Check for updates

Received: 11 January 2026 | Accepted: 8 June 2026

DOI: 10.1002/agj2.70485

Agronomy Journal

ORIGINAL ARTICLE

Biometry, Modeling, and Statistics

Detecting cotton Verticillium wilt within and across multiple fields with Sentinel-2 imagery and machine learning

Abhasha Joshi[1] | **Thomas Bishop**[1,2] | **Tim Weaver**[1] | **Ratneel Deo**[1] | **Patrick Filippi**[1]

[1]Sydney Institute of Agriculture, School of Life and Environmental Sciences, The University of Sydney, Eveleigh, New South Wales, Australia

[2]Sydney Informatics Hub, The University of Sydney, Newtown, New South Wales, Australia

Correspondence
Abhasha Joshi, Sydney Institute of Agriculture, School of Life and Environmental Sciences, The University of Sydney, Eveleigh, New South Wales 2015, Australia.
Email: abhasha.joshi@sydney.edu.au

Assigned to Associate Editor Vijaya Raj Joshi.

Funding information
SmartSat CRC, Cotton Research and Development Corporation

Abstract

Verticillium wilt, caused by *Verticillium dahliae*, poses a major threat to cotton production worldwide. This study presents a framework for predicting Verticillium wilt incidence across multiple cotton fields by combining Sentinel-2 imagery, preseason inoculum, and weather variables with machine learning models. Three machine-learning models: random forest, support vector machine, and extreme gradient boosting were evaluated using data from seven commercial cotton fields in northern New South Wales, Australia. Model performance was assessed under two validation schemes: fivefold cross-validation to assess within-field interpolation performance and leave-one-field-out validation to evaluate extrapolation across fields. Random forest achieved the highest accuracy (mean overall accuracy = 0.63, mean weighted F1 = 0.64 for interpolation; mean overall accuracy = 0.58, mean weighted F1 = 0.56 for extrapolation) when all data sources were combined. Model interpretability was assessed using Shapley additive explanations, which identified preseason inoculum, near-infrared and red-edge bands are the most influential predictors. Results highlight the advantage of integrating multisource data for disease prediction and the importance of rigorous validation to avoid overestimation of model performance—something that is commonly overlooked in published disease studies. The approach offers a scalable, nondestructive framework for operational monitoring of Verticillium wilt in cotton.

Plain Language Summary

Verticillium wilt is a soil-borne disease that damages cotton plants by blocking the flow of water and nutrients, causing leaf yellowing, wilting, and major yield losses. Reliable large-scale prediction of this disease is important for supporting farm management decisions. This study tested whether combining satellite imagery, soil

Abbreviations: Clr, chlorophyll red edge index; GDD, growing degree days; KNN, K-nearest neighbors; NDVI, normalized difference vegetation index; NIR, near-infrared; RF, random forest; SFS, sequential feature selector; SHAP, Shapley additive explanations; SVM, support vector machine; SWIR, short-wave infrared; UAV, unmanned aerial vehicle; XGBoost, extreme gradient boosting.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2026 The Author(s). *Agronomy Journal* published by Wiley Periodicals LLC on behalf of American Society of Agronomy.

pathogen levels measured before planting, and weather data could improve the prediction of Verticillium wilt in cotton fields. Three machine-learning models were applied to data from seven commercial cotton fields, and their performance was evaluated both within known fields and across new fields. Results showed that prediction accuracy was highest when all three data sources were used together, with a mean weighted F1 score of 0.64. The study also showed that commonly reported validation approaches can lead to inflated accuracy estimates. Overall, the proposed framework provides a scalable, nondestructive approach for operational monitoring of Verticillium wilt in cotton.

1 | INTRODUCTION

Upland cotton (*Gossypium hirsutum*) is a crop of major global economic importance and is the most widely used natural fiber, with an estimated annual production of 26 million metric ton worldwide (Lanclos et al., 2025). In Australia, it ranks as the fourth most valuable agricultural export, contributing $4.9 billion in export revenue in 2022–2023 (Department of Agriculture, Fisheries and Forestry, 2024). Verticillium wilt, a soilborne disease is a major global threat to cotton production. This disease can cause severe reductions in cotton yield and fiber quality, with reported losses of up to 80% (Wei et al., 2015). In Australia, yield losses in infected fields ranged from 10% to 62% (Holman et al., 2016).

Verticillium wilt is caused by the soilborne fungus *Verticillium dahliae* that affects more than 200 agricultural and horticultural crop species worldwide (Pegg & Brady, 2002). *V. dahliae* infects plants by penetrating the root system and spreading in the vascular tissues, where it blocks the transport of water and nutrients from the roots to the foliage. This disruption leads to reduced leaf water content, shrinkage of leaf cells, and chlorophyll degradation, causing symptoms such as vascular discoloration, yellowing and wilting of leaves, leaf mottle and necrosis, and in severe cases complete defoliation depending on the strain (EFSA Panel on Plant Health, 2014). *V. dahliae* is a host-adapted pathogen known for its ability to expand its host range and evolve new pathotypes to overcome plant resistance (EFSA Panel on Plant Health, 2014).

Currently, once established, the control of Verticillium wilt remains highly challenging (Wang et al., 2016). The pathogen is highly persistent and can survive in a dormant life stage in soil for over a decade (NSW Department of Primary Industries, 2023). Nevertheless, prediction of Verticillium wilt is crucial for mitigating its impact. Understanding the variability of disease presence within and across fields is essential for estimating potential yield and quality losses and for supporting informed management decisions, including crop selection for the following seasons.

The conventional approach to detecting Verticillium wilt typically relies on field sampling, involving visual inspection of leaves or cutting stems to assess symptoms, followed by laboratory-based testing to confirm the presence of *V. dahliae* in plant tissue or soil (M C Fadden et al., 2001; Wheeler & Woodward, 2016). Laboratory testing typically includes pathogen isolation and culturing or polymerase chain reaction–based assays for detecting and quantifying pathogen DNA (Usami et al., 2002). Although this approach can provide accurate diagnoses, it is destructive and resource and time intensive, which limits their applicability for large-scale monitoring and understanding of spatial variation within fields.

Risk modeling based on pathogen levels in soil and environmental conditions is another promising strategy. Several studies have attempted to forecast Verticillium wilt risk from canopy to regional scale using key drivers such as weather variables (Requena-Mullor et al., 2020; T. Zhang et al., 2024; T. Zhang, Zheng, et al., 2025), soil pathogen inoculum levels (Bejarano-Alcazar, 1995; Harris & Yang, 1996), and soil type and irrigation practices (Land et al., 2017). However, reliable prediction of risk using these drivers remains challenging, and the difficulty increases when extending prediction across larger areas. This is due to the complex interactions among the multiple environmental and biological factors, coupled with limited data availability. Additional factors such as crop physiology, biophysical conditions, management practices, and cultivar differences complicate efforts to obtain the necessary data and reliably model these relationships, particularly at scale.

Remote sensing provides spatially continuous information on vegetation biophysical characteristics and has emerged as a nondestructive, scalable approach for detecting and monitoring responses of different plant disease. In recent years, several studies have demonstrated the use of remote sensing for detecting, monitoring, and quantifying *Verticillium* wilt in cotton. For instance, Ma et al. (2024) achieve accurate detection of cotton Verticillium wilt severity in a field trial using unmanned aerial vehicle (UAV) multi-

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

spectral data, combining vegetation indices, color indices derived from visible bands, and texture features computed from multispectral bands. Similarly, Li et al. (2024) applied UAV hyperspectral data to distinguish spectral separability between varying degrees of Verticillium wilt severity in a plot trial. A study using hyperspectral data and chlorophyll fluorescence derived from hyperspectral measurements demonstrated promising potential for early detection of Verticillium wilt during the asymptomatic phase. There have also been attempts to integrate UAV and satellite imagery for Verticillium wilt detection in cotton. Nie et al. (2024) combined UAV data with Sentinel-1 and Sentinel-2 satellite imagery to detect Verticillium wilt in cotton fields, showing that while satellites can provide spatially extensive coverage for identifying potentially diseased areas, UAV imagery refines the detection and classification at finer scales.

However, in most remote sensing–based crop disease and stress detection studies including the abovementioned cotton verticillium wilt monitoring studies, both model development and validation are performed on the datasets from the same experimental area often using random train-test splits. This practice overlooks the assumption of independence between training and testing sets, as nearby samples share similar agro-environmental conditions. Evidence from agricultural and ecological remote sensing research shows that random data partitioning leads to overly optimistic estimates of model accuracy (Ferraciolli et al., 2019; Filippi et al., 2025; Karasiak et al., 2022; Meyer & Pebesma, 2022; Stehman & Foody, 2019). Spatial autocorrelation can mask limited model generalization. Alternative validation strategies, such as leave-one-field-out, provide more realistic and trustworthy estimates of generalization error (Roberts et al., 2017; Siegmann & Jarmer, 2015). Since the ultimate goal of these models is to support agronomic decision-making, adopting robust validation schemes is essential to ensure that predictive models perform reliably under real-world conditions.

Furthermore, disease development is governed by complex interactions within the host–pathogen–environment triangle, and reliable prediction ideally requires comprehensive, spatially explicit data for all three components. However, collecting such information at scale is often unfeasible. Integrating remote sensing with key disease drivers such as weather variables and preseason inoculum density offers a practical proxy for improving prediction accuracy. To the best of our knowledge, no previous study has integrated these specific data sources to detect Verticillium wilt in cotton.

This study addresses these gaps by integrating remote sensing data, weather variables and preseason inoculum density to predict Verticillium wilt incidence in cotton across multiple commercial fields in Australia. These data sources collectively represent the environmental, and pathogen components of the disease system. The validation strategies are designed to provide realistic deployment scenarios in the cotton industry, where models are required to perform reliably both within field interpolation and across new, unseen fields for disease forecasting. In addition, Shapley additive explanation (SHAP) analysis was employed to interpret model behavior and understand the key drivers. The specific objectives are (1) to evaluate the predictive performance of different data combinations (remote sensing, weather, and preseason inoculum) using multiple machine learning (ML) models, (2) to test and compare two validation strategies—5-fold cross-validation (5FCV) and leave-one-field-out—in order to quantify within-field interpolation and across-field generalization, and (3) to determine the relative importance of different predictors through SHAP-based interpretability analysis.

2 | METHODOLOGY

2.1 | Study area

The study was conducted across seven irrigated commercial cotton fields (F1–F7) located in northern New South Wales, Australia, a region with a known background presence of Verticillium wilt (Figure 1). The field sizes ranged from 27 to 91 ha, with an average field size of 61 ha. Two fields were irrigated using overhead center-pivot systems, while the remaining five fields were flood irrigated. Data were collected over a 3-year period: Year 1 (2020–2021), Year 2 (2021–2022), and Year 3 (2022–2023). All fields were sampled once, except field F-2, which was sampled in Years 1 and 3. Cotton in this region is typically sown around October and harvested in May of the following year.

Field sampling followed a grid-based strategy to ensure systematic and spatially representative coverage within each field. Each study field was divided into 2-ha spatial grid using QGIS (Graser et al., 2025), with one sampling location established at the center of each grid cell. The total number of sampling locations per field ranged from 14 to 52.

2.2 | Data

The data used in this study are summarized in Table 1 and described in the following subsections. Disease incidence and preseason inoculum data were measured in field following the grid-based sampling plan. Weather and remote sensing data were obtained from publicly available sources. The datasets used in this study have varying resolutions, as shown in Table 1, and were used at their native resolutions.

2.2.1 | Verticillium wilt incidence

The incidence of Verticillium wilt was used as the target variable for model development and testing and was assessed at each predefined sampling grid point across the study fields.

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

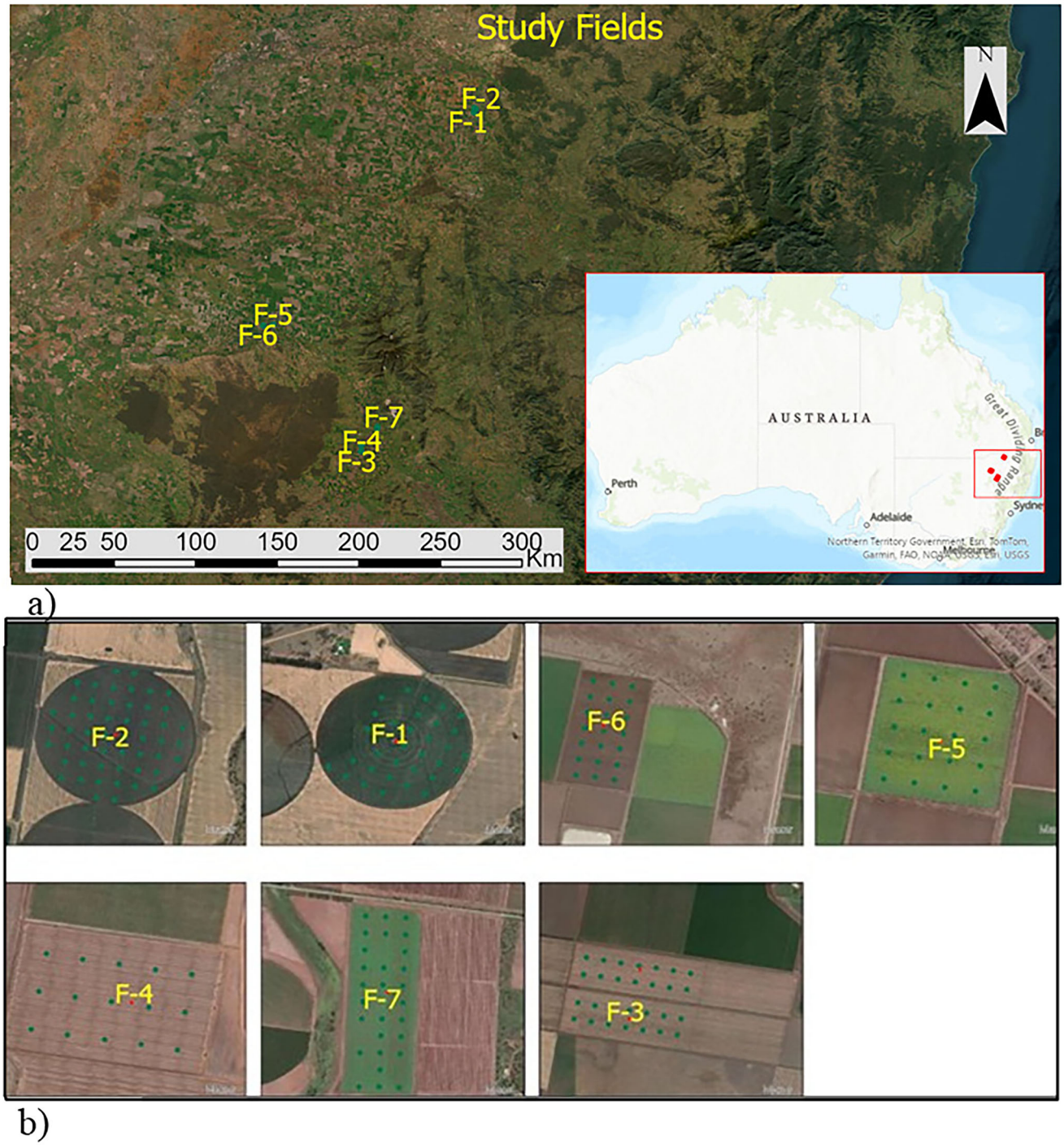


FIGURE 1 Study area. (a) Geographic distribution of the seven cotton fields (F1–F7) located in northern New South Wales, Australia. (b) Distribution of sampling locations within each study field (F1–F7) marked as green dots.

The sampling points established at center of every 2-ha grid were located using GPS. At each sampling site, 40 plants were inspected, with 10 plants selected from each of four adjacent rows covering an approximate area of 3 m × 2 m. Each plant was inspected for characteristic symptoms of Verticillium wilt and proportion of the infected plants at each site was recorded. This disease incidence is the proportion of plants showing symptoms of Verticillium infection within each sampling site.

Field assessments were conducted around the cotton-picking period using stem cuts to confirm the presence or absence of Verticillium wilt symptoms. Based on these proportion, each sampling site was classified into one of four discrete categories. The classification thresholds were defined in consultation with industry experts. Sites with no infected plants were assigned to Class 0. Sites with more than 0% but less than 25% of plants infected were categorized as Class 1, rep-

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Summary of data used in the study.

Category	Data	Approach	Sampling		
Field Data	Incidence	Stem cut	Grid based with grid size of 2 ha		
	Preseason inoculum	qPCR test	Grid based with grid size of 2 ha		
		Source	Spatial resolution	Temporal resolution	Time periods
Weather	Rainfall	SILO (Jeffrey et al., 2001)	~5 km	Daily	June 1–September 30, October 1–March 30
	Growing degree days (GDD)	SILO (for temperature data)	~5 km	Daily	October 1–March 30
	Evapotranspiration (ET)	CMRSET (McVicar et al., 2022)	30 m	Monthly	October 1–January 30 October 1–March 30
Remote sensing	Sentinel 2—spectral bands and vegetation indices	Drusch et al. (2012)	10, 20, and 60 m	5 days	Febuary 15–March 15

Abbreviations: CMRSET, CSIRO MODIS reflectance-based evapotranspiration; qPCR, quantitative polymerase chain reaction.

resenting low incidence. Those with 25% to less than 50% incidence were labeled as Class 2 (moderate incidence), while sites with 50% or more infected plants were assigned to Class 3, indicating high disease incidence. This grouping approach reflects agronomic disease incidence threshold relevant for management decisions.

2.2.2 | Preseason inoculum

Preseason soil inoculum levels were included as an input variable in the model to account for baseline pathogen pressure prior to crop establishment. In commercial cotton production, growers typically measure inoculum through preplant soil sampling, followed by laboratory-based diagnostic assays. This information provides an indication of disease risk and guide field management decisions.

In this study, soil sampling was conducted prior to planting at each predefined sampling grid point, corresponding to those later used for wilt incidence assessment. At each sampling location, soil cores were collected from the top 0–10 cm of soil. Sampling was conducted in two adjacent crop rows, with eight cores per row (16 cores total) taken using a 12-mm hand corer. The 16 cores from each location were bulked into a single composite sample to provide a representative measure of inoculum density for that site.

The collected soil samples were analyzed by Crown Analytical Laboratories using quantitative polymerase chain reaction which detects and quantifies specific pathogen DNA in soil (Dan et al., 2001; Fu et al., 2024). Inoculum population densities were expressed as picograms of pathogen DNA per gram of soil (pg DNA/g).

2.2.3 | Weather data

Weather conditions play an important role in both Verticillium wilt development and cotton growth. The most important variable that contribute to proliferation and spread of Verticillium wilt are temperature and moisture (Arbogast et al., 1999; Y. Zhang, Zhao, et al., 2025; Zhao et al., 2016). To capture these environmental influences, rainfall, evapotranspiration (ET), and growing degree days (GDD) were used. These variables provide insights into moisture levels, irrigation, and temperature across different cotton fields.

Rainfall and temperature data were obtained from the SILO climate database (Jeffrey et al., 2001) (spatial resolution: ~5 km; daily temporal resolution), while ET data were sourced from the CMRSET (CSIRO MODIS reflectance-based evapotranspiration) product (McVicar et al., 2022), which provides 30 m spatial and monthly temporal resolution.

Cumulative rainfall were calculated for two periods: the pre-sowing period (June 1–September 30) and sowing to peak

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

vegetative growth period, as determined by the timing of maximum normalized difference vegetation index (NDVI) (October 1–March 30 of the following year). Similarly, cumulative GDD were calculated for the sowing to peak vegetative growth period (October 1–March 30). GDD was calculated using the new method named "DD 1532," which is specifically designed for Australian cotton-growing regions and improves upon traditional GDD models by providing a more accurate estimate of cotton development (Bange et al., 2022). ET data were aggregated over two growing-season periods: sowing to the early growth stage (October 1–January 31) and sowing to peak vegetative growth stage (October 1–March 30).

2.2.4 | Remote sensing data

Sentinel-2 imagery was used to capture canopy biophysical and physiological responses potentially associated with Verticillium wilt. Sentinel-2 imagery is freely available for any location worldwide. It provides visible, near-infrared (NIR), and short-wave infrared (SWIR) bands, which are effective in detecting verticillium wilt proxies such as such as changes in pigment, canopy water content, or biomass loss in the case of defoliation.

Data was selected from peak vegetative stage, which also corresponds to flowering and boll fill stage of cotton growth. To find the exact date of the image, the greenest pixel composite approach was applied (Z. Zhang et al., 2021). A greenest pixel composite is an image created from a time series of satellite observations in which, for each pixel location, the observation with the highest NDVI value is selected to represent that pixel. Because NDVI peaks when vegetation is most photosynthetically active, the resulting composite effectively corresponds to peak canopy condition. This would both minimize the impact of cloud cover and differences caused by varying crop development stages between farms. A window from February 15 to March 15 was used to prepare the composite. This period was identified using NDVI time series data as the peak vegetative stage across all fields. This approach ensured that imagery from all fields represented a similar phenological stage. A previous study has also suggested that remote sensing data acquired during the flowering and boll-fill stages are particularly sensitive to canopy structural and physiological variations caused by Verticillium wilt (Ma et al., 2024).

A range of vegetation and water stress indices were computed from Sentinel-2 multispectral imagery and used as model input. These indices have been used to assess vegetation vigor, chlorophyll content, and water status, and have been shown to be effective for crop health and stress monitoring. Table 2 summarizes the indices used in this study, and their formula.

2.3 | Feature selection

To identify the most influential predictors for the model, a sequential feature selector (SFS) was implemented using a K-nearest neighbors (KNN) classifier as the estimator. The algorithm followed a backward elimination approach, in which all variables were initially included in the model and subsequently removed one by one. At each iteration the approach removes the least important feature based on cross-validation performance. KNN was used within the SFS process instead of the final classifiers to keep feature selection unbiased. Otherwise, the selected features could be tailored to a specific model's inductive bias, leading to overfitting and unfair comparisons between models.

The initial feature set comprised 22 predictor variables, including 10 Sentinel-2 spectral bands (blue, green, red, red edge 1, red edge 2, red edge 3, NIR, narrow NIR, SWIR 1, and SWIR 2), six derived vegetation indices (NDVI, triple red-edge index, Clr, disease water stress index, normalized difference moisture index, and modified normalized difference water index [MNDWI]), five meteorological variables, and preseason disease inoculum levels. Feature selection performance at each iteration was evaluated using 5FCV. Mean classification accuracy, averaged across folds, was used as the selection criterion to rank and retain features.

2.4 | Classification model

Verticillium wilt incidence was modeled as a classification problem in the study where disease incidence was categorized into four classes based on proportion of infected plant. Three ML algorithms were applied to classify Verticillium wilt incidence: random forest (RF), extreme gradient boosting (XGBoost), and support vector machine (SVM). RF is an ensemble method that constructs multiple decision trees using different bootstrap samples of the dataset and subsets of predictor variables (Breiman, 2001). Predictions are made by aggregating the outputs of all decision trees, which improves generalization. RF has been widely applied in agricultural monitoring, including crop disease and stress detection, due to its robustness and ability to handle complex, nonlinear relationships (Adelabu et al., 2014; Dutta et al., 2024). XGBoost is also a tree-based ensemble method and uses gradient boosting to sequentially build decision trees, with each tree correcting the errors of its predecessors (T. Chen & Guestrin, 2016). By incorporating regularization and efficient optimization, XGB achieves high accuracy, speed, and scalability. SVM is a supervised classification algorithm that identifies the optimal hyperplane to separate classes in a feature space (Cortes & Vapnik, 1995). By applying kernel functions, SVM can effectively model both linear and nonlinear decision boundaries, making it suitable for classi-

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Vegetation indices used in the study.

Indices	Source	Formula
Normalized difference vegetation index (NDVI)	(Tucker, 1979)	$\frac{NIR-Red}{NIR+Red}$
Triple red-edge index (TREI)	(Al-Shammari et al., 2025)	$(RE_2 - RE_1) \cdot (\frac{RE_3}{RE_1+1e-10})$
Chlorophyll red edge (Clr)	(Gitelson et al., 2003)	$\frac{NIR}{RE_1} - 1$
Disease water stress index (DSWI)	(Galvão et al., 2005)	$\frac{NIR+Green}{SWIR+Red}$
Normalized difference moisture index (NDMI)	(Wilson & Sader, 2002)	$\frac{NIR-SWIR_1}{NIR+SWIR_1}$
Modified normalized difference water index (MNDWI)	(Skakun et al., 2003)	$\frac{Green-SWIR_1}{Green+SWIR_1}$

Abbreviations: NIR, near-infrared; RE, red edge; SWIR, short-wave infrared.

fication tasks where class separation is complex (Negrisoli et al., 2022).

To mitigate imbalanced disease class distributions across fields, class-balanced sample weights were applied during model fitting. These weights were computed based on the inverse frequency of each class. During training, samples with higher weights have more influence on the loss function thereby assigning greater importance to underrepresented classes.

All analyses were conducted in Python (version 3.12) using the scikit-learn (Pedregosa et al., 2011) and XGBoost library (T. Chen et al., 2015).

2.5 | Model validation and hyperparameter tuning

Two validation strategies were adopted to evaluate the performance of the proposed approach: (1) 5FCV and (2) one-field-out cross-validation (1FOCV).

In 5FCV scheme, the data from all fields were randomly partitioned into five folds of approximately equal size using a stratified sampling approach to preserve class distribution. At each iteration, one fold was reserved for testing and the remaining folds were used for model training, such that all observations were validated once. The 5FCV strategy assesses the model's interpolation performance within fields because samples from the same field can appear in both the training and test sets.

In 1FOCV scheme, all observation from one field were withheld for testing while the models were trained on data from the remaining fields. This process was repeated until every field had been used once as the validation set. This approach provides an assessment of the ability of the models to generalize between different fields, noting that they collectively cover a large region.

Following feature selection, model hyperparameters were tuned prior to final model training and validation. Hyperparameter tuning was performed using a nested cross-validation approach to avoid information leakage and ensure unbiased performance estimation (Wainer & Cawley, 2021). Specifically, an inner threefold cross-validation was applied exclusively to the training data within each outer 5FCV or 1FOCV split. This inner cross-validation involved a new reshuffling and repartitioning of the outer training data. A grid search was used to evaluate all candidate hyperparameter combinations. For each combination, model performance was assessed by computing the mean scoring metric (accuracy and weighted F1-score) across the three inner folds. The hyperparameter set that achieved the highest mean inner-fold score was selected as optimal. After identifying the optimal hyperparameters, the model was retrained on the entire outer training dataset using these parameters and then evaluated on the held-out outer test fold. This nested design ensures that hyperparameter optimization does not use any test data and provides unbiased estimates of model performance.

2.6 | Evaluation metrics

Model performance was assessed using overall accuracy (OA) and weighted F1-score. OA represents the proportion of correctly classified samples relative to the total number of samples. Weighted F1-score is the harmonic mean of precision and recall, computed for each class and averaged according to number of samples per class. This metric provides a balanced evaluation by accounting for both false positives and false negatives, while also addressing class imbalance in the dataset. Together, these metrics provide a comprehensive evaluation of classification performance. In addition, confusion matrices were used to provide a detailed assessment of classification performance across individual classes. A confusion matrix is a table that summarizes the model's performance where rows represent true classes, columns represent predicted classes, and each entry shows the number of samples assigned to each class combination.

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.7 | Explainability

Explainability methods offer valuable insights into how ML models make decisions, thereby improving transparency and trust in data-driven decision-making. To interpret the predictions generated by the models, this study employed SHAP. SHAP has been increasingly used in the explainable AI domain to develop transparent and interpretable models, including in agricultural modeling (Filippi et al., 2024; Joshi et al., 2025).

SHAP is a post hoc, model-agnostic framework based on the concept of Shapley values (Lundberg & Lee, 2017). Originally introduced in coalitional game theory, Shapley values were developed to fairly distribute reward among players in a cooperative game based on their contribution (Lundberg & Lee, 2017). These values are computed by averaging a participant's marginal contribution across all possible coalitions. In the context of ML, each feature's contribution to the final model output is estimated by averaging its marginal impact across all possible feature combinations. It provides both local and global interpretability. Because evaluating all feature subsets can be computationally intensive, especially for models with a large number of predictors, SHAP reformulates the Shapley value computation as an optimization problem and employs various explainers to efficiently approximate these values. In this study, TreeSHAP was applied. All SHAP values were computed using the SHAP python library.

2.8 | Summary of model, data, and validation experimental setup

The experimental setup involved testing different combinations of data sources, models, and validation strategies to evaluate cotton Verticillium wilt incidence prediction across multiple fields. Three data configurations were considered: (1) remote sensing, weather variables, and preseason inoculum; (2) remote sensing with weather variables; and (iii) weather variables with preseason inoculum. For each configuration, three ML models—RF, SVM, and XGBoost—were trained. The models were evaluated using two validation strategies: 5FCV and 1FOCV. Each experiment was repeated 10 times and mean performance metrics were reported. Model interpretability was further assessed using SHAP analysis on the full dataset to identify the most influential features.

3 | RESULTS

3.1 | Disease incidence across fields

The distribution of Verticillium wilt incidence levels across the fields is shown in Figure 2. Both the number of samples per field and the proportion of incidence classes varied substantially across locations. Field F-2 had the highest number of samples overall, due to its larger field area, and was characterized by a strong dominance of Class 3. Field F-1 also had a relatively large number of samples but exhibited a more balanced distribution across all four incidence classes, including a notable proportion of no incidence class. Across most fields, Class 3 was the dominant category, except in Field F-7, where Class 1 was more prevalent.

3.2 | Spectral reflectance

Figure 3 shows the mean spectral reflectance curves for all fields combined as observed by Sentinel-2A. Differences among disease classes are minimal in the visible and SWIR regions but become more pronounced in the red-edge and NIR regions. As expected, healthy samples exhibit higher reflectance, while diseased samples show consistently lower reflectance, with the separation being most distinct for Class 3 (most diseased class). This pattern is explained by wilting and chlorosis, which reduce chlorophyll content, canopy structure, and overall plant vigor. These factors lower reflectance in the red-edge and NIR wavelengths where healthy vegetation typically reflects strongly.

Figure 4 presents the mean spectral curves averaged for each individual field. In general, similar spectral patterns are observed at the field scale, though not across all fields, with minimal variation in the visible and SWIR regions and greater discrimination between disease classes in the red-edge and NIR regions. Fields F-2, F-4, and F-6 particularly show higher reflectance for healthier samples and lower reflectance for high-incidence samples for NIR and red edge bands. However, this relationship is not consistently observed across all fields. These results indicate that simple models using only remote sensing data are unlikely to reliably distinguish between incidence classes across all fields.

3.3 | Model performance: Interpolation

Figure 5 summarizes the 5FCV results averaged across 10 independent runs for testing model interpolation within the fields using different models and data combinations. OA and weighted F1-score are shown for each model. The bar represents mean performance score of each model while the error bars indicate the corresponding standard deviations, capturing variability across runs. When models were trained using the full set of remote sensing, weather, and preseason inoculum data, all the ML models achieved their highest performance. Among them, RF showed the best overall results (OA = 0.63; weighted F1 = 0.64), closely followed by XGBoost, which exhibited comparable mean OA of 0.6

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

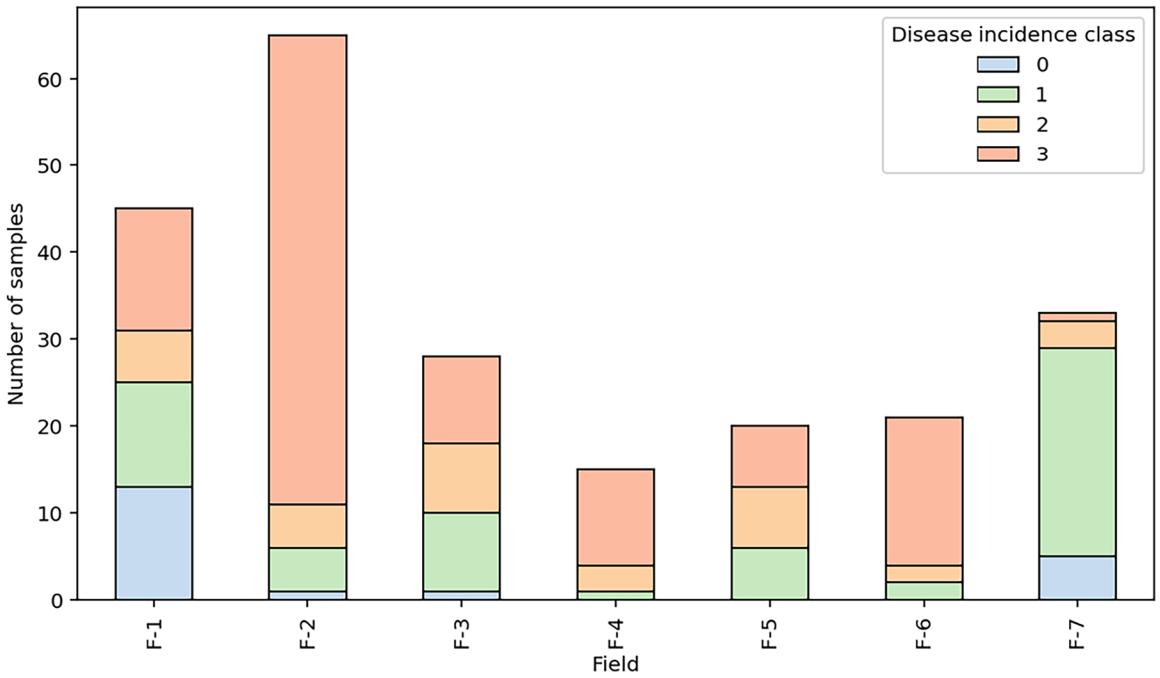


FIGURE 2 Distribution of samples by disease incidence level across seven fields (F-1 to F-7).

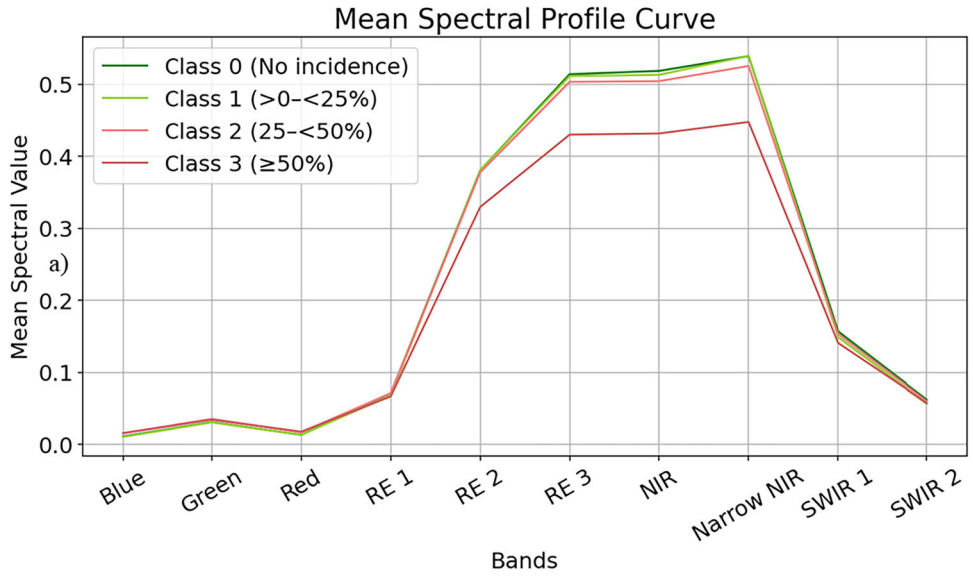


FIGURE 3 Mean spectral reflectance profiles of Verticillium wilt incidence classes across all fields. NIR, near-infrared; RE, red edge; SWIR, short-wave infrared.

and mean weighted F1-score of 0.61. SVM did the poorest with mean OA of 0.51 and mean weighted F1-score of 0.53.

Using only remote sensing and weather data led to a reduction in performance relative to the full data combination (Figure 5b). In this setting, RF and XGBoost continued to outperform SVM, which showed lower OA and weighted F1-scores.

When trained with weather and preseason inoculum data only, RF (OA = 0.60; weighted F1 = 0.62) and XGBoost (OA = 0.59; weighted F1 = 0.61) achieved performance levels comparable to those obtained with the full data set, indicating that these variables alone still provide substantial predictive information. However, SVM again showed limited performance across both metrics. Overall, the within-field interpolation results using 5-FCV indicate a modest decline

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

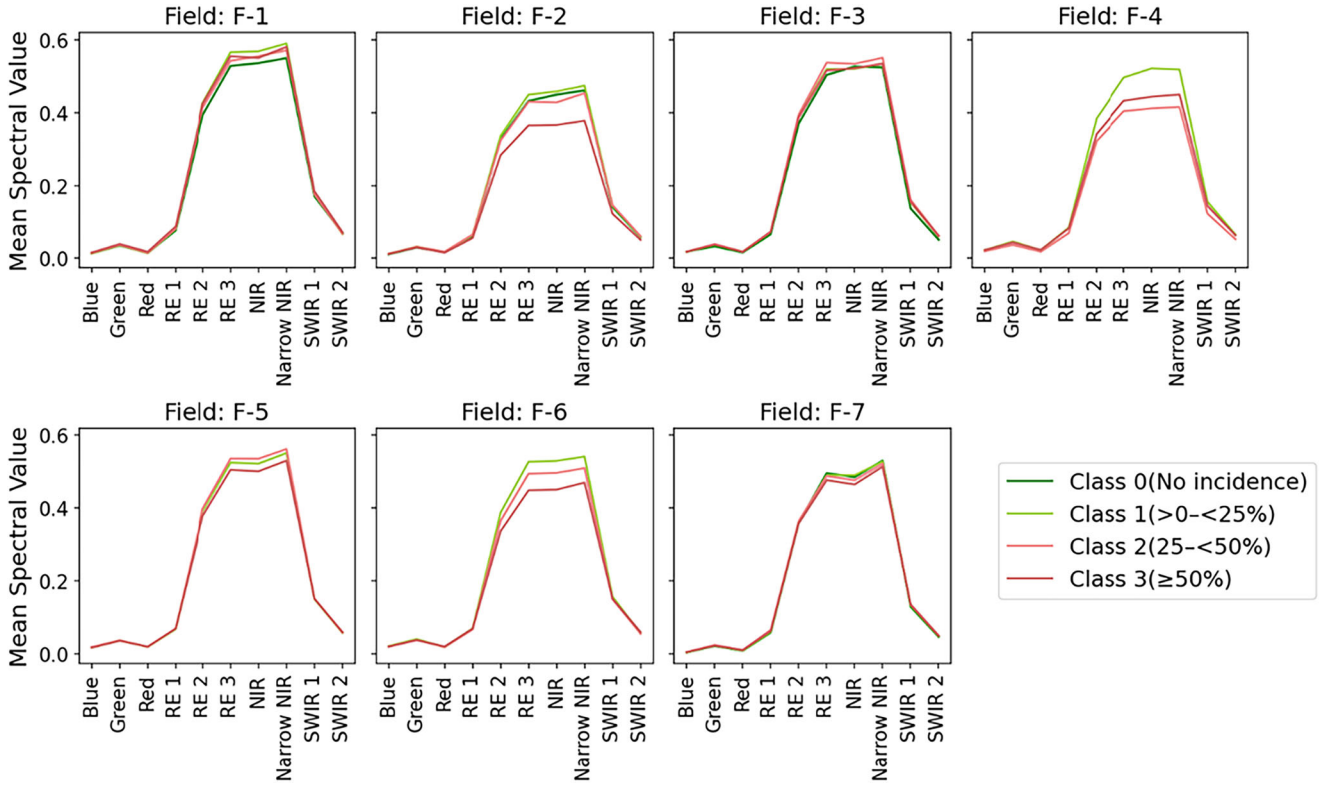


FIGURE 4 Spectral reflectance curves of different Verticillium wilt incidence classes in different fields. NIR, near-infrared; SWIR, short-wave infrared; RE, red edge.

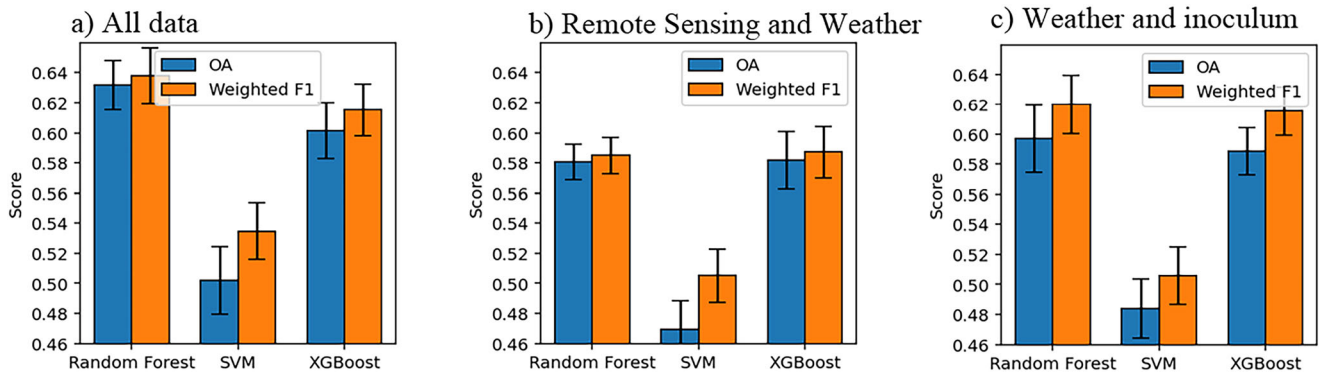


FIGURE 5 Accuracy and weighted F1-score of fivefold cross-validation (5FCV) across different data combinations. (a) Models trained with remote sensing, weather, and preseason inoculum data. (b) Models trained with remote sensing and weather data. (c) Models trained using only weather and preseason inoculum data. OA, overall accuracy; SVM, support vector machine; XGBoost, extreme gradient boosting.

in predictive performance when all three data sources are not used together. Across all data combinations, RF and XGBoost consistently outperformed SVM, achieving higher OA and weighted F1-scores.

The confusion matrices in Figure 6 provide class-wise insight into how each model discriminates between disease severity levels under different input scenarios. Values are normalized by true class, such that diagonal elements represent per-class recall, while off-diagonal elements indicate misclassification patterns. SVM showed particularly low recall for Class 2. In general, most approaches achieved recall values above 0.25, which is the approximates expected recall of a random uniform classifier. The main exception was the SVM model trained without remote sensing data (Figure 6c2), where recall for Class 2 fell well below this threshold (0.12). Overall, the confusion matrices indicated that classification errors were dominated by confusion between neighboring incidence classes. The consistently higher recall for severe

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

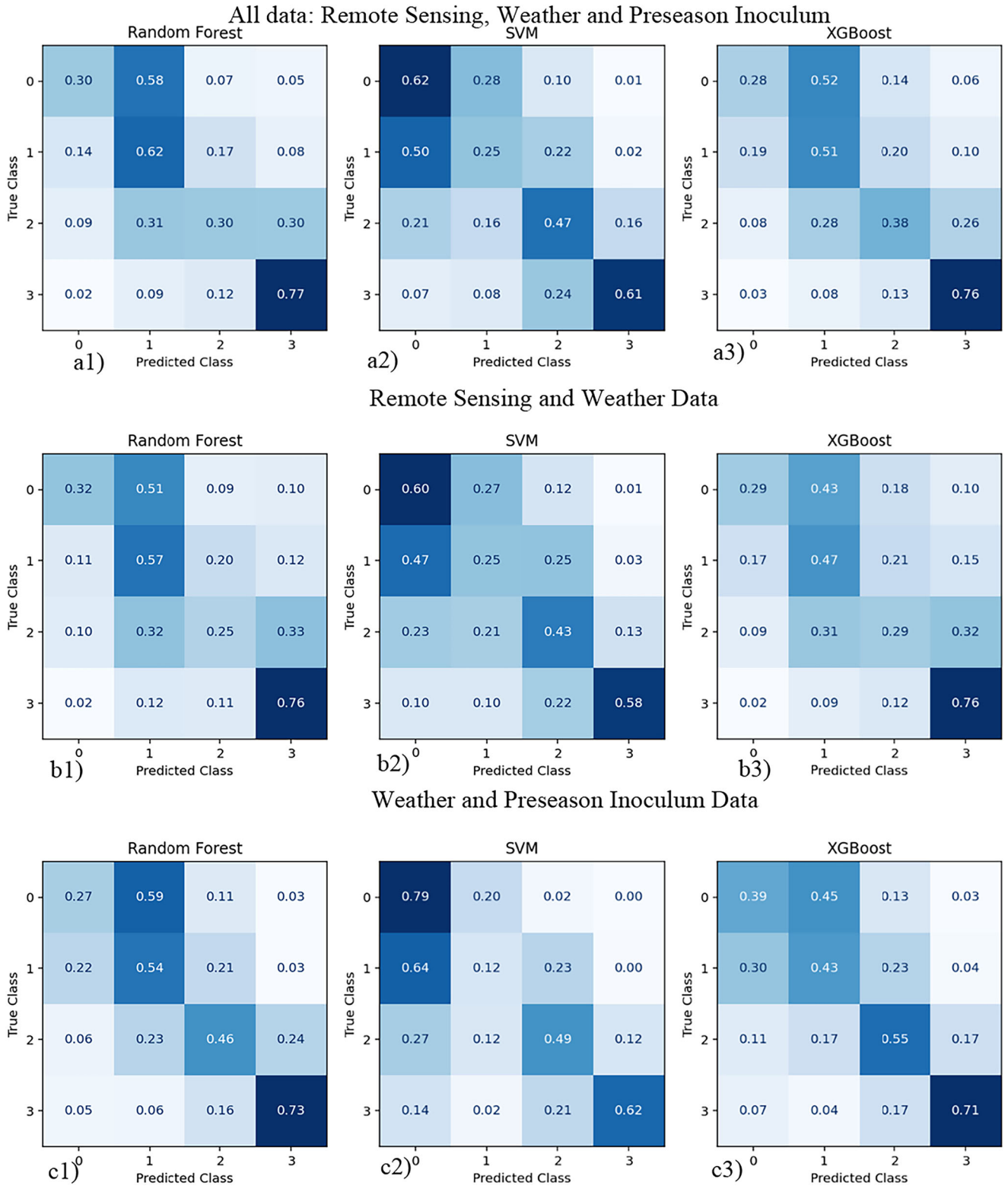


FIGURE 6 Confusion matrices showing the performance of three machine learning models—Random forest, support vector machine (SVM), and extreme gradient boosting (XGBoost)—under different data input scenarios: with remote sensing, preseason inoculum, and weather (a1–a3); without preseason inoculum (b1–b3); and with weather and preseason inoculum (c1–c3).

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

incidence classes achieved by RF and XGBoost highlighted their suitability for operational disease monitoring, particularly where accurate identification of high-risk areas is critical.

3.4 | Model performance: Extrapolation

To evaluate the generalizability of the models between fields, a 1FOCV was conducted where models were trained using data from all but one field and tested on the held-out field, using three input data combinations. The results are presented in Table 3.

RF showed the strongest and most consistent performance across fields for all data combinations, particularly when remote sensing data were included. XGBoost followed a similar pattern but with lower overall performance. SVM models were particularly unreliable for field-level extrapolation. They showed consistently lower and more variable accuracy and weighted F1-scores across fields.

When all data sources were included, RF achieved a mean OA of 0.58 and a mean weighted F1-score of 0.56, while XGBoost reached a mean OA of 0.49 and a mean weighted F1-score of 0.49. Field-wise performance varied widely, for instance for RF model ranging from as low as 0.28 OA and 0.19 weighted F1 (Field 1) to as high as 0.80 OA and weighted F1 (Field 6).

Removing preseason inoculum data reduced classification performance across all models. For RF, average OA dropped to 0.53 and weighted F1 to 0.51. XGBoost also showed declines, with averages OA of 0.42 and weighted F1 to 0.45. These findings suggest that preseason inoculum data provide important discriminative information that enhances classification reliability.

When using only weather and preseason inoculum data, performance was generally weak, RF yielded the best results among the three models but still with limited reliability (0.35 OA accuracy; 0.36 weighted F1). This represents a substantial reduction of more than one-third in both OA and weighted F1-score compared to the approach that includes remote sensing data, highlighting the critical contribution of remote sensing features to model performance. At the field scale, only F-4 and F-7 showed moderate performance all other fields remained below 0.4 in accuracy.

It is also important to note that when remote sensing variables were included, the RF and XGB models achieved F1 scores and overall accuracies above the null uniform benchmark (0.25) for all fields except field F-1.

Overall, the 1FOCV results highlight substantial inter-field variability and confirm that models generalize best when trained with the full set of predictors. Across all data combinations, RF consistently achieved the highest accuracy and F1-scores, followed by XGBoost. The result also underscores the challenge of transferring disease incidence models across heterogeneous fields.

3.5 | SHAP analysis

SHAP values were computed for the RF model, which was identified as the best-performing model. The SHAP summary plot, based on mean absolute SHAP values (Figure 7), highlights the most influential features contributing to model predictions. Among these, preseason inoculum, MNDWI, NIR, red-edge reflectance, chlorophyll red edge index (Clr), and rainfall emerged as the most important predictors of Verticillium wilt incidence. SHAP also enables the exploration of individual features for specific incidence classes. For example, a scatter plot of preseason inoculum values and their corresponding SHAP values for Class 3 is shown in Figure 8. In this plot, the *x*-axis represents preseason inoculum values, while the *y*-axis represents the associated SHAP values. Higher SHAP values indicate a stronger positive contribution to predicting the incidence for Class 3, which is high incidence. Higher preseason inoculum levels are associated with higher SHAP values, indicating a stronger positive contribution to predicting high Verticillium wilt incidence, with evidence of a nonlinear relationship. This underscores its critical role as a driver of Verticillium wilt.

4 | DISCUSSION

This study was undertaken to develop robust prediction of Verticillium wilt in cotton by integrating remote sensing, preseason inoculum, and weather variables, addressing the limitations of field-specific models that are often shown in the literature. The results demonstrated that combining all three data sources yielded the most accurate predictions for within-field interpolation and extrapolation across fields. At the same time, the variation in performance across individual fields highlights the importance of rigorous validation and the need for caution when applying a model to fields different from those in which it was trained.

The spectral reflectance curves (Figure 3) showed a separation in the red-edge and NIR bands with classes of higher disease incidence exhibiting lower reflectance values. SHAP analysis also confirmed the importance of red-edge and NIR bands and related indices (e.g., Clr) in determining incidence classes. These findings are consistent with previous leaf-level studies, where cotton leaves with higher Verticillium wilt severity demonstrated reduced reflectance in the red-edge and NIR regions (700–900 nm) (B. Chen et al., 2012). In that study, reflectance in the visible range (400–700 nm) was reported to increase with severity level. However, in our study at the canopy scale, as captured by Sentinel-2, the

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Results of leave-one-field-out experiment for different machine learning (ML) models and data configurations.

	Remote sensing, weather and preseason inoculum		Remote sensing and weather		Weather and preseason inoculum		
Field	**Accuracy**	**F1-score**	**Accuracy**	**F1-score**	**Accuracy**	**F1-score**	
F-1	0.28	0.19	0.22	0.17	0.25	0.16	RF
F-2	0.69	0.70	0.84	0.78	0.3	0.38	
F-3	0.51	0.48	0.37	0.31	0.36	0.19	
F-4	0.73	0.62	0.73	0.62	0.73	0.62	
F-5	0.46	0.42	0.35	0.30	0.39	0.27	
F-6	0.80	0.80	0.41	0.45	0.27	0.22	
F-7	0.67	0.62	0.58	0.54	0.41	0.45	
Average	0.58	0.56	0.53	0.51	0.35	0.36	
F-1	0.12	0.04	0.15	0.08	0.18	0.09	SVM
F-2	0.18	0.26	0.34	0.38	0.23	0.23	
F-3	0.38	0.28	0.38	0.29	0.35	0.35	
F-4	0.54	0.49	0.48	0.43	0.66	0.66	
F-5	0.19	0.18	0.32	0.28	0.1	0.1	
F-6	0.4	0.49	0.01	0.01	0.49	0.49	
F-7	0.37	0.28	0.69	0.61	0.53	0.48	
Average	0.26	0.29	0.33	0.34	0.32	0.36	
F-1	0.26	0.15	0.15	0.13	0.25	0.17	XGB
F-2	0.44	0.54	0.50	0.57	0.22	0.27	
F-3	0.41	0.32	0.41	0.30	0.36	0.19	
F-4	0.73	0.62	0.71	0.61	0.73	0.62	
F-5	0.47	0.44	0.44	0.37	0.3	0.25	
F-6	0.83	0.81	0.32	0.39	0.69	0.71	
F-7	0.68	0.63	0.59	0.56	0.28	0.26	
Average	0.49	0.49	0.42	0.45	0.34	0.35	

Abbreviations: RF, random forest; SVM, support vector machine; XGBoost, extreme gradient boosting.

1435064s, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

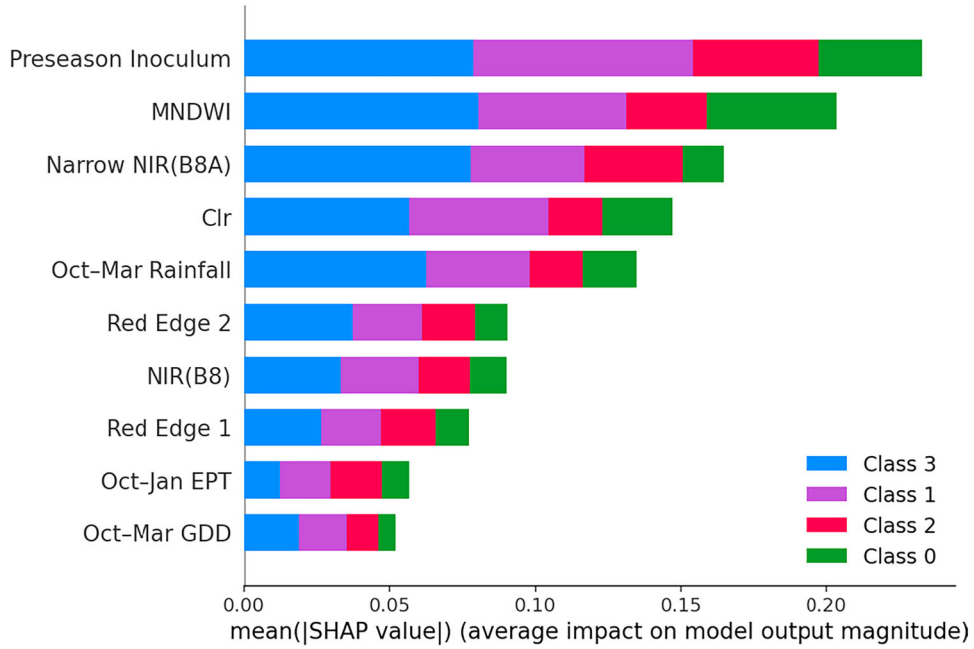


FIGURE 7 Important features as depicted by Shapley additive explanations (SHAP) analysis. The important features include MNDWI (modified normalized difference water index), near-infrared (NIR) and narrow NIR (NIR bands), Clr (chlorophyll red-edge index), red edge 1–2 (Sentinel-2 red-edge bands), Oct–Mar rainfall (cumulative rainfall), Oct–Mar growing degree days (GDD) (cumulative GDD), and Oct–Jan EPT (cumulative evapotranspiration), and preseason inoculum.

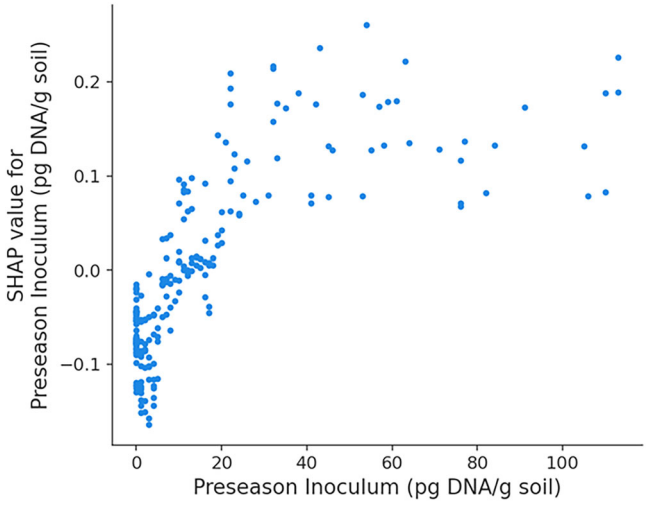


FIGURE 8 Shapley additive explanations (SHAP) dependence plot showing the relationship between preseason inoculum values and their corresponding SHAP values for Class 3.

visible bands (red, green, blue) showed minimal differences between incidence classes. At the field scale (Figure 4), the relationship was not consistent in all the fields likely due to differences in disease severity and its impact on reflectance patterns.

The integration of remote sensing, preseason inoculum, and weather data produced the highest accuracy and F1-scores, whereas excluding either inoculum or remote sensing data reduced performance. The decline in performance was particularly evident when the model was required to generalize to new fields, as shown by the low metrics in 1FOCV (Table 3). Inoculum pressure provides a baseline risk, while weather variables such as rainfall and ET influence pathogen proliferation and plant stress, creating conditions favorable for infection (Arbogast et al., 1999; Y. Zhang, Zhao, et al., 2025; Zhao et al., 2016). Remote sensing adds complementary information by capturing canopy-level physiological responses, including chlorophyll degradation, water stress, and defoliation (J. Zhang et al., 2019). The synergy between these variables explains why combining all data sources yielded the most reliable predictions.

It is also important to note that in 5FCV, higher misclassification of samples from Class 0 (no incidence) to Class 1 (low incidence) was observed (Figure 6). For instance, in the RF model trained with remote sensing, weather, and preseason inoculum data, 58% of samples from Class 0 were misclassified as Class 1. This may be attributed to the smaller number of samples in Class 0 (Figure 2), resulting data imbalance, and the closer resemblance in feature space between Class 0 and Class 1.

Unlike many previous remote sensing–based disease studies that have relied on random train-test splits using data from the same experimental area reporting high predictive accuracy (Ma et al., 2024; Nie et al., 2024)—this study evaluated model performance using a leave-one-field-out validation framework, in which models were trained on a set of fields and

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

tested on a spatially independent field. This distinction is important, as studies that train and validate models using data from the same field implicitly exploit spatial autocorrelation, resulting in inflated performance estimates (Ferraciolli et al., 2019; Filippi et al., 2025; Karasiak et al., 2022; Stehman & Foody, 2019). When assessed under this more realistic and operationally relevant evaluation strategy, predictive accuracy was lower than typically reported in the literature and varied substantially among fields (Table 3). Importantly, this validation design more closely replicates how a real-world decision-support tool would be deployed, where predictions are required for new fields for which no labeled disease observations are available.

The observed variability likely reflects underlying differences in management practices, soil properties, cultivar resistance, severity differences, etc., which were not accounted for in this study and were not clearly captured by the Sentinel 2 data. Expanding the dataset to include more fields, incorporating additional explanatory variables such as those reflecting management practices, and integrating disease severity (extent of disease symptoms expressed in affected plants) alongside incidence may therefore be necessary to better capture this variability. This will be the subject of future research.

Preseason inoculum was identified as an important predictor through feature selection, SHAP analysis, and the observed decrease in prediction accuracy for both within-field interpolation and across-field extrapolation when it was excluded. While Bejarano-Alcazar (1995) reported no significant linear relationship between initial inoculum density and disease incidence, Paplomatas (1992) has shown an exponential relationship between soil inoculum density of *V. dahliae* and Verticillium wilt incidence in susceptible cotton cultivars such as Acala SJ-2. That study also highlighted the role of cultivar, with the tolerant cultivar Acala GC-510 maintaining relatively low incidence across a wide range of inoculum densities. Interestingly, SHAP values for Class 3 in our study also exhibited an exponential relationship with inoculum density (Figure 8). It should be noted that the inoculum–disease relationship is further highly influenced by environmental factors such as temperature and soil moisture (Erreguerena et al., 2019).

The current study highlights both the potential and the challenges of deploying predictive models for operational Verticillium wilt management. On one hand, the integration of inoculum, weather, and remote sensing provides a scalable framework for disease risk mapping at regional levels. On the other, inconsistent performance across fields signals the need for caution in translating models trained in one location to another. Robust validation approaches, such as leave-one-field-out, are therefore essential before applying predictions for decision support. The identified importance of preseason inoculum emphasizes the value of routine soil testing in disease forecasting.

A key strength of this study lies in the use of a multisource dataset from multiple fields to predict Verticillium wilt incidence in cotton, with performance evaluated using both 5FCV (within-field interpolation) and 1FOCV(across-field generalization). This study is among few if not first to integrate the preseason inoculum and weather in disease detection model for cotton. We applied SHAP to add interpretability by providing meaningful insights into the drivers of model predictions. Nonetheless, several limitations remain. First, the dataset is relatively small, with unbalanced representation across incidence classes, which affected model performance. Second, disease severity was not considered; severity is not always proportional to incidence and may provide complementary information for understanding disease dynamics. Third, several important management variables were not included in the modeling framework. Irrigation practices, in particular, may influence disease incidence in ways not captured by the current model. Previous research has shown that disease severity can be significantly greater under non-flooded conditions than under periodic flooding (Sanogo et al., 2008). Detailed information on irrigation scheduling and water application was not available for inclusion in this study. Other management factors, such as crop rotation history and tillage practices, were also not considered, despite their known influence on inoculum levels and disease expression. These omissions may partly explain differences in model performance. In addition, the models may capture signals of general plant stress rather than Verticillium-specific symptoms. Many disease responses might not be unique to this pathogen and can arise from a range of biotic and abiotic stressors. For example, symptoms of *V. dahliae* can closely resemble those caused by other wilt pathogens, such as *Fusarium oxysporum*, the causal agent of Fusarium wilt (EFSA Panel on Plant Health, 2014). Therefore, background knowledge of field conditions is essential for accurate interpretation.

The approach developed in this study provides a scalable and nondestructive framework for large-scale monitoring of Verticillium wilt across both farm and regional levels. Because, the framework relies on freely available remote sensing and weather data, it can be readily implemented. Furthermore, many cotton growers already collect preseason inoculum data through routine soil sampling and diagnostic testing, providing a valuable input that can be easily integrated with these open-source datasets. A decision-support tool built on this framework could enable cotton growers and agronomists to identify high-risk zones early in the season, facilitating more informed management decisions such as crop rotation planning, selection of resistant cultivars, and prioritization of field interventions. In addition, within-season identification of disease risk could help estimate potential yield impacts supporting production planning.

Future work should expand the work by incorporating data from more fields to improve model generalizability.

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Increasing the representation of minority classes will help address class imbalance and enhance prediction robustness. The variation in field-wise performance also underscores the importance of incorporating diverse training data from multiple fields and seasons. Another promising direction is the integration of advanced remote sensing data sources. Hyperspectral (Cogliati et al., 2021; Guanter et al., 2015) and thermal imagery may provide early indicators of crop stress and disease onset, enabling earlier and more accurate detection (Calderón et al., 2013; Raza et al., 2015; J. Zhang et al., 2014).

Furthermore, coupling empirical ML models with radiative transfer models such as PROSAIL could improve the disease detection by linking spectral features directly to physiological parameters affected by Verticillium wilt (e.g., leaf chlorophyll, water content, and leaf area index). This hybrid approach could enhance model generalization across regions and growing conditions (Bai et al., 2025). Similarly, advances in foundational Earth Observation models (Blumenstiel et al., 2025; Zheng et al., 2024) present new opportunities for the disease monitoring. These models, trained on massive and diverse global datasets, enable the transfer of learned representations across environments, provide a foundation for developing disease monitoring models. Finally, extending the analysis to incorporate disease severity alongside incidence would provide a more comprehensive understanding of disease dynamics in cotton fields.

5 | CONCLUSION

This study presents a framework that integrates preseason inoculum, weather, and remote sensing data for predicting Verticillium wilt incidence in cotton using different ML models and evaluates its robustness using multiple validation strategies. Models that incorporated all three data sources achieved better predictive performance for both within field interpolation and extrapolation across most fields than those relying on subsets of these variables. The analysis also confirmed the critical role of spectral information particularly from the NIR and red-edge regions, as well as the importance of preseason inoculum, whose influence on disease incidence followed a nonlinear pattern. Among the evaluated approaches, the RF model, showed the most robust performance under both interpolation and extrapolation scenarios.

This integrated approach enables field-specific assessment of Verticillium wilt incidence, helping growers make informed decisions about yield loss expectations and targeted field management. High-risk areas can be prioritized for intensive treatment practices, while severely affected areas may be temporarily removed from cotton production to reduce inoculum buildup. The proposed approach has strong potential for scaling across diverse cotton-growing regions, particularly where similar multisource datasets are available.

Importantly, the study highlights the need for caution in operational deployment, as models trained and tested within the same field can overestimate the model performance, the evaluation approach that has been commonly adopted in previous studies. In contrast, 1FOCV experiment resulted in relatively lower predictive performance than typically reported in the literature, underscoring the importance of robust validation strategies to have reliable estimate of model generalization. Future work should aim to expand datasets to include more fields and minority classes, integrate disease severity in addition to incidence, and leverage advanced sensing technologies such as hyperspectral and thermal imagery for earlier and more sensitive prediction of Verticillium wilt.

AUTHOR CONTRIBUTIONS

Abhasha Joshi: Conceptualization; data curation; formal analysis; investigation; methodology; software; validation; visualization; writing—original draft; writing—review and editing. **Thomas Bishop**: Conceptualization; funding acquisition; methodology; project administration; resources; supervision; writing—review and editing. **Tim Weaver**: Methodology; resources; writing—review and editing. **Ratneel Deo**: Resources; writing—review and editing. **Patrick Filippi**: Conceptualization; funding acquisition; methodology; project administration; resources; supervision; writing—review and editing.

ACKNOWLEDGMENTS

Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australasian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Abhasha Joshi https://orcid.org/0000-0002-1422-465X

REFERENCES

Adelabu, S., Mutanga, O., & Adam, E. (2014). Evaluating the impact of red-edge band from Rapideye image for classifying insect defoliation levels. *ISPRS Journal of Photogrammetry and Remote Sensing*, *95*, 34–41.

Al-Shammari, D., Whelan, B. M., Wang, C., Bramley, R. G., & Bishop, T. F. (2025). Assessment of red-edge based vegetation indices for crop yield prediction at the field scale across large regions in Australia. *European Journal of Agronomy*, *164*, 127479.

Arbogast, M., Powelson, M., Cappaert, M., & Watrud, L. (1999). Response of six potato cultivars to amount of applied water and *Verticillium dahliae*. *Phytopathology*, *89*(9), 782–788.

Bai, Y., Zarco-Tejada, P. J., Peñuelas, J., McCabe, M. F., Hawkesford, M. J., Atzberger, C., Poblete, T., Kumar, L., Reynolds, M. P., & Nie, C. (2025). Hyperspectral remote sensing for monitoring crop dis-

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ease: Applications, challenges, and perspectives. *IEEE Geoscience and Remote Sensing Magazine*, *14*(1), 96–120.

Bange, M., Nunn, C., Mahan, J., Payton, P., Milroy, S., Finger, N., Caton, J., Dodge, W., & Quinn, J. (2022). Improving temperature-based predictions of the timing of flowering in cotton. *Agronomy Journal*, *114*(5), 2728–2742.

Bejarano-Alcazar, J. (1995). Influence of inoculum density of defoliating and nondefoliating pathotypes of *Verticillium dahliae* on epidemics of Verticillium wilt of cotton in southern Spain. *Phytopathology*, *85*(12), 1474.

Blumenstiel, B., Fraccaro, P., Marsocci, V., Jakubik, J., Maurogiovanni, S., Czerkawski, M., Sedona, R., Cavallaro, G., Brunschwiler, T., & Moreno, J. B. (2025). Terramesh: A planetary mosaic of multimodal earth observation data. In *Proceedings of the computer vision and pattern recognition conference* (pp. 2394–2402). IEEE Xplore.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Calderón, R., Navas-Cortés, J. A., Lucena, C., & Zarco-Tejada, P. J. (2013). High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. *Remote Sensing of Environment*, *139*, 231–245.

Chen, B., Li, S., Wang, K., Zhou, G., & Bai, J. (2012). Evaluating the severity level of cotton Verticillium using spectral signature analysis. *International Journal of Remote Sensing*, *33*(9), 2706–2724.

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). ACM Digital Library.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., & Zhou, T. (2015). Xgboost: Extreme Gradient Boosting (R Package Version 0.4-2) [Computer software]. CRAN.

Cogliati, S., Sarti, F., Chiarantini, L., Cosi, M., Lorusso, R., Lopinto, E., Miglietta, F., Genesio, L., Guanter, L., Damm, A., Pérez-López, S., Scheffler, D., Tagliabue, G., Panigada, C., Rascher, U., Dowling, T. P. F., Giardino, C., & Colombo, R. (2021). The PRISMA imaging spectroscopy mission: Overview and first performance analysis. *Remote Sensing of Environment*, *262*, 112499.

Cortes, C., & Vapnik, V. (1995). Support vector machine. *Machine learning*, *20*(3), 273–297.

Dan, H., Ali-Khan, S., & Robb, J. (2001). Use of quantitative PCR diagnostics to identify tolerance and resistance to *Verticillium dahliae* in potato. *Plant Disease*, *85*(7), 700–705.

Department of Agriculture, Fisheries and Forestry. (2024). *Overview – Agriculture, fisheries, and forestry exports in 2022–23*. Australian Government. https://www.agriculture.gov.au/about/news/overview-aff-exports-2022-23

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, *120*, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Dutta, A., Tyagi, R., Chattopadhyay, A., Chatterjee, D., Sarkar, A., Lall, B., & Sharma, S. (2024). Early detection of wilt in *Cajanus cajan* using satellite hyperspectral images: Development and validation of disease-specific spectral index with integrated methodology. *Computers and Electronics in Agriculture*, *219*, 108784.

EFSA Panel on Plant Health (PLH). (2014). Scientific opinion on the pest categorisation of Verticillium dahliae Kleb. *EFSA Journal*, *12*(12), 3928. https://doi.org/10.2903/j.efsa.2014.3928

Erreguerena, I. A., Rojo, R., Quiroz, F. J., Clemente, G. E., & Escande, A. R. (2019). Relationship between *Verticillium dahliae* inoculum and sunflower wilt in Argentina. *Canadian Journal of Plant Pathology*, *41*(4), 576–584.

Ferraciolli, M. A., Bocca, F. F., & Rodrigues, L. H. A. (2019). Neglecting spatial autocorrelation causes underestimation of the error of sugarcane yield models. *Computers and Electronics in Agriculture*, *161*, 233–240.

Filippi, P., Han, S. Y., & Bishop, T. F. (2025). On crop yield modelling, predicting, and forecasting and addressing the common issues in published studies. *Precision Agriculture*, *26*(1), 8.

Filippi, P., Whelan, B. M., & Bishop, T. F. (2024). Explainable machine learning to map the impact of weather and soil on wheat yield and revenue across the eastern Australian grain belt. *Agriculture*, *14*(12), 2318.

Fu, H., Yang, Y., Jiang, J., Daniels, G. C., Hill, B., Xue, S., Zahr, K., Stellar, L., Harding, M. W., Feindel, D., Bvindi, C., Fernando, W. G. D., Wang, L., & Feng, J. (2024). Development of a qPCR and a lamp assay for *Verticillium longisporum* detection and a triplex qPCR assay for simultaneous detection of *V. longisporum*, *Leptosphaeria biglobosa*, and *L. maculans* from canola samples. *PhytoFrontiers*, *4*(4), 643–650.

Galvão, L. S., Formaggio, A. R., & Tisot, D. A. (2005). Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. *Remote Sensing of Environment*, *94*(4), 523–534.

Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., & Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical Research Letters*, *30*(5), 1248.

Graser, A., Sutton, T., & Bernasocchi, M. (2025). The QGIS project: Spatial without compromise. *Patterns*, *6*(7), 101265.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., … Sang, B. (2015). The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sensing*, *7*(7), 8830–8857.

Harris, D., & Yang, J. (1996). The relationship between the amount of *Verticillium dahliae* in soil and the incidence of strawberry wilt as a basis for disease risk prediction. *Plant Pathology*, *45*(1), 106–114.

Holman, S., Kirkby, K., Smith, L., & Hartnett, H. (2016). *Vert update: The latest in vert research* [Fact sheet]. CottonInfo. https://www.cottoninfo.com.au/sites/default/files/documents/Vert%20update%20%28long%29%20-%20August%202016%20v3.pdf

Jeffrey, S. J., Carter, J. O., Moodie, K. B., & Beswick, A. R. (2001). Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling & Software*, *16*(4), 309–330.

Joshi, A., Pradhan, B., Chakraborty, S., Varatharajoo, R., Alamri, A., Gite, S., & Lee, C.-W. (2025). An explainable Bi-LSTM model for winter wheat yield prediction. *Frontiers in Plant Science*, *15*, 1491493.

Karasiak, N., Dejoux, J.-F., Monteil, C., & Sheeren, D. (2022). Spatial dependence between training and test sets: Another pitfall of classi-

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

fication accuracy assessment in remote sensing. *Machine Learning*, *111*(7), 2715–2740.

Lanclos, K., Meyer, L., Dew, T., Jarrell, P., Soley, G., Galbraith, S., & Bein, O. (2025). *Cotton outlook*. U.S. Department of Agriculture. https://www.usda.gov/sites/default/files/documents/2025AOF-cotton-outlook.pdf

Land, C., Lawrence, K., Burmester, C., & Meyer, B. (2017). Cultivar, irrigation, and soil contribution to the enhancement of Verticillium wilt disease in cotton. *Crop Protection*, *96*, 1–6.

Li, W., Guo, Y., Yang, W., Huang, L., Zhang, J., Peng, J., & Lan, Y. (2024). Severity assessment of cotton canopy verticillium wilt by machine learning based on feature selection and optimization algorithm using UAV hyperspectral data. *Remote Sensing*, *16*(24), 4637.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4766–4777.

Ma, R., Zhang, N., Zhang, X., Bai, T., Yuan, X., Bao, H., He, D., Sun, W., & He, Y. (2024). Cotton Verticillium wilt monitoring based on UAV multispectral-visible multi-source feature fusion. *Computers and Electronics in Agriculture*, *217*, 108628.

M C Fadden, H. G., Chapple, R., De Feyter, R., & Dennis, E. (2001). Expression of pathogenesis-related genes in cotton stems in response to infection by *Verticillium dahliae*. *Physiological and Molecular Plant Pathology*, *58*(3), 119–131.

McVicar, T., Vleeshouwer, J., Van Niel, T., & Guerschman, J. (2022). *Actual evapotranspiration for Australia using CMRSET algorithm*. Terrestrial Ecosystem Research Network.

Meyer, H., & Pebesma, E. (2022). Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications*, *13*(1), 2208.

Negrisoli, M. M., Negrisoli, R., Da Silva, F., Lopes, L. S., Souza Júnior, F. S. D., Velini, E. D., Carbonari, C. A., Rodrigues, S. A., & Raetano, C. G. (2022). Soybean rust detection and disease severity classification by remote sensing. *Agronomy Journal*, *114*(6), 3246–3262.

Nie, J., Jiang, J., Li, Y., Li, J., Chao, X., & Ercisli, S. (2024). Efficient detection of cotton verticillium wilt by combining satellite time series data and multi-view UAV images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 13547–13557.

NSW Department of Primary Industries. (2023). *Verticillium wilt*. NSW Government. https://www.dpi.nsw.gov.au/dpi/climate/climate-vulnerability-assessment/biosecurity-risks2/verticillium-wilt

Paplomatas, E. J. (1992). Incidence of Verticillium wilt and yield losses of cotton cultivars (*Gossypium hirsutum*) based on soil inoculum density of *Verticillium dahliae*. *Phytopathology*, *82*(12), 1417.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Pegg, G. F., & Brady, B. L. (2002). *Verticillium wilts*. CABI.

Raza, S.-A., Prince, G., Clarkson, J. P., & Rajpoot, N. M. (2015). Automatic detection of diseased tomato plants using thermal and stereo visible light images. *PLoS ONE*, *10*(4), e0123262.

Requena-Mullor, J. M., García-Garrido, J. M., García, P. A., & Rodríguez, E. (2020). Climatic drivers of *Verticillium dahliae* occurrence in Mediterranean olive-growing areas of southern Spain. *PLoS ONE*, *15*(12), e0232648.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929.

Sanogo, S., El-Sebai, O. I., & Sanderson, R. (2008). Severity of Verticillium wilt, plant growth, and spectral reflectance indices of chile pepper under periodic flooding and no-flooding conditions. *Hortscience*, *43*(2), 414–419.

Siegmann, B., & Jarmer, T. (2015). Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data. *International Journal of Remote Sensing*, *36*(18), 4519–4534.

Skakun, R. S., Wulder, M. A., & Franklin, S. E. (2003). Sensitivity of the thematic mapper enhanced wetness difference index to detect mountain pine beetle red-attack damage. *Remote Sensing of Environment*, *86*(4), 433–443.

Stehman, S. V., & Foody, G. M. (2019). Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment*, *231*, 111199.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150.

Usami, T., Abiko, M., Shishido, M., & Amemiya, Y. (2002). Specific detection of tomato pathotype of *Verticillium dahliae* by PCR assays. *Journal of General Plant Pathology*, *68*(2), 134–140.

Wainer, J., & Cawley, G. (2021). Nested cross-validation when selecting classifiers is overzealous for most practical applications. *Expert Systems with Applications*, *182*, 115222.

Wang, Y., Liang, C., Wu, S., Zhang, X., Tang, J., Jian, G., Jiao, G., Li, F., & Chu, C. (2016). Significant improvement of cotton Verticillium wilt resistance by manipulating the expression of Gastrodia antifungal proteins. *Molecular Plant*, *9*(10), 1436–1439.

Wei, F., Fan, R., Dong, H., Shang, W., Xu, X., Zhu, H., Yang, J., & Hu, X. (2015). Threshold microsclerotial inoculum for cotton Verticillium wilt determined through wet-sieving and real-time quantitative PCR. *Phytopathology*, *105*(2), 220–229.

Wheeler, T. A., & Woodward, J. E. (2016). Field assessment of commercial cotton cultivars for Verticillium wilt resistance and yield. *Crop Protection*, *88*, 1–6.

Wilson, E. H., & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, *80*(3), 385–396.

Zhang, J., Huang, Y., Pu, R., Gonzalez-Moreno, P., Yuan, L., Wu, K., & Huang, W. (2019). Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and Electronics in Agriculture*, *165*, 104943.

Zhang, J., Yuan, L., Pu, R., Loraamm, R. W., Yang, G., & Wang, J. (2014). Comparison between wavelet spectral features and conventional spectral features in detecting yellow rust for winter wheat. *Computers and Electronics in Agriculture*, *100*, 79–87.

Zhang, T., Xie, Z., Zhou, J., Feng, H., & Zhang, T. (2024). Temperature impacts on cotton yield superposed by effects on plant growth and verticillium wilt infection in China. *International Journal of Biometeorology*, *68*(2), 199–209.

Zhang, T., Zheng, B., Xie, Z., Zhang, T., Feng, H., Zhou, J., & Ouyang, F. (2025). Temperature-driven cotton verticillium wilt: A beta model for risk assessment from laboratory insights to climate scenarios. *Pest Management Science*, *81*(3), 1571–1578.

14530645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Zhang, Y., Zhao, L., Li, D., Li, Z., Feng, H., Feng, Z., Wei, F., Zhou, J., Ma, Z., Yang, J., & Zhu, H. (2025). A comprehensive review on elucidating the host disease resistance mechanism from the perspective of the interaction between cotton and *Verticillium dahliae*. *Journal of Cotton Research*, *8*(1), 5.

Zhang, Z., Wei, M., Pu, D., He, G., Wang, G., & Long, T. (2021). Assessment of annual composite images obtained by google earth engine for urban areas mapping using random forest. *Remote Sensing*, *13*(4), 748.

Zhao, Y.-L., Zhou, T.-T., & Guo, H.-S. (2016). Hyphopodium-specific VdNoxB/VdPls1-dependent ROS-Ca^{2+} signaling is required for plant infection by *Verticillium dahliae*. *PLoS Pathogens*, *12*(7), e1005793.

Zheng, Z., Ermon, S., Kim, D., Zhang, L., & Zhong, Y. (2024). Changen2: Multi-temporal remote sensing generative change foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *47*, 725–741.

How to cite this article: Joshi, A., Bishop, T., Weaver, T., Deo, R., & Filippi, P. (2026). Detecting cotton Verticillium wilt within and across multiple fields with Sentinel-2 imagery and machine learning. *Agronomy Journal*, *118*, e70485. https://doi.org/10.1002/agj2.70485

14350645, 2026, 4, Downloaded from https://acsess.onlinelibrary.wiley.com/doi/10.1002/agj2.70485 by University Of Technology Sydney, Wiley Online Library on [08/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License