

Artificial intelligence in soil microbiome analysis: a potential application in predicting and enhancing soil health—a review

Roberta Pace^{1,2} · Vincenzo Schiano Di Cola³ · Maurilia Maria Monti² · Antonio Affinito⁴ · Salvatore Cuomo³ · Francesco Loreto¹ · Michelina Ruocco²

Received: 26 August 2024 / Accepted: 25 November 2024

Published online: 16 January 2025

© The Author(s) 2024 OPEN

Abstract

Soil is a depletable and non-renewable resource essential for food production, crop growth, and supporting ecosystem services, such as the retaining and cycling of various elements, including water. Therefore characterization and preservation of soil biological health is a key point for the development of sustainable agriculture. We conducted a comprehensive review of the use of Artificial Intelligence (AI) techniques to develop forecasting models based on soil microbiota data able to monitor and predict soil health. We also investigated the potentiality of AI-based Decision Support Systems (DSSs) for improving the use of microorganisms to enhance soil health and fertility. While available studies are limited, potential applications of AI seem relevant to develop predictive models for soil fertility, based on its biological properties and activities, and implement sustainable precision agriculture, safeguarding ecosystems, bolstering soil resilience, and ensuring the production of high-quality food.

Article Highlights

- Maintaining healthy soil is crucial for sustainable agriculture to ensure long-term food production and environmental protection.
- Soil microbiota plays a main role in preserving soil health, therefore a correct analysis of soil microorganisms will help farmers understand and improve soil quality.
- Although research is limited, integrating AI technologies to monitor soil health holds great promise for sustainable precision agriculture.

Keywords Soil health · Microbial soil community · Artificial Intelligence · Forecasting models · Decision Support Systems

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s42452-024-06381-4>.

✉ Maurilia Maria Monti, mauriliamaria.monti@cnr.it | ¹Dipartimento di Biologia, Complesso Universitario Monte S. Angelo, Università degli Studi di Napoli Federico II, Edif.7, Via Cinthia, 80126 Napoli, Italia. ²Istituto per la Protezione Sostenibile delle Piante, Consiglio Nazionale delle Ricerche, Piazzale Enrico Fermi, 1, 80055 Portici (Napoli), Italia. ³Dipartimento di Matematica e Applicazioni R. Caccioppoli, Complesso Universitario Monte S. Angelo, Università degli Studi di Napoli Federico II, Via Cinthia, 80126 Napoli, Italia. ⁴EVJA s.r.l., Via Benedetto Brin, 63, 80142 Napoli, Italia.



1 Introduction

Soil is a depletable and not renewable resource. Soil degradation, erosion, and contamination cause soil loss and can rapidly occur and significantly impact ecosystem services. Whereas, the formation of new layers of soil occurs at an often slower rate than soil loss, resulting in a net reduction of soil availability over time, and thus compromising agricultural productivity, biodiversity, water, and nutrient cycles [1, 2]. This highlights the importance of soil conservation and sustainable soil management practices, essential for maintaining soil biological health and ensuring its continued ability to support life on Earth [3].

The current systems of food production are among the main causes of the depletion of natural resources and environmental degradation. Notably, they affect the quality and resilience of the soil, making agriculture unsustainable [4, 5]. Even if soil quality and soil health are terms often used interchangeably [6], here we will use the term soil health. Soil quality is often associated with specific functions that soil performs, such as supporting plant growth, filtering water, and cycling nutrients. It focuses on the soil characteristics that contribute to these functions, assessing how well the soil can perform its roles in the ecosystem [7]. Soil quality is influenced by both natural and anthropogenic factors, including land management practices [8]. On the other hand, soil health is a broader concept that encompasses soil quality and is not just about current physical and chemical qualities but also about the soil overall fertility and sustainability over time [9].

Soil health recognizes that the soil is a living resource, harboring a diverse community of micro and macro-organisms [10–12]. These communities are part of the soil biodiversity and their differences in quality and quantity directly affect soil processes, such as the decomposition of organic matter, the solubilization of chemical elements, and the absorption and degradation of heavy metals and other pollutants [13, 14].

Intensive farming, monoculture, and limited crop rotations are the primary causes of loss of biodiversity, a main indicator of soil biological health [15, 16]. Sustainable land management practices, soil conservation techniques, and efforts to reduce soil pollution when farming are therefore crucial for preserving soil health and biodiversity and ensuring the long-term productivity of agriculture [17].

Nowadays, governments and international agencies are offering more support for sustainable agriculture through policies and research funding, recognizing the role of agriculture in achieving sustainability and development goals. For instance, under the EU's new policy framework "Food 2030," the European Commission advocates a novel development approach, aligned with the 17 Sustainable Development Goals of the United Nations, where sustainable land management is a pillar [18, 19].

Soil health also supports plant health [20–23] and this is the basis of the current EU "One Health" policies (https://health.ec.europa.eu/one-health_en). In agricultural soils, the dynamic interplay among biological, physical, and chemical properties is crucial. The microbial biomass plays a pivotal role in this process, influencing soil health, nutrient cycling, and overall agricultural productivity [24, 25].

In the recent past, agriculture has neglected soil multi-functionality and ecosystem services, leading to practices that have negatively impacted soil health [4]. Conventional agriculture, which uses chemical fertilization, pesticides, and intensive cultivation, can boost short-term productivity but damage soil by eroding, depleting nutrients, and reducing biodiversity [26, 27].

Although since 1970 the theoretical approach to agriculture has always had a focus on agronomic practices deemed sustainable, only from 2010 began to appear a better consciousness of the importance of agricultural management aimed at restoring ecosystem balance [28]. The intention now is to preserve soil fertility avoiding the use of synthetic agrochemicals, and minimizing the environmental impact of human activities with practices such as soft tillage, integrated management of the crops, and other environmentally friendly practices aimed to improve ecosystem services. These practices are preferred over conventional systems to maintain the optimum level of agroecosystem health status [8, 29–31]. Despite the growing attention on soil science, studying soils can be quite difficult due to their functional complexity [30–32], and finding the right indicators to assess whether the soil has suitable levels of quality and health is challenging [33–35].

The role of microorganisms in improving ecosystem soil health and promoting optimal plant development was largely discussed and deeply studied [36–41], and soil metagenome is now considered a "second genome" of plants [42]. The use of beneficial soil microorganisms, as biological pesticides or as biostimulants, for example, can be considered a sustainable agronomic practice to be implemented together with integrated management [43]. Augmentation of beneficial microorganisms to increase soil fertility can consequently be instrumental for improved crops and for other successful agricultural practices [44]. An ideal agricultural ecosystem should support microbial biodiversity with practices aimed at

maintaining local biological and ecological cycles increasing soil fertility, improving soil structure (e.g. by crop rotation), and preferring integrated pest management (IPM) with a limited use of chemicals [45–47].

1.1 Soil microbiota

Microbial soil communities have a fundamental role in establishing a network of positive interaction belowground [48, 49]. There are thousands of species of soil microorganisms that interact with each other and compete for nutrients, each of them with a specific role [50–52]. Through many of the enzymes and other substances they produce, microorganisms, especially bacteria and fungi, can stimulate plant growth, protect them from the attacks of pathogens and pests, and promote the mineralization of organic matter releasing nutrients necessary for plant development [53–55]. Many microorganisms form symbiotic associations with plants at the root level, playing an essential role in plant development by providing numerous benefits such as producing antibiotics, modulating the plant immune system, and acting as biostimulants [23, 56, 57], but the mechanisms that take place within the soil after the inoculum of such microorganisms is still unclear.

For hundreds of millions of years, a co-evolutionary phenomenon occurred, in which plants and microorganisms established strong symbiotic relationships [58] and influenced each other: plants can produce organic substances, especially under biotic or abiotic stresses, changing the surrounding chemical and physical environment, activating some groups of microorganisms, and removing others, and vice versa. Understanding and harnessing the complex interactions between soil, microorganisms and plants is essential for sustainable agriculture, ecosystem restoration, and environmental conservation efforts. It underscores the importance of maintaining healthy soil ecosystems for the well-being of agricultural systems. Using beneficial microbes, modern agriculture is not in contradiction with co-evolution: it is exploiting knowledge of the co-evolutionary phenomenon to optimize crop health and productivity under current conditions. We aim to integrate these principles, along with new technologies, to advance the development of management systems and soil treatment methods in open field conditions.

1.2 Methods for studying soil microorganisms

Microorganisms and their communities in soil can be assessed, both in terms of diversity and abundance, through different laboratory techniques: from the traditional techniques as culture on selective media to the most recent applications of multi-omics approaches [59], the latter largely improving precise and accurate overviews [7] and detection of microbial changes under different conditions and management practices, which may in turn affect soil properties [48]. Combining different techniques (the “multimethod” or “mixed-methods” approach) may also help cross-verify results, mitigating the limitations of individual methods, and gaining a more comprehensive understanding of the dynamics in the soil [60]. Although individual microorganisms are tiny and work within microenvironments, such as around a single root or within soil aggregates, their processes, such as nutrient cycling, organic matter breakdown, and pathogen suppression, affect larger, visible aspects of the environment like plant health, soil fertility, and carbon storage. We should better understand how to enhance soil bacterial processes, scaling them up from the microscale to levels that are significant for soil management. Sample size, along with replication, sampling design, DNA extraction, and molecular analysis methods, significantly influence measures of community structure under field conditions [61]. Scaling up results, from small sample size to represent a large field, requires careful consideration of these factors to avoid misrepresentation of microbial diversity. For this reason, it is important to consider standardized protocols like ISO 11063, which aim to optimize soil DNA extraction for reproducibility and comparability, though variations in methods and sample sizes can still impact the assessment of microbial abundance and diversity [62]. Also, modeling the dynamics of a microbial soil community can help understand belowground interactions that have repercussions on the soil ecosystem [63]. Among approaches based on soil total DNA analysis to provide taxonomic profiles of microbial communities, one of the most common methods is the metabarcoding: it is based on the amplification of specific DNA regions, such as 16S rDNA (ribosomal DNA) for bacteria and ITS (Internal Transcribed Spacer) for fungi, using PCR (Polymerase Chain Reaction), followed by high-throughput sequencing, e.g. with NGS (Next Generation Sequencing) methods [64]. Although this approach has the undeniable advantage of enabling the simultaneous analysis of many samples with a good level of precision and in a relatively short time, it does have some limitations that can affect the accuracy of taxonomic assignments. These drawbacks include potential biases introduced during DNA extraction and amplification [65], difficulties in distinguishing between live and dead organisms [66], resolution usually limited to family or genus level that does not permit differentiation between closely related species or sub-species [67], and challenges associated with incomplete

reference databases [65]. Additionally, the presence of PCR inhibitors in soil samples can complicate the process and lead to the underrepresentation of certain taxa. To address these limitations, researchers are exploring and developing alternative approaches and technologies. For example, to enhance taxonomic resolution, multiple genetic markers could be combined, and thanks to sequencing technologies developed by PacBio (Pacific Bioscience) and Oxford Nanopore, longer reads compared to traditional NGS methods are provided, enabling a better assembly of genomes and an accurate characterization of complex microbial communities [68]. To explore the genetic diversity and functional potential of complex microbial communities in a wide range of environments, metagenomics provides a powerful tool [69]. Instead of targeting specific genetic markers, Metagenomic Shotgun Sequencing involves the sequencing of all DNAs present in a sample providing a more comprehensive view, allowing the discovery of new species and functional genes.

Investigating the genera or the species of microorganisms present in soil can unveil valuable perspectives about the environment, as specific microorganisms can serve as “indicator species” or “bioindicators” [70, 71].

At a community level, the physiological profiling method performed using the Biolog® system is based on patterns of single carbon source utilization by microorganisms present in the soil. This analysis permits obtaining information on how microbial communities respond to changes of the environment over both spatial and temporal scales [72].

Enzymes produced by microorganisms are key factors in soil processes such as decomposing organic matter, cycling nutrients, and enhancing soil structure [73]. Measuring enzymatic activities of microorganisms in soil provides insights into microbial activity, and can serve as indirect indicators of soil health. Higher enzymatic activity levels often indicate greater microbial diversity, especially in the active soil microorganisms, essential for maintaining soil health [74–76]. An example of a soil sensor able to give information on biological enzyme activity was developed by DIGIT-soil [77]. This biosensor works through a customizable membrane targeting several soil enzymes, providing in-situ measurements and automated analysis of soil enzyme activity in few minutes. These sensors can be integrated into soil monitoring systems, allowing for continuous monitoring and timely interventions to maintain, restore, or improve soil fertility and health, providing real-time or near-real-time data on soil biological parameters. All the methods mentioned are summarized in Table 1.

1.3 Harnessing technologies for smarter agriculture

Agriculture 4.0 represented a significant transformation driven by the integration of cutting-edge technologies [85] to optimize crop yields while minimizing inputs such as water, fertilizer, and pesticides.

Big data analytics involves the examination of large and diverse datasets to identify hidden patterns, correlations, trends, and valuable insights, enabling farmers to make informed decisions based on a comprehensive understanding of factors influencing farming operations [86, 87]. Analysis of large datasets collected from sensors located proximally or distally from the cultivated area (e.g. drones or satellites), through the new computational technologies, can help farmers optimize resource allocation, predict crop yields, manage risks, and identify areas for improvement of farming practices [88].

Furthermore, devices, such as soil moisture sensors, weather stations, and automated irrigation systems, allow farmers to monitor and control agricultural processes in real-time. Connected sensors provide continuous data on environmental conditions and crop status, enabling precise management and timely interventions.

By integrating these innovations, farmers can advance beyond conventional methods, adopting a more data-driven and efficient approach that enables real-time informed decisions [89, 90]. Today, the decision-making process is facilitated by Decision Support Systems (DSSs) that gradually evolved to incorporate more sophisticated tools and techniques [91, 92]. With the incorporation of machine learning (ML) algorithms and predictive models, indeed, DSSs could analyze complex datasets, offering insights and recommendations for optimizing outcomes [93]. One of the most important questions in agriculture is to predict soil suitability for specific crops before planting [94]. In this context, a balanced microbiota supports optimal production yield while maintaining the ability of the soil to sustainably produce, which is a relevant entry of the DSS algorithm [95, 96], DSS capable of monitoring soil health must utilize many data inputs including soil nutrient levels, moisture content, temperature, microbial diversity, to feed ML algorithms aiming at predicting crop performances. DSSs integrating soil microbiology monitoring can also ensure that farmers can take action to restore microbial balance when necessary, by using beneficial microorganisms or compost amendments. This approach supports sustainable crop productivity and the maintenance and enrichment of soil health.

Data that allow DSS elaboration often need to be examined by Artificial Intelligence (AI). AI technologies, such as Machine Learning (ML) and Deep Learning (DL), have found numerous applications in agriculture, offering solutions to various challenges [97–103], and could result fundamental to analyze and model large amounts of data detected

Table 1 A glossary of the main methods mentioned in the paper used for studying soil microorganisms

Method	Basic principle	References
Metabarcoding	is a molecular biology technique used to identify and quantify species within a complex mixture of DNA, including environmental samples. It combines DNA barcoding, which involves sequencing a short, standardized region of DNA ('barcode') that is unique to a species, with high-throughput sequencing (HTS) technologies, allowing for the simultaneous analysis of multiple species from a single sample	[78]
High-Throughput Sequencing (HTS)	refers to a set of advanced technologies that allow the rapid and large-scale sequencing of DNA or RNA. Unlike traditional Sanger sequencing, which sequences DNA one fragment at a time, HTS can sequence millions to billions of DNA fragments simultaneously, making it highly efficient and capable of generating vast amount of data in a relatively short time	[79]
Polymerase Chain Reaction (PCR)	is a widely used molecular biology technique that allows the amplification of specific DNA sequence, through key components: two primers, one forward and one reverse, short single-stranded DNA sequences that bind to opposite strands of the DNA template; an enzyme that synthesizes new DNA strands by adding nucleotides to the primers (polymerase); Deoxyribonucleotide Triphosphates (dNTPs); the building blocks (adenine [A], cytosine [C], guanine [G], and thymine [T]) used by DNA polymerase to synthesize the new DNA strands; a buffer solution that maintains the optimal pH and ionic conditions for the PCR reaction, ensuring that the DNA polymerase functions efficiently; the template, the DNA sequence that will be amplified. PCR involves three main steps: each cycle of denaturation, annealing, and extension doubles the amount of the target DNA sequence. After 20–40 cycles, the target DNA can be amplified millions to billions of times	[80]
Next Generation Sequencing (NGS)	is a collection of advanced sequencing technologies that enables the rapid and comprehensive sequencing of DNA or RNA. Unlike traditional Sanger sequencing, which sequences DNA one fragment at a time, NGS can sequence millions to billions of DNA fragments simultaneously, providing a massive amount of genetic information quickly and at a lower cost per base	[81]
PacBio sequencing	is an advanced DNA and RNA sequencing technology developed by Pacific Biosciences known as Single-Molecule Real-Time (SMRT) Sequencing. This technology is distinct from other sequencing methods due to its ability to produce extremely long reads, providing more comprehensive and accurate insights into genetic sequences, including complex regions that are difficult to sequence with other technologies	[81]
Oxford Nanopore sequencing	is an advanced DNA and RNA sequencing technology developed by Oxford Nanopore Technologies. This technology is known for its ability to produce ultra-long reads by directly sequencing nucleic acids as they pass through a nanopore, a tiny hole, embedded in a membrane. It is unique among sequencing technologies because it can generate real-time data, sequence very long fragments of DNA or RNA, and analyze a wider range of molecule types	[81]
Metagenomics	is the study of genetic material recovered directly from complex samples, rather than a single organism. This field allows scientists to analyze the collective genomes of entire communities of microorganisms, such as bacteria, archaea, viruses, and fungi, that inhabit a particular environment. The key advantage of metagenomics is that it enables the study of organisms that are difficult or impossible to culture in the lab, thus providing a more comprehensive understanding of microbial diversity, ecology, and function	[82]
Metagenomic Shotgun Sequencing	is a comprehensive and unbiased approach to study the genetic material within entire communities (including microbiota) directly from environmental samples. Unlike targeted sequencing methods, which focus on specific genes, shotgun metagenomics sequences all the DNA in a sample, capturing the full diversity of organisms and their functional potential	[83]
Biolog [®]	is a technology used for the identification and characterization of microorganisms based on their metabolic activities. It is commonly used in microbiology laboratories for the identification of bacteria, yeast, and fungi, as well as for assessing their metabolic profiles. It assesses the metabolic capabilities of microorganisms by measuring their ability to utilize various carbon sources and other substrates. It profiles different species or strains based on their metabolic fingerprints	[71]
Biosensors	are sophisticated analytical tools that integrate biological recognition elements with physical or chemical detectors to measure specific substances. Their applications span across medical diagnostics, environmental monitoring, food safety, biotechnology, and more, providing valuable insights and real-time data for various fields	[84]

in the field. AI-driven DSSs are instrumental for precision farming and its applications for sustainable agricultural practices [104], aimed at preserving ecosystems, increasing the resilience of soils, and ensuring the quality and quantity of food production (Agriculture 5.0) [105]. AI technologies are increasingly being utilized in agriculture to enhance productivity, efficiency, and sustainability of crops, and play a leading role in this context by enabling machines to perform tasks related to human intelligence [106]. Moreover, the integration of Internet of Things (IoT), Decision Support Systems (DSSs), Cloud Computing, Big data analytics and Remote Sensing has indeed brought significant advancements in agriculture [107, 108]. The complementation with automation and digitization systems enables both the collection of high-throughput data and the optimization of field-level practices such as planting [109–111], fertilization [112–114], irrigation [115–119], and pest control [120–122] based on real-time data acquisition and analysis, and often are also used for features such as data visualization, modeling, and simulation [123].

Numerous publications on generic applications of AI techniques for agricultural sustainability are available, which will not be reviewed here. Bhagat et al. (2022) made a bibliometric analysis, finding 465 papers only in the period 2000–2021 on this subject [124]. However, models developed using AI techniques to differentiate soil health status based on soil microbial composition, which serves as a fundamental input for sustainable agriculture, are scant in the literature.

1.4 AI empowers microbial community studies

For this review, we intended to consider only papers dealing with the prediction of soil health metrics through microbiome data as input, but this search was very restricting. So, we added studies that utilized AI techniques to predict outcomes based on microbiome composition, as well as those that employed AI to determine microbiome composition as an output. We consulted publications from "Web of Science", "SCOPUS" and "Google Scholar" databases until 1st October 2023, using the following keywords: "artificial intelligence", "soil microbial community", "soil health prediction", "soil microbiome monitoring", "soil health assessment", "soil microbiome", "soil microbial indicators", alone and using the Boolean operator "AND" to refine the research (e.g. "soil microbial community AND soil health prediction"), in all possible combination, excluding the redundant ones. Results are shown in Table S1. Out of 3508 papers found, only 12 encountered the requested criteria (Table 2), and their classification based on the AI approach applied is shown in Fig. 1.

As already mentioned, only a few studies have been published on how to predict soil health through information on microbial soil communities using AI techniques and forecasting models, despite the high potential of these applications [125]. In the context of studying soil microbiomes, researchers have proposed microbial-based metrics to evaluate soil health [76]. Indeed, Correa-Garcia et al. (2022) [125] claim that considering the ability of microorganisms to reflect past and present conditions of the habitats in which they are found, and are directly involved in the ecosystem processes, they also have some predictive capacity for future processes. Microbiomes are ideal for modeling and predicting the evolution of ecosystem processes as they participate directly in them. De Andrade et al. (2021) [105] propose guidelines for developing an Artificially Intelligent Soil Quality Index, incorporating input data for the ML model that includes recent land management practices, current genomic features, present and projected plant cover, as well as current soil biological, chemical, and physical parameters. Wilhelm et al. (2022) [126] aimed to develop a predictive model using soil bacterial community composition (in terms of 16S DNA sequences) to estimate 12 key soil health metrics, including biological, chemical, and physical properties. The study employed supervised machine learning algorithms, specifically Random Forest (RF) and Support Vector Machines (SVM), to classify and predict soil health metrics based on microbiome data collected from 949 soil samples from North American farmlands. Biological health metrics were the most accurately predicted, while the models struggled more with chemical and physical metrics, such as aggregate stability. However, the study suggests that larger and more region-specific datasets are needed to improve accuracy further. Hermans et al. (2020) [127] applied RF models to assess the degree to which bacterial communities in cultivated soils can be used to forecast physio-chemical soil characteristics: they verified a robust correlation between the composition of bacterial communities and land use, asserting that bacteria sensitivity to changes in specific soil parameters highlights the potential of bacterial DNA analysis as a tool for assessing soil quality. Chang et al. (2017) [128] applied a RF machine learning approach to predict crop productivity based on microbiome composition, revealing that differences in crop productivity were linked to the bulk soil microbiome composition and highlighting that several taxa are associated with nitrogen utilization.

Other publications were found on the use of AI techniques to study the relationship between soil microbial characteristics and the physical and chemical properties of soils [129, 130], to predict soil microbial biomass [131], to predict soil enzyme activity [132], and to assess soil microbiological characteristics or bacterial community structure [133–135].

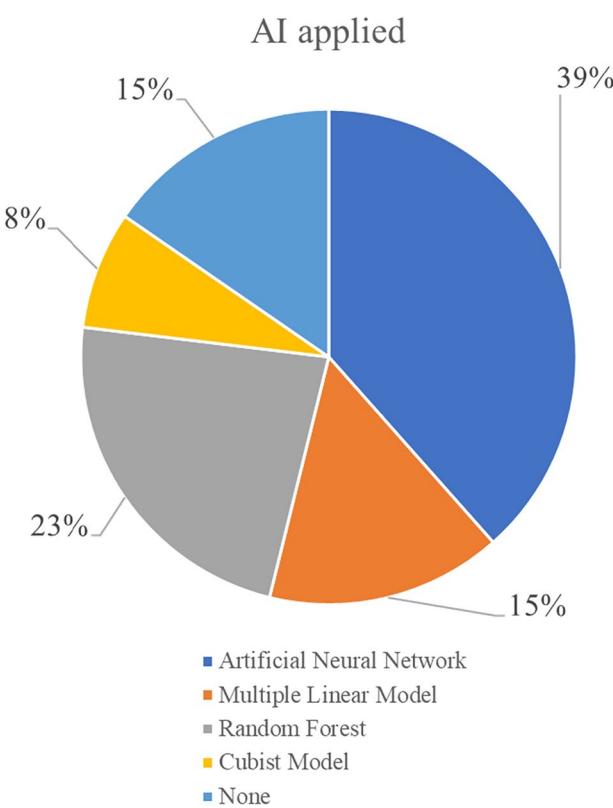
Table 2 Reviewed papers

Research reviewed		Ref	Year	Authors	Principle aspect	Category	Principle input	AI techniques applied
[105]	2021	de Andrade et al.	the authors propose guidelines for assembling an Artificially Intelligent Soil Quality Index, giving as input to the ML model data on recent land managements, current genomic features, current and projected plant-cover, time and current soil biological, chemical and physical parameters	Rewiev/Opinion paper	Microbiome data	None	None	
[125]	2022	Correa-Garcia et al.	the authors discuss key criteria that differentiate the various statistical learning approaches and propose Supervised and Unsupervised learning methods to unlock forecasting potential of the microbiome for ecosystem processes	Rewiev/Opinion paper	Microbiome data	None	None	
[126]	2022	Wilhelm et al.	the authors use 16S rRNA gene information to predict 12 measures of soil health through the application of supervised Machine Learning techniques	Original research paper	Microbiome data	Random Forest and Support Vector Machine regression and classification models	Random Forest and Support Vector Machine regression and classification models	
[127]	2020	Hermans et al.	the authors assess the degree to which bacterial communities in cultivated soils can forecast soil physico-chemical characteristics	Original research paper	Microbiome data	Random Forest model	Random Forest model	
[128]	2017	Chang et al.	the authors show that crop productivity differences were associated with bulk soil microbiome composition and highlighted several nitrogen utility-related taxa	Original research paper	Microbiome data	Random Forest model	Random Forest model	
[129]	2023	Sadeghi et al.	the authors used soil physical and chemical properties and total phospholipid fatty acid (PLFA) analysis to predict soil biological properties, and found a relationship between soil microbial characteristics and the physical and chemical properties of soil	Original research paper	Soil physical and chemical metrics	Cubist model algorithm	Cubist model algorithm	
[130]	2017	Ebrahim et al.	the authors evaluate the population of Azotobacter in soils with different land uses	Original research paper	Soil physical and chemical metrics	Artificial Neural Networks and Multivariate Linear Regression approaches	Artificial Neural Networks and Multivariate Linear Regression approaches	

Table 2 (continued)

Research reviewed		Principle aspect	Category	Principle input	AI techniques applied	
Ref	Year	Authors				
[131]	2021	Pellegrini et al.	the authors predict soil microbial biomass from a range of physical and chemical properties	Original research paper	Soil biological physical and chemical metrics	Artificial Neural Network model approach
[132]	2012	Tajik et al.	the authors use soil and topographic attributes to predict the activity of three soil enzymes, L-asparaginase, L-glutaminase, and urease	Original research paper	Digital Terrain Analysis	Artificial Neural Networks and Multiple Linear Regression approaches
[133]	2012	Larsen et al.	the authors describe a bioclimatic modeling approach to predict microbial community structure as a function of environmental parameters and microbial interactions	Original research paper	Environmental parameters and microbial interactions	Artificial Neural Network model approach
[134]	2014	Santos et al.	the authors identify patterns in soil chemical variables and microbial community structures across forest soils in three regions of Brazil's Atlantic Forest	Original research paper	Soil physical and chemical metrics	Artificial Neural Network model approach
[135]	2018	Ramirez et al.	the authors merge 30 independent bacterial data sets comprising 1998 soil samples from 21 countries and show that disparate amplicon sequence data can be combined at the taxonomy-based level to assess bacterial community structure	Original research paper	Bacterial community	Random Forest models

Fig. 1 Classification of the papers based on the AI techniques applied



All the above-mentioned papers are summarized in Table 2, with additional explanations of the main terms used in this paper in Table S2.

1.5 AI techniques applicable to soil microbiome analysis

The analysis of microbiome features can encompass a range of computational methods, from conventional statistical techniques to more sophisticated algorithms employed by AI. The first applications of AI to microbiota studies were used to investigate the role of human gut microbiota in human health: AI techniques such as ML and DL algorithms were applied for the diagnosis and therapy of critical diseases, with significant potential for improving public health [136–140]. An overview of ML methods for human microbiome analysis is given in Namkung (2020) [141].

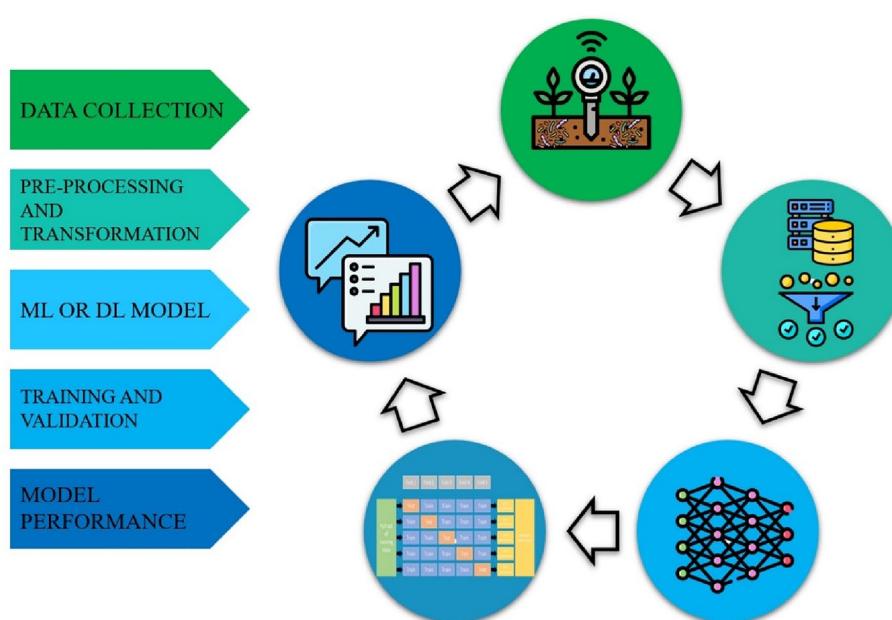
Oh and Zhang (2020) [137] used autoencoders to convert microbiome data into low-dimensional representations, thereby enhancing the ability to predict human diseases. Curry et al. (2021) [142] also investigated ML techniques for disease detection from the gut microbiome, highlighting further computational and biological challenges. Morton et al. (2019) [143] presented a training process that aims to estimate the likelihood of detecting a metabolite given the provided microbe sequence through an Artificial Neural Network (ANN) approach. Tataru and David (2020) [144] used word-embedding techniques to analyze microbiome data and to better understand microbial interactions in human diseases. Costello and Martin (2018) [145] demonstrated how ML can predict metabolic pathways from time-series multi-omics data, thereby eliminating the need for traditional kinetic models. Metwally et al. (2019) [146] used longitudinal microbiome data to demonstrate the advantages of whole genome shotgun sequencing over amplicon sequencing, providing more insight into the microbiome's impact over time. Iadanza et al. (2020) [136] explored the use of AI in the identification and management of diseases by analyzing the gut microbiota. They emphasized the importance of AI in interpreting complex data sets. Liu et al. (2022) [147] created a ML framework to identify gut microbiome biomarkers across multiple geographic regions, utilizing ensemble learning and reinforcement learning techniques for personalized modulation analysis. Li et al. (2022) [148] investigated the use of multi-omics data to improve the accuracy of disease prediction models. They used ML techniques to study the gut microbiota and its effects on host metabolism. In summary, autoencoders and other ANN architectures have proven to be efficient in handling the complex and diverse

characteristics of microbiome data in medical science. This is critical for accurately predicting phenotypes and identifying biomarkers, as demonstrated by Hernández Medina et al. (2022) [149].

Expanding the use of ML techniques from medical to agricultural microbiome analysis presents numerous challenges, including data collection, sparsity, and biomarker identification. Soil samples differ significantly from controlled medical settings due to many factors, e.g. weather, topography, and host organism diversity. Furthermore, soil microbiome datasets are frequently smaller and less dense than human ones, thus needing more sophisticated data augmentation and handling techniques to avoid overfitting. While many studies on the human microbiome focus on disease prediction, the goal of agricultural applications often is the identification of biomarkers associated with soil fertility or crop yield. Nevertheless, the aforementioned medical examples can serve as a reference for applying AI techniques to study the soil microbiome. ML algorithms can significantly contribute to the advancement of our knowledge about soil health by analyzing data on microbial communities and making predictions about different environmental indicators [150].

The procedure for analyzing soil microbiome data using AI techniques adheres to a systematic sequence of steps commonly employed in bioinformatics and computational biology research (Fig. 2). Data pre-processing is an essential step in microbiome analysis to adequately prepare the input data for accurate ML and DL modeling. Normalization of microbiome data is necessary to adjust for differences in sequencing depths among samples. Various techniques such as rarefaction, total-sum scaling, or more advanced methods can be employed for this purpose [151]. To address skewness, logarithmic or other transformations such as square root can be employed. Missing values can be managed through techniques like k-nearest neighbors (k-NN), multiple imputation, or by assigning zero or minimal values, depending on the specific circumstances. In addition, when analyzing data that encompasses different taxa, creating new features might help in capturing more complex relationships that may not be apparent from the original data alone. Techniques such as Principal Component Analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP) can be used to reduce the number of features and uncover structure in the data that may be related to the phenotype of interest. Dimensionality reduction is accomplished by using autoencoders to address the issue of high dimensionality in microbiome data. This method greatly accelerates model training and hyperparameter optimization [137]. After pre-processing, the next step is to choose and train an appropriate ML or DL model based on the hypothesis and nature of the data. Random Forests, Gradient Boosting Machines, and Support Vector Machines are common models for microbiome studies, as well as ANN architectures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Autoencoders for DL tasks. DL models, particularly CNNs and RNNs, are used to analyze complex microbiome datasets and predict diseases or phenotypic traits based on microbiome composition. These models are specifically intended to handle the massive amounts of data generated by high-throughput sequencing technologies [137]. Implementing cross-validation is fundamental because it helps ensure that the model does not just perform well on the training data but also generalizes well to new, unseen data, using techniques such as k-fold cross-validation, grid search, random search, or Bayesian optimization, that can help find the optimal model

Fig. 2 Principal phases of developing a Machine Learning (ML) or Deep Learning (DL) model



settings [152]. Additionally, overfitting can be avoided by training the model on the training dataset and evaluating its performance on a validation set. The outcomes of the model evaluation, which occurs after it has been trained, can be interpreted within the framework of microbiome research. Metrics such as accuracy, precision, recall, F1-score for classification tasks, and mean squared error (MSE) or root mean squared error (RMSE) for regression tasks are computed to evaluate the performance of the model, depending on the specific task at hand. Models such as feature importance from tree-based models, SHAP values (Shapley Additive exPlanations), or permutation importance are among the commonly used to ascertain the taxa, features, or functions that are influencing the predictions of the model. By integrating diverse omics data, including metabolomics and metagenomics, and applying ML methodologies to this information, predictive models become more accurate. By adopting this workflow researchers can gain comprehensive insights into how microorganisms interact with one another and how these interactions influence their broader environment, whether this is a specific ecosystem, agricultural setting, or human body [137]. An overview of ML approaches useful for analyzing microbiome data is available in the paper by Papoutsoglou et al. (2023) [153]. A comprehensive review of utilizing DL techniques for analyzing environmental microbiome data is also provided by Deng et al. (2021) [154]. This last review outlines the advancements in DL techniques and elucidates their advantages over classical ML methods. The insights gained can be applied to predict outcomes or classify new samples in environmental settings. This workflow must be modified to reflect specific project goals, available data, and computational resources. Each step should be carefully planned and carried out to ensure the conclusions are reliable and valid.

A significant step forward would be the creation of AI-based forecasting models that integrate physical, chemical, and biological soil characteristics as well as agrometeorological data, allowing for the correlation of variation in microbial community composition with different levels of soil health. Most of the bases are in place to complete this step, but much work remains to be done on collecting data, soil parameter monitoring, parameter correlation, and mathematical model training for prediction. The initial step involves systematically collecting detailed information over time at regular intervals, focusing on soil properties such as pH, nutrient levels, organic matter content, and microbial composition. Scheduling and tailoring the samplings at consistent intervals (e.g., monthly, seasonally) to capture temporal variations, collecting them from multiple locations within the study area to account for spatial variability. This information could be obtained through classical soil sampling and analysis, and with the help of remote sensing technologies. Also, it is important to use a centralized database to store all collected data and record contextual information about sampling methods, locations, and conditions to provide context for data interpretation. Understanding the interactions between soil microorganisms and their surroundings is critical: the roles of various microbial species in soil health, their functions, and how they interact with plants can provide useful information. Using advanced molecular techniques such as metagenomics, researchers can identify beneficial microorganisms found in healthy soils, such as nitrogen-fixing bacteria, mycorrhizal fungi, and other symbiotic microbes that improve soil fertility and plant growth. Algorithms can be developed based on the collected data and understanding of microbial ecology to recommend specific microbial inoculants or management practices that are tailored to the soil's unique characteristics and the crops grown. By incorporating these elements into a DSS, farmers will be able to make more informed decisions that promote sustainable agriculture and productivity. Furthermore, farmers' feedback on the effectiveness of recommended practices and observed changes in soil health can be used to fine-tune the algorithms over time, which is critical for the DSS ongoing improvement. This entire process will take a long time to develop, but the results will be significant for farmers and the global agricultural sector.

The use of AI in microbiome research could revolutionize our understanding and control over microbial communities.

2 Concluding remarks

Globally, soils face severe degradation due to unsustainable agricultural practices, deforestation, and climate change impacts, leading to declines in soil fertility, erosion, and reductions in biodiversity (FAO on soil recarbonization practices: The Nature Conservancy <https://www.nature.org/en-us>).

The use of AI to predict soil health through microbiome information could represent a significant advancement in soil science. Microbiome data is inherently complex, with numerous microbial species interacting in intricate ways. Traditional analysis methods can be time-consuming and may overlook the full depth of these interactions. Therefore, microbiome data manipulation relies on key techniques to manage, analyze, and interpret the complex, high-dimensional datasets produced by sequencing technologies.

AI models can efficiently process vast amounts of data and identify patterns that may be difficult for humans to discern. Currently, the main examples of data manipulation on the microbiome are in medicine, while soil science

appears to be a new and still open field for AI applications. ML techniques, including RFs, SVMs, and ANNs, are widely used in microbial community classification, clustering, and prediction modeling. CNNs excel at interpreting microbial abundance data by transforming it into structured formats, whereas RNNs are especially effective at analyzing time-series data, such as modeling microbial succession dynamics. One significant advantage of AI models is their ability to constantly learn and adapt as new data becomes available. This adaptability is critical for studying soil microbiomes, which live in constantly changing environments. As new information becomes available, models can be refined to provide more accurate predictions of soil health. Furthermore, as discussed by Cuomo et al. (2024) [155], novel techniques such as Physics-Informed Neural Networks (PINNs) show promise in modeling the growth of microbial populations while accounting for environmental factors, providing a powerful tool for studying microbial dynamics.

Certainly, the advent of AI-based tools has the potential to significantly reduce the time and human resources needed. Future developments should prioritize collaboration between researchers, farmers, and industries for developing standardized datasets and protocols. This can improve the reliability and interoperability of AI models across different regions and agricultural systems, ensuring broader applicability and adoption. Understanding how well AI predicts soil health is essential for gaining the trust of farmers and other stakeholders, and is crucial for the widespread acceptance and adoption of AI-driven technologies.

In conclusion, the integration of AI for the prediction of soil health through microbiome information holds great promise for sustainable agriculture. Continued research, technological advancements, and collaborative efforts are essential to unlock the full potential of AI in optimizing soil management practices for a more resilient and productive agricultural future.

Acknowledgements This study was carried out within a National PhD in "Artificial Intelligence—Agrifood and Environment Area" co-funded by EVJA s.r.l.

Author contributions Conceptualisation: R.P., M.M.M. and M.R.; writing—original draft preparation: R.P., M.M.M. and M.R.; writing—review and editing: R.P., V.S.C., M.M.M., A.A., S.C., F.L. and M.R.; supervision: M.R.; funding acquisition: M.M.M. and M.R. All authors have read and agreed to the published version of the manuscript.

Funding This study was carried out within the Agritech National Research Center and received funding from the European Union Next-Generation EU (Piano Nazionale di Ripresa e Resilienza (PNRR)—missione 4 componente 2, investimento 1.4—d.d. 1032 17/06/2022, cn00000022); and within the MISURA 16 "Cooperazione" SOTTOMISURA 16.2 – Tipologia di operazione 16.2.1 "Sostegno a progetti pilota e allo sviluppo di nuovi prodotti, pratiche, processi e tecnologie", CUP F59H23000010009. This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

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

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Montgomery DR. Soil erosion and agricultural sustainability. PNAS. 2007. <https://doi.org/10.1073/pnas.0611508104>.
2. Lal R. Restoring soil quality to mitigate soil degradation. Sustainability. 2015. <https://doi.org/10.3390/su7055875>.
3. Yin C, Zhao W, Pereira P. Soil conservation service underpins sustainable development goals. GECCO. 2022. <https://doi.org/10.1016/j.gecco.2021.e01974>.
4. Springmann M, Clark M, Mason-D'Croz D, Wiebe K, Bodirsky BL, Lassaletta L, de Vries W, Vermeulen SJ, Herrero M, Carlson KM, Jonell M, Troell M, DeClerck F, Gordon LJ, Zurayk R, Scarborough P, Rayner M, Loken B, Fanzo J, Charles HJG, Tilman D, Rockström J, Willett W. Options for keeping the food system within environmental limits. Nature. 2018. <https://doi.org/10.1038/s41586-018-0594-0>.

5. Giller KE, Delaune T, Silva JV, Descheemaeker K, van de Ven G, Schut AG, van Wijk M, Hammond J, Hochman Z, Taulya G, Chikowo R, Narayanan S, Kishore A, Bresciani F, Mancini Teixeira H, Andersson AJ, van Ittersum MK. The future of farming: who will produce our food? *Food Secur.* 2021. <https://doi.org/10.1007/s12571-021-01184-6>.
6. Büinemann EK, Bongiorno G, Bai Z, Creamer RE, De Deyn G, de Goede R, Fleskens L, Geissen V, Kuyper TW, Mäder P, Pulleman M, Sukkel W, Groenigen JW, Brussaard L. Soil quality—a critical review. *Soil Biol Biochem.* 2018. <https://doi.org/10.1016/j.soilbio.2018.01.030>.
7. Laishram J, Saxena KG, Maikhuri RK, Rao KS. Soil quality and soil health: a review. *Int J Ecol Environ Sci.* 2012;38(1):19–37.
8. Allen WJ, Bufford JL, Barnes DA, Barratt BIP, Deslippe JR, Dickie IA, Goldson SL, Howlett BG, Hulme PE, Lavorel S, O'Brien SA, Waller LP, Tylianakis JM. A network perspective for sustainable agroecosystems. *Trends Plant Sci.* 2022. <https://doi.org/10.1016/j.tplants.2022.04.002>.
9. Lehmann J, Bossio DA, Kögel-Knabner I, Rillig MC. The concept and future prospects of soil health. *Nat Rev Earth Environ.* 2020. <https://doi.org/10.1038/s43017-020-0080-8>.
10. Doran JW, Zeiss MR. Soil health and sustainability: managing the biotic component of soil quality. *Appl Soil Eco.* 2000. [https://doi.org/10.1016/S0929-1393\(00\)00067-6](https://doi.org/10.1016/S0929-1393(00)00067-6).
11. Schloter M, Nannipieri P, Sørensen SJ, van Elsas JD. Microbial indicators for soil quality. *Biol Fertil Soils.* 2018. <https://doi.org/10.1007/s00374-017-1248-3>.
12. Ferreira CS, Seifollahi-Aghmuni S, Destouni G, Ghajarnia N, Kalantari Z. Soil degradation in the European Mediterranean region: processes, status and consequences. *Sci Total Environ.* 2022. <https://doi.org/10.1016/j.scitotenv.2021.150106>.
13. Wagg C, Bender SF, Widmer F, Van Der Heijden MG. Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proc Natl Acad Sci.* 2014. <https://doi.org/10.1073/pnas.1320054111>.
14. Zhang J, Van Der Heijden MG, Zhang F, Bender SF. Soil biodiversity and crop diversification are vital components of healthy soils and agricultural sustainability. *FASE.* 2020. <https://doi.org/10.5167/uzh-193808>.
15. McLaughlin A, Mineau P. The impact of agricultural practices on biodiversity. *Agric Ecosyst Environ.* 1995. [https://doi.org/10.1016/0167-8809\(95\)00609-V](https://doi.org/10.1016/0167-8809(95)00609-V).
16. Bertola M, Ferrarini A, Visioli G. Improvement of soil microbial diversity through sustainable agricultural practices and its evaluation by-omics approaches. A perspective for the environment, food quality, and human safety. *Microorganisms.* 2021. <https://doi.org/10.3390/microorganisms9071400>.
17. Cárcelés Rodríguez B, Durán-Zuazo VH, Soriano Rodríguez M, García-Tejero IF, Gálvez Ruiz B, Cuadros TS. Conservation agriculture as a sustainable system for soil health: a review. *Soil Syst.* 2022. <https://doi.org/10.3390/soilsystems6040087>.
18. European Commission and Directorate-General for Research and Innovation and Bizzo G, Fabbri K, Gajdzinska M, Haentjens W, Lueth D, Lutzeyer H, Luvarà G, Meloni G, Molina Villanueva J, Sainz Lopez N, Schoumacher C, Zampoukas N. Food 2030—Pathways for action 2.0—R&I policy as a driver for sustainable, healthy, climate resilient and inclusive food systems. Publications Office of the European Union. 2023; <https://doi.org/10.2777/365011>.
19. European Commission and European Research Council Executive Agency. Mapping ERC frontier research sustainable food production and consumption. Publications Office of the European Union. 2023; <https://doi.org/10.2828/106806>.
20. Ortiz A, Sansinenea E. The role of beneficial microorganisms in soil quality and plant health. *Sustainability.* 2022. <https://doi.org/10.3390/su14095358>.
21. Mohanty LK, Singh NK, Raj P, Prakash A, Tiwari AK, Singh V, Sachan P. Nurturing crops, enhancing soil health, and sustaining agricultural prosperity worldwide through agronomy. *J Exp Agric Int.* 2024. <https://doi.org/10.9734/JEAI/2024/v46i22308>.
22. Creamer RE, Brennan F, Fenton O, Healy MG, Lalor ST, Lanigan GJ, Regan JT, Griffiths BS. Implications of the proposed Soil Framework Directive on agricultural systems in Atlantic Europe—a review. *Soil Use Manag.* 2010. <https://doi.org/10.1111/j.1475-2743.2010.00288.x>.
23. Chaparro JM, Sheflin AM, Manter DK, Vivanco JM. Manipulating the soil microbiome to increase soil health and plant fertility. *Biol Fertil Soils.* 2012. <https://doi.org/10.1007/s00374-012-0691-4>.
24. Shahane AA, Shivay YS. Soil health and its improvement through novel agronomic and innovative approaches. *Front Agron.* 2021. <https://doi.org/10.3389/fagro.2021.680456>.
25. Kumar G, Baweja P, Gandhi PB. Impact of anthropogenic activities on soil patterns and diversity. *Pedosphere.* 2022. https://doi.org/10.1007/978-981-16-8770-9_13.
26. Norris CE, Congreves KA. Alternative management practices improve soil health indices in intensive vegetable cropping systems: a review. *Front Environ Sci.* 2018. <https://doi.org/10.3389/fenvs.2018.00050>.
27. Montoya D, Gaba S, de Mazancourt C, Bretagnolle V, Loreau M. Reconciling biodiversity conservation, food production, and farmers' demand in agricultural landscapes. *Ecol Model.* 2020. <https://doi.org/10.1016/j.ecolmodel.2019.108889>.
28. Dönmez D, Isak MA, İzgür T, Şimşek Ö. Green horizons: navigating the future of agriculture through sustainable practices. *Sustainability.* 2024. <https://doi.org/10.3390/su16083505>.
29. Liu B, Tu C, Hu S, Gumpertz M, Ristaino JB. Effect of organic, sustainable, and conventional management strategies in grower fields on soil physical, chemical, and biological factors and the incidence of Southern blight. *Appl Soil Ecol.* 2007. <https://doi.org/10.1016/j.apsoil.2007.06.007>.
30. Rodrigo-Comino J, López-Vicente M, Kumar V, Rodríguez-Seijo A, Valkó O, Rojas C, Pourghasemi HR, Salvati L, Bakr N, Vaudour E, Brevik EC, Radziemska M, Pulido M, Di Prima S, Dondini M, de Vries W, Santos ES, de Lourdes M-S, Yu Y, Panagos P. Soil science challenges in a new era: a transdisciplinary overview of relevant topics. *Air Soil Water Res.* 2020. <https://doi.org/10.1177/11786221209774>.
31. Lehmann J, Hansel CM, Kaiser C, Kleber M, Maher K, Manzoni S, Nunan N, Reichstein M, Schimel JP, Torn MS, Wieder WR, Kögel-Knabner I. Persistence of soil organic carbon caused by functional complexity. *Nat Geosci.* 2020. <https://doi.org/10.1038/s41561-020-0612-3>.
32. Hou D, Bolan NS, Tsang DC, Kirkham MB, O'connor D. Sustainable soil use and management: an interdisciplinary and systematic approach. *Sci Total Environ.* 2020. <https://doi.org/10.1016/j.scitotenv.2020.138961>.
33. Rinot O, Levy GJ, Steinberger Y, Svoray T, Eshel G. Soil health assessment: a critical review of current methodologies and a proposed new approach. *Sci Total Environ.* 2019. <https://doi.org/10.1016/j.scitotenv.2018.08.259>.
34. Vogel HJ, Bartke S, Daedlow K, Helming K, Kögel-Knabner I, Lang B, Rabot E, Russell D, Stößel B, Weller U, Wiesmeier M, Wollschläger U. A systemic approach for modeling soil functions. 2018. *Soil.* <https://doi.org/10.5194/soil-4-83-2018>.

35. Evangelista SJ, Field DJ, McBratney AB, Minasny B, Ng W, Padarian J, Dobarco MR, Wadoux AMC. A proposal for the assessment of soil security: Soil functions, soil services and threats to soil. *Soil Security*. 2023. <https://doi.org/10.1016/j.soisec.2023.100086>.
36. Welbaum GE, Sturz AV, Dong Z, Nowak J. Managing soil microorganisms to improve productivity of agro-ecosystems. *Crit Rev Plant Sci*. 2004. <https://doi.org/10.1080/07352680490433295>.
37. de Salamone IEG. Use of soil microorganisms to improve plant growth and ecosystem sustainability. In: The molecular basis of plant genetic diversity. 2012. pp. 233–258.
38. Buscot F, Varma A. Microorganisms in soils: roles in genesis and functions. 2005; <https://doi.org/10.1007/b137872>.
39. Aislabie J, Deslippe JR, Dymond J. Soil microbes and their contribution to soil services. *Ecosystem services in New Zealand—conditions and trends*. 2013;1(12): 143–161.
40. Maron PA, Sarr A, Kaisermann A, Lévéque J, Mathieu O, Guigue J, Karimi B, Bernard L, Dequiedt S, Terrat S, Chabbi A, Ranjard L. High microbial diversity promotes soil ecosystem functioning. *Appl Environ Microbiol*. 2018. <https://doi.org/10.1128/AEM.02738-17>.
41. Aguilar-Paredes A, Valdés G, Nuti M. Ecosystem functions of microbial consortia in sustainable agriculture. *Agronomy*. 2020. <https://doi.org/10.3390/agronomy10121902>.
42. Banerjee S, van der Heijden MG. Soil microbiomes and one health. *Nat Rev Microbiol*. 2023. <https://doi.org/10.1038/s41579-022-00779-w>.
43. Hendriksen NB. Microbial biostimulants—the need for clarification in EU regulation. *Trends Microbiol*. 2022. <https://doi.org/10.1016/j.tim.2022.01.008>.
44. Bonanomi G, De Filippis F, Zotti M, Idbella M, Cesaroni G, Al-Rowaili S, Abd-ElGawad A. Repeated applications of organic amendments promote beneficial microbiota, improve soil fertility and increase crop yield. *Appl Soil Ecol*. 2020. <https://doi.org/10.1016/j.apsoil.2020.103714>.
45. Ball BC, Bingham I, Rees RM, Watson CA, Litterick A. The role of crop rotations in determining soil structure and crop growth conditions. *Can J Soil Sci*. 2005. <https://doi.org/10.4141/S04-078>.
46. Hartmann M, Six J. Soil structure and microbiome functions in agroecosystems. *Nat Rev Earth Environ*. 2023. <https://doi.org/10.1038/s43017-022-00366-w>.
47. Trivedi P, Mattupalli C, Eversole K, Leach JE. Enabling sustainable agriculture through understanding and enhancement of microbiomes. *New Phytol*. 2021. <https://doi.org/10.1111/nph.17319>.
48. Dubey A, Malla MA, Khan F, Chowdhary K, Yadav S, Kumar A, Sharma S, Pramod K, Khare PK, Khan ML. Soil microbiome: a key player for conservation of soil health under changing climate. *Biodivers Conserv*. 2019. <https://doi.org/10.1007/s10531-019-01760-5>.
49. Jing J, Cong WF, Bezemer TM. Legacies at work: plant–soil–microbiome interactions underpinning agricultural sustainability. *Trends Plant Sci*. 2022. <https://doi.org/10.1016/j.tplants.2022.05.007>.
50. Bardgett RD, Van Der Putten WH. Belowground biodiversity and ecosystem functioning. *Nature*. 2014. <https://doi.org/10.1038/nature13855>.
51. Bardgett RD, Caruso T. Soil microbial community responses to climate extremes: resistance, resilience, and transitions to alternative states. *Philos Trans R Soc B*. 2020. <https://doi.org/10.1098/rstb.2019.0112>.
52. Ali S, Moon YS, Hamayun M, Khan MA, Bibi K, Lee IJ. Pragmatic role of microbial plant biostimulants in abiotic stress relief in crop plants. *J Plant Interact*. 2022. <https://doi.org/10.1080/17429145.2022.2091801>.
53. Berendsen RL, Pieterse CM, Bakker PA. The rhizosphere microbiome and plant health. *Trends Plant Sci*. 2012. <https://doi.org/10.1016/j.tplants.2012.04.001>.
54. Tahat MM, Alananbeh KM, Othman YA, Leskovar DI. Soil health and sustainable agriculture. *Sustainability*. 2020. <https://doi.org/10.3390/su12124859>.
55. Iftikhar A, Farooq R, Akhtar M, Khalid H, Hussain N, Ali Q, Ul Malook S, Ali D. Ecological and sustainable implications of phosphorous-solubilizing microorganisms in soil. *Discover Appl Sci*. 2024. <https://doi.org/10.1007/s42452-024-05683-x>.
56. Balestrini R, Chitarra W, Fotopoulos V, Ruocco M. Potential role of beneficial soil microorganisms in plant tolerance to abiotic stress factors. In: *Soil biological communities and ecosystem resilience*. Springer International Publishing. 191–207, 2017; https://doi.org/10.1007/978-3-319-63336-7_12.
57. Caporale AG, Vitaglione P, Troise AD, Pigna M, Ruocco M. Influence of three different soil types on the interaction of two strains of *Trichoderma harzianum* with *Brassica rapa* subsp. *sylvestris* cv. *esculentana*, under soil mineral fertilization. *Geoderma*. 2019. <https://doi.org/10.1016/j.geoderma.2019.05.003>.
58. Wallenstein MD. Managing and manipulating the rhizosphere microbiome for plant health: a systems approach. *Rhizosphere*. 2017. <https://doi.org/10.1016/j.rhisph.2017.04.004>.
59. Enespa, Chandra P. Tool and techniques study to plant microbiome current understanding and future needs: an overview. *Commun Integr Biol*. 2022. <https://doi.org/10.1080/19420889.2022.2082736>.
60. Manfredini A, Malusà E, Costa C, Pallottino F, Mocali S, Pinzari F, Canfora L. Current methods, common practices, and perspectives in tracking and monitoring bioinoculants in soil. *Front Microbiol*. 2021. <https://doi.org/10.3389/fmicb.2021.698491>.
61. Philippot L, Ritz K, Pandard P, Hallin S, Martin-Laurent F. Standardisation of methods in soil microbiology: progress and challenges. *FEMS Microbiol Ecol*. 2012. <https://doi.org/10.1111/j.1574-6941.2012.01436.x>.
62. Penton CR, Gupta VV, Yu J, Tiedje JM. Size matters: assessing optimum soil sample size for fungal and bacterial community structure analyses using high throughput sequencing of rRNA gene amplicons. *Front Microbiol*. 2016. <https://doi.org/10.3389/fmicb.2016.00824>.
63. König S, Vogel HJ, Harms H, Worrlich A. Physical, chemical and biological effects on soil bacterial dynamics in microscale models. *Front Ecol Evol*. 2020. <https://doi.org/10.3389/fevo.2020.00053>.
64. Satam H, Joshi K, Mangrolia U, Waghoor S, Zaidi G, Rawool S, Thakare RP, Banday S, Alok K, Mishra AK, Das G, Malonia SK. Next-generation sequencing technology: current trends and advancements. *Biology*. 2023. <https://doi.org/10.3390/biology12070997>.
65. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol Ecol*. 2012. <https://doi.org/10.1111/j.1365-294X.2012.05470.x>.
66. Wutkowska M, Vader A, Mundra S, Cooper EJ, Eidesen PB. Dead or alive; or does it really matter? Level of congruency between trophic modes in total and active fungal communities in high arctic soil. *Front Microbiol*. 2019. <https://doi.org/10.3389/fmicb.2018.03243>.

67. Rubiola S, Macori G, Civera T, Fanning S, Mitchell M, Chiesa F. Comparison between full-length 16S rRNA metabarcoding and whole metagenome sequencing suggests the use of either is suitable for large-scale microbiome studies. *Foodborne Pathog Dis.* 2022. <https://doi.org/10.1089/fpd.2022.0027>.
68. Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. *Front Genet.* 2015. <https://doi.org/10.3389/fgene.2015.00348>.
69. Zapata JDD, Florez JEM, Alvarez DL. Metagenomics approaches to understanding soil health in environmental research-a review. *Soil Sci.* 2023. <https://doi.org/10.37501/soilsa/163080>.
70. Blagodatskaya E, Kuzyakov Y. Active microorganisms in soil: critical review of estimation criteria and approaches. *Soil Biol Biochem.* 2013. <https://doi.org/10.1016/j.soilbio.2013.08.024>.
71. Sofo A, Ricciuti P. A standardized method for estimating the functional diversity of soil bacterial community by Biolog® EcoPlates™ assay—the case study of a sustainable olive orchard. *Appl Sci.* 2019. <https://doi.org/10.3390/app9194035>.
72. Ma F, Wang C, Zhang Y, Chen J, Xie R, Sun Z. Development of microbial indicators in ecological systems. *Int J Environ Res Public Health.* 2022. <https://doi.org/10.3390/ijerph192113888>.
73. Das SK, Varma A. Role of enzymes in maintaining soil health. *Soil Biol.* 2011. https://doi.org/10.1007/978-3-642-14225-3_2.
74. Klose S, Tabatabai MA. Urease activity of microbial biomass in soils as affected by cropping systems. *Biol Fert Soils.* 2000. <https://doi.org/10.1007/s003740050645>.
75. Ekenler M, Tabatabai M. β -Glucosaminidase activity of soils: effect of cropping systems and its relationship to nitrogen mineralization. *Biol Fert Soils.* 2022. <https://doi.org/10.1007/s00374-002-0541-x>.
76. Fierer N, Wood SA, de Mesquita CPB. How microbes can, and cannot, be used to assess soil health. *Soil Biol Biochem.* 2021. <https://doi.org/10.1016/j.soilbio.2020.108111>.
77. Iven H, Meller S, Luster J, Frossard E. A novel in-situ soil enzymatic activity sensor-expanding soil precision measurements to indicator of soil health. In: EGU General Assembly Conference Abstracts. 2021; <https://doi.org/10.5194/egusphere-egu21-10708>.
78. Ruppert KM, Kline RJ, Rahman MS. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. *GECCO.* 2019. <https://doi.org/10.1016/j.gecco.2019.e00547>.
79. Dalca AV, Brudno M. Genome variation discovery with high-throughput sequencing data. *Brief Bioinform.* 2010. <https://doi.org/10.1093/bib/bbp058>.
80. Dey P. Polymerase Chain Reaction: Principle, Technique and Applications in Pathology. In: Basic and Advanced Laboratory Techniques in Histopathology and Cytology. 2018; https://doi.org/10.1007/978-981-10-8252-8_20.
81. Marudamuthu B, Sharma T, Purru S, Soam SK, Rao CS. Next-generation sequencing technology: a boon to agriculture. *Genet Resour Crop Evol.* 2023. <https://doi.org/10.1007/s10722-022-01512-5>.
82. Nazir A. Review on metagenomics and its applications. *Imp J Intersd Res.* 2016;2:277–86.
83. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol.* 2017. <https://doi.org/10.1038/nbt.3935>.
84. Ramesh M, Janani R, Deepa C, Rajeshkumar L. Nanotechnology-enabled biosensors: a review of fundamentals, design principles, materials, and applications. *Biosens.* 2022. <https://doi.org/10.3390/bios13010040>.
85. Javaid M, Haleem A, Singh RP, Suman R. Enhancing smart farming through the applications of Agriculture 4.0 technologies. *Int J Intell Netw.* 2022. <https://doi.org/10.1016/j.ijin.2022.09.004>.
86. Ang KLM, Seng JKP. Big data and machine learning with hyperspectral information in agriculture. *IEEE Access.* 2021. <https://doi.org/10.1109/ACCESS.2021.3051196>.
87. Bhat SA, Huang NF. Big data and AI revolution in precision agriculture: survey and challenges. *IEEE Access.* 2021. <https://doi.org/10.1109/ACCESS.2021.3102227>.
88. Cravero A, Pardo S, Sepúlveda S, Muñoz L. Challenges to use machine learning in agricultural big data: a systematic literature review. *Agronomy.* 2022. <https://doi.org/10.3390/agronomy12030748>.
89. Purnama S, Sejati W. Internet of things, big data, and artificial intelligence in the food and agriculture sector. *Int Trans Artif Intell.* 2023. <https://doi.org/10.33050/italic.v1i2.274>.
90. Oliveira LF, Moreira AP, Silva MF. Advances in agriculture robotics: a state-of-the-art review and challenges ahead. *Robotics.* 2021. <https://doi.org/10.3390/robotics10020052>.
91. Talari G, Cummins E, McNamara C, O'Brien J. State of the art review of Big Data and web-based Decision Support Systems (DSS) for food safety risk assessment with respect to climate change. *Trends Food Sci Technol.* 2022. <https://doi.org/10.1016/j.tifs.2021.08.032>.
92. Fenu G, Mallocci FM. DSS LANDS: a decision support system for agriculture in Sardinia. *HighTech Innov J.* 2020. <https://doi.org/10.28991/HIJ-2020-01-03-05>.
93. Folorunso O, Ojo O, Busari M, Adebayo M, Joshua A, Folorunso D, Okechukwu Ugwunna C, Olabanjo O, Olabanjo O. Exploring machine learning models for soil nutrient properties prediction: a systematic review. *BDCC.* 2023. <https://doi.org/10.3390/bdcc7020113>.
94. Araújo SO, Peres RS, Ramalho JC, Lidon F, Barata J. Machine learning applications in agriculture: current trends, challenges, and future perspectives. *Agronomy.* 2023. <https://doi.org/10.3390/agronomy13122976>.
95. Smircina DN, Bailey VL, Hofmockel KS. Micro on a macroscale: relating microbial-scale soil processes to global ecosystem function. *FEMS Microbiol Ecol.* 2021. <https://doi.org/10.1093/femsec/fiab091>.
96. Wagg C, Hautier Y, Pellkofer S, Banerjee S, Schmid B, van der Heijden MG. Diversity and asynchrony in soil microbial communities stabilizes ecosystem functioning. *Elife.* 2021. <https://doi.org/10.7554/elife.62813>.
97. Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: a review. *Sensors.* 2018. <https://doi.org/10.3390/s18082674>.
98. Kamilaris A, Prenafeta-Boldú FX. Deep learning in agriculture: a survey. *Comput Electron Agr.* 2018. <https://doi.org/10.1016/j.compag.2018.02.016>.
99. Arvanitis KG, Symeonaki EG. Agriculture 4.0: the role of innovative smart technologies towards sustainable farm management. *Open Agric.* 2020. <https://doi.org/10.2174/1874331502014010130>.

100. Ren C, Kim DK, Jeong D. A survey of deep learning in agriculture: techniques and their applications. *J Inf Process Syst.* 2020. <https://doi.org/10.3745/JIPS.04.0187>.
101. Santos L, Santos FN, Oliveira PM, Shinde P. Deep learning applications in agriculture: a short review. *Robot.* 2020. https://doi.org/10.1007/978-3-030-35990-4_12.
102. Benos L, Tagarakis AC, Dolias G, Berruto R, Kateris D, Bochtis D. Machine learning in agriculture: a comprehensive updated review. *Sensors.* 2021. <https://doi.org/10.3390/s21113758>.
103. Meshram V, Patil K, Meshram V, Hanchate D, Ramkuteke SD. Machine learning in agriculture domain: a state-of-art survey. *Artif Intell Life.* 2021. <https://doi.org/10.1016/j.ailsci.2021.100010>.
104. Naman A, Tiyndel GJ, Bhatia JK, Bhardwaj N, Rahul. Utilizing artificial intelligence (AI) for sustainable agriculture: precision farming as a catalyst for environmental conservation. *Int J Agric Extension Soc Dev.* 2024. <https://doi.org/10.33545/26180723.2024.v7.i3e.441>.
105. de Andrade VHGZ, Redmile-Gordon M, Barbosa BHG, Andreote FD, Roesch LFW, Pylro VS. Artificially intelligent soil quality and health indices for 'next generation' food production systems. *Trends Food Sci.* 2021. <https://doi.org/10.1016/j.tifs.2020.10.018>.
106. Camaréna S. Artificial intelligence in the design of the transitions to sustainable food systems. *J Clean Prod.* 2020. <https://doi.org/10.1016/j.jclepro.2020.122574>.
107. Bharman P, Saad SA, Khan S, Jahan I, Ray M, Biswas M. Deep learning in agriculture: a review. *Asian J Res Comput Sci.* 2022. <https://doi.org/10.9734/AJRROS/2022/v13i230311>.
108. Alahmad T, Neményi M, Nyéki A. Applying IoT sensors and big data to improve precision crop production: a review. *Agronomy.* 2023. <https://doi.org/10.3390/agronomy13102603>.
109. Raj R, Aravind A, Akshay VS, Chandy M, Sharun ND. A seed planting robot with two control variables. *ICOEI.* 2019. <https://doi.org/10.1109/ICOEI.2019.8862541>.
110. Srinivasan N, Prabhu P, Smruthi SS, Sivaraman NV, Gladwin SJ, Rajavel R, Natarajan AR. Design of an autonomous seed planting robot. *IEEE R10-HTC.* 2016. <https://doi.org/10.1109/R10-HTC.2016.7906789>.
111. Ku HH, Liu CH, Wang WC. Design of an artificial intelligence of things based indoor planting model for mentha spicata. *Processes.* 2022. <https://doi.org/10.3390/pr10010116>.
112. Fiorentini M, Schillaci C, Denora M, Zenobi S, Deligios PA, Santilocchi R, Perniola M, Ledda L, Orsini R. Fertilization and soil management machine learning based sustainable agronomic prescriptions for durum wheat in Italy. *Precis Agric.* 2024. <https://doi.org/10.1007/s11119-024-10153-w>.
113. Gao J, Zeng W, Ren Z, Ao C, Lei G, Gaiser T, Srivastava AK. A fertilization decision model for maize, rice, and soybean based on machine learning and swarm intelligent search algorithms. *Agronomy.* 2023. <https://doi.org/10.3390/agronomy13051400>.
114. Firmansyah E, Pardamean B, Ginting C, Mawandha HG, Putra DP, Suparyanto T. Development of artificial intelligence for variable rate application based oil palm fertilization recommendation system. *ICIMTech.* 2021. <https://doi.org/10.1109/ICIMTech53080.2021.9535082>.
115. Talaviya T, Shah D, Patel N, Yagnik H, Shah M. Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. *Artif Intell Agric.* 2020. <https://doi.org/10.1016/j.aiia.2020.04.002>.
116. Tace Y, Elfilali S, Tabaa M, Leghris C. Implementation of smart irrigation using IoT and artificial intelligence. *Math Model Comput.* 2023. <https://doi.org/10.23939/MMC2023.02.575>.
117. Benhmad T, Rhaimi CB, Alomari S, Aljuhani L. Design and implementation of an integrated iot and artificial intelligence system for smart irrigation management. *Int J Adv Soft Comput Appl.* 2024. <https://doi.org/10.15849/IJASCA.240330.12>.
118. Biswas S, Sharma LK, Ranjan R, Saha S, Chakraborty A, Banerjee JS. Smart farming and water saving-based intelligent irrigation system implementation using the internet of things. In *Recent trends in computational intelligence enabled research.* 2021; <https://doi.org/10.1016/B978-0-12-822844-9.00043-8>.
119. Klyushin D, Tymoshenko A. Optimization of drip irrigation systems using artificial intelligence methods for sustainable agriculture and environment. *Artificial intelligence for sustainable development: Theory, practice and future applications.* 2021; https://doi.org/10.1007/978-3-030-51920-9_1.
120. Bhoi SK, Jena KK, Panda SK, Long HV, Kumar R, Subbulakshmi P, Jebreen HB. An Internet of Things assisted Unmanned Aerial Vehicle based artificial intelligence model for rice pest detection. *Microprocess Microsyst.* 2021. <https://doi.org/10.1016/j.micpro.2020.103607>.
121. Balaska V, Adamidou Z, Vryzas Z, Gasteratos A. Sustainable crop protection via robotics and artificial intelligence solutions. *Machines.* 2023. <https://doi.org/10.3390/machines11080774>.
122. Li C, Wang M. Pest and disease management in agricultural production with artificial intelligence: innovative applications and development trends. *Adv Res.* 2024. https://doi.org/10.50908/arr.4.3_381.
123. Zhai Z, Martínez JF, Beltran V, Martínez NL. Decision support systems for agriculture 4.0: survey and challenges. *Comput Electron Agric.* 2020. <https://doi.org/10.2174/1874331502014010130>.
124. Bhagat PR, Naz F, Magda R. Artificial intelligence solutions enabling sustainable agriculture: a bibliometric analysis. *PLoS ONE.* 2022. <https://doi.org/10.1371/journal.pone.0268989>.
125. Correa-Garcia S, Constant P, Yergeau E. The forecasting power of the microbiome. *Trends Microbiol.* 2022. <https://doi.org/10.1016/j.tim.2022.11.013>.
126. Wilhelm RC, van Es HM, Buckley DH. Predicting measures of soil health using the microbiome and supervised machine learning. *Soil Biol Biochem.* 2022. <https://doi.org/10.1016/j.soilbio.2021.108472>.
127. Hermans SM, Buckley HL, Case BS, Curran-Cournane F, Taylor M, Lear G. Using soil bacterial communities to predict physico-chemical variables and soil quality. *Microbiome.* 2020. <https://doi.org/10.1186/s40168-020-00858-1>.
128. Chang HX, Haudenschild JS, Bowen CR, Hartman GL. Metagenome-wide association study and machine learning prediction of bulk soil microbiome and crop productivity. *Front Microbiol.* 2017. <https://doi.org/10.3389/fmicb.2017.00519>.
129. Sadeghi S, Petermann BJ, Steffan J, Brevik EC, Gedeon C. Predicting microbial responses to changes in soil physical and chemical properties under different land management. *Appl Soil Ecol.* 2023. <https://doi.org/10.1016/j.apsoil.2023.104878>.
130. Pellegrini E, Rovere N, Zaninotti S, Franco I, De Nobili M, Contin M. Artificial neural network (ANN) modelling for the estimation of soil microbial biomass in vineyard soils. *Biol Fertil Soils.* 2021. <https://doi.org/10.1007/s00374-020-01498-1>.

131. Tajik S, Ayoubi S, Nourbakhsh F. Prediction of soil enzymes activity by digital terrain analysis: comparing artificial neural network and multiple linear regression models. *Environ Eng Sci.* 2012. <https://doi.org/10.1089/ees.2011.0313>.
132. Larsen PE, Field D, Gilbert JA. Predicting bacterial community assemblages using an artificial neural network approach. *Nat Methods.* 2012. <https://doi.org/10.1038/nmeth.1975>.
133. Santos EC, Armas ED, Crowley D, Lambais MR. Artificial neural network modeling of microbial community structures in the Atlantic Forest of Brazil. *Soil Biol Biochem.* 2014. <https://doi.org/10.1016/j.soilbio.2013.10.049>.
134. Ebrahimi M, Sinegan AAS, Sarikhani MR, Mohammadi SA. Comparison of artificial neural network and multivariate regression models for prediction of Azotobacteria population in soil under different land uses. *Comput Electron Agric.* 2017. <https://doi.org/10.1016/j.compag.2017.06.019>.
135. Ramirez KS, Knight CG, De Hollande M, Brearley FQ, Constantinides B, Cotton A, Creer S, Crowther TW, Davison J, Delgado-Baquerizo M, Dorrepaal E, Elliott DR, Fox G, Griffiths RI, Hale C, Hartman K, Houlden A, Jones DL, Krab EJ, Maestre FT, McGuire KL, Monteux S, Orr CH, van der Putten WH, Roberts IS, Robinson DA, Rocca JD, Rowntree J, Schlaepi K, Shepherd M, Singh BK, Straathof AL, Bhatnagar JM, Thion C, van der Heijden MGA, De Vries FT. Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nat Microbiol.* 2018. <https://doi.org/10.1038/s41564-017-0062-x>.
136. Iadanza E, Fabbri R, Bašić-Čišak D, Amedei A, Telalovic JH. Gut microbiota and artificial intelligence approaches: a scoping review. *Health Technol.* 2020. <https://doi.org/10.1007/s12553-020-00486-7>.
137. Oh M, Zhang L. DeepMicro: deep representation learning for disease prediction based on microbiome data. *Sci Rep.* 2020. <https://doi.org/10.1038/s41598-020-63159-5>.
138. Zeng T, Yu X, Chen Z. Applying artificial intelligence in the microbiome for gastrointestinal diseases: a review. *J Gastroenterol Hepatol.* 2021. <https://doi.org/10.1111/jgh.1550>.
139. Kumar P, Sinha R, Shukla P. Artificial intelligence and synthetic biology approaches for human gut microbiome. *Crit Rev Food Sci Nutr.* 2022. <https://doi.org/10.1080/10408398.2020.1850415>.
140. Vilne B, Kibilds J, Siksnas I, Lazda I, Valciņa O, Krūmiņa A. Could artificial intelligence/machine learning and inclusion of diet-gut microbiome interactions improve disease risk prediction? Case study: coronary artery disease. *Front Microbiol.* 2022. <https://doi.org/10.3389/fmicb.2022.627892>.
141. Namkung J. Machine learning methods for microbiome studies. *J Microbiol.* 2020. <https://doi.org/10.1007/s12275-020-0066-8>.
142. Curry KD, Nute MG, Treangen TJ. It takes guts to learn: machine learning techniques for disease detection from the gut microbiome. *Emerg Top Life Sci.* 2021. <https://doi.org/10.1042/ETLS20210213>.
143. Morton JT, Aksenen AA, Nothias LF, Foulds JR, Quinn RA, Badri MH, Swenson TL, Van Goethem MW, Northen TR, Vazquez-Baeza Y, Wang M, Bokulich NA, Watters A, Song SJ, Bonneau R, Dorrestein PC, Knight R. Learning representations of microbe–metabolite interactions. *Nat Methods.* 2019. <https://doi.org/10.1038/s41592-019-0616-3>.
144. Tataru CA, David MM. Decoding the language of microbiomes using word-embedding techniques, and applications in inflammatory bowel disease. *PLoS Comput Biol.* 2020. <https://doi.org/10.1371/journal.pcbi.1007859>.
145. Costello Z, Martin HG. A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *NPJ Syst Biol Appl.* 2018. <https://doi.org/10.1038/s41540-018-0054-3>.
146. Metwally AA, Yu PS, Reiman D, Dai Y, Finn PW, Perkins DL. Utilizing longitudinal microbiome taxonomic profiles to predict food allergy via long short-term memory networks. *PLoS Comput Biol.* 2019. <https://doi.org/10.1371/journal.pcbi.1006693>.
147. Liu Y, Zhu J, Wang H, Lu W, Lee YK, Zhao J, Zhang H. Machine learning framework for gut microbiome biomarkers discovery and modulation analysis in large-scale obese population. *BMC Genomics.* 2022. <https://doi.org/10.1186/s12864-022-09087-2>.
148. Li P, Luo H, Ji B, Nielsen J. Machine learning for data integration in human gut microbiome. *Microb Cell Fact.* 2022. <https://doi.org/10.1186/s12934-022-01973-4>.
149. Hernández Medina R, Kutuzova S, Nielsen KN, Johansen J, Hansen LH, Nielsen M, Rasmussen S. Machine learning and deep learning applications in microbiome research. *ISME Commun.* 2022. <https://doi.org/10.1038/s43705-022-00182-9>.
150. Ghannam RB, Techtmann SM. Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. *Comput Struct Biotechnol J.* 2021. <https://doi.org/10.1016/j.csbj.2021.01.028>.
151. Xia Y. Statistical normalization methods in microbiome data with application to microbiome cancer research. *Gut microbes.* 2023. <https://doi.org/10.1080/19490976.2023.2244139>.
152. Hossain MR, Timmer D. Machine learning model optimization with hyper parameter tuning approach. *Glob J Comput Sci Technol D Neural Artif Intell.* 2021;21(2):31.
153. Papoutsoglou G, Tarazona S, Lopes MB, Klammsteiner T, Ibrahim E, Eckenberger J, Novielli P, Tonda A, Simeon A, Shigdel R, Béreux S, Vitali G, Tangaro S, Lahti L, Temko A, Claesson MJ, Berland M. Machine learning approaches in microbiome research: challenges and best practices. *Front Microbiol.* 2023. <https://doi.org/10.3389/fmicb.2023.1261889>.
154. Deng Z, Zhang J, Li J, Zhang X. Application of deep learning in plant–microbiota association analysis. *Front Genet.* 2021. <https://doi.org/10.3389/fgene.2021.697090>.
155. Cuomo S, De Rosa M, Piccialli F, Pompameo L, Vocca V. A numerical approach for soil microbiota growth prediction through physics-informed neural network. *Appl Numer Math.* 2024. <https://doi.org/10.1016/j.apnum.2024.08.025>.