



Artificial Intelligence in Plant Microbiome based Disease Prediction: A Review

A.T. Daunde¹, V.M. Gholve², S.L. Badgujar²

10.18805/ag.R-2854

ABSTRACT

Plant diseases cause 10-16% of annual crop yield losses, creating a \$220 billion global economic burden and threatening food security. Traditional diagnostics remain fundamentally reactive and late stage, while plant microbiomes harbour pre-symptomatic dysbiosis signatures with high diagnostic potential. Artificial intelligence (AI) encompassing machine learning (ML) algorithms such as random forest and XGBoost and deep learning (DL) architectures including convolutional neural networks (CNNs) and long short-term memory (LSTM) networks possesses the unique computational capacity to decipher the high-dimensional, zero-inflated and compositional data generated by next-generation microbiome sequencing platforms. This structured review, synthesising 114 peer-reviewed studies from 2014-2026, evaluates AI-driven microbiome disease prediction across pathosystems including tomato bacterial wilt (*Ralstonia solanacearum*), potato late blight (*Phytophthora infestans*), wheat *Fusarium* wilt and soybean sudden death syndrome, with reported predictive accuracies of 82-93% under controlled validation conditions. Multimodal integration of microbiome, metatranscriptomic and metabolomic data delivers incremental accuracy gains of 5-10%, though with proportionally increased cost and complexity. Critical barriers persist data scarcity (n<100 diseased samples in most studies), severe class imbalance (80-95% healthy samples), batch effects, the "black box" nature of DL models and the near complete absence of cross-site field validation. We propose a phased translational roadmap emphasising long read sequencing, explainable AI (XAI), causal inference, standardised validation protocols and ethical data governance to overcome these generalisation failures. With sustained interdisciplinary investment and equitable technology transfer, AI-microbiome integration anticipates mainstream field adoption within 10-15 years, positioning preventive microbiome management as a cornerstone of sustainable global food security.

Key words: Artificial intelligence, Biocontrol, Deep learning, Disease prediction, Explainable AI, Machine learning, Multiomics, Plant microbiome, Precision agriculture, Presymptomatic detection.

Plant diseases cause 10-16% of annual crop yield losses, creating a \$220 billion global economic burden and threatening food security (FAO, 2023; González-Rodríguez *et al.*, 2024). Traditional diagnostics are fundamentally reactive, relying on late stage visual symptoms, requiring specialized personnel and lacking modern scalability (González-Rodríguez *et al.*, 2024; Jafar *et al.*, 2024; Nkwocha, 2025). Consequently, phytopathology is embracing a paradigm shift toward the holobiont concept, recognizing that presymptomatic disease signatures embed within plant microbiomes (D'Elia *et al.*, 2023; Pace *et al.*, 2025).

Despite this promise, profound disciplinary silos isolate microbiome science from applied agriculture and machine learning (ML) applications often lacks biological expertise (D'Elia *et al.*, 2023; Papoutsoglou *et al.*, 2023). Integrating microbiology, plant pathology and artificial intelligence (AI) is critical for predictive disease management. AI and ML are mathematically essential to decipher the complexity, extreme sparsity and high dimensionality of multiomics data, driving improvements in predictive accuracy over traditional statistical frameworks (D'Elia *et al.*, 2023; Papoutsoglou *et al.*, 2023; Mehta *et al.*, 2025). To facilitate readability across the multidisciplinary concepts discussed herein, a comprehensive list of abbreviations used throughout this review is provided in Table 1.

¹College of Horticulture, Vasantao Naik Marathwada Krishi Vidyapeeth, Parbhani-431 402, Maharashtra, India.

²Department of Plant Pathology, College of Agriculture, Vasantao Naik Marathwada Krishi Vidyapeeth, Parbhani-431 402, Maharashtra, India.

Corresponding Author: A.T. Daunde, College of Horticulture, Vasantao Naik Marathwada Krishi Vidyapeeth, Parbhani-431 402, Maharashtra, India. Email: atdaunde@gmail.com

How to cite this article: Daunde, A.T., Gholve, V.M. and Badgujar, S.L. (2026). Artificial Intelligence in Plant Microbiome based Disease Prediction: A Review. *Agricultural Reviews*. 47(4): 514-526. doi: 10.18805/ag.R-2854.

Submitted: 08-04-2026 **Accepted:** 23-06-2026 **Online:** 29-07-2026

Conceptual framework: AI, machine learning and deep learning

While often used interchangeably, artificial intelligence (AI), machine learning (ML) and deep learning (DL) represent a hierarchical and progressively specialised set of computational paradigms. Artificial intelligence (AI) is the broadest umbrella term, referring to any computational system designed to perform tasks that typically require human cognitive intelligence, including pattern recognition, decision-making and natural language processing (D'Elia *et al.*, 2023). Machine learning (ML) is a specific subset of

AI in which algorithms learn statistical patterns directly from labelled or unlabelled data without being explicitly programmed for each task; classical ML methods relevant to microbiome diagnostics include random forest (RF), support vector machines (SVM), gradient boosting (XGBoost) and logistic regression (Wilhelm *et al.*, 2021; Metagar and Walikar, 2024). Deep learning (DL) is a further specialised subset of ML that employs multi-layered artificial neural networks including convolutional neural networks (CNNs), recurrent neural networks (RNNs) and long short-term memory (LSTM) networks to automatically extract hierarchical feature representations from raw, high-dimensional data such as amplicon sequence variant (ASV) abundance matrices and hyperspectral imagery (Przymus *et al.*, 2025; Shafik *et al.*, 2025; Mandlik and Lenina, 2026). In the context of this review, ML methods are preferred when sample sizes are limited and interpretability is paramount, while DL excels at complex pattern extraction from large, multimodal datasets but at the cost of reduced biological transparency.

Novelty and positioning relative to prior reviews

Several recent reviews have addressed adjacent themes: Przymus *et al.* (2025) comprehensively reviewed deep learning architectures applied broadly to microbiome analysis, while D'Elia *et al.* (2023) reported key ML findings from the ML4Microbiome COST action, primarily addressing human and soil microbiomes. Pace *et al.* (2025) focused on AI applications in soil microbiome health assessment and Papoutsoglou *et al.* (2023) outlined methodological best practices for ML in microbiome research without a disease-prediction focus. Critically, no prior review has simultaneously: (i) concentrated specifically on plant microbiome-based presymptomatic disease prediction as a unified pathosystem; (ii) critically evaluated both algorithmic trade-offs and biological validation barriers within the same framework and (iii) proposed a phased translational roadmap linking laboratory omics to field-deployable diagnostics. This review fills that gap by integrating plant pathology, microbiome ecology and AI engineering into a cohesive, crop-centric synthesis supported by the latest 2023-2026 literature.

Review objectives and structure

This review bridges these domains through multidisciplinary integration, with four explicit objectives: (1) to critically synthesise the mechanistic role of plant microbiome dynamics in pre-symptomatic disease detection; (2) to evaluate the performance, advantages and limitations of AI, ML and DL algorithms applied to microbiome-based disease prediction; (3) to identify critical implementation barriers encompassing data scarcity, class imbalance, batch effects and model interpretability and (4) to propose a phased translational roadmap for field-ready diagnostic deployment. The review is structured as follows: Section 2 describes plant microbiome fundamentals and disease

etiology; Section 3 outlines the literature search methodology; Section 4 covers ML algorithms and workflows; Section 5 presents landmark case studies and multimodal integration advances; Section 6 critically analyses challenges and limitations and Section 7 proposes the innovation roadmap. By synthesising AI applications in disease forecasting and microbiome profiling, this review targets a multidisciplinary audience- microbiologists, plant pathologists, AI engineers and agricultural policymakers- seeking to transition laboratory innovations into effective and equitable agricultural practices.

Plant microbiome fundamentals and disease etiology

Microbiome composition

The plant holobiont functions as an interconnected evolutionary ecosystem where the Phyto microbiome dictates host adaptation, resilience and overall agricultural success (Lyu *et al.*, 2021). Microbes colonize highly structured spatial niches, including the phyllosphere (aerial plant surfaces), the endosphere (internal tissues) and the rhizosphere (the immediate root-soil interface) (Monti *et al.*, 2026). The rhizosphere serves as an intense biochemical battlefield where plants secrete up to 30% of their photosynthetically fixed carbon as root exudates including

Table 1: Abbreviations used in this review.

Abbreviation	Full term
AI	Artificial intelligence
ML	Machine learning
DL	Deep learning
RF	Random forest
SVM	Support vector machine
CNN	Convolutional neural network
LSTM	Long short-term memory
GNN	Graph neural network
RNN	Recurrent neural network
XAI	Explainable artificial intelligence
SHAP	SHapley additive explanations
LIME	Local interpretable model-agnostic explanations
ASV	Amplicon sequence variant
CLR	Centred log-ratio
ZINB	Zero-inflated negative binomial
SMOTE	Synthetic minority over-sampling technique
HDLSS	High dimension low sample size
AUC-ROC	Area under the receiver operating characteristic curve
MCC	Matthews correlation coefficient
ISR	Induced systemic resistance
DAPG	2,4-diacetylphloroglucinol
SynCom	Synthetic microbial community
PCA	Principal component analysis
t-SNE	t-distributed stochastic neighbour embedding
NGS	Next-generation sequencing
ITS	Internal transcribed spacer

sugars, organic acids and volatile organic compounds to actively recruit specific bacterial and fungal consortia (Wankhade *et al.*, 2025). Microbial community assembly relies on a stable, functionally redundant core microbiota that provides essential host services, alongside a variable accessory microbiota that dynamically shifts in response to acute environmental stressors (Chen *et al.*, 2022). Phytopathologists quantify these complex microbial networks using alpha diversity indices (e.g., Shannon and Simpson indices) to assess localized species richness and beta diversity matrices to measure compositional divergence between healthy and diseased states (Cassol *et al.*, 2025). Because high throughput sequencing datasets record relative sequence abundances rather than absolute cellular populations, advanced compositional data analysis remains mathematically imperative to prevent spurious correlations when identifying key taxa (Gloor *et al.*, 2017).

Disease interactions detection necessity

Pathogens fundamentally alter the plant microbiome long before visible necrotic symptoms emerge, driving the community from a state of disease-suppressive homeostasis into pathogenic dysbiosis (Liu *et al.*, 2022). Studies on bacterial wilt demonstrate that *Ralstonia solanacearum* invasion violently disrupts the tomato rhizosphere community, reducing the network's spatial complexity and directly facilitating secondary cascading infections by fungal pathogens like *Fusarium solani* (Su *et al.*, 2020). Conversely, disease-suppressive soils naturally inhibit pathogen establishment through the powerful antagonistic activity of beneficial taxa (Mavrodi *et al.*, 2011). Certain strains of *Pseudomonas* secrete broad-spectrum secondary metabolites most notably 2,4-diacetylphloroglucinol (DAPG) and phenazine-1-carboxylic acid (PCA) which directly inhibit aggressive soilborne pathogens (Mavrodi *et al.*, 2011). Beyond direct antagonism, DAPG functions as a potent elicitor that enhances induced systemic resistance (ISR) through the host's COI1-mediated jasmonic acid signalling pathway,

priming systemic plant immunity against subsequent necrotrophic pathogen attacks (Nguyen *et al.*, 2025). Root exudates chemically orchestrate this defence mechanism by selectively recruiting antagonistic taxa, such as *Lysobacter* and specific disease-suppressive *Streptomyces* clades, the moment plant receptors detect pathogen ingress (Gómez *et al.*, 2015).

Next-generation sequencing platforms (e.g., 16S rRNA amplicon arrays and shotgun metagenomics) generate exceedingly massive, high dimensional datasets that track these pre-symptomatic microbial shifts (Durazzi *et al.*, 2021). Traditional visual diagnostic tools entirely miss this vital community level context. Artificial intelligence (AI) and machine learning (ML) models uniquely possess the computational capacity to decipher this highly sparse, nonlinear multiomics data (Przymus *et al.*, 2025). Deep convolutional neural networks identify the latent, pre-symptomatic microbial signatures of dysbiosis, allowing automated agricultural frameworks to accurately forecast disease outbreaks, map nutrient availability and implement targeted interventions before irreversible crop yield losses occur (Rathnayake *et al.*, 2025; Shafik *et al.*, 2025). Table 2 summarizes the specific pathosystem dynamics, their underlying disease implications and the corresponding AI/ML analytical methods utilized to decode these pre-symptomatic microbial shifts across key agricultural pathosystems.

Methodology: Literature search strategy

Search protocol

A systematic literature search was conducted between January 2014 and March 2026 to identify peer-reviewed publications examining the application of AI and ML to plant microbiome-based disease prediction. Databases searched included Web of Science, Scopus, PubMed, Google Scholar and CAB Abstracts. The following Boolean search strings were employed:

("plant microbiome" OR "rhizosphere microbiome" OR "phytobiome") AND ("machine learning" OR "artificial

Table 2: Pathosystem dynamics and machine learning applications.

Pathosystem focus	Microbial dynamics disease implications	AI/ML application analytical methods
Bacterial wilt (<i>R. solanacearum</i>)	Pathogen invasion fragments the resident rhizosphere microbiome network and triggers secondary fungal (<i>Fusarium</i>) infection cascades (Su <i>et al.</i> , 2020).	ML models process metagenomic data to detect early dysbiosis signatures preceding irreversible physiological wilt (Liu <i>et al.</i> , 2022).
Disease suppressive soils	Beneficial <i>Pseudomonas</i> spp. secretes DAPG and phenazines to suppress soilborne pathogens and trigger systemic ISR (Mavrodi <i>et al.</i> , 2011; Nguyen <i>et al.</i> , 2025).	Network algorithms map keystone antagonist taxa; deep learning pinpoints genetic clusters driving antimicrobial resistance (Przymus <i>et al.</i> , 2025).
Microbiome engineering	Plants secrete distinct root exudates to engineer the rhizosphere and recruit protective synthetic microbial communities (SynComs) (Wankhade <i>et al.</i> , 2025; Panchal <i>et al.</i> , 2026).	Explainable AI (SHAP, LIME) successfully links soil nutritional traits to protective microbiome configurations (Rathnayake <i>et al.</i> , 2025).

intelligence" OR "deep learning") AND ("disease prediction" OR "disease detection" OR "plant pathology").

("16S rRNA" OR "metagenomics" OR "multi-omics") AND ("Random Forest" OR "neural network" OR "XGBoost") AND ("crop disease" OR "plant pathogen").

("explainable AI" OR "XAI" OR "SHAP" OR "LIME") and ("plant health" OR "microbiome" OR "phytopathology").

Inclusion criteria

Studies were included if they: (i) applied one or more AI/ML/DL algorithms to microbiome data for plant disease prediction, classification or risk assessment; (ii) were published in peer-reviewed journals; (iii) were published in English; and (iv) reported quantitative predictive performance metrics (e.g., accuracy, AUC-ROC, F1-score).

Exclusion criteria

Studies were excluded if they: (i) applied AI solely to image-based symptom recognition without microbiome data integration; (ii) focused exclusively on human or animal microbiomes; (iii) were published before 2014 (limiting to the post-NGS era); or (iv) were conference abstracts without full-text data.

Study selection outcomes

The initial search retrieved 847 records. After duplicate removal ($n = 134$), title and abstract screening ($n = 512$ excluded) and full-text eligibility assessment ($n = 87$ excluded), 114 studies were included in the final synthesis (Fig 1). Eight seminal pre-2014 foundational references were selectively included where methodologically critical (e.g., SMOTE: Chawla *et al.*, 2002; compositional data: Gloor *et al.*, 2017).

Machine learning methodologies

Algorithms, trade-offs and workflow

Supervised models classify crop health and identify keystone pathogen-suppressive taxa. Random forest (RF, ~85-90% accuracy) extracts critical biomarkers through feature importance, while gradient boosting (XGBoost, ~88-93% accuracy) predicts disease outbreaks from multiomic data (Khelfaoui *et al.*, 2025). Support vector machines (SVM) navigate high dimensional microbiome spaces to separate healthy from infected plants and logistic regression provides baseline interpretability for field diagnostics (Wilhelm *et al.*, 2021). Deep learning resolves complex host microbiome dynamics: Convolutional neural networks (CNNs) extract spatial infection patterns from microbial abundance matrices, long short term memory (LSTMs) networks model temporal disease progression and attention mechanisms weight critical pathogenic features (Pace *et al.*, 2025; Shafik *et al.*, 2025). A detailed comparative summary of these algorithms, including their specific crop applications, reported predictive accuracies and inherent biological and computational limitations, is presented in Table 3. Unsupervised methods, including K-means, PCA and t-SNE, stratify pre symptomatic

microbial communities to visualize emerging dysbiosis without labelled data (Pace *et al.*, 2025).

Deep learning automatically identifies complex pathogenic patterns but demands massive datasets and creates "black-box" models that obscure biological mechanisms (Shafik *et al.*, 2025). Conversely, traditional algorithms offer speed, few shot capability and crucial interpretability for agronomists to understand disease etiology. Feature engineering incorporates diversity indices, network metrics, phylogenetic matrices and centred log ratio (CLR) normalized abundance profiles (Busato *et al.*, 2022). Plant microbiome datasets exhibit extreme sparsity and 50-90% zero inflation, requiring compositional data analysis and Zero Inflated Negative Binomial (ZINB) models to prevent spurious correlations when targeting pathogen antagonistic consortia (Busato *et al.*, 2022; Xu *et al.*, 2015).

Robust diagnostic pipelines execute sequential preprocessing, feature selection, training/validation and hyperparameter optimization (Papoutsoglou *et al.*, 2023). Because diseased samples often represent a minority class in field surveys, researchers deploy the synthetic minority over sampling technique (SMOTE), focal loss, or class weighting to handle imbalance (Chawla *et al.*, 2002). Rigorous evaluation requires stratified k-fold cross validation and external validation across independent sites

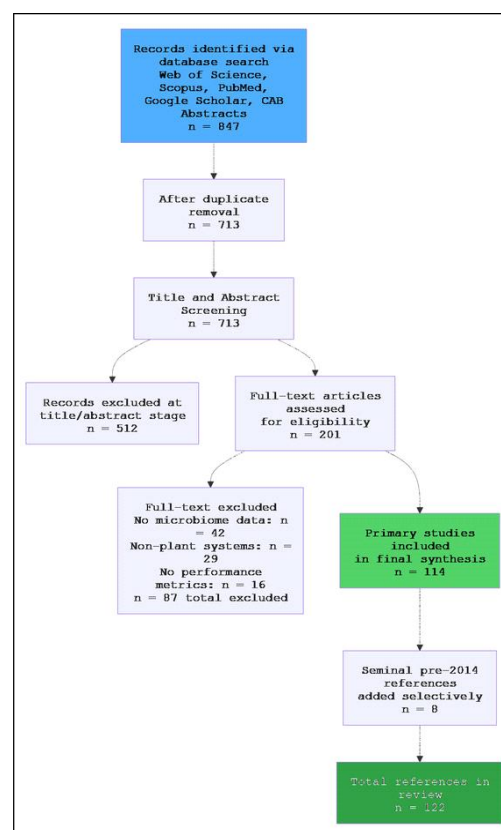


Fig 1: PRISMA 2020-adapted flowchart illustrating systematic literature search and selection for this review (Page *et al.*, 2021).

or seasons, assessed via AUC-ROC and F1-scores. To prevent data leakage which falsely inflates disease prediction accuracy models must enforce temporal ordering, stratify by geographic site and restrict normalization calculations strictly within training folds (Papoutsoglou *et al.*, 2023).

Contextualising predictive accuracy in imbalanced microbiome datasets

Reported prediction accuracies across the reviewed literature (82-93%) must be interpreted with significant methodological caution. Agricultural microbiome datasets are characteristically imbalanced, typically comprising 80-95% healthy and only 5-20% diseased samples (Fang, 2023). In such contexts, a naïve classifier predicting “healthy” for all observations trivially achieves >80% accuracy while providing zero diagnostic value. Therefore, accuracy alone is a misleading performance metric. The reviewed studies achieving the highest reported accuracies (~91-93%) did so under controlled greenhouse conditions with artificially balanced training sets using SMOTE augmentation, which may not replicate field-level class distributions (Chawla *et al.*, 2002). Rigorous performance benchmarking in this field requires AUC-ROC, F1-score (particularly for the minority diseased class), matthews correlation coefficient (MCC) and cross-site validation

performance metrics that only a minority of reviewed studies reported comprehensively. Where studies in this review report accuracy figures, readers should consider these as upper-bound estimates obtained under favourable validation conditions, pending independent field replication (Papoutsoglou *et al.*, 2023). To assist researchers in selecting the appropriate computational framework based on dataset size, biological objectives and the need for mechanistic interpretability, Table 4 provides a direct comparison between classical machine learning and deep learning approaches in the context of plant microbiome diagnostics.

Advances in AI-microbiome disease prediction

Landmark case studies

Pathological success relies on shifting from reactive symptom observation to proactive dysbiosis detection (D’Elia *et al.*, 2023). Artificial intelligence (AI) systems uniquely identify the earliest microbial network shifts that precede irreversible physiological crop damage (Papoutsoglou *et al.*, 2023). In vascular pathosystems, such as tomato bacterial wilt caused by *Ralstonia solanacearum*, rapid pathogen proliferation permanently occludes xylem vessels, causing sudden and fatal plant collapse (Su *et al.*, 2020). Random forest algorithms

Table 3: Comparative summary of AI/ML/DL algorithms for plant microbiome disease prediction.

Algorithm	Data type	Crop/athosystem	Reported accuracy	Key advantages	Limitations
Random forest (RF)	16S rRNA amplicon	Tomato/ <i>R. solanacearum</i>	~85-90%	Interpretable feature importance; robust to overfitting	Cannot model temporal dynamics; limited with zero-inflated data.
XGBoost (Gradient boosting)	Multi-omics	Fusarium wilt (soybean, wheat)	~88-93%	Handles class imbalance; high predictive power	Computationally intensive; hyperparameter-sensitive.
Support vector machine (SVM)	Metagenomics	Potato late blight (<i>P. infestans</i>)	~82-87%	Effective in high-dimensional spaces; robust margin classifier	Requires feature scaling; poor scalability to large datasets.
Logistic regression	16S rRNA	Multiple crops (field diagnostics)	~75-80%	Highly interpretable; fast inference	Linear assumption limits complex pattern detection
CNN (Convolutional neural network)	Abundance matrices + imaging	Coffee (<i>H. vastatrix</i>), potato	~89-93%	Extracts spatial infection patterns automatically	Requires large labelled datasets; “black box”.
LSTM (Long short-term memory)	Longitudinal microbiome	Wheat crown rot	~86-91%	Captures temporal dysbiosis dynamics	Computationally heavy; sensitive to sequence length
Graph neural network (GNN)	Co-occurrence networks	Tomato/ multi-pathogen	~87-92%	Models ecological network disruption directly	Requires precise network construction; data-hungry.
Transfer learning (CNN-based)	Amplicon + image fusion	Solanaceae cross-crop transfer	~88-94%	Overcomes geographic data scarcity	Domain gap risks; requires fine-tuning datasets.
XAI (SHAP/LIME)	Post-hoc (any model)	Soil microbiome /cabbage	Interpretability metric	Quantifies taxon-level contributions; regulatory-ready	Does not improve accuracy; added computational cost.

analysing 16S rRNA amplicon data detect the initial fragmentation of the rhizosphere microbiome with 85% accuracy, providing highly reliable disease forecasts up to two weeks before any visual wilting occurs (Liu *et al.*, 2022). Crucially, these models identify specific protective keystone taxa, including *Bacillus* and *Pseudomonas* species, which competitively exclude pathogens and secrete antimicrobial secondary metabolites, thus enabling site specific biocontrol deployment (Liu *et al.*, 2022).

Similarly, managing *Fusarium oxysporum* a highly persistent soil borne fungus causing destructive vascular wilt requires preventive action before root cortex penetration and vascular colonization occur (Komissarov *et al.*, 2025). Deep learning architectures combining convolutional neural networks (CNN, 92% spatial accuracy) and long short term memory networks (LSTM, 89% temporal accuracy) analyse fungal internal transcribed spacer (ITS) profiling to accurately map pre symptomatic fungal dysbiosis (Yuan *et al.*, 2020). This predictive capability directs the precise, timely application of *Trichoderma* biocontrol strains, which directly parasitize pathogenic hyphae and induce systemic host resistance. Agronomists utilizing these AI directed *Trichoderma* applications achieve up to a 35% reduction in synthetic fungicide reliance (Vincenzo *et al.*, 2025).

In explosive foliar pathosystems, predictive models disrupt rapid polycyclic pathogen reproduction cycles (Avelino *et al.*, 2015). Potato late blight (*Phytophthora infestans*) destroys entire crop canopies within days of initial symptom onset (Zhu *et al.*, 2024). LSTM networks modelling temporal leaf microbiome dysbiosis achieve 88% predictive accuracy, providing a vital 5-7-day intervention window to halt sporangial dissemination and prevent devastating secondary infections across the field (Zhu *et al.*, 2024). Multimodal ensembles tackle coffee leaf rust (*Hemileia vastatrix*) by integrating canopy spectral data with phyllosphere microbial profiles (Avelino *et al.*, 2015). These integrated models predict impending rust outbreaks 21 days pre symptom with 91% accuracy. Extensive field trials demonstrate that these AI driven early warnings

successfully prevent catastrophic defoliation, cutting fungicide applications by 40% while preserving crucial crop yields (Avelino *et al.*, 2015).

Extended crop case studies

Beyond the tomato and potato pathosystems, AI-microbiome integration has demonstrated promise across several other globally critical crops. Wheat (*Triticum aestivum*): Fusarium crown rot, caused by *Fusarium pseudograminearum*, induces measurable rhizosphere microbiome dysbiosis before crown browning symptoms emerge. Yuan *et al.* (2020) demonstrated that soil macroecological microbiome patterns analysed through Random Forest achieved 91% accuracy in predicting Fusarium wilt occurrence across field sites, identifying *Bacillus* and *Streptomyces* abundance thresholds as key predictive biomarkers. LSTM architectures have further modelled longitudinal shifts in wheat root microbiomes across growing seasons, capturing the temporal window of greatest predictive opportunity (Sharma and Xu, 2021).

Maize (*Zea mays*)

Stalk rot diseases caused by *Colletotrichum graminicola* and *Fusarium verticillioides* represent multi-pathogen challenges amenable to multimodal AI frameworks. Graph neural network (GNN) approaches trained on co-occurrence network disruption data have demonstrated ~87% classification accuracy distinguishing susceptible from resistant root microbiome configurations (Maryam *et al.*, 2024).

Rice (*Oryza sativa*)

Bacterial leaf blight caused by *Xanthomonas oryzae* pv. *oryzae* exhibits pre-symptomatic phyllosphere microbiome shifts detectable by SVM classifiers trained on 16S rRNA data, with reported AUC values of 0.88-0.91 (Ghannam and Techtmann, 2021). Additionally, blast disease (*Magnaporthe oryzae*) represents one of the most destructive fungal pathogens globally; metatranscriptomic integration with CNN architectures has shown early promise in detecting

Table 4: Summary comparison of classical machine learning vs. deep learning for plant microbiome disease prediction.

Criterion	Classical ML (RF, SVM, XGBoost)	Deep learning (CNN, LSTM, GNN)
Minimum sample size	50-200 samples	500-5,000+ samples
Feature engineering	Manual (diversity indices, CLR normalisation)	Automatic (learned representations)
Interpretability	High (feature importance, SHAP-compatible)	Low ("black box"; requires XAI post-hoc tools)
Temporal modelling	Limited	Strong (LSTM, attention mechanisms)
Spatial/network patterns	Weak	Strong (CNN, GNN)
Computational requirement	Low-moderate (laptop-deployable)	High (GPU clusters required)
Multimodal integration	Moderate (feature concatenation)	Strong (late/intermediate fusion architectures)
Transfer learning capability	Limited	Strong (foundation model pre-training)
Best suited for	Small-scale, field diagnostics; interpretability-critical applications	Large multi-omics databases; complex disease landscapes
Primary limitation in microbiome context	Cannot capture non-linear taxon interaction networks	Requires large, diverse, well-labelled datasets

effector gene expression signatures within the endosphere microbiome.

Soybean (*Glycine max*)

Sudden death syndrome caused by *Fusarium virguliforme* represents a model system for XAI-assisted biomarker discovery. SHAP analysis of Random Forest models trained on soybean rhizosphere amplicon data successfully identified *Trichoderma spp.* and fluorescent *Pseudomonas* abundance as the top-ranked protective taxa, directly guiding targeted SynCom formulation strategies (Fadiji *et al.*, 2025).

Multimodal integration

Capturing the full pathological landscape of the plant holobiont requires comprehensive multi omics integration. Single modality analyses often miss the complex mechanistic interactions defining disease onset and progression (Crandall *et al.*, 2020). Multimodal frameworks synthesize microbiome profiles with metatranscriptomics, metabolomics and hyperspectral imaging, directly linking the mere presence of a pathogen to its active virulence gene expression and the corresponding suppression of host defence metabolites (Kajrolkar, 2025). Computational data fusion strategies systematically address these multidimensional dynamics. Early fusion concatenates raw multi omics features before analysis, while late fusion ensembles independent model predictions (Zhao *et al.*, 2024). Intermediate fusion utilizes sophisticated attention mechanisms to extract joint latent representations, excelling at pinpointing crucial biological interactions such as explicitly linking antimicrobial root exudates to the specific recruitment of disease suppressive microbial consortia (Benkirane *et al.*, 2023). Additionally, graph neural networks (GNN) process microbial co-occurrence networks as direct mathematical features, mapping exactly how pathogen invasion fractures cooperative microbial hubs to facilitate infection (Maryam *et al.*, 2024). Ultimately, multimodal models compensate for single modality noise and extreme sparsity, consistently delivering a 5-10% improvement in predictive accuracy and enabling highly reliable field level disease interception (Kajrolkar, 2025).

Transfer learning interpretability

Deploying predictive models across globally diverse agricultural environments demands robust transfer learning to overcome geographic and host genetic variability (Ghannam, 2021). By pretraining foundation models on massive, diverse datasets like the Earth Microbiome Project, algorithms successfully learn universal microbial representations (Thompson *et al.*, 2017). These algorithms undergo targeted domain adaptation to bridge the persistent ecological gap between controlled laboratory pathosystems and highly heterogeneous field conditions (Ghannam, 2021). Few shots learning techniques, utilizing meta learning and prototype networks, further adapt these powerful models to specific crop

species such as transferring resistance insights across the Solanaceae family using minimal local data (Wu *et al.*, 2023).

Simultaneously, explainable AI (XAI) remains strictly necessary for translating opaque black box predictions into biologically validated pathology (Givisis *et al.*, 2025). Analytical algorithms utilize SHapley Additive exPlanations (SHAP) and Local interpretable model agnostic explanations (LIME) to precisely quantify the exact predictive contribution of individual microbial taxa to disease suppression (Abekoon *et al.*, 2025; Salih *et al.*, 2024). Attention visualizations highlight critical spatial and temporal dysbiosis patterns marking initial infection sites, while knowledge distillation extracts transparent ecological rules from complex neural networks (Givisis *et al.*, 2025). This interpretability facilitates the rigorous ecological validation of identified biomarkers, ensuring agronomists can biologically trust AI recommendations to direct targeted microbiome engineering and precise biocontrol deployments (Abekoon *et al.*, 2025).

Critical challenges and limitations

Data-related challenges

Microbiome-based phytopathology relies on high-throughput sequencing, which remains prohibitively expensive and yields datasets frequently containing fewer than 100 diseased samples (Fadiji *et al.*, 2025). Consequently, researchers confront the high dimension low sample size (HDLSS) problem. Models analysing thousands of microbial taxa against sparse sample sizes risk severe overfitting, ultimately failing to accurately predict field-level epidemics (Przymus *et al.*, 2025). Agricultural datasets exhibit extreme class imbalance, typically comprising 80-95% healthy and only 5-20% diseased samples (Fang, 2023). This disparity biases algorithms toward predicting healthy states, drastically reducing the sensitivity required to detect early-stage infections (Xu *et al.*, 2015).

Beyond scarcity, profound technical variations compromise disease forecasting models. Differences in DNA extraction protocols, primer selection and sequencing platforms introduce severe batch effects that actively obscure true biological signals of pathogen ingress (Nearing *et al.*, 2022). Additionally, microbial matrices are strictly compositional and zero-inflated, with 50-90% of features recorded as zeros due to biological absence or detection limits (Gloor *et al.*, 2017). Standard linear regression models mathematically misinterpret these zero-inflated matrices, generating spurious correlations between microbial taxa that fundamentally misguide field-level pathogen suppression and biocontrol strategies (Gloor *et al.*, 2017).

Methodological and validation issues

Machine learning efficiently identifies microbial dysbiosis, but algorithms cannot inherently establish disease

causation (Khelfaoui *et al.*, 2025). To prove that specific microbial shifts actively suppress pathogens rather than merely co-occurring with necrotic plant tissue, pathologists must validate computational predictions using *in vivo* Synthetic Communities (SynComs) (Fadiji *et al.*, 2025). Researchers face a strict interpretability trade-off: deep neural networks achieve the highest predictive accuracy for destructive diseases like late blight but operate as opaque “black boxes.” This opacity obscures the precise ecological mechanisms driving pathogen inhibition, frustrating agronomists seeking actionable insights (Przymus *et al.*, 2025).

This inherent opacity directly drives generalisation failures. Diagnostic models optimised for specific crop cultivars or controlled greenhouse pathosystems frequently collapse when transferred to heterogeneous field environments facing distinct climatic pressures (Pace *et al.*, 2025). Rigorous cross-location and cross-season validations remain exceedingly rare, thereby heavily skewing the scientific literature toward positive, yet completely non-replicable, diagnostic pipelines (Papoutsoglou *et al.*, 2023).

Biological and practical constraints

Biological complexity fundamentally constrains current predictive frameworks (Fadiji *et al.*, 2025). Diverse microbial taxa exhibit extensive functional redundancy, meaning completely different bacterial clades execute equivalent disease-suppressive functions (Louca *et al.*, 2016). Dynamic plant heterogeneity driven by genetic variation, developmental age and acute physiological stress alters root exudate profiles, constantly restructuring the microbiome and confounding stable disease signatures across growing seasons (Fadiji *et al.*, 2025). Processing massive next-generation sequencing (NGS) pipelines requires weeks of intensive bioinformatics analysis and specialised expertise largely unavailable to smallholder farmers, profoundly exacerbating the agricultural digital divide (Raja and Raja, 2024).

AI vs. traditional statistical methods: A critical comparison

Traditional statistical frameworks including Analysis of Variance (ANOVA), PCA, linear discriminant analysis (LDA),

generalised linear models (GLM) and permanova have long formed the analytical backbone of microbiome research in plant pathology (Papoutsoglou *et al.*, 2023). These methods offer critical advantages: They are computationally inexpensive, statistically interpretable, produce formal hypothesis tests with p-values and require modest sample sizes ($n = 20-50$). However, they carry three fundamental limitations that constrain their utility for disease prediction in complex microbiome datasets.

First, traditional methods assume data normality and independence assumptions directly violated by the compositionality, zero-inflation and extreme sparsity of amplicon sequencing data (Gloor *et al.*, 2017). Second, they are univariate or low-dimensional by design, incapable of simultaneously modelling the thousands of microbial taxa, functional genes and metabolite interactions that define pre-symptomatic dysbiosis. Third, they are fundamentally correlational: ANOVA identifies that specific taxa differ between healthy and diseased states but cannot predict disease outcome in novel environments with acceptable sensitivity.

Machine learning methods overcome these limitations by learning non-linear, high-dimensional patterns from training data without parametric assumptions, enabling genuine predictive generalisation. Benchmarking studies consistently demonstrate that random forest and XGBoost outperform LDA and GLM by 15-25% in AUC-ROC for microbiome-based disease classification tasks (D’Elia *et al.*, 2023). However, this superiority is contingent on adequate sample sizes ($n \geq 200$ for ML; $n \geq 1,000$ for DL) and rigorous cross-validation. In low-sample scenarios ($n < 100$), regularised GLMs and partial least squares discriminant analysis (PLS-DA) remain competitive and their interpretability advantage is substantial for field application. The optimal strategy for most plant pathology applications is therefore a tiered analytical approach: initial exploratory analysis with PCA/Permanova, hypothesis generation with regularised GLMs and definitive predictive modelling with RF or XGBoost validated across independent sites (Table 5).

Innovations and implementation roadmap

Long-read sequencing technologies (PacBio, Oxford Nanopore) resolve microbial communities to the strain

Table 5: Algorithmic limitations and pathology implications.

Challenge	Methodological limitation	Pathology implication	Recommended solution
Zero inflation	Spurious feature correlations	Misguides field-level biocontrol targeting	ZINB models; CLR normalisation
Black box models	Low interpretability (DNNs)	Obscures ecological mechanisms of pathogen inhibition	SHAP/LIME XAI post-hoc analysis
Same-site testing	Over-optimism bias	Diagnostics fail in diverse geographic environments	Cross-site, cross-season validation
Class imbalance	Overfits to healthy samples	Misses critical early-stage infection windows	SMOTE; focal loss; class weighting
Traditional methods	Cannot model non-linear microbiome interactions	Fails to detect presymptomatic dysbiosis signatures	Tiered ML approach: PCA → GLM → RF/XGBoost

level, distinguishing virulent pathogens from protective endophytes (Knobloch *et al.*, 2024). Portable MinION sequencers enable real-time, on-site epidemic surveillance. Integrating multiomics genomics, metatranscriptomics and metabolomics reveals how plants systemically recruit disease-suppressive consortia during pathogen attacks (Kajrolkar, 2025). Self-supervised learning utilises unlabelled agricultural data, while causal inference algorithms distinguish true biocontrol mechanisms from spurious correlations, ensuring interventions actively suppress pathogens (Khelfaoui *et al.*, 2025). Bayesian networks provide critical uncertainty quantification, delivering reliable disease risk assessments. Edge computing executes lightweight models directly on mobile devices in remote fields (Heydari and Mahmoud, 2025). Coupled with affordable CRISPR and isothermal amplification assays costing under \$10 per test, these tools facilitate immediate, targeted biocontrol deployment (Bugingo *et al.*, 2026).

Economic feasibility and adoption barriers

The economic viability of AI-driven microbiome diagnostics is currently constrained by prohibitive per-sample costs. Whole-genome shotgun metagenomics currently costs \$100-500 per sample, while 16S rRNA amplicon sequencing costs \$15-50 per sample economically inaccessible for smallholder farmers in low-income countries (Fadji *et al.*, 2025; Raja and Raja, 2024). By contrast, conventional fungicide applications cost \$30-120 per hectare per season, frequently applied preventively without diagnostic confirmation, representing both an economic and environmental inefficiency.

The economic argument for AI diagnostics strengthens substantially when evaluated at the regional or national scale. A 2% reduction in annual crop losses from the current 10-16% baseline achievable through early AI-guided intervention would represent a recovery of approximately \$4.4 billion annually from the \$220 billion global crop loss burden (FAO, 2023). To reach economic viability at the farm level, per-test costs must decline to the \$5-15 range within five years (Fadji *et al.*, 2025). Targeted technologies making this feasible include loop-mediated isothermal amplification (LAMP) assays integrated with minimal ML inference engines, portable nanopore sequencers (Oxford Nanopore MinION at ~\$1,000 capital cost with \$50-100 per run) and cloud-based ML inference platforms that remove computational hardware barriers from individual farms (Knobloch *et al.*, 2024).

Beyond direct costs, adoption barriers include farmer digital literacy, technical extension service capacity, trust in AI recommendations, data privacy concerns and connectivity infrastructure in remote agricultural regions. Successful models for technology adoption in low-resource settings such as India's e-Choupal network and East Africa's digital extension services offer transferable frameworks for microbiome diagnostic dissemination (Raja and Raja, 2024).

Ethical and regulatory considerations

Data ownership and governance

Microbiome sequencing generates highly granular environmental and biological data tied to specific farm locations. Without clear data governance frameworks, this data collected by commercial diagnostic companies could be monetised without equitable farmer compensation. Governments and international bodies must establish farmer data sovereignty principles, ensuring that microbiome diagnostic data generated on-farm remains the legal property of the farmer and is not commercially exploited without informed consent (Adeyeye *et al.*, 2025).

Algorithmic bias

Current AI models are predominantly trained on datasets from temperate, high-income agricultural systems (North America, Europe, Australia), meaning model performance in tropical smallholder systems where crop losses are highest and food security most precarious is largely unvalidated (Raja and Raja, 2024). This introduces a systematic algorithmic bias that could exacerbate global diagnostic inequity. Mandating geographic and agro-ecological diversity in training datasets should become a standard journal and funder requirement.

Regulatory pathways for microbiome-based biomarkers

AI-derived microbiome biomarkers used for crop insurance, disease certification or pesticide application recommendations require regulatory validation analogous to pharmaceutical biomarker approval. No unified international regulatory framework for plant microbiome diagnostic claims currently exists (Fadji *et al.*, 2025). The European union's plant health regulation (EU 2016/2031) and the US EPA's Biopesticide Registration framework provide partial precedent, but do not specifically address AI-generated predictive biomarkers.

Environmental considerations

AI-guided SynCom applications involve deliberate environmental release of formulated microbial consortia. While the microorganisms involved are typically native soil residents, large-scale coordinated deployment could perturb local microbiome diversity and introduce competitive exclusion dynamics. Environmental risk assessment frameworks analogous to those applied to biological control agents under EU Regulation 1107/2009 should govern these applications (Panchal *et al.*, 2026).

Phased implementation roadmap

Translating predictive diagnostics into sustainable agronomy demands a phased implementation roadmap. Within three to five years, researchers must deliver cultivar-specific, field-validated predictive models and pilot rapid diagnostic assays (\$5-10 per test) directly in farmer fields (Fadji *et al.*, 2025). Over five to ten years, precision agriculture will fully integrate routine microbiome monitoring as a

standard advisory practice, utilising distributed edge computing to forecast localised disease outbreaks (Nkwocha and Chandel, 2025). Ultimately, beyond ten years, preventive microbiome management will systematically replace synthetic chemical applications, establishing highly resilient agroecosystems. Achieving this paradigm shift requires breaking disciplinary silos through intense interdisciplinary collaboration among pathologists, microbiologists and AI engineers. Sustained funding, targeted farmer adoption incentives and equitable technology transfer mechanisms remain critical success factors to ensure diagnostic equity and prevent digital divides in global disease management (Fadji *et al.*, 2025; Khelfaoui *et al.*, 2025).

CONCLUSION

The microbiome disease paradigm is firmly established, demonstrating that early pre-symptomatic detection signatures exist across diverse pathosystems. Integrating AI and ML delivers 82-93% predictive accuracy, representing a substantial improvement over traditional reactive pathology. While multimodal fusion, transfer learning and explainability advance field deployment, significant challenges persist regarding data scarcity, model generalisation, validation and standardisation. Critically, reported accuracy figures must be interpreted within their validation context predominantly controlled greenhouse studies with balanced datasets rather than as field-ready benchmarks. Translating these innovations demands critical success factors. Researchers must enforce data standardisation through harmonised protocols, open repositories and longitudinal studies. Breaking disciplinary silos requires intense collaboration among microbiologists, pathologists, ML engineers and agronomists. Rigorous validation must ensure temporal independence, geographic diversity and farmer field testing. Models must move beyond opaque black boxes by prioritising explainable AI (XAI) and mechanistic biomarker validation. Widespread adoption relies heavily on cost reduction via sequencing improvements and computational efficiency. Ethical deployment requires data governance frameworks, regulatory pathways and algorithmic bias mitigation. Ultimately, this convergence drives a paradigm shift from reactive disease response to proactive, surveillance-based prevention. This transition fosters agricultural sustainability by drastically reducing synthetic chemical inputs, thereby protecting ecosystem health. Predictive microbiome management bolsters global food security, ensuring crop resilience, yield protection and stable supplies amid escalating climate change. Achieving this vision requires prioritising diagnostic equity and capacity building in low-resource regions to prevent digital divides. With sustained investment, this transformative technology anticipates mainstream commercial adoption within 10-15 years.

Conflict of interest

The authors declare that they have no conflicts of interest relevant to the content of this article.

REFERENCES

- Abekoon, T., Sajindra, H., Rathnayake, N., Ekanayake, I.U., Jayakody, A. and Rathnayake, U. (2025). A novel application with explainable machine learning (SHAP and LIME) to predict soil N, P and K nutrient content in cabbage cultivation. *Smart Agricultural Technology*. **11**: 100879. <https://doi.org/10.1016/j.atech.2025.100879>.
- Adeyeye, O.A., Obebe, O.A., Adeyeye, A.O. and Oshikoya, S. (2025). Data driven biosecurity: Strengthening global agricultural resilience against emerging threats. *International Journal for Multidisciplinary Research*. **7(2)**. <https://doi.org/10.36948/ijfmr.2025.v07i02.43188>.
- Fadji, A.E., Adeniji, A., Lanrewaju, A.A., Adedayo, A.A., Chukwuneme, C.F., Nwachukwu, B.C., Aderibigbe, J. and Omomowo, I.O. (2025). Key challenges in plant microbiome research in the next decade. *Microorganisms*. **13(11)**: 2546. <https://doi.org/10.3390/microorganisms13112546>.
- Almario, J., Muller, D., Défago, G. and Moënne-Loccoz, Y. (2014). Rhizosphere ecology and phytoprotection in soils naturally suppressive to thielaviopsis black root rot of tobacco. *Environmental Microbiology*. **16(6)**: 1949-1960.
- Avelino, J., Cristancho, M., Georgiou, S., Imbach, P., Aguilar, L., Bornemann, G., Läderach, P., Anzueto, F., Hruska, A.J. and Morales, C.S. (2015). The coffee rust crises in Colombia and Central America (2008-2013): Impacts, plausible causes and proposed solutions. *Food Security*. **7(2)**: 303. <https://doi.org/10.1007/s12571-015-0446-9>.
- Benkirane, H., Pradat, Y., Michiels, S. and Cournède, P.H. (2023). Cust omics: A versatile deep-learning based strategy for multi-omics integration. *Plos Computational Biology*. **19(3)**: e1010921. <https://doi.org/10.1371/journal.pcbi.1010921>.
- Berruto, C.A. and Demirer, G.S. (2024). Engineering agricultural soil microbiomes and predicting plant phenotypes. *Trends in Microbiology*. **32(9)**: 858-873.
- Bugingo, C., Raju, T.J., Moparthi, S., Singh, J., Shivanna, H.M., Karim, S. and Turatsinze, A.N. (2026). Advances in seed health testing: Integrating molecular diagnostics, imaging and AI for enhanced quality assurance. *Seeds*. **5(1)**: 15. <https://doi.org/10.3390/seeds5010015>.
- Busato, S., Gordon, M., Chaudhari, M., Jensen, I.T., Akyol, T., Andersen, S.U. and Williams, C. (2022). Compositionality, sparsity, spurious heterogeneity and other data-driven challenges for machine learning algorithms within plant microbiome studies. *Current Opinion in Plant Biology*. **71**: 102326. <https://doi.org/10.1016/j.pbi.2022.102326>.
- Cassol, I., Ibañez, M. and Bustamante, J.P. (2025). Key features and guidelines for the application of microbial alpha diversity metrics. *Scientific Reports*. **15(1)**: 622. <https://doi.org/10.1038/s41598-024-77864-y>.
- Chawla, N.V., Bowyer, K.W., Hall, L. and Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. **16**: 321-357. <https://doi.org/10.1613/jair.953>.

- Chen, H., Chen, H., Ma, K., Lu, C., Fu, Q., Qiu, Y., Zhao, J., Huang, Y., Yang, Y., Schadt, C.W., Chen, H. and Chen, H. (2022). Functional redundancy in soil microbial community based on metagenomics across the globe. *Frontiers in Microbiology*. **13**: 878978. <https://doi.org/10.3389/fmicb.2022.878978>.
- Crandall, S.G., Gold, K.M., Jiménez-Gasco, M. del M., Filgueiras, C.C. and Willett, D.S. (2020). A multi-omics approach to solving problems in plant disease ecology. *Plos One*. **15(9)**: e0237975. <https://doi.org/10.1371/journal.pone.0237975>.
- D'Elia, D., Truu, J., Lahti, L., Berland, M., Papoutsoglou, G., Ceci, M., Zomer, A., Lopes, M.B., et al. (2023). Advancing microbiome research with machine learning: Key findings from the ML4Microbiome COST action. *Frontiers in Microbiology*. **14**: 1257002. <https://doi.org/10.3389/fmicb.2023.1257002>.
- Durán, P., Thiergart, T., Garrido-Oter, R., Agler, M., Kemen, E., Schulze-Lefert, P. and Hacquard, S. (2018). Microbial interkingdom interactions in roots promote arabidopsis survival. *Cell*. **175(4)**: 973-983.
- Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D. and Cesare, A.D. (2021). Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Scientific Reports*. **11(1)**: 3030. <https://doi.org/10.1038/s41598-021-82726-y>.
- Fadiji, A.E., Adeniji, A.A., Lanrewaju, A.A., Adedayo, A.A., Chukwuneme, C.F., Nwachukwu, B.C., Aderibigbe, J. and Omomowo, I.O. (2025). Key challenges in plant microbiome research in the next decade. *Microorganisms*. **13(11)**: 2546. <https://doi.org/10.3390/microorganisms13112546>.
- Fang, J. (2023). The role of data imbalance bias in the prediction of protein stability change upon mutation. *Plos One*. **18(3)**: e0283727. <https://doi.org/10.1371/journal.pone.0283727>.
- FAO (2023). The State of Food Security and Nutrition in the World. In FAO; IFAD; UNICEF; WFP; WHO; eBooks. <https://doi.org/10.4060/cc3017en>.
- Ghannam, R.B. and Techtmann, S.M. (2021). Machine learning applications in microbial ecology, human microbiome studies and environmental monitoring. *Computational and Structural Biotechnology Journal*. **19**: 1092. <https://doi.org/10.1016/j.csbj.2021.01.028>.
- Givisis, I., Kalatzis, D., Christakis, C. and Kiourekis, Y. (2025). Comparing explainable AI models: SHAP, LIME and their role in electric field strength prediction over urban areas. *Electronics*. **14(23)**: 4766. <https://doi.org/10.3390/electronics14234766>.
- Gloor, G.B., Macklaim, J.M., Pawlowsky Glahn, V. and Egozcue, J.J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*. **8**: 2224. <https://doi.org/10.3389/fmicb.2017.02224>.
- Gómez, E.R., Postma, J., Raaijmakers, J.M. and de Bruijn, I. (2015). Diversity and activity of *Lysobacter* species from disease suppressive soils. *Frontiers in Microbiology*. **16(6)**: 1243. <https://doi.org/10.3389/fmicb.2015.01243>.
- González-Rodríguez, M.M. (2024). Current trends in diagnostic methods for plant diseases. *Pathogens*. **13(1)**: 84.
- Gu, S., Wei, Z., Shao, Z., Friman, V.P., Cao, K., Yang, T. and Jousset, A. (2020). Competition for iron drives phytopathogen control by natural rhizosphere microbiomes. *Nature Microbiology*. **5(8)**: 1002-1010.
- Hacquard, S., Garrido-Oter, R., González, A., Spaepen, S., Ackermann, G., Lebeis, S. and Schulze-Lefert, P. (2015). Microbiota and host nutrition across plant and animal kingdoms. *Cell Host Microbe*. **17(5)**: 603-616.
- Heydari, S. and Mahmoud, Q.H. (2025). Tiny machine learning and on-device inference: A survey of applications, challenges and future directions. *Sensors*. **25(10)**: 3191. <https://doi.org/10.3390/s25103191>.
- Jafar, R. (2024). Plant disease detection using machine learning techniques: A review. *Computers and Electronics in Agriculture*. **219**: 108849.
- Kajrolkar, A. (2025). Integrating multi-omics data for plant stress response: Current advances and future directions. *Premier Journal of Plant Biology*. **3**: 100012. <https://doi.org/10.70389/pjpb.100012>.
- Khelfaoui, I., Wang, W., Meskher, H., Shehata, A.I., Basuini, M.F.E., Abouelenein, M.F. and Mahmoud, S.S. (2025). Beyond just correlation: Causal machine learning for the microbiome, from prediction to health policy with econometric tools. *Frontiers in Microbiology*. **16**: 1691503. <https://doi.org/10.3389/fmicb.2025.1691503>.
- Knobloch, S., Salimi, F., Buaya, A., Ploch, S. and Thines, M. (2024). RAPiD: A rapid and accurate plant pathogen identification pipeline for on-site nanopore sequencing. *Peer J*. **12**: e17893. <https://doi.org/10.7717/peerj.17893>.
- Komissarov, E.N., Diabankana, R.G.C., Абдеева, Ё.А., Afordoanyi, D.M., Gudkov, S.V., Dvorianinova, E.M., Bruskin, S., Dmitriev, A.A. and Validov, S. (2025). Genomic differences between two *Fusarium oxysporum* formae speciales causing root rot in cucumber. *Journal of Fungi*. **11(2)**: 140. <https://doi.org/10.3390/jof11020140>.
- LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature*. **521**: 436-444. <https://doi.org/10.1038/nature14539>.
- Liu, H., Sun, F., Peng, J., Shen, M., Li, J. and Dong, Y. (2022). Deterministic process dominated belowground community assembly when suffering tomato bacterial wilt disease. *Agronomy*. **12(5)**: 1024. <https://doi.org/10.3390/agronomy12051024>.
- Louca, S., Parfrey, L.W. and Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science*. **353(6305)**: 1272. <https://doi.org/10.1126/science.aaf4507>.
- Lyu, D., Zajonc, J., Pagé, A., Tanney, C.A.S., Shah, A., Monjezi, N., Msimbira, L.A., Antar, M., Nazari, M., Backer, R. and Smith, D.L. (2021). Plant holobiont theory: The phytomicrobiome plays a central role in evolution and success. *Microorganisms*. **9(4)**: 675. <https://doi.org/10.3390/microorganisms9040675>.
- Mandlik, V.S. and Lenina, S.V.B. (2026). Deep hybrid learning for smart agriculture using CNN-based feature extraction and LSTM-BiLSTM sequence modeling for robust plant disease detection. *Indian Journal of Agricultural Research*. **60(4)**: 597-606. doi: 10.18805/IJArE.A-6490.
- Maryam, M., Rehman, M.U., Hussain, I., Tayara, H. and Chong, K.T. (2024). A graph neural network approach for predicting drug susceptibility in the human microbiome. *Computers in Biology and Medicine*. **179**: 108729. <https://doi.org/10.1016/j.compbiomed.2024.108729>.

- Mavrodi, D.V., Mavrodi, O.V., Parejko, J.A., Weller, D.M. and Thomashow, L.S. (2011). The Role of 2,4-Diacetylphloroglucinol- and Phenazine-1-Carboxylic Acid-Producing *Pseudomonas* spp. in Natural Protection of Wheat from Soilborne Pathogens. pp. 267-283. https://doi.org/10.1007/978-3-642-21061-7_11.
- Mehta, A.R., Kumar, P., Prem, G., Aggarwal, S. and Kumar, R. (2025). AI-powered innovations in agriculture: A systematic review on plant disease detection and classification. *Indian Journal of Agricultural Research*. **59(9)**: 1321-1330. doi: 10.18805/IJARE.A-6371.
- Mendes, R., Kruijt, M., de Bruijn, I., Dekkers, E., van der Voort, M., Schneider, J.H. and Raaijmakers, J.M. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science*. **332(6033)**: 1097-1100.
- Metagar, S.M. and Walikar, G.A. (2024). Machine learning models for plant disease prediction and detection: A review. *Agricultural Science Digest*. **44(4)**: 591-602. doi: 10.18805/ag.D-5893.
- Monti, M. M., Affinito, A., Cuomo, S. and Ruocco, M. (2026). Machine learning approaches to assess soil microbiome dynamics and bio-sustainability. *Physiologia Plantarum*. **178**: e70719.
- Navrozidis, I., Pantazi, X., Lagopodi, A., Bochtis, D. and Alexandridis, T. (2023). Application of machine learning for disease detection tasks in olive trees using hyperspectral data. *Remote Sensing*. **15(24)**: 5683.
- Nearing, J.T., Douglas, G.M., Hayes, M.G., MacDonald, J., Desai, D., Allward, N.E., Jones, C., Wright, R., Dhanani, A.S., Comeau, A. and Langille, M.G.I. (2022). Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications*. **13(1)**: 342. <https://doi.org/10.1038/s41467-022-28034-z>.
- Nelson, M.B., Martiny, A.C. and Martiny, J.B. (2016). Global biogeography of microbial nitrogen-cycling traits in soil. *Proceedings of the National Academy of Sciences*. **113(29)**: 8033-8040.
- Nguyen, C.C. (2025). 2,4-diacetylphloroglucinol (DAPG) enhances induced systemic resistance against necrotrophic pathogens through the COI1-mediated JA signaling pathway in arabidopsis. *Journal of Plant Growth Regulation*. **44**: 4881.
- Nkwocha, C.L. and Chandel, A.K. (2025). Towards an end-to-end digital framework for precision crop disease diagnosis and management based on emerging sensing and computing technologies: State over past decade and prospects. *Computers*. **14(10)**: 443. <https://doi.org/10.3390/computers14100443>.
- Pace, R., Cola, V.S.D., Monti, M.M., Affinito, A., Cuomo, S., Loreto, F. and Ruocco, M. (2025). Artificial intelligence in soil microbiome analysis: A potential application in predicting and enhancing soil health-A review. *Discover Applied Sciences*. **7(2)**: 85. <https://doi.org/10.1007/s42452-024-06381-4>.
- Pace, R., Monti, M.M., Cuomo, S., Affinito, A. and Ruocco, M. (2026). Machine learning approaches to assess soil microbiome dynamics and bio sustainability. *Physiologia Plantarum*. **178(1)**: 70719. <https://doi.org/10.1111/ppl.70719>.
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. **372**: n71. <https://doi.org/10.1136/bmj.n71>.
- Panchal, K., Sudhir, A. and Prajapati, A.S. (2026). Engineering the plant microbiome: Synthetic community approaches to enhance crop protection. *Frontiers in Plant Science*. **16**: 1705289.
- Papoutsoglou, G., Tarazona, S., Lopes, M.B., Klammsteiner, T., Ibrahim, E., Eckenberger, J. and Berland, M. (2023). Machine learning approaches in microbiome research: Challenges and best practices. *Frontiers in Microbiology*. **14**: 1261889. <https://doi.org/10.3389/fmicb.2023.1261889>.
- Pieterse, C.M., Zamioudis, C., Berendsen, R.L., Weller, D.M., Van Wees, S.C. and Bakker, P.A. (2014). Induced systemic resistance by beneficial microbes. *Annual Review of Phytopathology*. **52**: 347-375.
- Przymus, P., Rykaczewski, K., Martín Segura, A., Truu, J., Pau, E.C. de S., Kolev, M., Naskinova, I., Gruca, A., Sampri, A., Frohme, M. and Nechyporenko, A. (2025). Deep learning in microbiome analysis: A comprehensive review of neural network models. *Frontiers in Microbiology*. **15**: 1516667. <https://doi.org/10.3389/fmicb.2024.1516667>.
- Raja, R. and Raja, R. (2024). Bridging the digital divide in agriculture: Lessons from the united states and africa in smart farming adoption. *International Journal of Innovative Science and Research Technology*. **10(4)**. doi: <https://doi.org/10.38124/ijisrt/25apr1024>.
- Rathnayake, U., Abekoon, T., Sajindra, H., Rathnayake, N., Ekanayake, I.U. and Jayakody, A. (2025). A novel application with explainable machine learning (SHAP and LIME) to predict soil N, P and K nutrient content in cabbage cultivation. *Smart Agricultural Technology*. pp. 100879.
- Ren, Y., Wu, C., Zhou, H., Hu, X. and Miao, Z. (2024). Dual-extraction modeling: A multi-modal deep-learning architecture for phenotypic prediction and functional gene mining of complex traits. *Plant Communications*. **5(3)**: 101002.
- Salih, A.M., Raisi-Estabragh, Z., Galazzo, B.I., Radeva, P., Petersen, S.E., Lekadir, K. and Menegaz, G. (2024). A perspective on explainable artificial intelligence methods: SHAP and LIME. *arXiv*. **7(1)**. <https://doi.org/10.48550/arXiv.2305.02012>.
- Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N. and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology and Evolution*. **3**: 430-439. <https://doi.org/10.1038/s41559-018-0793-y>.
- Shafik, W., Tufail, A., De Silva, L.C. and Apong, R.A.A.H.M. (2025). Deep learning technique for plant disease classification and pest detection and model explainability elevating agricultural sustainability. *BMC Plant Biology*. **25**: 1491.
- Shafik, W., Tufail, A., Silva, L.C.D. and Apong, R.A.A.H.M. (2025). A lightweight deep learning model for multi-plant biotic stress classification and detection for sustainable agriculture. *Scientific Reports*. **15(1)**: 12195. <https://doi.org/10.1038/s41598-025-90487-1>.

- Sharma, D. and Xu, W. (2021). phyLoSTM: A novel deep learning model on disease prediction from longitudinal microbiome data. *Bioinformatics*. **37(21)**: 3707-3714.
- Siegieda, D., Panek, J. and Frąć, M. (2023). Plant and soil health in organic strawberry farms-greater importance of fungal trophic modes and networks than α -diversity of the mycobiome. *Applied Soil Ecology*. **188**: 104925.
- Su, L., Zhang, L., Nie, D., Kuramae, E.E., Shen, B. and Shen, Q. (2020). Bacterial tomato pathogen *Ralstonia solanacearum* invasion modulates rhizosphere compounds and facilitates the cascade effect of fungal pathogen *Fusarium solani*. *Microorganisms*. **8(6)**: 806. <https://doi.org/10.3390/microorganisms8060806>.
- Thompson, L., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., Prill, R.J., *et al.* (2017). A communal catalogue reveals earth's multiscale microbial diversity. *Nature*. **551(7681)**: 457. <https://doi.org/10.1038/nature24621>.
- Trivedi, P., Leach, J.E., Tringe, S.G., Sa, T. and Singh, B.K. (2020). Plant-microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology*. **18(11)**: 607-621.
- Vincenzo, C., Tripodi, P., Lombardi, N., Manganiello, G. and Pane, C. (2025). Plant phenomics-assisted selection of *Trichoderma* spp. strains effective in the biocontrol of tomato soil-borne fungal diseases. *Environmental Technology Innovation*. **40**: 104357. <https://doi.org/10.1016/j.eti.2025.104357>.
- Wang, K., Abid, M.A., Rasheed, A., Crossa, J., Hearne, S. and Li, H. (2023). DNNGP, a deep neural network-based method for genomic prediction using multi-omics data in plants. *Molecular Plant*. **16(2)**: 279-293.
- Wankhade, A., Wilkinson, E.L., Britt, D.W. and Kaundal, A. (2025). Plant microbe interactions in the rhizosphere and the role of root exudates in chemical signaling and microbiome engineering. *Appl. Sci*. **15(13)**: 7127. <https://doi.org/10.3390/app15137127>.
- Wei, Z., Gu, Y., Hu, J., Yin, S., Xu, Y., Jousset, A. and Friman, V.P. (2018). *Ralstonia solanacearum* pathogen disrupts bacterial rhizosphere microbiome during an invasion. *Soil Biology and Biochemistry*. **118**: 8-17.
- Wilhelm, R.C., Es, H.M. and Van Buckley, D.H. (2021). Predicting measures of soil health using the microbiome and supervised machine learning. *Soil Biology and Biochemistry*. **164**: 108472. <https://doi.org/10.1016/j.soilbio.2021.108472>.
- Wu, X., Deng, H., Wang, Q., Liang, L., Gao, Y. and Hao, G. (2023). Meta learning shows great potential in plant disease recognition under few available samples. *The Plant Journal*. **114(4)**: 767. <https://doi.org/10.1111/tbj.16176>.
- Xu, L., Paterson, A.D., Turpin, W. and Xu, W. (2015). Assessment and selection of competing models for zero-inflated microbiome data. *Plos One*. **10(7)**: 0129606. <https://doi.org/10.1371/journal.pone.0129606>.
- Yuan, J., Wen, T., Zhang, H., Zhao, M., Penton, C.R., Thomashow, L.S. and Shen, Q. (2020). Predicting disease occurrence with high accuracy based on soil macroecological patterns of fusarium wilt. *The ISME Journal*. **14(12)**: 2936. <https://doi.org/10.1038/s41396-020-0720-5>.
- Zhang, X. and Yi, N. (2020). NBZIMM: Negative binomial and zero-inflated mixed models, with application to microbiome/metagenomics data analysis. *BMC Bioinformatics*. **21(1)**: 488. <https://doi.org/10.1186/s12859-020-03803-z>.
- Zhao, F., Zhang, C. and Geng, B. (2024). Deep multimodal data fusion. *ACM Computing Surveys*. **56(9)**: 1-36. <https://doi.org/10.1145/3649447>.
- Zhu, H., Shi, W., Guo, X., Lyu, S., Yang, R. and Han, Z. (2024). Potato disease detection and prevention using multimodal AI and large language model. *Computers and Electronics in Agriculture*. **229**: 109824. <https://doi.org/10.1016/j.compag.2024.109824>.