L. cv. Chandragiri, which was found more susceptible to stem necrosis and leaf spot diseases caused by the pathogen *Myrothecium roridum,* during the survey period was used for the study. Seedlings of above-mentioned coffee cultivar are raised in many coffee nurseries in Chikkamagaluru district of Karnataka state and are recommended for commercial cultivation by the Coffee Board of India (Anonymous, 2014).

2.2 Experimental design

An experiment was laid out in net house conditions at CCRI, with the arabica coffee cultivar "Chandragiri" as test material. The experiment was conducted for two consecutive years 2018 and 2019. Seeds of Chandragiri coffee were sown in the germination bed comprising the nursery mixture, i.e., Jungle soil; Farmyard Manure; sand in the ratio of 3:2:1 (Anonymous, 2023). The germination bed was covered

10.3389/fclim.2026.1733203

with paddy straw to retain moisture and to maintain humidity. After 45 days, the 'topee' stage seedlings were transplanted to the poly bags (secondary nursery bed), containing the same nursery mixture (one kg per poly bag; Coffee Guide, 2023). Fifty numbers of poly bags (22.5 × 12.5 cm) planted with 'topee' stage coffee seedlings were arranged in the secondary beds with the plot size of 3 × 12 feet.

Because the objective of the study was observational epidemiological monitoring under naturally occurring disease pressure, no artificial inoculation treatment was imposed. Weekly disease observations were repeatedly recorded from the same seedling population over time. Consequently, the dataset represents repeated time-series measurements rather than independent experimental replicates. The study consisted of a single observational nursery cohort without replicated treatment plots because no experimental treatments were imposed.

2.3 Period of observation

After 90 days of sowing, the coffee seedlings attained three pairs of leaves. The observations were recorded on first appearance of the symptoms, leaf spot disease incidence and severity were recorded at weekly intervals, on all the seedlings for a continuous period of 22 weeks starting from 22nd standard week until 43rd standard week during the years 2018 and 2019. After the 43rd week the seedlings sent to the field for planting. Effect of weather parameters on *Myrothecium* leaf spot disease incidence and severity in arabica coffee (cv. Chandragiri) data recorded in 2018 and 2019 season are provided in Supplementary Table S1.

2.4 Assessment of disease incidence and severity

Leaf spot disease incidence and severity were recorded at weekly intervals (late May to late October Month and this was coincides with the monsoon), on all the Chandragiri coffee seedlings between 22nd and 43rd standard weeks, during the years 2018 and 2019, as per the standard methodology. Percent incidence of the disease and severity was calculated by using formulas mentioned below (Santoshreddy et al., 2019).

$$\text{Disease incidence}\,(\%) = \frac{\text{Number of infected leaves}}{\text{Total number of leaves observed}} \times 100.$$

Observation on disease severity was recorded by following 0–9 point scale as described by Mayee and Datar (1986) and percent disease severity was expressed using the formula as mentioned below by Wheeler (1969).

Disease rating scale:

Grade	Per cent leaf area infected
0	0
1	1–10
3	11–25
5	26–50
7	51–75
9	>75

The description of Wheeler (1969) was used to calculate percent disease severity.

$$\text{Disease severity}\,(\%) = \frac{\text{Sum of all numerical disease ratings}}{\text{Total No.of samples observed} \times \text{Maximum disease rating grade}} \times 100$$

In addition to disease incidence (DI) and severity, spore dissemination (Sv) potential was assessed as an indicator of pathogen dissemination potential under prevailing environmental conditions. Sv was quantified based on the relative intensity of visible sporulation on infected leaves, recorded concurrently with disease severity observations using a standardized ordinal scoring approach (Santoshreddy et al., 2019; Singh et al., 2003). This parameter reflects the potential for secondary spread of *Myrothecium roridum* under prevailing environmental conditions.

2.5 Collection of data on weather parameters

Thermo-hygro-clock fixed at the experimental site was used to record minimum and maximum temperature, relative humidity in the morning and evening. The Indian Metrological Department (IMD) approved weather station, located near the experimental site was used for collecting rainfall data, throughout the experimental period. The mean temperature and relative humidity of the week and cumulative rainfall was calculated.

Observations recorded during the years 2018 and 2019, on leaf spot disease incidence and severity was correlated with the weather variables. The linear regression equations were developed for estimation leaf spot incidence and severity by taking weather parameters as input variables and linear mathematical models were developed using standard statistical procedure (Gomez and Gomez, 1984; Panse and Sukhatme, 1985).

2.6 Descriptive statistical analysis

Descriptive statistics including mean, standard deviation, minimum and maximum were computed for all variables to summarize central tendency and dispersion. Year-wise comparisons were made to identify interannual variability in climatic conditions and disease indicators. Visualizations such as bar plots with error bars (mean ± SD) were generated using the *ggplot2* package (Wickham, 2016) to facilitate interpretation.

2.7 Correlation analysis

Pearson's correlation coefficients (r) were calculated to examine the linear relationships among climatic variables, DI and Sv. A heatmap was created to visualize the strength and direction of these associations. Statistical significance of correlations was evaluated using standard p-value thresholds $p < 0.05$, $p < 0.01$ and $p < 0.001$ following best practices in environmental statistics (Dormann et al., 2013; Tomar et al., 2009). However, since the data consist of weekly time-series observations, temporal autocorrelation may violate the independence assumption of Pearson correlation. Therefore, these correlations are interpreted as exploratory associations rather than definitive relationships.

To address temporal dependencies more rigorously, Granger causality analysis within a VAR framework was performed, which explicitly accounts for lagged relationships among variables.

2.8 Modeling and forecasting

To assess the predictive performance of short-term disease forecasting models, six different approaches were implemented for one-week-ahead prediction of the disease index:

1. XGBoost (Extreme Gradient Boosting; Chen and Guestrin, 2016)
2. HMM (Hidden Markov Model)
3. FDA-FPCA (Functional Data Analysis with Functional Principal Component Analysis; Ramsay and Silverman, 2005)
4. DTW-KNN (Dynamic Time Warping with K-Nearest Neighbors; Petitjean et al., 2011)
5. DLNM (Distributed Lag Nonlinear Model; Gasparrini et al., 2011)
6. VAR (Vector AutoRegression; Lutkepohl, 2005)

Model performance was evaluated using three standard error metrics (a) Mean Absolute Error (MAE), (b) Root Mean Square Error (RMSE) and (c) Coefficient of Determination (R^2).

The above each model was selected based on its suitability for time-series prediction and its prior application in environmental and epidemiological studies. XGBoost is an ensemble learning method known for handling nonlinear relationships and interactions efficiently. HMM captures hidden state transitions in temporal processes. FDA-FPCA reduces dimensionality in functional data, while DTW-KNN enables similarity-based classification of time-series patterns. DLNM is particularly useful for capturing lagged nonlinear effects of climatic variables, and VAR models represent multivariate linear interdependencies.

Model performance was evaluated using RMSE, MAE, and R^2. The model with the lowest MAE and RMSE and highest R^2 was deemed the most accurate and robust for disease forecasting under variable climatic conditionsHyperparameter tuning was performed where applicable to optimize predictive performance. Due to dataset limitations, external validation was not conducted, which is acknowledged as a limitation.

It is important to note that the dataset consisted of 44 weekly observations (22 per year), which may limit the robustness of machine learning models. To reduce the risk of overfitting, model complexity was controlled through careful hyperparameter tuning and performance was evaluated using multiple metrics (RMSE, MAE and R^2). However, due to the limited sample size, the predictive results should be interpreted as preliminary rather than fully generalizable. Due to the relatively small dataset, formal k-fold cross-validation was not feasible without substantially reducing training data. Therefore, model evaluation was conducted using available data and this limitation is acknowledged as a constraint on model generalizability. For one-week-ahead forecasting, models were trained using sequential historical observations and evaluated on subsequent unseen weekly observations. Because of the limited dataset size ($n = 44$), a formal rolling-origin or k-fold cross-validation strategy was not feasible without substantially reducing training data. Because the dataset was limited to 44 weekly observations collected from a single experimental nursery system, independent external validation using geographically distinct datasets was not feasible. Therefore, the present forecasting models should be interpreted as proof-of-concept predictive frameworks requiring future validation across multiple locations and seasons.

2.9 Granger causality analysis

To determine the temporal influence of climatic variables on DI and Sv, Granger causality tests were performed using a Vector AutoRegressive (VAR) framework with optimal lag selection ($p = 2$), based on the Akaike Information Criterion (AIC). These tests assessed whether historical values of each climate variable could significantly predict future values of DI or Sv.

Resulting p-values were transformed into $-\log_{10}(p)$ values for visualization, with a threshold of 1.3 (corresponding to $p = 0.05$) used to determine statistical significance, as recommended in time-series causality studies (McRoberts et al., 2003; Alex et al., 2018).

2.10 Software and tools

All analyses were performed using R statistical software (*version 4.3.1*) within the RStudio environment. The following R packages were employed:

tidyverse (v2.0.0) for data manipulation and wrangling (Wickham et al., 2019).

ggplot2 (v3.4.4) for creating all figures including plots and heatmaps (Wickham, 2016).

xgboost (v1.7.5) for implementing gradient boosting models (Chen and Guestrin, 2016).

vars (v1.5–8) for fitting VAR models and conducting Granger tests.

fda(v5.5.1) for functional data analysis and FPCA.

dtwclust (v5.5.13) for DTW-based time series classification.

dlnm (v2.4.9) for modeling distributed lag effects (Gasparrini, 2011).

forecast (v8.22) for time series model diagnostics and evaluation.

reshape2 (v1.4.4) for data reshaping.

All figures were exported using the ggsave function with 300 dpi resolution in JPEG formats. R projects and versioned scripts were used to ensure reproducibility throughout the analysis.

2.11 Statistical analysis

Year-wise comparisons of climatic variables and disease parameters were evaluated using Welch's t-test after assessing variance heterogeneity. Statistical significance was considered at $p < 0.05$. Confidence intervals (95%) were calculated for regression coefficients and model evaluation metrics wherever applicable.

Because repeated weekly observations were collected from the same experimental units, temporal autocorrelation was expected. Therefore, Pearson correlation analysis was interpreted cautiously, while Granger causality analysis was used to better account for lagged temporal relationships. Given the relatively small sample size and repeated temporal observations, the statistical analyses were intended primarily for exploratory epidemiological inference and comparative model evaluation rather than definitive predictive generalization.

3 Results

3.1 Effect of weather factors on the development of incidence and severity of the leaf spot disease of coffee seedlings

Variations in the climate has impacted significantly widespread damage in-terms of occurrence of pests and diseases in turn on crop production. The environmental factors or weather conditions, most seriously affect the initiation and development of infectious plant disease are temperature, humidity of the air, amount of rainfall, soil temperature, moisture etc. Present study indicated that, during first season, 2018 first expression of symptoms of leaf spot disease were observed after 90 days of transplanting the 'topee' stage coffee seedlings into poly bags. The weather conditions during the 43rd standard week of the year showed minimum disease incidence and severity (4.52 and 0.79% respectively) during which period the average minimum and maximum temperature was 26.84 and 27.18 °C, respectively, (Figure 1). Morning and evening relative humidity was 70.64 and 64.71% correspondingly and weekly cumulative rainfall was 2.8 mm. Whereas, during 32nd standard week the weather conditions were very congenial for the buildup of the disease and recorded maximum leaf spot disease incidence and severity (24.74% and 9.14%) was recorded during the week, minimum and maximum temperature was 15.33 and 32.14 °C, respectively. Morning and evening relative humidity was 95.14 and 90.2% correspondingly. The cumulative weekly rainfall of the week was 563.6 mm. It is observed that, the disease spread rapidly when there was continuous rains coupled with high humidity.

The period of minimum and maximum leaf spot incidence and severity recorded in 2019 season was slightly different from those of 2018. The minimum level of disease incidence and severity (6.98% and 2.89% respectively) was recorded in 22nd standard week of the year during which period, the average minimum and maximum temperature was 19.79 °C and 25.18 °C, respectively. Morning and evening relative humidity was 87.71 and 75.81% correspondingly and the weekly cumulative rainfall was 51.8 mm. Maximum level of disease incidence and severity (34.62% and 12.52% respectively) was recorded during 31st standard week, wherein the average weekly minimum and maximum temperature were 17 °C and 32.93 °C, respectively. Morning and evening relative humidity was 96.14 and 90.78% correspondingly and the cumulative weekly rainfall was 656.1 mm (Figure 2). It was observed that, the disease incidence was high when the rainfall was maximum during the period of the study (656.1 mm).

3.2 Descriptive statistics

The comparative analysis of climatic variables and disease indicators between 2018 and 2019 is presented in Figure 3. The disease index (DI) related to *Myrothecium* infection was notably higher in 2019, with a mean value exceeding that of 2018 by a substantial margin. Similarly, spore dissemination potential (Sv) also showed an upward trend in 2019. Rainfall demonstrated high variability, with a marked increase in mean values during 2019 and large standard deviations in both years, suggesting irregular precipitation events. Relative humidity, both minimum and maximum, exhibited a slight increase in 2019. In contrast, temperature parameters declined average minimum and maximum temperatures were lower in 2019 compared to 2018. The mean disease incidence and rainfall values observed during 2019 were significantly greater than those recorded during 2018 (Welch's t-test, $p < 0.05$), indicating that wetter environmental conditions were associated with greater disease development.

3.3 Pearson correlation heatmap

The Pearson correlation heatmap (Figure 4) illustrates the strength and direction of associations among climatic variables (Rainfall, Temperature, Relative Humidity), spore spread potential (Sv) and disease incidence (DI), potentially due to *Myrothecium roridum* infection in coffee. Strong positive correlations were observed between DI and both Rainfall ($r = 0.76$) and Minimum

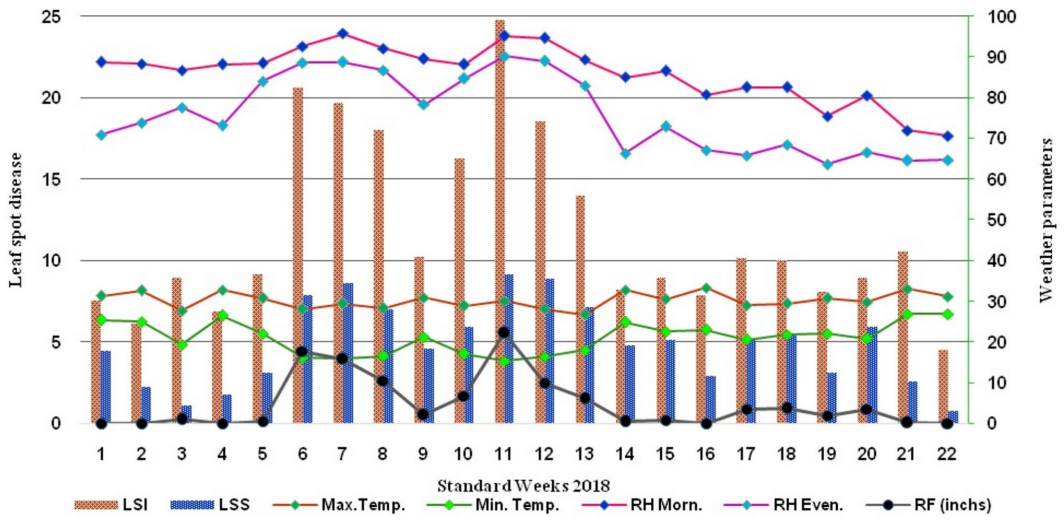


FIGURE 1
Graphical representation of weather parameters on *Myrothecium* leaf spot disease incidence and severity in arabica coffee (cv. Chandragiri) - data recorded in 2018 season.

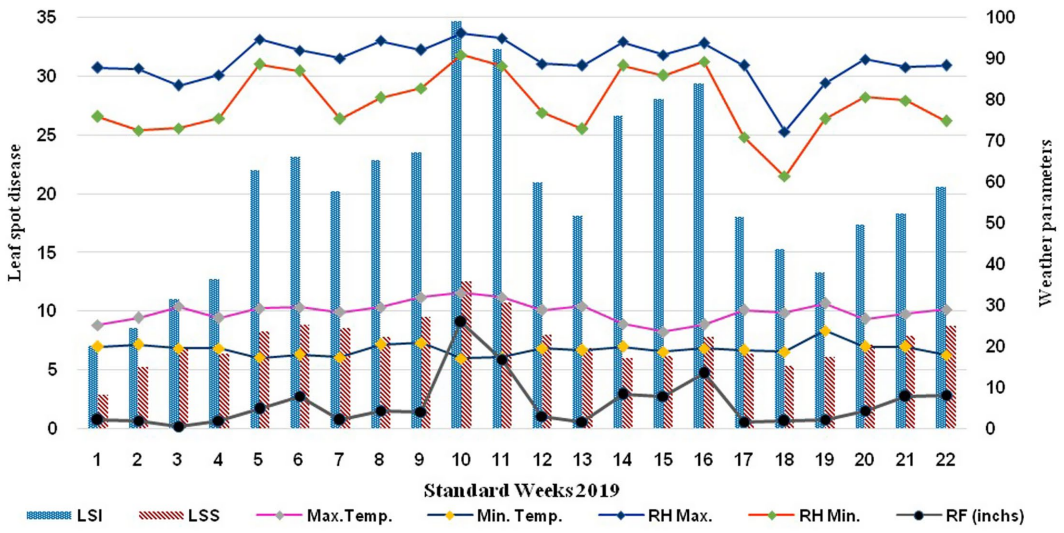


FIGURE 2
Graphical representation of weather parameters on *Myrothecium* leaf spot disease incidence and severity in arabica coffee (cv. Chandragiri) - data recorded in 2019 season.

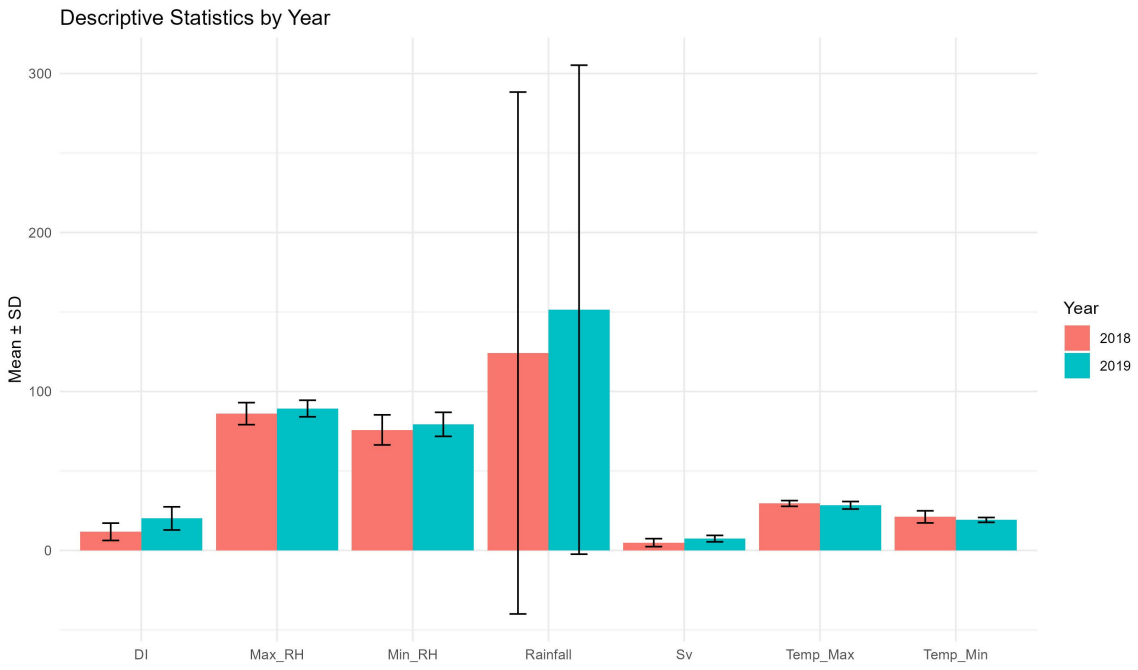


FIGURE 3
Interannual comparison of mean (±standard deviation) values for disease index (DI), spore dissemination potential (Sv), and key weather parameters maximum relative humidity (Max_RH), minimum relative humidity (Min_RH), rainfall, maximum temperature (Temp_Max), and minimum temperature (Temp_Min) for the years 2018 and 2019. The year 2019 exhibited notably higher DI, rainfall, and humidity levels, along with lower minimum temperatures, suggesting that shifts in climatic conditions may have intensified *Myrothecium* leaf spot severity. These patterns align with emerging evidence of climate variability influencing disease dynamics in tropical plantation systems.

RH ($r = 0.74$), indicating that higher precipitation and humidity were significantly associated with increased disease levels. Similarly, Sv was strongly correlated with Rainfall ($r = 0.72$) and DI ($r = 0.84$), suggesting a close link between moisture availability, disease incidence and fungal spread. All reported Pearson correlation coefficients were statistically significant at $p < 0.05$ unless otherwise indicated in Figure 4.

Maximum RH was also positively correlated with DI ($r = 0.66$) and Sv ($r = 0.66$), reinforcing the importance of humidity in disease and spore proliferation. In contrast, Temperature Minimum (T-Min) showed strong negative correlations with Sv ($r = -0.73$), DI ($r = -0.67$) and Rainfall ($r = -0.67$), suggesting that cooler minimum temperatures may promote favorable conditions for disease development.



FIGURE 4
Pearson correlation heatmap illustrating the relationships among weather variables (rainfall, maximum and minimum temperature, maximum and minimum relative humidity), disease index (DI), and spore dissemination potential (Sv) in coffee seedlings. Strong positive correlations were observed between DI and rainfall (r = 0.76), minimum RH (r = 0.74), and Sv (r = 0.84), suggesting that elevated humidity and rainfall conditions increasingly associated with climate variability intensify disease expression and pathogen dissemination. Conversely, minimum temperature was negatively correlated with DI and Sv, indicating a potential suppressive role of cooler night temperatures under warming trends.

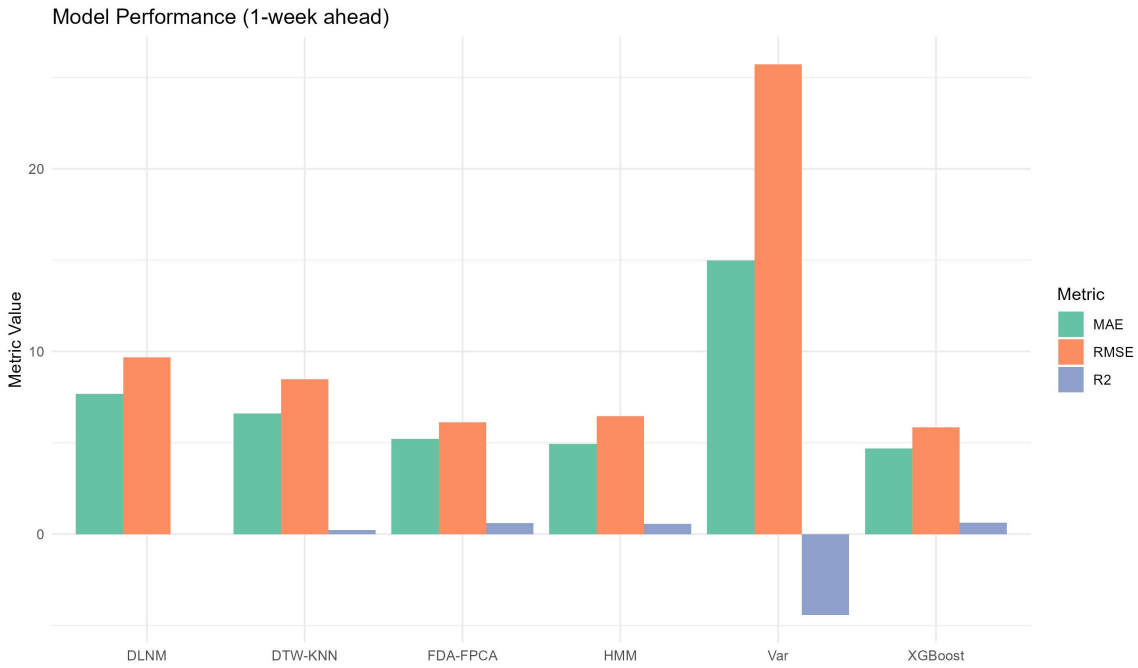


FIGURE 5
Comparison of forecasting model performance for 1-week-ahead prediction of disease index (DI) in coffee seedlings using various metrics: mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). Among the six tested models DLNM, DTW-KNN, FDA + FPCA, HMM, VAR, and XGBoost—XGBoost outperformed others, achieving the lowest error values and the highest R^2 (~0.63), while the VAR model performed poorly with negative R^2. These results highlight the utility of machine learning approaches, particularly XGBoost, in developing accurate, short-term disease forecasts under climate-sensitive conditions, highlighting potential utility for future early warning system development in plantation crop management.

3.4 Model performance

Figure 5 compares the predictive performance of six modeling approaches for 1-week ahead forecasting of disease index (DI), related to *Myrothecium* infection. Performance was evaluated using three metrics Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and the coefficient of determination (R^2).

Among all models, XGBoost achieved the best performance with the lowest RMSE (~5.85) and MAE (~4.70) and the highest R^2 value (~0.63), indicating a strong ability to capture the underlying data structure. However, the moderate coefficient of determination indicates that approximately 37% of the variability in disease incidence remained unexplained. Therefore, the predictive models should presently be interpreted as preliminary forecasting tools rather than operational early-warning systems. The HMM (Hidden Markov Model) and FDA-FPCA (Functional Data Analysis with Functional Principal Component Analysis) followed, with moderate R^2 values (~0.55–0.60) and slightly higher RMSEs (~6.1–6.5).

The DTW-KNN (Dynamic Time Warping with K-Nearest Neighbors) performed less effectively, showing a lower R^2 (~0.23). DLNM (Distributed Lag Nonlinear Model) yielded poor predictive power with R^2 near zero and the VAR model (Vector AutoRegression) exhibited the worst performance, with a negative R^2 (~ $-$ 4.4) and significantly higher RMSE (~25.7) and MAE (~15), indicating it failed to generalize from the data. Confidence intervals (95%) for model performance metrics are provided in Supplementary Table S2.

3.5 Granger causality significance

Granger causality analysis (Figure 6) was conducted to assess the temporal influence of climatic variables on disease incidence (DI) and spore dissemination potential (Sv) associated with *Myrothecium* in coffee. Statistical significance was visualized as $-\log_{10}(p\text{-value})$, with the red dashed line at 1.3 representing the $p = 0.05$ threshold.

Among the predictors, Rainfall showed a highly significant Granger-causal effect on DI [$-\log_{10}(p) \approx 3.4$], indicating strong predictive temporal influence. Maximum RH was the only significant predictor of Sv [$-\log_{10}(p) \approx 2.6$], suggesting that maximum humidity plays a key role in determining spore dissemination dynamics or spore dissemination potential.

Other climatic variables including Minimum RH, Temperature Maximum and Temperature Minimum did not surpass the significance threshold for either DI or Sv. Temperature Maximum was marginally predictive of DI [$-\log_{10}(p)$ just below 1.3], indicating a possible weak or delayed effect. The estimated VAR coefficients and lag structure used in the Granger causality analysis are provided in Supplementary Table S3.

4 Discussion

There is growing evidence that climate change is altering the incidence, distribution and severity of plant diseases globally, including in coffee-based agroecosystems (Garrett et al., 2006; Jaramillo et al., 2011; Bunn et al., 2015; Patil et al., 2025; Bebber et al., 2013; IPCC, 2021). In coffee, shifts in temperature and precipitation patterns have already been linked to changes in pest and disease dynamics, particularly in tropical production systems. The present study further supports this evidence by demonstrating that interannual climatic variability, especially increased rainfall and humidity coupled with lower minimum temperatures, significantly intensified *Myrothecium* leaf spot outbreaks.

In this study, strong positive correlations were observed between disease incidence and both cumulative rainfall and relative humidity,

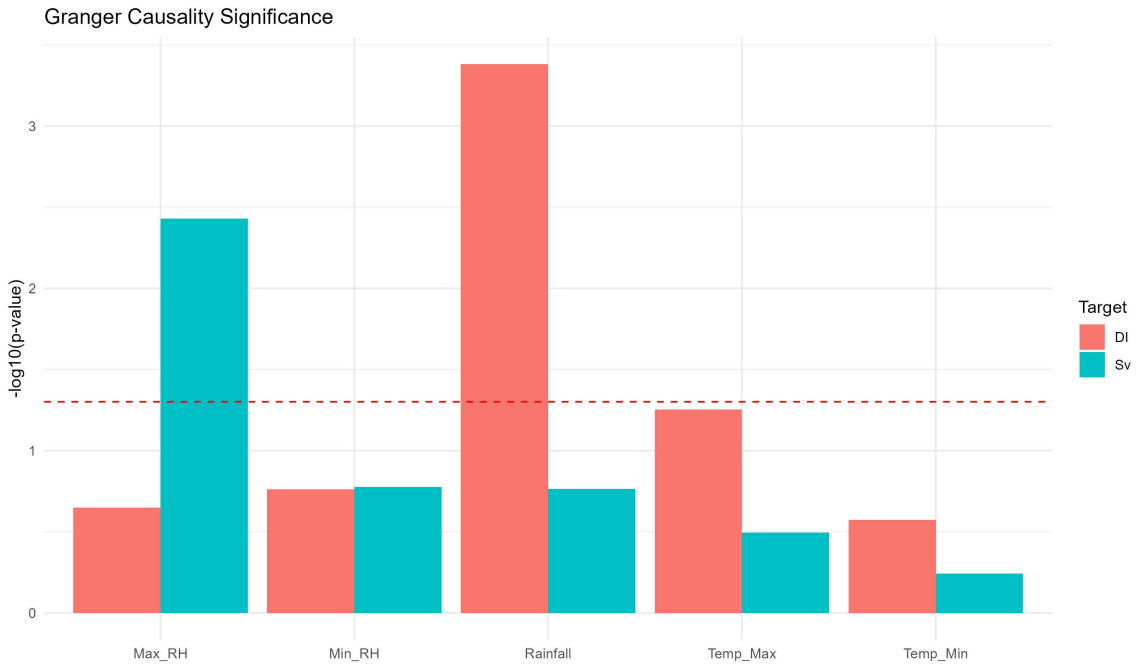


FIGURE 6
Granger causality analysis showing the statistical significance of meteorological variables maximum relative humidity (Max_RH), minimum RH (Min_RH), rainfall, maximum temperature (Temp_Max), and minimum temperature (Temp_Min) as predictors of disease index (DI) and spore dissemination potential (Sv) in coffee seedlings. Bars represent the negative log-transformed *p*-values log10(*p*), with the dotted red line indicating the significance threshold ($p = 0.05$). Rainfall and Max_RH emerged as significant causal drivers for DI and Sv, respectively, indicating that fluctuations in these climate-sensitive variables precede disease outbreaks. These findings support the use of rainfall and humidity as early-warning indicators in climate-responsive disease forecasting models.

while minimum temperature was negatively correlated with disease severity. These patterns are consistent with the notion that warmer, wetter conditions, coupled with cooler nights, create favorable microclimates for *M. roridum* infection and spread. Importantly, the greater disease burden in 2019, a year with anomalously high precipitation, may reflect early indicators of long-term climate-driven shifts in disease epidemiology (Pautasso et al., 2012; Chakraborty, 2005). While correlation analysis provided initial insights into weather-disease relationships, correlation alone does not establish causality. To overcome this limitation, the present study incorporated Granger causality analysis, which identifies temporal cause-effect relationships between climatic variables and disease dynamics. This approach strengthens the inference beyond simple association and supports the identification of rainfall and relative humidity as predictive drivers of disease outbreaks.

The use of machine learning and time series forecasting models especially XGBoost, which outperformed traditional models demonstrates how data-driven tools can enhance our capacity to predict and mitigate climate-related disease risks (Jeong et al., 2016; de Oliveira et al., 2020). The success of XGBoost ($R^2 \approx 0.63$) suggests that models trained on historical weather and disease data can provide short-term disease forecasts to inform nursery-level interventions. In contrast, traditional models like Vector AutoRegression (VAR), which assume linear relationships and stationarity, performed poorly under the dynamic, nonlinear influence of climate variability.

Additionally, Granger causality analysis identified rainfall as a climate-sensitive causal driver of disease incidence, with maximum RH influencing spore dissemination potential (Sv). These results underscore the importance of real-time monitoring of key climate variables, which can be used to trigger preemptive disease control measures under changing climatic regimes (Cooper et al., 2019).

This study suggests that interannual climate variability may contribute to *Myrothecium* outbreaks in coffee seedlings. The integration of high-resolution weather monitoring, predictive modeling and early warning systems could be pivotal in developing climate-smart disease management strategies for plantation crops (Challinor et al., 2014). Future research should focus on expanding this approach to other climatic regions, incorporating long-term datasets and exploring synergistic effects of temperature, CO_2 and humidity under projected climate scenarios. The identification of rainfall and humidity as key drivers enables the development of weather-based early warning systems for coffee nurseries. Such systems can support timely interventions, including optimized fungicide application, improved irrigation management and nursery microclimate regulation.

A key strength of this study is the integration of field-based epidemiological observations with multiple forecasting approaches, allowing comparative evaluation of model performance under real-world conditions. The use of machine learning alongside traditional statistical models provides a comprehensive framework for disease prediction.

However, the study has certain limitations. The dataset is limited to 2 years and a single geographic location, which may affect generalizability. A major limitation of the present study is the absence of independent external validation using geographically or temporally distinct datasets. Consequently, model transferability and operational applicability across different nursery environments remain uncertain and should be evaluated in future multi-location studies. Additionally, while XGBoost performed best, the moderate R^2 value (~0.63) indicates that further improvements are possible by incorporating additional variables such as leaf wetness, soil moisture or pathogen inoculum load.

Several limitations should be acknowledged. First, the study was conducted at a single geographic location over only two growing seasons, limiting broader inference regarding long-term climatic effects. Second, the relatively small dataset ($n = 44$ weekly observations) restricts the robustness of machine learning models and increases the possibility of overfitting. Third, independent external validation was not available. Consequently, the presented forecasting models should be considered proof-of-concept demonstrations rather than deployable operational systems. Additional variables including inoculum load, leaf wetness duration, irrigation practices and nursery management conditions may also have influenced disease development and should be incorporated in future studies.

Future studies should additionally incorporate SHAP or partial dependence analyses to improve interpretability of machine learning models and should evaluate sensitivity across alternative train-test splits and larger datasets. Accordingly, the present study should be interpreted as a preliminary epidemiological and forecasting framework rather than a fully operational disease prediction system.

5 Conclusion

This study provides compelling evidence that the development and progression of *Myrothecium* leaf spot in coffee nurseries are strongly influenced by weather variability particularly, rainfall, humidity and minimum temperature. The substantial increase in disease incidence during 2019 compared to 2018 suggests that short-term climatic fluctuations, associated with short-term weather variability, may be intensifying disease pressure in coffee-growing regions.

By integrating weather monitoring with advanced forecasting models, particularly XGBoost, the study provides preliminary evidence supporting the practical feasibility of predicting disease outbreaks with reasonable accuracy. The identification of rainfall and maximum relative humidity as causal predictors highlights the need for their inclusion in disease surveillance systems. These insights support the development of early warning tools that can enhance nursery-level disease preparedness and reduce seedling losses. A key limitation of this study is the relatively small dataset, which restricts the statistical power and may increase the risk of overfitting in machine learning models. Additionally, weekly observations from the same experimental units are inherently autocorrelated, which may influence statistical estimates. Although Granger causality partially addresses temporal dependence, future studies should incorporate longer time series, multi-location datasets and independent validation to improve model robustness and generalizability.

Despite these limitations, the study provides a proof-of-concept framework demonstrating how weather-driven epidemiological data can be integrated with predictive modeling for disease forecasting in plantation crops. Further, the methodology presented in this study can serve as a foundation for future research and operational disease forecasting systems because it integrates real-time weather monitoring, statistically validated predictors, and scalable machine learning models within a unified framework. This approach is adaptable to other crops and regions, particularly where disease dynamics are strongly influenced by climatic variability. By combining epidemiological understanding with predictive analytics, the framework offers

a robust basis for developing climate-resilient plant disease management strategies.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

AP: Data curation, Supervision, Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization, Visualization, Resources, Validation, Formal analysis. SM: Resources, Visualization, Validation, Supervision, Methodology, Writing – review & editing, Conceptualization. SP: Methodology, Data curation, Resources, Validation, Software, Writing – review & editing, Formal analysis. MG: Writing – review & editing, Formal analysis, Visualization, Validation, Supervision. RN: Writing – review & editing, Supervision. SS: Writing – review & editing, Supervision. CB: Writing – review & editing, Supervision. MS: Writing – review & editing, Supervision.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Acknowledgments

The authors gratefully acknowledge the Coffee Board of India and the Central Coffee Research Institute (CCRI), Chikkamagaluru, Karnataka, INDIA for providing the necessary facilities, institutional support and an enabling environment for the preparation of this article. Their continued encouragement were instrumental in completing this work.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fclim.2026.1733203/full#supplementary-material

References

Alex, T., Ian, C., Nicholas, J. F., Ali, S., and Emily, B. F. (2018). Neural granger causality for nonlinear time series. *arXiv preprint*. doi: 10.48550/arXiv.1802.05842

Anonymous. (2014). *Coffee Guide: A Manual of Coffee Cultivation*. Bengaluru, India: Coffee Board of India.

Anonymous (2023). *Coffee Guide - A Manual of Coffee Cultivation and Processing*. Bengaluru, India: Published by Coffee Board of India, 339.

Avelino, J., Cristancho, M., Georgiou, S., Imbach, P., Aguilar, L., Bornemann, G., et al. (2015). The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. *Food Secur.* 7, 303–321. doi: 10.1007/s12571-015-0446-9

Bebber, D. P., Ramotowski, M. A. T., and Gurr, S. J. (2013). Crop pests and pathogens move polewards in a warming world. *Nat. Clim. Chang.* 3, 985–988. doi: 10.1038/nclimate1990

Bunn, C., Laderach, P., Rivera, O. O., and Kirschke, D. (2015). A bitter cup: climate change profile of global production of Arabica and Robusta coffee. *Clim. Chang.* 129, 89–101. doi: 10.1007/s10584-014-1306-x

Chakraborty, S. (2005). Potential impact of climate change on plant–pathogen interactions. *Australas. Plant Pathol.* 34, 443–448. doi: 10.1071/ap05084

Chakraborty, S., and Newton, A. C. (2011). Climate change, plant diseases and food security: an overview. *Plant Pathol.* 60, 2–14. doi: 10.1111/j.1365-3059.2010.02411.x

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., and Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nat. Clim. Chang.* 4, 287–291. doi: 10.1038/nclimate2153

Chen, T., and Guestrin, C. (2016). "XGBoost: A scalable tree boosting system," In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*, (New York, NY: Association for Computing Machinery), 785–794. doi: 10.1145/2939672.2939785

Coffee Guide (2023). A Manual of Coffee Cultivation, 339.

Cohen, J. M., Civitello, D. J., Venesky, M. D., McMahon, T. A., and Rohr, J. R. (2020). An interaction between climate change and infectious disease drove the worldwide decline of amphibians. *Glob. Change Biol.* 26, 3743–3758.

Cooper, I., Petty, I. T., and McBeath, J. H. (2019). Predictive analytics in plant disease management: challenges and opportunities. *Annu. Rev. Phytopathol.* 57, 223–239.

Daivasikamani, S., Ranjini, A. P., and Raghuramulu, Y. (2016). "Diseases of coffee and their management," in *Plant Pathogens and Their Management*, ed. P. C. Trivedi (Jaipur, Rajasthan, India: Avaishkar Publishers, Distributors), 19–50.

de Oliveira, R. T., Sant'Anna, A. M. O., and Silva, M. A. S. (2020). Predictive models of plant disease occurrence based on weather variables: a systematic review. *Agric. Syst.* 182:102843.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C. M., Carl, G., Carre, G., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46. doi: 10.1111/j.1600-0587.2012.07348.x

Dutta, S. K., Dutta, A. K., Ghosh, M. K., and Nirmal Kumar, S. (2013). Forewarning of *Myrothecium* leaf spot disease of mulberry (*Morus alba* L.) in Malda district of West Bengal and its prophylactic measure. *Sci. Res. Rep.* 3, 204–207.

Elisaye, B., Gebiremariam, M. T., and Salau, A. (2023). Development of a coffee wilt disease identification model using deep learning. *Inf. Med. Unlocked* 42:101344. doi: 10.1016/j.imu.2023.101344

Elith, J., Leathwick, J. R., and Hastie, T. (2008). A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. doi: 10.1111/j.1365-2656.2008.01390.x

Garrett, K. A., Dendy, S. P., Frank, E. E., Rouse, M. N., and Travers, S. E. (2006). Climate change effects on plant disease: genomes to ecosystems. *Annu. Rev. Phytopathol.* 44, 489–509. doi: 10.1146/annurev.phyto.44.070505.143420

Gasparrini, A. (2011). Distributed lag linear and non-linear models in R: the package DLNM. *J. Stat. Softw.* 43, 1–20. doi: 10.18637/jss.v043.i08

Gasparrini, A., Armstrong, B., and Kenward, M. G. (2011). Distributed lag non-linear models. *Statistics in Medicine*, 29, 2224–2234. doi: 10.1002/sim.3940

Gomez, K. A., and Gomez, A. A. (1984). *Statistical Procedures for Agricultural Research*. New York: John Wiley and Sons Inc., 680.

IPCC (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Jaramillo, J., Muchugu, E., Vega, F. E., Davis, A., Borgemeister, C., and Chabi-Olaye, A. (2011). Some like it hot: the influence and implications of climate change on coffee berry borer (*Hypothenemushampei*) and coffee production in East Africa. *PLoS One* 6:e24528. doi: 10.1371/journal.pone.0024528

Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., et al. (2016). Random forests for global and regional crop yield predictions. *PLoS One* 11:e0156571. doi: 10.1371/journal.pone.0156571

Lutkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin, Germany: Springer doi: 10.1007/978-3-540-27752-1.

Martinez, F., Montiel, H., and Martinez, F. (2022). Machine learning model for the diagnosis of coffee diseases. *Int. J. Adv. Comput. Sci. Appl.* 13. doi: 10.14569/IJACSA.2022.01304110

Mayee, C. D., and Datar, V. V. (1986). Phytopathometry, Technical Bulletin-1 (Special Bulletin-3) Marathwada Agricultural University, Parbhani, Maharashtra, India, 95.

McRoberts, N., Hughes, G., and Madden, L. V. (2003). The theoretical basis and practical application of relationships between different disease intensity measurements in plants. *Annals of Applied Biology*, 142, 191–211. doi: 10.1111/j.1744-7348.2003.tb00242.x

Nawaz, M., Nazir, T., Javed, A., Amin, S. T., Jeribi, F., and Tahir, A. (2024). CoffeeNet: a deep learning approach for coffee plant leaves diseases recognition. *Expert Syst. Appl.* 237:121481. doi: 10.1016/j.eswa.2023.121481

Panse, V. G., and Sukhatme, P. V. (1985). *Statistical Methods for Agricultural Workers*. New Delhi, India: ICAR Publications, 359.

Patil, S., Rudragouda, C. S., Govindappa, M., Roobakkumar, A., Machenahalli, S., Jingade, P., et al. (2025). Mitigating climate risks in coffee production through agroforestry: global evidence from a systematic review and meta-analysis. *Front. Clim.* 7:1699037. doi: 10.3389/fclim.2025.1699037

Pautasso, M., Doring, T. F., Garbelotto, M., Pellis, L., and Jeger, M. J. (2012). Impacts of climate change on plant diseases-opinions and trends. *Eur. J. Plant Pathol.* 133, 295–313. doi: 10.1007/s10658-012-9936-1

Petitjean, F., Ketterlin, A., and Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44, 678–693. doi: 10.1016/j.patcog.2010.09.013

Ramsay, J. O., and Silverman, B. W. (2005). *Functional Data Analysis*. 2nd Edn. New York, NY, USA: Springer doi: 10.1007/b98888.

Ranjini, A. P., and Rajanaika (2018). Leaf spot and stem necrosis disease of coffee seedlings caused by *Myrothecium roridum* Tode ex Fr. In India. *J. Mycopathol. Res.* 56, 5–10.

Ranjini, A. P., Uma, M. S., Daivasikamani, S., and Raghuramulu, Y. (2017). Coffee Nursery: Pest and Disease management. *Indian Coffee. LXXXI*, 22–24.

Santoshreddy, M., Ranjini, A. P., Giri, M. S., Sudha, M., and Daivasikamani, S. (2019). Efficacy of fungicides against coffee brown eye spot disease caused by *CercosporaCoffeicola*. *J. Plant Dis. Sci.* 14, 124–128.

Singh, B. K., Singh, J., and Narain, U. (2003). Role of environmental factors for development of *Myrothecium* leaf spot of pigeon pea. *Ann. Plant Prot. Sci.* 11, 401–402.

Sudhakar Bhat, S., Hanumantha, B. T., Daivasikamani, S., Kiran Kumar, K. C., Shyam, S., Sudha, M., et al. (2011). Occurrence of stem necrosis and leaf spot disease of coffee seedlings. *Indian Coffee, LXXV* 7&8, 29–32.

Tomar, D. S., Shastry, P. P., Srivastav, A. K., Nayak, M. K., and Pathak, R. K. (2009). Effect of epidemiological factors on *Myrothecium roridum* of cotton. *J. Oilseeds Res.* 26, 440–441.

Topal, A., Tunga, B., and Tirkolaee, E. (2024). DeepMPR: coffee leaf disease detection with deep learning and enhanced multivariance product representation. *PeerJ Computer Science* 10:e2406. doi: 10.7717/peerj-cs.2406

Velasquez, A. C., Castroverde, C. D., and He, S. Y. (2018). Plant–pathogen warfare under changing climate conditions. *Curr. Biol.* 28, R619–R634. doi: 10.1016/j.cub.2018.03.054

Wheeler, B. E. J. (1969). *An Introduction to plant Diseases*. London: John Wiley and Sons Limited, 301.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag. doi: 10.1007/978-3-319-24277-4

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4:1686. doi: 10.21105/joss.01686