





Review

Machine Learning Applications in Agriculture: Current Trends, Challenges, and Future Perspectives

Sara Oleiro Araújo ^{1,2,*} , Ricardo Silva Peres ^{1,3,4} , José Cochicho Ramalho ^{5,6} , Fernando Lidon ^{2,5} and José Barata ^{1,3,4,*} 

- ¹ UNINOVA—Centre of Technology and Systems (CTS), FCT Campus, Monte de Caparica, 2829-516 Caparica, Portugal; ricardo.peres@uninova.pt
- ² Earth Sciences Department (DCT), School of Sciences and Technology (NOVA-SST), NOVA University of Lisbon, 2829-516 Caparica, Portugal; fjl@fct.unl.pt
- ³ Electrical and Computer Engineering Department (DEEC), School of Sciences and Technology (NOVA-SST), 2829-516 Caparica, Portugal
- ⁴ Intelligent Systems Associate Laboratory (LASI), 4800-058 Guimarães, Portugal
- ⁵ GeoBioSciences, GeoTechnologies and GeoEngineering Unit (GeoBiotec), School of Sciences and Technology (NOVA-SST), 2829-516 Caparica, Portugal; cochichor@mail.telepac.pt
- ⁶ PlantStress and Biodiversity Lab, Forest Research Center (CEF), Associate Laboratory TERRA, School of Agriculture (ISA), University of Lisbon (ULisboa), 2784-505 Oeiras, Portugal
- * Correspondence: s.araujo@uninova.pt (S.O.A.); jab@uninova.pt (J.B.)

Abstract: Progress in agricultural productivity and sustainability hinges on strategic investments in technological research. Evolving technologies such as the Internet of Things, sensors, robotics, Artificial Intelligence, Machine Learning, Big Data, and Cloud Computing are propelling the agricultural sector towards the transformative Agriculture 4.0 paradigm. The present systematic literature review employs the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology to explore the usage of Machine Learning in agriculture. The study investigates the foremost applications of Machine Learning, including crop, water, soil, and animal management, revealing its important role in revolutionising traditional agricultural practices. Furthermore, it assesses the substantial impacts and outcomes of Machine Learning adoption and highlights some challenges associated with its integration in agricultural systems. This review not only provides valuable insights into the current landscape of Machine Learning applications in agriculture, but it also outlines promising directions for future research and innovation in this rapidly evolving field.

Keywords: Agriculture 4.0; machine learning; PRISMA; systematic reviews and meta analytics



Citation: Araújo, S.O.; Peres, R.S.; Ramalho, J.C.; Lidon, F.; Barata, J. Machine Learning Applications in Agriculture: Current Trends, Challenges, and Future Perspectives. *Agronomy* **2023**, *13*, 2976. <https://doi.org/10.3390/agronomy13122976>

Academic Editors: Gniewko Niedbala and Maofang Gao

Received: 16 October 2023
Revised: 31 October 2023
Accepted: 28 November 2023
Published: 1 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Agriculture 4.0 [1–5], also known as “Digital Agricultural Revolution” [6], represents a paradigm shift in agriculture, leveraging cutting-edge technologies to optimise various aspects of farming operations. These technologies encompass the Internet of Things (IoT), Artificial Intelligence (AI), Big Data, cloud computing, Decision Support System (DSS), advanced sensing technology, and autonomous robots [1,6,7]. Sensors and robotics play a crucial role in collecting essential field data, which is then transmitted to a local or cloud server via IoT technology for storage, processing, and analysis. Big data and AI-based techniques can be used to convert these data into valuable insights. To facilitate user interaction and informed decision making, a DSS equips users with the necessary tools to optimise the agricultural system and undertake appropriate actions.

Machine Learning (ML), a subset of AI, has shown great potential in enhancing various aspects of Agriculture 4.0. It can be defined as a computer program or system that can learn specific tasks without being explicitly programmed to do so [8–10]. It is a process that involves the use of a computer to make decisions based on multiple data inputs [8]. In this case, data mean a set of examples. Labeled data is often used for supervised learning tasks

(where the model learns from labeled examples), and unlabeled data might be used for unsupervised learning tasks (where the model finds patterns and structures in the data) [9].

ML indeed benefit from large amounts of data to achieve meaningful accuracy in their tasks. In the context of agriculture, obtaining vast and diverse data can be sometimes challenging yet pivotal for the success of ML models. IoT sensors are instrumental in collecting a diverse range of agricultural data as they can be strategically deployed across fields to capture relevant information regarding, for instance, soil conditions, climate variables, crop health, and livestock metrics [1]. The widespread adoption of IoT technology facilitates continuous and real-time data acquisition, enabling the generation of extensive datasets over time. However, it is essential to consider that the data should be collected with sufficient quality to ensure its representativeness in the specific case study at hand. For instance, in crop management, studying the different stages of the crop is important for developing models that are accurate and applicable to real-world scenarios. Obtaining such representative datasets may take time, but it is a necessary investment for the effectiveness and reliability of ML applications in agriculture. Furthermore, collaborative initiatives and partnerships with farmers, agricultural institutions, and research organisations can contribute to the pooling of data resources.

A general flow for the creation of ML models and their deployment in agriculture is illustrated in Figure 1. The initial phase involves the retrieval of agricultural data from diverse sources, forming the foundational input for subsequent ML processes. These data are then divided into ‘training’ and ‘testing’ datasets. The training dataset becomes the substrate for instructing the ML model, while the testing dataset serves as an evaluation mechanism, assessing the model’s performance and ensuring its accuracy and reliability. The outcome of these processes is a robust ML model capable of making classifications, predictions, or decisions tailored to specific agricultural contexts. Subsequently, the validated model is ready for deployment across various agricultural domains, including crop (i.e., optimising crop yields and health), water (i.e., ensuring efficient utilisation of irrigation resources), soil (i.e., maintaining soil health and fertility), and animal management (i.e., monitoring and improving livestock health and productivity).

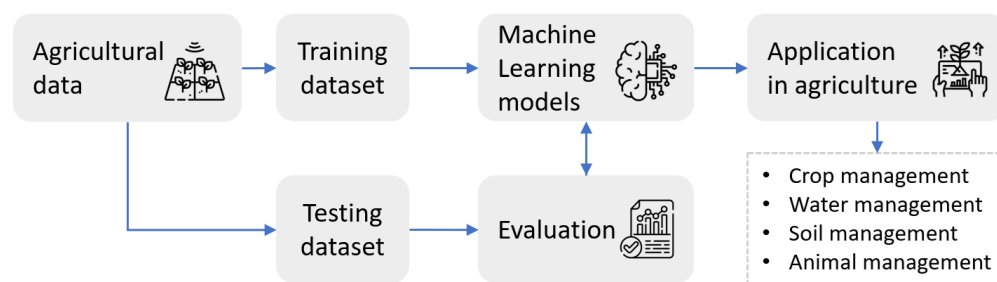


Figure 1. General flow for the creation of Machine Learning models and their application in agriculture.

Several prevalent ML algorithms have emerged within the context of Agriculture 4.0. These encompass well-known methods such as Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and an array of Deep Learning (DL) variations [1]. These algorithms play a key role in reshaping the agricultural domain, taking innovation and efficiency to new heights. While there has been extensive literature discussing the potential of ML in agriculture, the existing body of work often lacks a systematic and consolidated overview of the applications, impacts, outcomes, and challenges of ML integration in this dynamic field. A review made by [9] concluded that 61% of the analysed articles used ML techniques for crop management (22% disease detection, 20% yield prediction, 8% weed detection, 8% crop quality, and 3% species recognition), 19% for livestock management (12% livestock production and 7% animal welfare), 10% for soil management, and 10% for water management. Inspired by this study, the search for a current and comprehensive understanding of the landscape of ML applications in agriculture motivated the undertaking of a Systematic Literature Review (SLR) in 2023.

This effort seeks to elucidate the latest advances, trends, and challenges in this dynamic field, with the aim of contributing valuable insights to the agricultural research community.

The present SLR, conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology [11,12], addresses this research gap by providing a meticulous analysis of the state-of-the-art applications of ML in agriculture. By doing so, this review not only contributes to the existing body of knowledge, but it also offers valuable insights to researchers, practitioners, and stakeholders aiming to leverage ML technologies for sustainable and efficient agricultural practices. PRISMA is internationally recognised for its systematic framework, which effectively mitigates bias and increases the reliability of systematic reviews by providing a structured protocol for the identification, selection, and synthesis of studies. Adhering to the PRISMA guidelines was fundamental to maintaining the highest standards of methodological rigor, leading to an overall improvement in the validity and reliability of the results. To help guide this review, the following research questions were formulated:

- **RQ1:** What are the most used ML algorithms in agriculture?
- **RQ2:** What are the impacts and outcomes of integrating ML in agriculture?
- **RQ3:** What are the challenges and future directions associated with integrating ML in agriculture and agricultural systems?

The present document is organised as follows: Section 2 (Principles and Methods) details the PRISMA methodology, including the search strategy, inclusion, and exclusion criteria, data extraction process, and quality assessment of selected studies. Section 3 (Results and Discussion) discusses the results of the SLR. Section 4 (Machine Learning Trends) presents the most used ML models and algorithms of the SLR. Section 5 (Machine Learning in Agriculture) provides an overview of the ML techniques used, as well as specific applications, according to the domains outlined in Section 3. Section 6 (Challenges and Research Opportunities) discusses the challenges associated with implementing ML in agricultural systems. Finally, Section 7 (Conclusions, Limitations, and Future Work) summarises the main findings of the present SLR and how they contribute to the current understanding of ML in agriculture.

2. Principles and Methods

This SLR adheres to the well-established PRISMA guidelines, which describes how to collect and analyse data from available studies. The PRISMA Statement is a systematic framework encompassing 27 items in the form of a checklist, along with a four-phase flow diagram that serves as an invaluable tool for guiding researchers in the preparation of reviews and meta-analyses [11]. In the present section, each phase of the systematic review process is detailed, seamlessly aligned with the four fundamental phases outlined in the PRISMA methodology. These include the identification (Section 2.1), screening (Section 2.2), eligibility (Section 2.3), and inclusion phase (Section 2.4). Lastly, an overview of the PRISMA framework itself (Section 2.5) is provided, offering a holistic understanding of the structured approach adopted in this study.

2.1. Identification Phase

The identification phase represents the initial identification of records via various sources. To facilitate this step, a search string was formulated (Table 1). **Final string:** (“Agricultur*” OR “Farm*”) AND (“Machine Learning”) AND (“application” OR “implementation” OR “case study” OR “experimental” OR “practical”).

The search encompassed records indexed in the specified repositories (Web of Science (WoS) and Scopus) up until the 1st of July, 2023, and the search string was adjusted to the syntax of each digital repository. The wildcard symbol (“*”) was incorporated at the end of certain words in the search string, enabling the retrieval of all possible variants of the respective term. The boolean “AND” was employed to connect keywords originating from different groups within the search string, while “OR” was used to link keywords within the same group.

Table 1. Search keywords to be used to collect records from different digital databases.

Group 1	Group 2	Group 3
"agricultur*" "farm*"	"machine learning"	"application" "implementation" "case study" "experimental" "practical"

Additionally, some inclusion criteria were considered (Table 2) in order to meet what is intended in this study.

Table 2. Inclusion criteria for the Identification phase of the survey.

ID	Criteria	Description
IC1	Repositories	Web of Science and Scopus
IC2	Search period	From 2019 to 2023, both years included
IC3	Search within	Article title, abstract, and keywords
IC4	Document type	Articles
IC5	Language	English

In IC1, the digital repositories WoS and Scopus were selected as they are highly valued databases, due to their scientific and technical content, but also because they are closely related to the areas of knowledge associated with the objective of this article. The chosen search period, in IC2, was determined based on the significant progress and contributions to the field of integrating ML in the agricultural sector. Regarding IC3, the restriction to examining article titles, abstracts, and keywords was made with careful deliberation. This focused approach is predicated on the premise that these sections encapsulate succinct yet crucial information regarding the content and relevance of the articles, thereby facilitating the efficient identification of the pertinent literature. In IC4, the emphasis on scientific journal articles stems from a strategic assessment of the academic landscape in the specific domain of interest as they generally have a greater impact than other types of documents. Lastly, in IC5, the selection of English as the exclusive language criterion arises from its ubiquitous adoption as the lingua franca of global scientific discourse.

2.2. Screening Phase

In the screening phase, a systematic process is employed to evaluate the identified records against the predetermined inclusion and exclusion criteria (Table 3). The decision to focus exclusively on Q1 journal articles reflects a commitment to ensure the highest standard of research quality.

Table 3. Inclusion criteria for the Screening phase of the survey.

ID	Criteria	Description
SC1	Duplicates	Duplicate records were removed
SC2	Records Screening	Must include the title, year, abstract, and DOI
SC3	Journal rank	Must include only Q1 journals (Clarivate [13])
SC4	Article type	Must not include reviews/surveys without original research

2.3. Eligibility Phase

The eligible studies are determined via a detailed assessment of the full texts of the remaining records after the initial screening. This phase involves a thorough evaluation of the studies against the pre-defined inclusion and exclusion criteria (Table 4).

Table 4. Inclusion criteria for the Eligibility phase of the survey.

ID	Criteria	Description
EC1	Content	Publications specifically for the agricultural sector
EC2	Content	Publications with ML-related content
EC3	Full-text	Must be available

2.4. Inclusion Phase

The inclusion phase represents the studies that meet all the inclusion criteria and are included in the SLR. This screening was made manually by the same author.

2.5. PRISMA Overview

Following the PRISMA guidelines, an initial screening process involved searching the repositories WoS and Scopus, resulting in a total of 4580 articles matching the search string from Table 1. Through a rigorous screening and eligibility phase (Tables 3 and 4, respectively), a final selection of 272 articles was identified for in-depth analysis. Figure 2 illustrates the PRISMA flowchart reporting the different phases of the systematic review and the respective results.

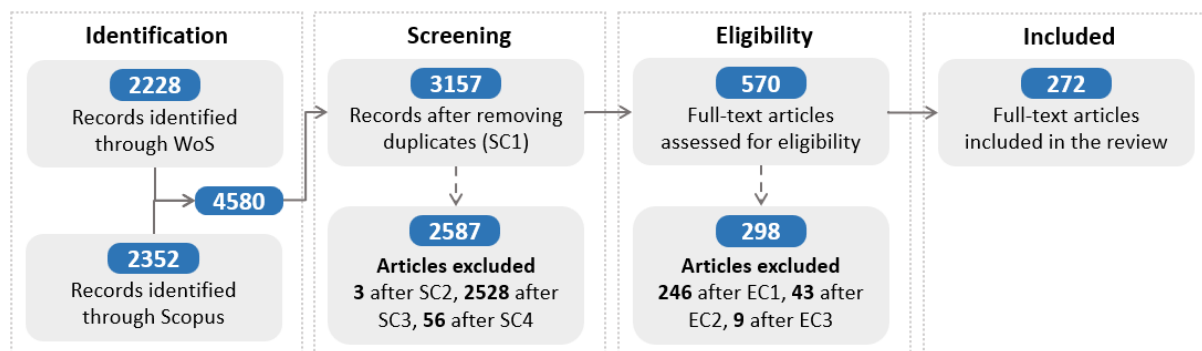


Figure 2. The flowchart illustrating the study inclusions and exclusions for the systematic literature review adheres to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Specific criteria were applied at various stages of the review process, including removal of duplicates (SC1), record screening (SC2), journal rank (SC3), document type (SC4), prioritisation of content relevant to agriculture (EC1) and Machine Learning applications (EC2), and inclusion of full-text versions (EC3).

3. Results and Discussion

3.1. Statistical Analysis

The number of publications by year and country are represented in Figures 3 and 4, respectively. According to Figure 3, the number of publications in the WoS and Scopus databases has generally increased over the years, with a significant jump from 2019 to 2020 and a consistent upward trend until 2022.

In Figure 3, the pre-PRISMA analysis of WoS publications showed steady growth, from 274 in 2019 to a peak of 862 in 2022. This trend indicates a growing interest in research efforts of ML science in agriculture during this period. Scopus publications exhibited a parallel increase, starting at 408 in 2019 and culminating at 1127 in 2022. This upward trend can be attributed to significant advancements in advanced technologies that have emerged within the agricultural sector, owing to the emergence and widespread acceptance of the Agriculture 4.0 paradigm [1]. However, it is important to note that in 2023, there was a decrease in the number of publications compared to the previous year. This decrease can be attributed to the fact that the search was conducted in the beginning of July 2023 and the data do not capture the full extent of publications for the entire year.

By examining these trends, we have uncovered a compelling narrative of how research efforts have evolved in tandem with the technological landscape. Technological advances

and their adoption in the agricultural sector have played a key role in driving this increase in interest and research activity. Notable advances in IoT, sensors, and robotics have catalysed the adoption of ML in various facets of agriculture. In addition, political changes, economic fluctuations, and global events may have influenced the trajectory of research in this field. Lastly, the application of the PRISMA methodology has resulted in a noticeable reduction in the overall count of publications for each year. While this might initially appear as a decrement, it signifies an intentional effort to ensure a higher and more reliable quality of studies. This approach reinforces the credibility and validity of the selected research, ultimately enhancing the robustness of the systematic review.

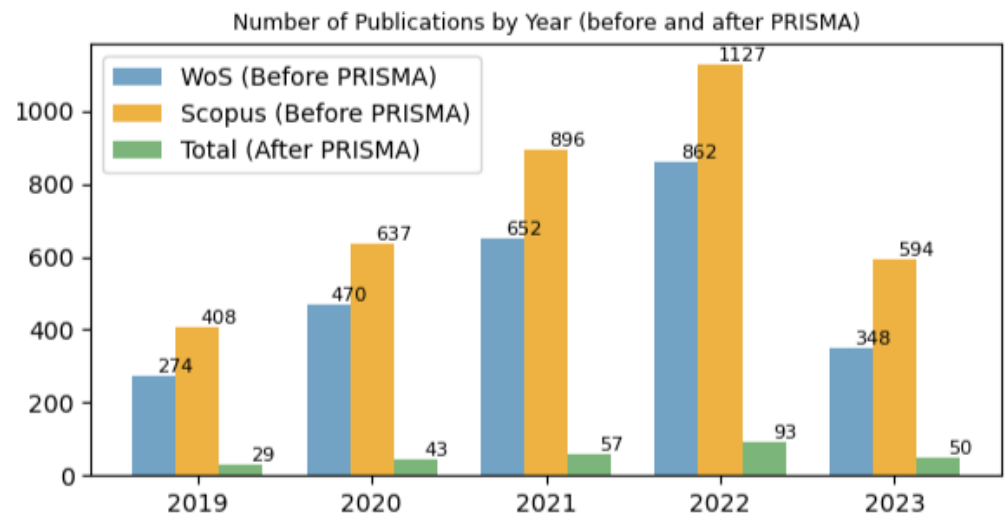


Figure 3. Distribution of the selected publications by year (before and after PRISMA).

Furthermore, Figure 4 highlights the top seven countries that have made substantial contributions in terms of published articles. Before PRISMA, the number of publications in the WoS database varies for each country, with China having the highest count of 619, followed by the United States of America (USA) with 552 and India with 305. The number of publications in the Scopus database also varies for each country, with India having the highest count of 882, followed by China with 654 and the United States with 541. After PRISMA, China and the United States lead the pack in terms of publication output, with 96 and 30 publications, respectively. On the other hand, countries such as India, Australia, the United Kingdom (UK), Germany, and Italy have relatively lower publication counts. These nuanced insights into the geographical distribution of research contributions enrich the understanding of the global landscape of ML applications in agriculture, underlining the need for international collaboration and knowledge exchange in this dynamic field.

Regarding the journals, the present SLR included 272 papers from 62 different journals. Figure 5 illustrates the top 10 journals, along with the corresponding counts of publications from each journal. From this analysis, “Remote Sensing” had the highest count of publications with 73 publications, followed by “Computers and Electronics in Agriculture” with 46 publications and “Agricultural Water Management” with 14 publications. It is further stated that the mean impact factor (Clarivate [13]) is 6.22, indicating the average influence of the journals. The range of impact factors spans from a minimum of 2.4 (Journal: “Mathematics”) to a maximum of 16.6 (Journal: “Nature Communications”). This range highlights the diversity of influence of the selected journals, which signifies the breadth and depth of academic engagement in this field. These insights not only enrich the understanding of the academic landscape, but it also serves as a testimony to the rigor and quality of the research efforts encompassed in this analysis.

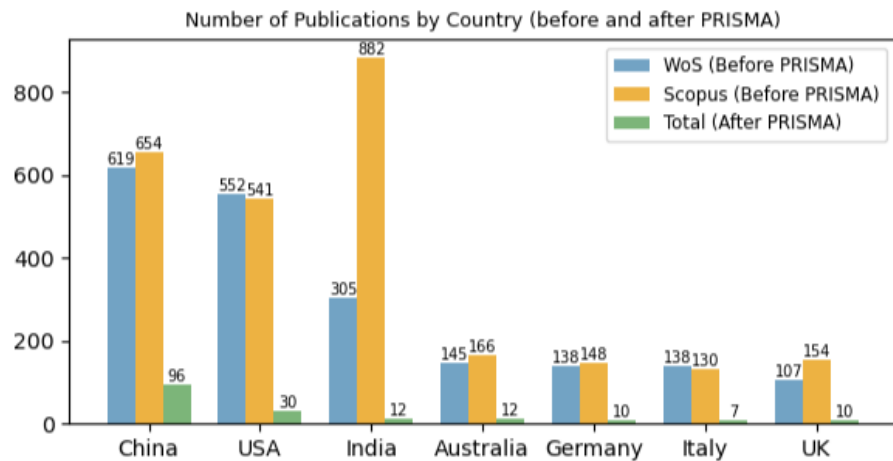


Figure 4. Distribution of the selected publications by country (before and after PRISMA).

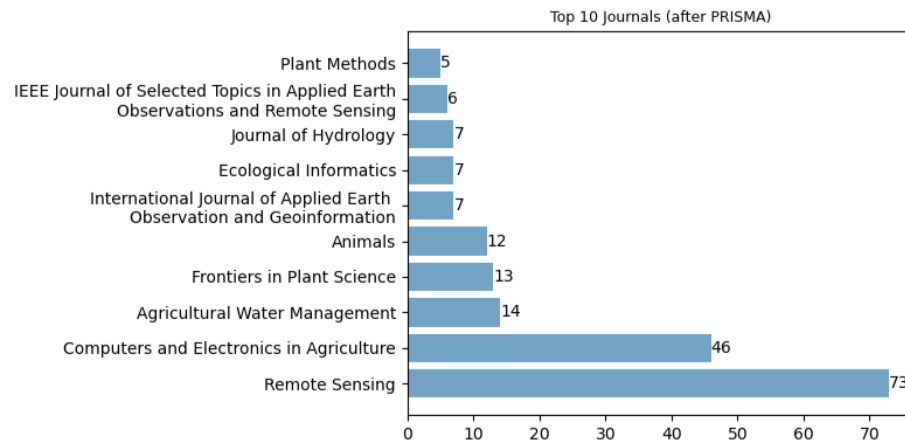


Figure 5. Distribution of the top 10 journals (after PRISMA).

At last, Figure 6 shows the count of publications in various categories (according to Clarivate [13]). The category “Geosciences” has the highest count of publications, with 82 publications. “Agriculture” is the second-highest category, with 55 publications, followed by “Agronomy” with 23 publications. This categorical distribution of publications provides valuable information on the various domains that intersect with ML applications in agriculture. It highlights the multifaceted nature of research efforts in this field and underlines the key role of interdisciplinary collaboration in the realisation of Agriculture 4.0 and the advancement of agricultural technologies.

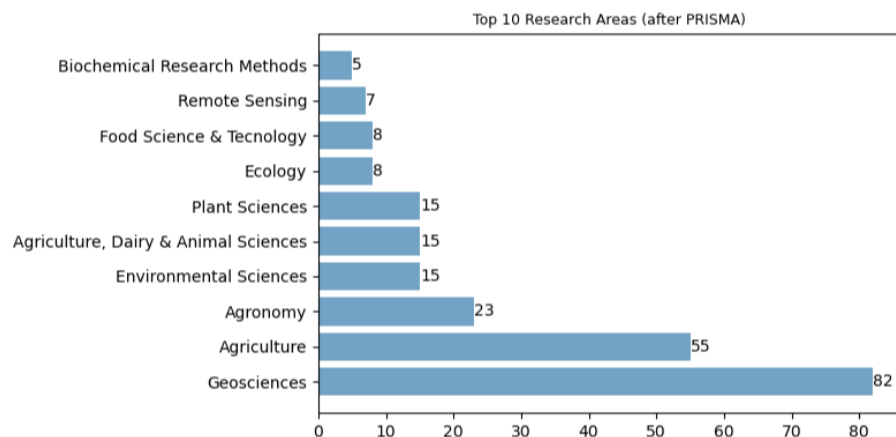


Figure 6. Distribution of the top 10 research areas (after PRISMA).

3.2. Application Domains in Agriculture

The distribution of application domains in agriculture, based on the SLR, is represented in Figure 7.

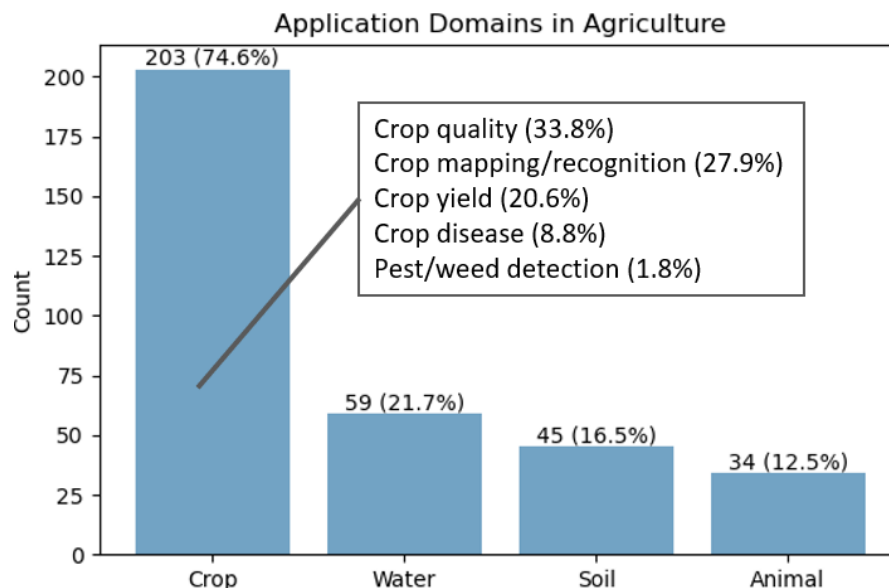


Figure 7. Distribution of application domains in agriculture (after PRISMA): crop, water, soil, and animal management.

As it is possible to see, the largest portion, accounting for 74.6%, is dedicated to *Crop management*. The *Water management* domain represents 21.7%, followed by the *Soil management* and *Animal management* domains, with 16.5% and 12.5%, respectively. The total percentage of each domain represents the proportion of articles that primarily focus on that specific domain. However, certain articles can be multidisciplinary in nature, addressing more than one domain within the agricultural context. Furthermore, five sub-domains were outlined, representing a distinct area of focus within the crop domain: *Crop quality* (33.8%), *Crop mapping/recognition* (27.9%), *Crop yield* (20.6%), *Crop disease* (8.8%), and *Pest/weed detection* (1.8%).

Moreover, it was possible to identify the most used crops and animals in this analysis:

- **Crop type:**
 - “Plants” with 46.36% of the total count. In this group, it is possible to find the following: wheat (13.91%), maize (12.17%), rice (6.09%), vineyards (3.04%), grass (3.04%), rapeseed (2.61%), sugarcane (2.17%), tea (1.74%), cotton (0.87%), peach leaf (0.87%), alfalfa (0.87%), bok choy (0.43%), barley (0.43%), Arabidopsis (0.43%), jujube (0.43%), parsley (0.43%), green coffee plant (0.43%), mushrooms (0.43%), oil palm leaf (0.43%), almond orchard (0.43%), and banana leaf (0.43%);
 - “Not Specified” with 13.59% of the total crop analysis;
 - “Vegetables” with 12.22% of the total count: soybean (3.04%), potato (2.61%), vegetables—not specified (1.74%), lettuce (0.87%), carrot (0.87%), sugar beet (0.87%), asparagus (0.43%), leek (0.43%), onions (0.43%), and cabbage (0.43%);
 - “Fruits” with 11.96% of the total count: tomato (2.61%), citrus (1.74%), pineapple (1.31%), watermelon (0.87%), mango (0.87%), banana (0.43%), strawberry (0.43%), date (0.43%), avocado (0.43%), muskmelon (0.43%), kiwi (0.43%), apricot (0.43%), durian (0.43%), peach (0.43%), grape (0.43%), guava (0.43%), and cucumber (0.43%);
 - “Trees” with 9.84% of the total count: apple tree (1.72%), olive tree (0.87%), pine tree (0.87%), gum tree (0.43%), Oriental beech tree (0.43%), Cinnamon tree (0.43%), Caribbean tree (0.43%), and shrub (0.43%);

- “Grain, seeds and nuts” with 6.04% of the total count: grains (3.01%), nuts—not specified (2.17%), pea seeds (0.43%), and radish seeds (0.43%).
- **Animal type:** cows and dairy cows (21.21%), chickens, broilers, and hens (18.18%), sheep (15.15%), fish (9.09%), pigs (9.09%), bees (6.06%), steers (3.03%), gorillas (3.03%), heifers (3.03%), horses (3.03%), lambs (3.03%), small ruminants (3.03%), and invasive insects—not specified (3.03%).

The meticulous analysis of crop and animal types provides valuable insights into prevailing trends and focus areas for ML applications in agriculture. The prominence of “Plants” as the most studied crop category underlines the crucial role of ML in optimising various plant-based agricultural practices. Wheat, maize, and rice emerge as the main areas of focus, reflecting the importance of staple crops in agricultural research. In addition, the inclusion of specialty crops, such as vineyards and tea, exemplifies the diversity of agricultural contexts in which ML-based interventions are making substantial contributions. In the animal context, cows dominate the category, followed by poultry and sheep. The prevalence of interdisciplinary approaches in Agriculture 4.0 is evident in the analysis. In addition to traditional crops and livestock, the inclusion of categories such as bees and invasive insects demonstrates the innovative application of ML in various agricultural contexts, from agriculture to pest management.

4. Machine Learning Trends

The present section provides crucial insights towards answering to RQ1 (*What are the most used ML algorithms in agriculture?*). As it is known, various ML algorithms can be used for statistical modeling, data analysis, classification, regression, and dimensionality reduction processes. In the context of this SLR, Figure 8 provides an overview of the most used ML algorithms employed in the agricultural scope.

From the analysis of Figure 8, RF [14,15] emerges as the most widely used ML algorithm, representing 19.2% of the overall distribution. Its versatility and robustness make it a favored choice for handling complex problems. SVM [16,17] ranks second with 15.9% as it is known for their effectiveness in both classification and regression tasks. The Gradient Boosted Tree (GBT) [18,19] (8.3%), and Convolutional Neural Network (CNN) [20,21] (7.3%) also demonstrate significant usage and adoption in the agricultural sector. In descending order of frequency, the ML approach categories present in the SLR, as well as their respective algorithms, are:

1. **Ensemble Learning:** this category has the largest percentage in the SLR with 35.6% of the total distribution. Ensemble Learning [22–24] emerges as a key force for improving the performance and generalisation of ML models, making them more robust and reliable. This category includes RF (19.2%, frequency: 127), GBT (8.3%, frequency: 55), Extreme Gradient Boosting (XGBoost) (4.5%, frequency: 30), AdaBoost (0.9%, frequency: 6), Bagging (0.8%, frequency: 5), CatBoost (0.6%, frequency: 4), Stacking (0.3%, frequency: 2), and “not specified” ensemble methods (0.3%, frequency: 2). Of all the algorithms within this category, the RF presents the highest frequency. While the Decision Tree (DT) [25] offers a simple and interpretable model, RF leverages the power of multiple DT to provide robust predictions and classifications that are crucial for the optimisation of various agricultural processes. These processes can include crop yield estimation, disease detection, and land cover classification based on remote sensing data.
2. **Artificial Neural Networks:** the second category within the scope of this SLR constitutes 24.9% of the overall content and encompasses a range of influential algorithms that fall under the domain of ANN [26,27]. The algorithms covered in this category include CNN (7.3%, frequency: 48), ANN-not specified (6.4%, frequency: 42), Long Short-Term Memory (LSTM) (3.0%, frequency: 20), Deep Neural Networks (DNN) (1.8%, frequency: 12), Multilayer Perceptron (MLP) (1.7%, frequency: 11), You Only Look Once (YOLO) (1.2%, frequency: 8), Extreme Learning Machines (ELM) (0.9%, fre-

- quency: 6), DL-not specified (0.6%, frequency: 4), Gated Recurrent Unit (GRU) (0.5%, frequency: 3), Recurrent Neural Network (RNN) (0.3%, frequency: 2), Generative Adversarial Networks (GAN) (0.2%, frequency: 1), and Encoder and Autoencoder (0.2%, frequency: 1). These algorithms are powerful tools capable of learning complex patterns, thereby facilitating accurate predictions and advancing the capabilities of numerous applications. Among these algorithms, the one that stands out the most is CNN, known to be specialised for image data analysis [28], making them valuable for tasks like crop disease identification, plant species recognition, and weed detection [1].
3. **Support Vector Machine:** the third most prominent category, accounting for 15.9% (frequency: 105), is the SVM. This algorithm holds significant popularity and widespread application for tasks encompassing both classification and regression procedures [29]. SVM underscores its significance in guiding informed decisions for bolstering agricultural productivity in the era of Agriculture 4.0 [1]. Through its adeptness in crop mapping, yield estimation, and disease detection, SVM contributes to the ongoing transformation of agriculture into a more precise, efficient, and resilient practice, aligning seamlessly with the evolving demands of a dynamic global food landscape.
 4. **Dimensionality Reduction:** this category represents 6.2% of the total distribution and includes three different algorithms that can aid in dimensionality reduction [30] and feature engineering [31] from agricultural datasets, namely Partial Least Squares (PLS) algorithm (3.9%, frequency: 26), Principal Component Analysis (PCA) (1.2%, frequency: 8), and Linear Discriminant Analysis (LDA) algorithm (1.1%, frequency: 7).
 5. **Generalised Linear Models:** comprising 6.0% of the total distribution, this category underscores the significance of statistical models that transcend the constraints of simple linear regression. Of this category are Multiple Linear Regression (MLR) (2.3%, frequency: 15), Logistic Regression (1.4%, frequency: 9), Ridge Regression (1.2%, frequency: 8), Cubist Regression (0.8%, frequency: 5), and Multivariate Adaptive Regression Splines (MARS) (0.3%, frequency: 2).
 6. **Nearest Neighbour:** this category exclusively employs the k-Nearest Neighbors (KNN) algorithm (4.5%, frequency: 30). Among the various algorithms in the field of ML, the KNN algorithm stands out as one of the simplest yet extensively employed methods for classification purposes [32,33]. Its adaptive and comprehensible design contributes to its popularity in various classification tasks.
 7. **Bayesian Models:** this category focuses on Gaussian distributions and probabilistic models. These methods leverage the principles of Gaussian processes [34] (2.0%, frequency: 13) and Naïve Bayes (NB) [35] (2.3%, frequency: 15). These techniques offer solutions that adapt to the complexities of diverse datasets and applications.
 8. **Decision Trees:** constituting 4.1% (frequency: 27) of the overall distribution, these tree-like structure algorithms are versatile tools that facilitate data-driven decisions [25].
 9. **Multi-task Learning:** represents 0.3% (frequency: 2) of the total distribution and its objective is to enhance the outcomes of several interconnected learning tasks by utilising valuable insights shared among them [36,37].

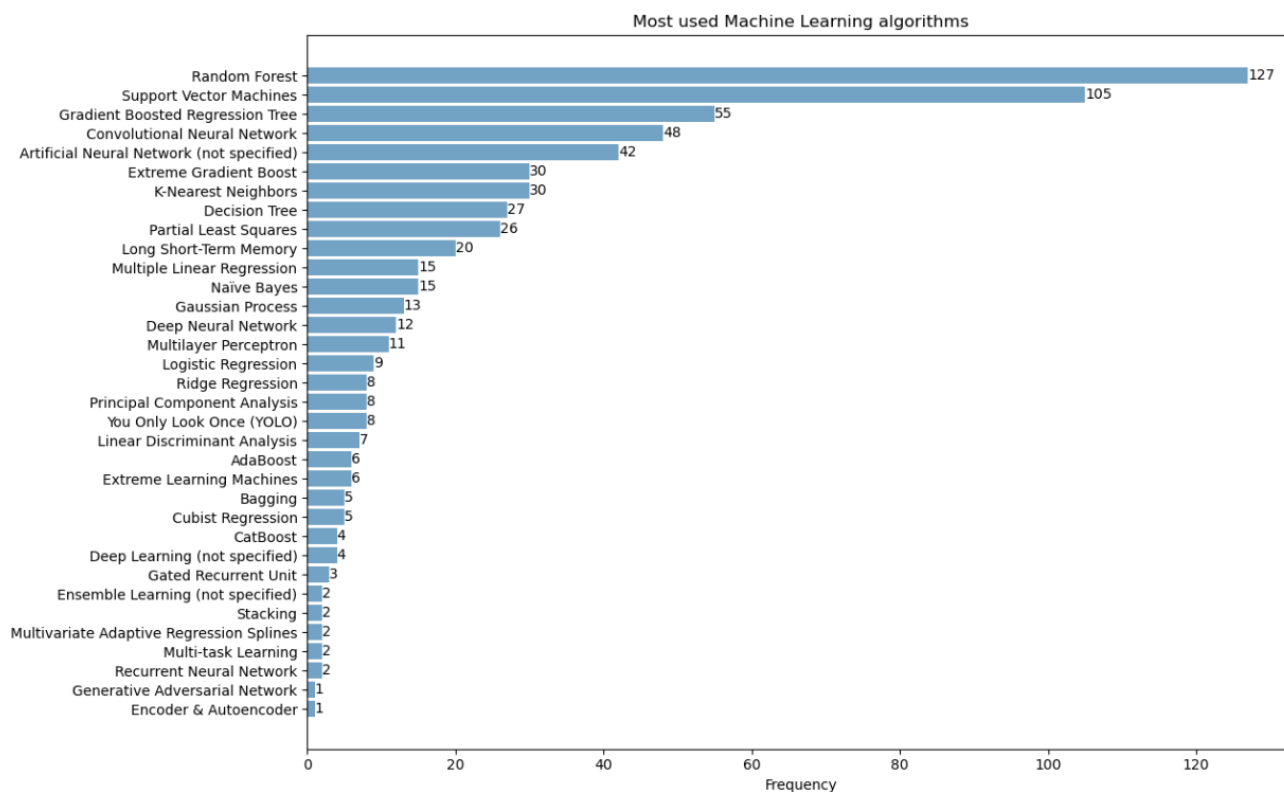


Figure 8. Distribution of the most used Machine Learning algorithms in the Systematic Literature Review.

5. Machine Learning in Agriculture

The current section is dedicated to addressing RQ2 (*What are the impacts and outcomes of integrating ML in agriculture?*), with a detailed analysis of the distribution of application domains in agriculture (as mentioned in Section 3.2). For each domain (and sub-domain, in the case of crop management), the authors selected five to seven articles based on their relevance, impact, and ability to provide insights that contribute to the overall understanding of the subject. This sample size is considered reasonable as it allows for the inclusion of key findings and trends within each domain without overwhelming the review with an exhaustive list of articles. Section 5.5 summarises the main findings from each domain, providing a brief overview of the impact that ML technologies have had on modernising agricultural practices.

5.1. Crop Management Domain

Crop management is associated to several agricultural practices that profoundly influence the growth and yield of cultivated crops. These practices encompass a wide range of activities, starting with the meticulous sowing process, extending to the vigilant maintenance of crops throughout their growth and development phases, and concluding with the phases of harvest [1]. The optimisation of crop management strategies is essential to increase agricultural productivity, thereby addressing the escalating global requisites for sustenance, textile fibers, energy sources, and fundamental raw materials [38].

According to Figure 7, the crop management domain has the largest portion, accounting for 74.6%, of the study. This finding indicates that the application of ML techniques in crop management has significantly revolutionised conventional farming practices, offering capabilities such as crop mapping and recognition, yield prediction, optimal irrigation scheduling, pest and weed management, and disease detection [1]. From the 272 articles included in the SLR, 203 articles are related to crop, where 92 are related to crop quality, 76 to crop mapping/recognition, 56 to crop yield, 24 to crop disease, and five to pest and

weed detection. We delve into each sub-domain to highlight their specific contributions to crop management and their impact on enhancing agricultural practices.

5.1.1. Crop Quality

Within this study, crop quality refers to the characteristics of crops that determine their value and suitability. Improving crop quality via ML involves monitoring and managing crop's growth, nutrient levels, organoleptic characteristics, and others parameters.

By examining Table 5, it becomes evident that ML-based techniques have harnessed their computational prowess to effectively manage complex datasets encompassing a wide range of crop attributes (such as spanning size, appearance, and sensory characteristics). The synergy between cutting-edge ML algorithms and real-time data, including images and meteorological information, has propelled substantial advancements in the agricultural sector. This convergence has unlocked remarkable progress, allowing for more precise evaluations of crop quality based on current conditions and attributes. Furthermore, ML methods demonstrate their adaptability by excelling in the prediction and evaluation of crop quality using non-destructive approaches. This innovative strategy obviates the need for intrusive testing while simultaneously facilitating seamless real-time quality control throughout the supply chain. This paradigm shift enhances the efficiency of crop management and distribution, underscoring the transformative potential of ML in optimising agricultural processes.

Table 5. Machine learning applications in crop quality sub-domain.

Ref.	Crop Type	Models Used	Summary
[39]	Lettuce	CNN, DNN	AirSurf platform developed for ultra-scale aerial phenotyping, crop counting, and crop quality assessment. AirSurf-Lettuce achieves high accuracy (>98%) in scoring and categorising iceberg lettuces and provides novel analysis functions for mapping lettuce size distribution to enhance precision agricultural practices.
[40]	Kiwifruit	ANN, SVM, Gaussian Process, Ensemble learning	Non-destructive tactile sensing approach for estimating the stiffness of kiwifruits, achieving accurate ripeness estimation with regression-based ML, showcasing potential applications in real-time quality control and sorting of fruits throughout the supply chain.
[41]	Watermelon	NB, Logistic regression, KNN, DT, RF, ANN, SVM, GBT	Proposes a fusion non-destructive method for classifying watermelon sweetness based on acoustic signals, image processing, and weight features. ML is used to develop sweetness classification models. GBT obtained the highest classification accuracy of 92%.
[42]	Peach flower	RF, SVM, KNN, NB	Assesses different ML methods for estimating and monitoring peach flowering phenological stages using real-time flower images and meteorological data. RF has the highest F1 score of 98.82% on the testing set, demonstrating its potential for real-time monitoring and applications in peach breeding, heat stress management, and irrigation scheduling.
[43]	Rice	Stepwise MLR, RF	Accurate and non-destructive in-season nitrogen (N) diagnosis and recommendation for rice crops. The study uses active canopy sensor data and combines it with environmental and agronomic variables to develop N status diagnosis and recommendation models. This approach can significantly enhance N management strategies in rice cultivation, contributing to sustainable development and food security.
[44]	Green coffee	SVM, RF, XGBoost, CatBoost, PCA	Focuses on distinguishing between special and traditional green coffee beans using an advanced multispectral imaging technique based on reflectance and autofluorescence data and combined with ML techniques. SVM achieves the highest accuracy (0.96) for the test dataset. This approach showcases its potential as a non-destructive and real-time tool for classifying green coffee beans in the food industry.
[45]	Wheat	Multi-task learning	Multi-task learning approach using a real-world agricultural dataset, showing superior accuracy and stability in fertilisation prediction, leading to the development of a precision fertilisation system for intelligent and personalised farm management.

5.1.2. Crop Mapping and Recognition

Crop mapping and recognition refers to the process of identifying and mapping different crop types within agricultural fields. It involves using various data sources (such as satellite imagery, aerial and/or proximal photography, and spectroscopy) to detect and classify different crops and their spatial distribution. With ML techniques, it is possible to create accurate and detailed crop maps and identify the unique characteristics of each crop, which can be valuable for agricultural planning, resource management, and yield estimation.

Drawing insights from Table 6, it becomes evident that the application of ML-based techniques extends its computational capabilities into the domain of crop mapping and recognition, revolutionising how agricultural landscapes are understood and managed. The ability to process intricate data, coupled with real-time insights, enhances the precision and efficiency with which crop types and distributions are identified. Furthermore, by harnessing methodologies such as DL and established ML algorithms, these studies underscore the potential to effectively distinguish specific crop varieties with a commendable level of accuracy.

Table 6. Machine learning applications in crop mapping and recognition sub-domain.

Ref.	Crop Type	Models Used	Summary
[46]	Apricot cultivars	DT, KNN, LDA, NB, SVM, BPNN	Demonstrates the feasibility of using ML to identify apricot cultivars based on their shape features, suggesting potential for non-destructive automatic identification systems. SVM achieved the best accuracy of 90.7% in the test set for classifying apricot cultivars.
[47]	Grapevines	SVM, CNN	The study demonstrated the feasibility of using spectroscopy, Big Data, and ML to distinguish specific grapevine varieties (Touriga Nacional or Touriga Franca) from a larger group of other varieties.
[48]	Various crops	RF	Compares various classification strategies for vegetation mapping over large-scale areas using Sentinel data within the Google Earth Engine platform and RF algorithms for classification.
[49]	Rapeseed	DL	Low-cost approach using DL (AlexNet, VGGNet16, ResNet18, ResNet50, and GoogLeNet) and UAV images for recognising freezing-tolerant rapeseed materials. The method achieves high accuracy (over 92%), with ResNet50 providing the best performance (93.3%), outperforming traditional ML methods.
[50]	Pineapple	ANN, SVM, RF, NB, DT, KNN	Method involving UAV-captured RGB images, image processing, and ML classifiers to identify pineapple crowns, classify them as fruit or non-fruit, and count them accurately. The process involves pre-processing and segmenting high spatial-resolution aerial images, extracting features based on shape, color, and texture, and optimising classifiers' performance via feature fusion using one-way analysis of variance (ANOVA).
[51]	Corn, soybean	DL, CNN	Innovative within-season emergence (WISE) phenology-normalised DL model for scalable within-season crop mapping using time-series remote sensing data. This approach accommodates spatiotemporal variations in crop phenological dynamics, yielding an over 90% overall accuracy for classifying corn and soybeans at the end of the season, as well as a satisfactory performance (85% overall accuracy) one to four weeks earlier than calendar-based approaches during the growing season.

5.1.3. Crop Yield

Crop yield refers to the quantity of agricultural produce obtained from a specific area of land during a growing season. Ensuring high crop yields is of utmost importance for addressing global food challenges and meeting the demands of a growing population [38]. There has been a growing application of ML methods to estimate crop yield, aiming to facilitate farming planning, resource allocation (such as water, fertilisers, and pesticides), enhance storage management and marketing strategies, and tackle the pressing challenges of food security in the forthcoming years [1].

Reflecting upon the compilation detailed in Table 7, it becomes apparent that the application of ML-based methodologies showcase the potential to predict crop yields with remarkable accuracy. By integrating diverse data sources like remote sensing imagery, meteorological data, and canopy geometric parameters, these models not only provide insight into crop yield, but it also highlights the interplay of various factors influencing the agricultural output.

Table 7. Machine learning applications in crop yield sub-domain.

Ref.	Crop Type	Models Used	Summary
[52]	Vineyard	ANN	Combines remote sensing, computer vision, and ML for vineyard yield estimation. By using VIs and vegetated fraction cover obtained from UAV multispectral imagery, along with ANN techniques, the approach provides accurate yield predictions with higher accuracy than traditional methods, supporting decision making in viticulture practices and harvest planning.
[53]	Rice	BPNN, RNN	Proposes an end-to-end model for rice yield prediction using DL fusion to learn deep spatial and temporal features from time-series meteorology and area data. The model achieves accurate predictions for both summer and winter rice yields.
[54]	Strawberry	RF, MLR, MARS, XGBoost, SVM, ANN	The combination of canopy geometric parameters and VIs obtained from UAV imagery proved effective for estimating strawberry dry biomass using ML models. ANN showed the highest accuracy in cross-validation, and red-edge-related VIs were found to be the most influential variables.
[55]	Apple tree	Ensemble learning, SVM, KNN	Develops an automatic processing channel to extract morphological and spectral features from UAV LiDAR and multispectral imagery data. The ensemble learning model outperforms other base learners (SVM and KNN) and provides accurate yield predictions for individual apple trees in the orchard.
[56]	Wheat grain	RF, SVM, MLR, generalised boosting regression	The research explores various VI, Sentinel-2 bands, and the biophysical parameter LAI retrieved from radiative transfer models (RTM) as input data for the models. RFRandom forest regression stands out as the most effective model.
[57]	Winter wheat	Linear regression, Ensemble learning, DT, SVM, Gaussian Process	The study employs ML and historical data to predict winter wheat yield and dry matter, with the Gaussian process model achieving the highest accuracy ($R^2 = 0.87$ and $R^2 = 0.86$, respectively). The results offer valuable insights into site-specific crop management and could aid in formulating water and nitrogen management strategies for global food security.

5.1.4. Crop Disease

Crop disease refers to the study and management of various diseases that affect agricultural crops, leading to reduced yields and economic losses for farmers and the agricultural industry as a whole.

The use of ML-based techniques have proven to be key strategies in crop disease management, as highlighted in Table 8. Several techniques are applied to discern disease patterns, anticipate outbreaks, and implement targeted interventions, thereby offering a promising avenue for detection, diagnosis, and control of crop diseases [1]. Through the fusion of ML models with diverse data sources, such as IoT-generated data and satellite and UAV imagery, these studies showcase the capacity to accurately categorise and identify diseases across various crops, enabling timely and effective responses to mitigate their impact.

Table 8. Machine learning applications in crop diseases sub-domain.

Ref.	Crop Type	Models Used	Summary
[58]	Banana plants	RF, PCA	Detects banana plants and their major diseases using satellite and UAV images and ML for classification. The developed model effectively categorised both healthy and diseased plants.

Table 8. Cont.

Ref.	Crop Type	Models Used	Summary
[59]	Tomato	YOLO (v3)	Employs a machine vision approach for early real-time detection of tomato diseases and pests in natural environments. The outcomes demonstrate an average recognition accuracy of 91.8%. The developed approach has been put into practice within real tomato cultivation settings, demonstrating its effectiveness in detecting small objects and leaves occlusion.
[60]	Not specified	SVM, CNN, KNN, NB	IoT-based that uses sensors and cameras to collect data from plants, which are then analysed via ML models. The system proposes ensemble classification and pattern recognition for crop monitoring system to identify plant diseases at the early.
[61]	Sugarcane	CNN, YOLO (v5)	Detects White Leaf Disease in sugarcane crops using UAV imagery and DL models. The proposed methodology provides technical guidelines for effective crop management and disease monitoring.
[62]	Watermelon	MLP, DT	Uses remote sensing, VIs, and ML for identifying and classifying different severity stages of Downy Mildew disease in watermelon. The highest classification accuracy was achieved via the MLP method.
[63]	Rice	MLP, SVM, NB, DT, KNN	Weather-based rice blast disease-forecasting system that uses an ensemble feature ranking approach to enhance predictive accuracy. By evaluating fifteen weather features, the proposed method identifies the most impactful ones. Among these features, average visibility, rainfall amount, sun exposure hours, maximum wind speed, and rainy days emerge as the most influential in rice blast prediction.
[64]	Potatoes	ANN	Innovative approach to the early detection of Verticillium wilt in potatoes using near-infrared spectroscopy and ANN models. The models accurately predict physiological responses to infection and classify infected plants within just two days after inoculation, even before visible symptoms appear.

5.1.5. Pest and Weed Detection

Instances of crop pest infestations, ranging from weeds, insects, pathogens, and rodents [65], have emerged as factors affecting global agricultural production. This sub-domain focuses on the utilisation of advanced technologies, such as sensors, imaging systems, and ML algorithms, to detect and mitigate the presence of unwanted organisms that can negatively impact crop growth and yield.

From Table 9, it is possible to understand that ML techniques can help analyse complex data from various sources (such as satellites, UAV, or sensors) and identify patterns and anomalies associated with pest and weed presence that may not be easily recognisable to the human eye. ML-powered systems can detect pests and weeds at their early stages, enabling swift intervention before infestations become widespread [1].

Table 9. Machine learning applications in pest and weed detection sub-domain.

Ref.	Crop Type	Models Used	Summary
[66]	Vineyard	DT with object-based image analysis	Innovative approach for mapping <i>Cynodon dactylon</i> (bermudagrass) infestations in vineyard cover crops using an automatic DT-OBIA algorithm combined with UAV imagery. This method is crucial due to the negative impacts of bermudagrass on vineyard productivity.
[67]	Wheat	SVM with Radial Basis Function	Assesses weed impact on wheat biomass using RGB images and proximal sensing techniques. The SVM model discriminates between crop and weeds and generates indicators like weed pressure and local wheat biomass production.
[68]	Wheat	DNN	Detection of Italian ryegrass in wheat fields using UAV imagery (RGB) and DNN, along with an extensive feature selection method to accurately detect ryegrass in wheat and estimate its canopy coverage. Predictive models were developed to relate early-season ryegrass canopy coverage with end-of-season ryegrass biomass and seed yield, as well as wheat biomass and grain yield reduction.

Table 9. Cont.

Ref.	Crop Type	Models Used	Summary
[69]	Wheat	DL with SVM, KNN, NN	Novel approach for classifying weed and wheat in drone-captured images, integrating an optimised voting classifier with NN, SVM, and KNN to classify features extracted using AlexNet via transfer learning.
[70]	Corn	YOLO (v7)	Identifies major pests (corn borer, armyworm, and bollworm) of corn using the YOLOv7 network combined with the Adam optimiser. The approach demonstrates the feasibility of using DL and advanced optimisation techniques for effective crop pests and disease identification, contributing to agricultural modernisation.

5.2. Water Management Domain

As water resources become increasingly finite and their management more complex, the fusion of cutting-edge technology with robust data analytics holds great promise in promoting more sustainable water management practices. IoT technology, sensors and actuators networks, data analytics, and predictive models have enabled farmers to monitor water quality, soil moisture levels, weather forecasts, and Crop Evapotranspiration (ETc) rates [71].

As demonstrated in this current study, the water management domain represents 21.7% of the study (Figure 7), highlighting its significance in agricultural applications. Table 10 exemplifies the utilisation of an array of ML algorithms, coupled with remote and proximal sensing techniques as well as innovative IoT technologies, to address diverse water-related challenges encompassing irrigation management, water quality surveillance, and ETc prediction.

Table 10. Machine learning applications in water management domain.

Ref.	Crop Field	Models Used	Summary
[72]	Maize	Linear regression, RF, Cubist, PLS, PCA, GBT	Uses remote sensing data and regression algorithms for predicting ETa and soil water content to enable remote irrigation management. The study employs VIs for training and phenology observations. Cubist showed slightly better performance for predicting ETa and RF for soil water content.
[73]	Cranberry	RF, XGBoost	Forecasts water table depth using DT-based modeling approaches for optimised irrigation management. XGBoost demonstrated superior predictive ability, accurately simulating water table depth fluctuations for longer periods than RF. Despite limitations with extrapolation and extreme events, the models hold potential with broader dataset ranges for practical applications.
[74]	Not applicable	KNN	Portable smart sensing system based on IoT for detecting nitrate, phosphate, pH, and temperature in water. KNN algorithm is used to enhance the accuracy of the system's analysis. The proposed system offers early hazard detection and promotes regular contaminant level evaluation.
[75]	Not specified	PCA, SVM, GBT	Focuses on accurately predicting crop ETc for efficient water resource management and irrigation. The research employs PCA techniques to identify key factors influencing ETc that are then used as inputs for prediction models. PSO was used to optimise SVM and GBT models. The PSO-GBT model exhibits the highest accuracy.
[76]	Maize	DT, RF, SVM, ANN, PLS	Uses UAV multispectral data and ML for estimating water content indicators, including equivalent water thickness, fuel moisture content, and specific leaf area of maize crops in smallholder farms. RF and SVM outperform others in predicting water content indicators. This approach offers accurate insights into drought-related water stress on smallholder farms.

Table 10. Cont.

Ref.	Crop Field	Models Used	Summary
[77]	Banana plants	KNN, GBT, LSTM	Employs IoT components to gather data (soil moisture, temperature, and weather conditions) and ML to optimise irrigation requirements and reduce energy consumption. The hybrid model predicts real-time and time-series water needs based on various observations. The work is demonstrated using banana cultivation, achieving up to a 31.4% water optimisation for a single banana tree.
[78]	Grains, vegetables, fruits, flowers	RF, NN, SVM	Predicts phosphorus concentrations in shallow groundwater in intensive agricultural regions. SVM achieved the highest accuracy ($R^2 = 0.60$). These findings support groundwater phosphorus monitoring, early warning, and pollution management decision making in intensive agricultural regions.

5.3. Soil Management Domain

Agricultural land is the extent of land considered suitable for agricultural production, covering both crop cultivation and livestock rearing [1]. By embracing the principles of Agriculture 4.0, the integration of IoT sensors for real-time parameter measurements, AI-driven data analysis techniques, and DSS for informed decision making equips farmers with the tools to effectively oversee their fields in a manner that is both efficient and sustainable [1,79]. ML-based techniques can process vast amounts of soil-related data (such as soil composition, texture, and moisture measurements) and generate insights into optimal irrigation schedules, nutrient management strategies, and soil health assessments.

In the present study, the soil management domain represents 16.5% of the entirety, as illustrated in Figure 7, and refers to the study and management of soil properties, composition, and conditions within agricultural systems. As is clear from Table 11, ML techniques possess the ability to predict soil properties and behaviours, empowering farmers to make well-informed choices pertaining to soil fertility, structure, moisture levels, and nutrient concentrations, all aimed at enhancing crop growth and yield. Additionally, by leveraging computer vision and the remote sensing data, ML simplifies the monitoring of both crops and soil conditions. This technological synergy allows for a comprehensive assessment of crop health, growth stages, and potential stressors. Beyond remote sensing, one particularly notable application of ML involves the utilisation of cell phone images, as demonstrated in the study by [80]. This innovative approach showcases the potential of ML to develop efficient proximal soil sensors capable of swiftly and accurately predicting crucial soil properties. By harnessing readily available technology, this advancement exemplifies the adaptability and practicality of ML solutions in modern soil management practices. This not only exemplifies the adaptability and practicality of ML solutions in modern soil management practices, but it also underscores the transformative impact that technology-driven approaches can have on agricultural sustainability.

Table 11. Machine learning applications in soil management domain.

Ref.	Crop Field	Models Used	Summary
[81]	Various soil samples	RF, SVM, Logistic Regression	Predicts disease occurrence with high accuracy by analysing soil macroecological patterns of Fusarium wilt, a destructive soil-borne plant disease. The research employs a ML approach using bacterial and fungal data sets from diseased and healthy soils across various countries and plant varieties. The results reveal distinct differences in bacterial and fungal communities between healthy and diseased soils.
[82]	Canola	RF	The research utilises a ML approach to determine key predictors of soil nitrous oxide (N_2O) emissions, including soil temperature, moisture, and nitrate availability. The results highlight that N_2O emissions were influenced by these factors, with emission factors being lower in high yield zones compared to low yield zones.

Table 11. Cont.

Ref.	Crop Field	Models Used	Summary
[80]	Maize, soybean	DT, RF, Cubist, Gaussian Process, SVM, ANN	Estimates soil organic matter (SOM) and soil moisture content (SMC) based on 22 color and texture features extracted from cell phone images. The study demonstrates the potential of using computer vision and ML to create an efficient proximal soil sensor for quick and accurate predictions of soil properties. Gaussian Process and Cubist models performed the best for SMC prediction, while ANN and Cubist showed satisfactory accuracy for SOM prediction.
[83]	Vineyard	NN regression, KNN, SVM with Linear Kernel, XGBoost, Cubist	Explores the potential of using soil protists as bioindicators to assess multiple stresses in agricultural soils. The findings indicate that changes in protist taxa occurrence and diversity metrics are effective predictors of key soil variables, with soil copper concentration, moisture, pH, and basal respiration being particularly well predicted.
[84]	Rice	CNN	A CNN model is developed to predict heavy metal (Cadmium, Lead, Chromium, Arsenic, and Mercury) concentrations in soil–rice system using 17 environmental factors. The model exhibits strong predictive accuracy, especially for Cadmium and Mercury. The study emphasises the model’s stability and robustness, particularly for quick predictions during emergencies.
[85]	Wheat, maize, peanut	RF, NN (regression, radial basis function), BPNN, ELM	Introduces a method for farmland surface soil moisture retrieval using feature (extracted from Sentinel-1/2 and Radarsat-2 remote sensing data) optimisation and ML. RF model exhibited the highest accuracy. The proposed method shows potential for accurate surface soil moisture retrieval and offers insights for future applications in other farmland surface types.
[86]	Not specified	ANN, KNN, SVM, RF, GBT, XGBoost, MLR, Cubist	Estimates soil water, salt contents, and bulk density from time domain reflectometry measurements using various ML algorithms. The research demonstrates that soil particle-size fractions are crucial predictors for all the targeted soil properties. XG-Boost is recommended for accurate soil gravimetric water content and bulk density estimation, while GBT is suggested for precise volumetric water content and soil salt content prediction.

5.4. Animal Management Domain

Animal (livestock and aquatic) production is a crucial part of agriculture, not only because it provides food and dairy products, but it also supplies other high-quality goods, such as wool and leather. Global demand for animal products is expected to increase further due to population growth [38], meaning that agrifood industries must optimise production practices by ensuring the welfare and safety of animals and increasing the capacity to prevent, detect, diagnose, and treat animal diseases. Considering this, there is a growing awareness that animal management can no longer be performed via traditional means and requires the adoption of new digital technologies.

The present SLR shows that the animal management domain comprises 12.5% of the research scope (Figure 7). The contents of Table 12 encompass a selection of seven articles dedicated to the utilisation of ML techniques in the domain of animal management. Smart animal monitoring systems have been viewed with great interest in the academic community, agrifood industries, and markets. Sensor-based animal wearables, computer vision systems, and other detection devices can capture the status of animals and environment in real time, which can be analysed afterwards with the aid of AI-based mechanisms to control and predict animals’ health, welfare, production, etc. Livestock monitoring includes information related to animals’ behaviour, physiology, clinical status, and performance [87], while in aquaculture, the desired information is more focused on water quality (water temperature, pH, dissolved oxygen content, ammonia, salt, etc.) [88,89].

Table 12. Machine learning applications in animal management domain.

Ref.	Animal	Models Used	Summary
[90]	Sheep	RF, SVM, LDA	Uses inertial motion sensors on 17 Merino sheep to collect behaviour data. Three ML approaches were employed to classify sheep behaviours accurately. Incorporating features from a range of time window sizes, spanning 2 to 15 s, significantly improved behaviour classification accuracy compared to a single window size. Among the methods, RF yielded the best results.
[91]	Dairy cows	DT, MLP, KNN, LSTM	Employs ML to detect abnormal behaviours in dairy cows with subacute ruminal acidosis (SARA), a condition known to induce behavioural changes. Monitoring 14 cows with SARA and 14 control cows involved tracking ruminal pH measurements and activity via stable-based positioning systems. KNN model exhibited the highest performance by identifying 83% of SARA cases. The study concludes that ML can successfully identify behaviour anomalies indicative of health issues.
[92]	Dairy cows	RF	This research offers insights into enhancing dairy farm management by predicting milk production trends under heat stress conditions, thereby increasing both productivity and animal welfare. The results demonstrate that the RF model is effective in detecting the impact of extreme heat conditions on milk yield, with an average relative error of about 18% for single daily yield predictions and 2% for total milk production.
[93]	Pigs	SVM, CNN	Automates the recognition and scoring of multiple postures of grouped pigs using depth images and a CNN-SVM model. The approach proves effective for detecting pig postures under commercial conditions, showing potential for improving pig welfare, health assessment, and behaviour analysis.
[94]	Invasive insects	DT, SVM, CNN, KNN, Ensemble learning (Boosted and Bagged Trees)	The proposed framework uses multi-modal data, including 3-D trajectories and infrared imagery, along with a multi-evidence approach to detect invasive insects near beehives. The framework achieves a high classification accuracy of 97.1% for <i>Vespa</i> hornets and honeybees, showing the potential to ensure the safety and smart monitoring of beehives against invasive species.
[95]	Fish	CNN, YOLO (v5)	The study uses a CNN for fish detection in recirculating aquaculture systems. The authors employ the one-stage YOLOv5 model and compare it with a two-stage Faster R-CNN model. The aim is to enhance fish production management via AI assistance.
[96]	Broilers	KNN, SVM, DT, RF, GBT	Identification of aflatoxin-poisoned broilers via wearable accelerometers and ML. Poisoned broilers exhibit distinct behavioural changes, such as reduced time spent on feeding, drinking, walking, and standing, as well as increased sitting behaviour. The study successfully demonstrated that the used ML models can accurately identify poisoned broilers, particularly those with higher aflatoxin concentrations, with GBT showing the best performance.

5.5. Main Findings

The study, development, and deployment of technologies stemming from the Agriculture 4.0 paradigm has revealed a multitude of transformative advances in the agricultural sector. By leveraging data-driven insights and advanced computational techniques, ML-based technologies are poised to further revolutionise the agricultural sector, driving efficiency, sustainability, and productivity to new heights [1].

5.5.1. Crop Management

ML techniques have demonstrated remarkable proficiency in evaluating crop quality attributes, enabling precise assessments without invasive testing. Additionally, they have revolutionised crop mapping and recognition, enhancing the accuracy of identifying specific crop varieties within agricultural landscapes. Moreover, ML-driven models exhibit exceptional capabilities in predicting crop yields by integrating diverse data sources, offering valuable insights into factors influencing the agricultural output. Additionally, ML-powered solutions have emerged as powerful tools for disease, pest, and weed detection. By leveraging satellite imagery and IoT-generated data, these models excel in accurately categorising and identifying diseases, pests, and weeds. This capability enables timely and effective interventions, minimising the impact of outbreaks on crop yield.

5.5.2. Water Management

Through the integration of advanced sensing techniques, coupled with IoT technologies, ML algorithms demonstrate exceptional proficiency in optimising water-related practices. Precision irrigation is a prominent application, where ML models suggest precise schedules based on data processed in real-time. In addition, these models excel at vigilantly monitoring water quality, ensuring that crops receive water with an optimal nutrient composition. Furthermore, ML-driven predictions of crop evapotranspiration rates offer valuable information on water requirements, facilitating a more sustainable approach to irrigation practices.

5.5.3. Soil Management

ML techniques have proven valuable in predicting soil properties, allowing farmers, researchers, and stakeholders to make informed decisions regarding soil fertility, moisture levels, and nutrient concentrations. By assimilating data from various sources, ML models provide valuable insights into the dynamic nature of soil behaviour, allowing for proactive adjustments in farming practices to ensure optimal conditions for crop growth and yield. Additionally, via the application of computer vision and remote sensing data, ML simplifies the monitoring of both crops and soil conditions by offering timely information on crop health, growth stages, and potential stressors.

5.5.4. Animal Management

The integration of ML with smart animal monitoring systems represents a significant leap forward in enhancing animal welfare and productivity. This innovative approach harnesses sensor-based wearables, computer vision systems, and other detection devices to capture real-time data on animal status and environmental conditions. ML algorithms, in tandem with these advanced technologies, enable the analysis of the captured data, providing valuable insights into animal health, behaviour, and overall wellbeing. This data can be processed and interpreted to control and predict various aspects of animal management, including health, welfare, and production.

6. Challenges and Research Opportunities

The present section focuses on answering RQ3 (*What are the challenges and future directions associated with integrating ML in agriculture and agricultural systems?*). The integration of ML in agriculture, although promising, still presents some challenges. In a study made by [1], various challenges were identified that need to be addressed to enable a successful transition towards Agriculture 4.0 paradigm. These are stratified into five main levels, namely device, data, network, application, and system. Of these levels, one that relates to the implementation of ML in agricultural systems is the data level. Table 13 provides an overview of some identified challenges covering a wide range of critical aspects related to the integration of ML in agriculture, along with possible solutions and further research.

Table 13. Challenges and proposed solutions for integrating Machine Learning in agriculture.

Challenge	Explanation	Proposed Solutions/Research Opportunities
Adaptability	Agricultural practices vary widely across regions, crops, and farming systems. Developing ML-based systems that are adaptable to diverse agricultural scenarios is a critical research frontier.	Developing adaptable models and algorithms that can be customised to suit diverse agricultural environments. Explore Transfer Learning techniques that allow models to leverage knowledge from one domain to another, making them more versatile and adaptable.

Table 13. Cont.

Challenge	Explanation	Proposed Solutions/Research Opportunities
Data accessibility	Encompass the efficient management of data, ensuring it is readily available to be used. For example, a delay in accessing data due to storage issues could hinder the real-time capabilities of ML applications.	Optimising data management systems and storage solutions, ensuring both efficiency and security.
Data accuracy	Accurate data are critical for training ML models. Inaccurate data can lead to incorrect predictions or recommendations.	Ensuring that data are accurate, credible, and trustworthy by exploring methods for data validation and quality assurance.
Data completeness	Incomplete data may result in biased or incomplete ML models. For example, missing data points in a crop monitoring dataset may hinder the model's ability to accurately predict crop yield.	Exploring techniques for data imputation/extrapolation to address missing data in agricultural datasets. Investigating methods for optimising models' performance in the presence of incomplete information (e.g., Feature Engineering).
Data consistency	Consistent data ensures that ML models are reliable and reproducible. For example, inconsistent labeling of images in a crop disease detection dataset could lead to incorrect classification.	Exploring data validation and cleaning techniques to ensure consistency in agricultural datasets. Developing techniques that can identify and rectify inconsistencies.
Data context	ML models need to be trained on data that are relevant to the specific agricultural task at hand. For example, using weather data from a different region may not provide accurate predictions for local farming conditions.	Investigating techniques for adapting ML models based on the specific agricultural context. A possible approach could be the use of Transfer Learning as it involves leveraging pre-trained models on similar tasks or domains and fine-tuning them using local data.
Data security and privacy	Agricultural data are often sensitive information that requires compliance with data protection regulations.	Exploring mechanisms that encompass data anonymisation, access control, and compliance with evolving data protection regulations will be crucial in building a foundation of trust for ML-driven agricultural solutions.
Data timeliness	Delayed/outdated data can lead to non-optimal results, impeding the potential benefits derived from ML-driven insights. However, it should be noted that there are scenarios in which historical data can be of significant use as it can offer invaluable insight into long-term trends, cyclical patterns, and the cumulative effects of farming practices.	Exploring methods for real-time data acquisition and processing that can adapt and make decisions based on the most up-to-date data, ensuring timely responses in ML applications. However, depending on the case at hand, a hybrid approach can be used, striking a balance between integrating real-time and historical data. This involves using real-time data for immediate decision making and integrating historical data for long-term strategic planning.
Human-machine collaboration	ML-based systems should enhance, rather than replace, human expertise in agriculture. Designing systems that facilitate seamless collaboration between stakeholders is an emerging area of research.	Designing collaborative decision making frameworks that seamlessly integrate ML insights with human expertise. Developing interfaces that empower users to interact with and guide ML models in agricultural tasks.
Interpretability and explainability	ML-based systems pose a significant challenge in gaining the trust and acceptance of farmers, stakeholders, and the agricultural industry. It is important to understand how models achieve their outputs.	Ensuring that ML models are transparent and that their inner workings are accessible. This means providing information on the features, variables, and algorithms that contribute to a model's results. Techniques such as SHAP values [97] or LIME [98] can be useful to identify which features are most influential in a model's predictions.
Limited literacy	Generally, aged workers may have limited literacy on digital technologies that could cause resistance or difficulties in adopting and effectively utilising technologies from the Agriculture 4.0.	Investing on training methods (e.g., workshops, courses), knowledge transfer, and skill-building in the context of ML-based technologies. Designing user-friendly interfaces tailored to older workers.
Resource constraints	ML-based systems often necessitate real-time processing and decision making. Remote regions or resource-constrained enterprises may lack the computational resources required for data processing.	Developing lightweight and efficient models that can operate effectively in low-resource scenarios. Investigating techniques for distributed and edge computing.

7. Conclusions, Limitations, and Future Work

7.1. Conclusions

Our study revealed notable findings when addressing RQ1 (*What are the most used ML algorithms in agriculture?*). As expected, RF emerged as the most prevalent ML algorithm, constituting 19.2% of the overall distribution. This popularity can be attributed to its versatility and robustness, which render it highly adept at handling intricate agricultural challenges. Following closely, SVM held the second position at 15.9%. Renowned for their efficacy in both classification and regression tasks, SVM have garnered substantial traction within the agricultural domain. GBT and CNN exhibited noteworthy adoption rates of 8.3% and 7.3%, respectively, further highlighting their significance in the agricultural sector.

By answering to RQ2 (*What are the impacts and outcomes of integrating ML in agriculture?*), it was possible to uncover a range of transformative effects arising from the integration of ML applications in the agricultural domain. Namely, we observed a substantial increase in the efficiency of agricultural production attributed to ML-based precision farming techniques, mainly with regard to advances in resource allocation strategies, ensuring that inputs such as water, fertilisers, and pesticides are used judiciously. Consequently, the environmental footprint of agricultural practices has been positively influenced by the incorporation of ML technologies. Through data-driven decision making, farmers can implement sustainable practices that reduce resource consumption and limit environmental impacts. For example, ML-powered precision irrigation systems can adaptively regulate water use based on real-time soil moisture data, promoting water conservation and maintaining optimal soil conditions. In addition, the integration of ML has substantially strengthened the agricultural sector's capabilities in managing diseases, pests, and weeds. ML algorithms have demonstrated remarkable accuracy in the early detection and classification of plant diseases, allowing for timely intervention and mitigation measures. This not only safeguards crop health, but it also mitigates potential yield losses. Overall, the integration of ML in agriculture represents a paradigm shift, propelling the sector towards a more efficient, sustainable, and technological future. The benefits go beyond mere productivity gains, encompassing a holistic transformation of agricultural practices to align with contemporary food safety and environmental management requirements. This bodes well for the resilience and adaptability of the agricultural sector in the face of evolving global challenges.

Given the growing potential of ML integration in agriculture, there are several open issues and promising avenues for future research. By answering RQ3 (*What are the challenges and future directions associated with integrating ML into agriculture and agricultural systems?*), it was possible to identify and explore some of these challenges and provide mitigation strategies. These include ensuring adaptable ML models, optimising data accessibility, and maintaining data accuracy, completeness, and consistency. Contextualising data usage, addressing security and privacy concerns, and ensuring timely data are also vital. Additionally, promoting human-machine collaboration, enhancing interpretability, and overcoming limited digital literacy among ordinary farmers are essential areas for attention. It is imperative to design ML applications with user-friendly interfaces that require minimal technical expertise. Incorporating intuitive visualisations and simple dashboards can enhance accessibility for farmers with limited literacy. Additionally, outreach programs and training initiatives tailored to the specific needs of agricultural communities can be implemented. Workshops, demonstrations, and educational campaigns can empower farmers with the knowledge and skills to effectively utilise ML-based technologies in their day-to-day practices. Moreover, in resource-constrained environments, developing lightweight models and exploring distributed computing methods are crucial steps toward a successful integration of ML in agriculture. Addressing these challenges will lead to a more effective and widespread implementation of ML technologies in the agricultural sector.

In summary, the integration of ML in agriculture produces substantial benefits, ranging from improved agricultural production and resource allocation to the better detection of diseases and pests and reduced environmental impacts. These advances pave the way for

a more sustainable and adaptable agricultural sector, ready to meet the demands of the future. As data-driven approaches continue to flourish, the agricultural landscape is on the brink of a more sophisticated and dynamic future, where technology and tradition converge harmoniously for the betterment of global agriculture.

7.2. Limitations

The extent of this review has been constrained to encompass Q1 articles, inadvertently leading to a decrease in the overall count of papers incorporated within the analysis. Due to the exclusion criteria used while performing the retrieval of identified research from the electronic databases, it is possible that some relevant publications might have been left out of the study. Hence, it is advisable that forthcoming researchers consider a broader range of literature sources (for example, ScienceDirect repository) to enrich the review's inclusiveness.

7.3. Future Work

Future research into the integration of ML in agriculture should focus on harnessing different data sources (such as satellite/drone imagery, IoT-based sensor data, and weather station information) for a better understanding of agricultural systems. In addition, the integration of ML with robotics and automation presents an opportunity for intelligent, self-learning systems capable of performing complex tasks for farmers or agricultural industries, such as the development of autonomous fruit-picking machines. Future efforts should focus on creating affordable and scalable ML solutions for regions with limited resources, ensuring that the benefits of the technology reach smallholder farmers and communities in developing areas. These research directions will move the field forward, leading to more sustainable, efficient, and resilient agricultural systems. Interdisciplinary collaboration between ML experts and professionals from specific fields, such as agronomy or chemistry, can lead to solutions tailored to agricultural challenges. Lastly, it would be interesting to carry out in-depth studies assessing the socio-economic impacts of the widespread adoption of ML in agriculture, including their effects on employment, economic viability, and equity in access to technological resources.

Author Contributions: Conceptualisation, S.O.A., R.S.P., J.B., F.L. and J.C.R.; methodology, S.O.A. and R.S.P.; formal analysis, S.O.A., R.S.P., J.B., F.L. and J.C.R.; investigation, S.O.A.; resources, S.O.A., R.S.P. and J.B.; data curation, S.O.A.; writing—original draft preparation, S.O.A.; writing—review and editing, S.O.A., R.S.P., J.B., F.L. and J.C.R.; visualisation, S.O.A.; supervision, J.B., F.L. and J.C.R.; project administration, J.B.; funding acquisition, J.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Fundação para a Ciência e a Tecnologia (FCT), Portugal, through the research units UNINOVA-CTS (UIDB/00066/2020), GeoBioTec (UIDP/04035/2020), CEF (UIDB/00239/2020), and the Associate Laboratory TERRA (LA/P/0092/2020).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript: AI (Artificial Intelligence); ANN (Artificial Neural Network); API (Application Programming Interface); BPNN (Backpropagation Neural Network); CNN (Convolutional Neural Network); DNN (Deep Neural Networks); DL (Deep Learning); DSS (Decision Support System); DT (Decision Tree); ELM (Extreme Learning Machines); ETa (Actual Evapotranspiration); ETc (Crop Evapotranspiration); ETo (Reference Evapotranspiration); FAO (Food and Agriculture Organisation); GBT (Gradient Boosted Tree); GRU (Gated Recurrent Unit); GAN (Generative Adversarial Networks); ICT (Information and Communication Technology); IoT (Internet of Things); KNN (k-Nearest Neighbors); LAI (Leaf Area Index); LDA (Linear Discriminant Analysis); LSTM (Long Short-Term Memory); MARS (Multivariate Adaptive Regression Splines); ML (Machine Learning); MLP (Multilayer Perceptron); MLR (Multiple Linear Regression); NB (Naïve Bayes); NLP (Natural Language Processing); NN (Neural Network); PCA (Principal Component

Analysis); PLS (Partial Least Squares); PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analysis); PSO (Particle Swarm Optimisation); RNN (Recurrent Neural Network); RF (Random Forest); RQ (Research Question); SLR (Systematic Literature Review); SVM (Support Vector Machine); UAV (Autonomous Unmanned Vehicle); VI (Vegetation index); WoS (Web of Science); XGBoost (Extreme Gradient Boosting); YOLO (You Only Look Once).

References

1. Araújo, S.O.; Peres, R.S.; Barata, J.; Lidon, F.; Ramalho, J.C. Characterising the Agriculture 4.0 Landscape—Emerging Trends, Challenges and Opportunities. *Agronomy* **2021**, *11*, 667. [CrossRef]
2. De Clercq, M.; Vats, A.; Biel, A. Agriculture 4.0: The future of farming technology. In Proceedings of the the World Government Summit, Dubai, United Arab Emirates, 11–13 February 2018; pp. 11–13.
3. Zambon, I.; Cecchini, M.; Egidi, G.; Saporito, M.G.; Colantoni, A. Revolution 4.0: Industry vs. agriculture in a future development for SMEs. *Processes* **2019**, *7*, 36. [CrossRef]
4. Liu, Y.; Ma, X.; Shu, L.; Hancke, G.P.; Abu-Mahfouz, A.M. From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges. *IEEE Trans. Ind. Inform.* **2020**, *17*, 4322–4334. [CrossRef]
5. Zhai, Z.; Martínez, J.F.; Beltran, V.; Martínez, N.L. Decision support systems for Agriculture 4.0: Survey and challenges. *Comput. Electron. Agric.* **2020**, *170*, 105256. [CrossRef]
6. Trendov, N.M.; Varas, S.; Zeng, M. *Digital Technologies in Agriculture and Rural Areas*; Briefing paper; FAO: Rome, Italy, 2019.
7. Rose, D.C.; Chilvers, J. Agriculture 4.0: Broadening responsible innovation in an era of smart farming. *Front. Sustain. Food Syst.* **2018**, *2*, 87. [CrossRef]
8. Ahmed, M.; Pathan, A.S.K. *Data Analytics: Concepts, Techniques, and Applications*; CRC Press: Boca Raton, FL, USA, 2018.
9. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef]
10. Mahesh, B. Machine learning algorithms—a review. *Int. J. Sci. Res. (IJSR)* **2020**, *9*, 381–386.
11. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Group, P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef]
12. PRISMA. Prisma Transparent Reporting of Systematic Reviews and Meta-Analyses. Available online: <http://www.prisma-statement.org/> (accessed on 6 July 2023).
13. Clarivate. Journal Citation Reports. Available online: <http://jcr.clarivate.com> (accessed on 6 July 2023).
14. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
15. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
16. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
17. Noble, W.S. What is a support vector machine? *Nat. Biotechnol.* **2006**, *24*, 1565–1567. [CrossRef]
18. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
19. Li, Q.; Wen, Z.; He, B. Practical federated gradient boosting decision trees. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 4642–4649. [CrossRef]
20. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6. [CrossRef]
21. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 6999–7019. [CrossRef] [PubMed]
22. Dietterich, T.G. Ensemble methods in machine learning. In Proceedings of the International Workshop on Multiple Classifier Systems, Cagliari, Italy, 21–23 June 2000; Springer: Berlin/Heidelberg, Germany, 2000; pp. 1–15.
23. Sagi, O.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249. [CrossRef]
24. Dong, X.; Yu, Z.; Cao, W.; Shi, Y.; Ma, Q. A survey on ensemble learning. *Front. Comput. Sci.* **2020**, *14*, 241–258. [CrossRef]
25. De Ville, B. Decision trees. *Wiley Interdiscip. Rev. Comput. Stat.* **2013**, *5*, 448–455. [CrossRef]
26. Krenker, A.; Bešter, J.; Kos, A. Introduction to the artificial neural networks. In *Artificial Neural Networks: Methodological Advances and Biomedical Applications*; InTech: London, UK, 2011; pp. 1–18.
27. Walczak, S. Artificial neural networks. In *Advanced Methodologies and Technologies in Artificial Intelligence, Computer Simulation, and Human-Computer Interaction*; IGI Global: Hershey, PA, USA, 2019; pp. 40–53. [CrossRef]
28. Liang, M.; Hu, X. Recurrent convolutional neural network for object recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3367–3375. [CrossRef]
29. Ray, S. A quick review of machine learning algorithms. In Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 14–16 February 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 35–39. [CrossRef]
30. Reddy, G.T.; Reddy, M.P.K.; Lakshmana, K.; Kaluri, R.; Rajput, D.S.; Srivastava, G.; Baker, T. Analysis of dimensionality reduction techniques on big data. *IEEE Access* **2020**, *8*, 54776–54788. [CrossRef]
31. Zheng, A.; Casari, A. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2018.

32. Cunningham, P.; Delany, S.J. k-Nearest neighbour classifiers-A Tutorial. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–25. [[CrossRef](#)]
33. Uddin, S.; Haque, I.; Lu, H.; Moni, M.A.; Gide, E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Sci. Rep.* **2022**, *12*, 6256. [[CrossRef](#)]
34. Rasmussen, C.E. Gaussian processes in machine learning. In *Summer School on Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 63–71.
35. Berrar, D. Bayes' theorem and naive Bayes classifier. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*; Elsevier: Amsterdam, The Netherlands, 2018; Volume 403, p. 412.
36. Thung, K.H.; Wee, C.Y. A brief review on multi-task learning. *Multimed. Tools Appl.* **2018**, *77*, 29705–29725. [[CrossRef](#)]
37. Zhang, Y.; Yang, Q. An overview of multi-task learning. *Natl. Sci. Rev.* **2018**, *5*, 30–43. [[CrossRef](#)]
38. FAO. *The Future of Food and Agriculture—Trends and Challenges*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2017.
39. Bauer, A.; Bostrom, A.G.; Ball, J.; Applegate, C.; Cheng, T.; Laycock, S.; Rojas, S.M.; Kirwan, J.; Zhou, J. Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: A case study of lettuce production. *Hortic. Res.* **2019**, *6*, 70. [[CrossRef](#)] [[PubMed](#)]
40. Manthou, E.; Lago, S.L.; Dargès, E.; Lianou, A.; Tsakanikas, P.; Panagou, E.Z.; Anastasiadi, M.; Mohareb, F.; Nychas, G.J.E. Application of spectroscopic and multispectral imaging technologies on the assessment of ready-to-eat pineapple quality: A performance evaluation study of machine learning models generated from two commercial data analytics tools. *Comput. Electron. Agric.* **2020**, *175*, 105529. [[CrossRef](#)]
41. Chawgien, K.; Kiattisin, S. Machine learning techniques for classifying the sweetness of watermelon using acoustic signal and image processing. *Comput. Electron. Agric.* **2021**, *181*, 105938. [[CrossRef](#)]
42. Zhu, Y.; Chen, M.; Gu, Q.; Zhao, Y.; Zhang, X.; Sun, Q.; Gu, X.; Zheng, K. Machine learning methods for efficient and automated in situ monitoring of peach flowering phenology. *Comput. Electron. Agric.* **2022**, *202*, 107370. [[CrossRef](#)]
43. Lu, J.; Dai, E.; Miao, Y.; Kusnierek, K. Improving active canopy sensor-based in-season rice nitrogen status diagnosis and recommendation using multi-source data fusion with machine learning. *J. Clean. Prod.* **2022**, *380*, 134926. [[CrossRef](#)]
44. Gomes, W.P.C.; Gonçalves, L.; da Silva, C.B.; Melchert, W.R. Application of multispectral imaging combined with machine learning models to discriminate special and traditional green coffee. *Comput. Electron. Agric.* **2022**, *198*, 107097. [[CrossRef](#)]
45. Zhang, Y.; Wang, X.; Liu, T.; Wang, R.; Li, Y.; Xue, Q.; Yang, P. Sustainable fertilisation management via tensor multi-task learning using multi-dimensional agricultural data. *J. Ind. Inf. Integr.* **2023**, *34*, 100461. [[CrossRef](#)]
46. Yang, X.; Zhang, R.; Zhai, Z.; Pang, Y.; Jin, Z. Machine learning for cultivar classification of apricots (*Prunus armeniaca* L.) based on shape features. *Sci. Hortic.* **2019**, *256*, 108524. [[CrossRef](#)]
47. Fernandes, A.M.; Utkin, A.B.; Eiras-Dias, J.; Cunha, J.; Silvestre, J.; Melo-Pinto, P. Grapevine variety identification using “Big Data” collected with miniaturized spectrometer combined with support vector machines and convolutional neural networks. *Comput. Electron. Agric.* **2019**, *163*, 104855. [[CrossRef](#)]
48. Del Valle, T.M.; Jiang, P. Comparison of common classification strategies for large-scale vegetation mapping over the Google Earth Engine platform. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *115*, 103092. [[CrossRef](#)]
49. Li, L.; Qiao, J.; Yao, J.; Li, J.; Li, L. Automatic freezing-tolerant rapeseed material recognition using UAV images and deep learning. *Plant Methods* **2022**, *18*, 1–13. [[CrossRef](#)] [[PubMed](#)]
50. Syazwani, R.W.N.; Asraf, H.M.; Amin, M.M.S.; Dalila, K.N. Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning. *Alex. Eng. J.* **2022**, *61*, 1265–1276. [[CrossRef](#)]
51. Yang, Z.; Diao, C.; Gao, F. Towards Scalable Within-Season Crop Mapping With Phenology Normalization and Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 1390–1402. [[CrossRef](#)]
52. Ballesteros, R.; Intrigliolo, D.S.; Ortega, J.F.; Ramírez-Cuesta, J.M.; Buesa, I.; Moreno, M.A. Vineyard yield estimation by combining remote sensing, computer vision and artificial neural network techniques. *Precis. Agric.* **2020**, *21*, 1242–1262. [[CrossRef](#)]
53. Chu, Z.; Yu, J. An end-to-end model for rice yield prediction using deep learning fusion. *Comput. Electron. Agric.* **2020**, *174*, 105471. [[CrossRef](#)]
54. Zheng, C.; Abd-Elrahman, A.; Whitaker, V.; Dalid, C. Prediction of Strawberry Dry Biomass from UAV Multispectral Imagery Using Multiple Machine Learning Methods. *Remote Sens.* **2022**, *14*, 4511. [[CrossRef](#)]
55. Chen, R.; Zhang, C.; Xu, B.; Zhu, Y.; Zhao, F.; Han, S.; Yang, G.; Yang, H. Predicting individual apple tree yield using UAV multi-source remote sensing data and ensemble learning. *Comput. Electron. Agric.* **2022**, *201*, 107275. [[CrossRef](#)]
56. Segarra, J.; Araus, J.L.; Kefauver, S.C. Farming and Earth Observation: Sentinel-2 data to estimate within-field wheat grain yield. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102697. [[CrossRef](#)]
57. Wang, Y.; Shi, W.; Wen, T. Prediction of winter wheat yield and dry matter in North China Plain using machine learning algorithms for optimal water and nitrogen application. *Agric. Water Manag.* **2023**, *277*, 108140. [[CrossRef](#)]
58. Selvaraj, M.G.; Vergara, A.; Montenegro, F.; Ruiz, H.A.; Safari, N.; Raymaekers, D.; Ocimati, W.; Ntamwira, J.; Tits, L.; Omondi, A.B.; et al. Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 110–124. [[CrossRef](#)]
59. Wang, X.; Liu, J.; Zhu, X. Early real-time detection algorithm of tomato diseases and pests in the natural environment. *Plant Methods* **2021**, *17*, 1–17. [[CrossRef](#)] [[PubMed](#)]

60. Nagasubramanian, G.; Sakthivel, R.K.; Patan, R.; Sankayya, M.; Daneshmand, M.; Gandomi, A.H. Ensemble classification and IoT-based pattern recognition for crop disease monitoring system. *IEEE Internet Things J.* **2021**, *8*, 12847–12854. [[CrossRef](#)]
61. Amarasingam, N.; Gonzalez, F.; Salgadoe, A.S.A.; Sandino, J.; Powell, K. Detection of White Leaf Disease in Sugarcane Crops Using UAV-Derived RGB Imagery with Existing Deep Learning Models. *Remote Sens.* **2022**, *14*, 6137. [[CrossRef](#)]
62. Abdulridha, J.; Ampatzidis, Y.; Qureshi, J.; Roberts, P. Identification and classification of downy mildew severity stages in watermelon utilizing aerial and ground remote sensing and machine learning. *Front. Plant Sci.* **2022**, *13*, 791018. [[CrossRef](#)] [[PubMed](#)]
63. Sriwana, K. Weather-based rice blast disease forecasting. *Comput. Electron. Agric.* **2022**, *193*, 106685. [[CrossRef](#)]
64. Shin, M.Y.; Viejo, C.G.; Tongson, E.; Wiechel, T.; Taylor, P.W.; Fuentes, S. Early detection of Verticillium wilt of potatoes using near-infrared spectroscopy and machine learning modeling. *Comput. Electron. Agric.* **2023**, *204*, 107567. [[CrossRef](#)]
65. Abbas, T.; Zahir, Z.A.; Naveed, M.; Kremer, R.J. Limitations of existing weed control practices necessitate development of alternative techniques based on biological approaches. *Adv. Agron.* **2018**, *147*, 239–280. [[CrossRef](#)]
66. de Castro, A.I.; Peña, J.M.; Torres-Sánchez, J.; Jiménez-Brenes, F.M.; Valencia-Gredilla, F.; Recasens, J.; López-Granados, F. Mapping cynodon dactylon infesting cover crops with an automatic decision tree-OBIA procedure and UAV imagery for precision viticulture. *Remote Sens.* **2019**, *12*, 56. [[CrossRef](#)]
67. Gée, C.; Denimal, E. RGB image-derived indicators for spatial assessment of the impact of broadleaf weeds on wheat biomass. *Remote Sens.* **2020**, *12*, 2982. [[CrossRef](#)]
68. Sapkota, B.; Singh, V.; Neely, C.; Rajan, N.; Bagavathiannan, M. Detection of Italian ryegrass in wheat and prediction of competitive interactions using remote-sensing and machine-learning techniques. *Remote Sens.* **2020**, *12*, 2977. [[CrossRef](#)]
69. El-Kenawy, E.S.M.; Khodadadi, N.; Mirjalili, S.; Makarovskikh, T.; Abotaleb, M.; Karim, F.K.; Alkahtani, H.K.; Abdelhamid, A.A.; Eid, M.M.; Horiuchi, T.; et al. Metaheuristic optimization for improving weed detection in wheat images captured by drones. *Mathematics* **2022**, *10*, 4421. [[CrossRef](#)]
70. Zhang, C.; Hu, Z.; Xu, L.; Zhao, Y. A YOLOv7 incorporating the Adan optimizer based corn pests identification method. *Front. Plant Sci.* **2023**, *14*, 1174556. [[CrossRef](#)]
71. Pereira, L.S.; Perrier, A.; Allen, R.G.; Alves, I. Evapotranspiration: Concepts and future trends. *J. Irrig. Drain. Eng.* **1999**, *125*, 45–51. [[CrossRef](#)]
72. Filgueiras, R.; Almeida, T.S.; Mantovani, E.C.; Dias, S.H.B.; Fernandes-Filho, E.I.; da Cunha, F.F.; Venancio, L.P. Soil water content and actual evapotranspiration predictions using regression algorithms and remote sensing data. *Agric. Water Manag.* **2020**, *241*, 106346. [[CrossRef](#)]
73. Brédy, J.; Gallichand, J.; Celicourt, P.; Gumiere, S.J. Water table depth forecasting in cranberry fields using two decision-tree-modeling approaches. *Agric. Water Manag.* **2020**, *233*, 106090. [[CrossRef](#)]
74. Akhter, F.; Siddiquei, H.R.; Alahi, M.E.E.; Jayasundera, K.P.; Mukhopadhyay, S.C. An IoT-enabled portable water quality monitoring system with MWCNT/PDMS multifunctional sensor for agricultural applications. *IEEE Internet Things J.* **2021**, *9*, 14307–14316. [[CrossRef](#)]
75. Zhao, L.; Zhao, X.; Zhou, H.; Wang, X.; Xing, X. Prediction model for daily reference crop evapotranspiration based on hybrid algorithm and principal components analysis in Southwest China. *Comput. Electron. Agric.* **2021**, *190*, 106424. [[CrossRef](#)]
76. Ndlovu, H.S.; Odindi, J.; Sibanda, M.; Mutanga, O.; Clulow, A.; Chimonyo, V.G.; Mabhaudhi, T. A comparative estimation of maize leaf water content using machine learning techniques and unmanned aerial vehicle (UAV)-based proximal and remotely sensed data. *Remote Sens.* **2021**, *13*, 4091. [[CrossRef](#)]
77. Vianny, D.M.M.; John, A.; Mohan, S.K.; Sarlan, A.; Ahmadian, A. Water optimization technique for precision irrigation system using IoT and machine learning. *Sustain. Energy Technol. Assessments* **2022**, *52*, 102307. [[CrossRef](#)]
78. Yang, H.; Wang, P.; Chen, A.; Ye, Y.; Chen, Q.; Cui, R.; Zhang, D. Prediction of phosphorus concentrations in shallow groundwater in intensive agricultural regions based on machine learning. *Chemosphere* **2023**, *313*, 137623. [[CrossRef](#)] [[PubMed](#)]
79. Na, A.; Isaac, W.; Varshney, S.; Khan, E. An IoT based system for remote monitoring of soil characteristics. In Proceedings of the 2016 International Conference on Information Technology (InCITE)—The Next Generation IT Summit on the Theme—Internet of Things: Connect your Worlds, Noida, India, 6–7 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 316–320. [[CrossRef](#)]
80. Taneja, P.; Vasava, H.K.; Daggupati, P.; Biswas, A. Multi-algorithm comparison to predict soil organic matter and soil moisture content from cell phone images. *Geoderma* **2021**, *385*, 114863. [[CrossRef](#)]
81. Yuan, J.; Wen, T.; Zhang, H.; Zhao, M.; Penton, C.R.; Thomashow, L.S.; Shen, Q. Predicting disease occurrence with high accuracy based on soil macroecological patterns of Fusarium wilt. *ISME J.* **2020**, *14*, 2936–2950. [[CrossRef](#)] [[PubMed](#)]
82. Glenn, A.J.; Moulin, A.P.; Roy, A.K.; Wilson, H.F. Soil nitrous oxide emissions from no-till canola production under variable rate nitrogen fertilizer management. *Geoderma* **2021**, *385*, 114857. [[CrossRef](#)]
83. Fournier, B.; Steiner, M.; Brochet, X.; Degruene, F.; Mammeri, J.; Carvalho, D.L.; Siliceo, S.L.; Bacher, S.; Peña-Reyes, C.A.; Heger, T.J. Toward the use of protists as bioindicators of multiple stresses in agricultural soils: A case study in vineyard ecosystems. *Ecol. Indic.* **2022**, *139*, 108955. [[CrossRef](#)]
84. Li, P.; Hao, H.; Mao, X.; Xu, J.; Lv, Y.; Chen, W.; Ge, D.; Zhang, Z. Convolutional neural network-based applied research on the enrichment of heavy metals in the soil–rice system in China. *Environ. Sci. Pollut. Res.* **2022**, *29*, 53642–53655. [[CrossRef](#)]
85. Zhao, J.; Zhang, C.; Min, L.; Guo, Z.; Li, N. Retrieval of farmland surface soil moisture based on feature optimization and machine learning. *Remote Sens.* **2022**, *14*, 5102. [[CrossRef](#)]

86. Wan, H.; Qi, H.; Shang, S. Estimating soil water and salt contents from field measurements with time domain reflectometry using machine learning algorithms. *Agric. Water Manag.* **2023**, *285*, 108364. [[CrossRef](#)]
87. Nasirahmadi, A.; Edwards, S.A.; Sturm, B. Implementation of machine vision for detecting behaviour of cattle and pigs. *Livest. Sci.* **2017**, *202*, 25–38. [[CrossRef](#)]
88. Raju, K.R.S.R.; Varma, G.H.K. Knowledge based real time monitoring system for aquaculture using IoT. In Proceedings of the 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, India, 5–7 January 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 318–321. [[CrossRef](#)]
89. Shi, X.; An, X.; Zhao, Q.; Liu, H.; Xia, L.; Sun, X.; Guo, Y. State-of-the-art internet of things in protected agriculture. *Sensors* **2019**, *19*, 1833. [[CrossRef](#)] [[PubMed](#)]
90. Hu, S.; Ingham, A.; Schmoelzl, S.; McNally, J.; Little, B.; Smith, D.; Bishop-Hurley, G.; Wang, Y.G.; Li, Y. Inclusion of features derived from a mixture of time window sizes improved classification accuracy of machine learning algorithms for sheep grazing behaviours. *Comput. Electron. Agric.* **2020**, *179*, 105857. [[CrossRef](#)]
91. Wagner, N.; Antoine, V.; Mialon, M.M.; Lardy, R.; Silberberg, M.; Koko, J.; Veissier, I. Machine learning to detect behavioural anomalies in dairy cows under subacute ruminal acidosis. *Comput. Electron. Agric.* **2020**, *170*, 105233. [[CrossRef](#)]
92. Bovo, M.; Agrusti, M.; Benni, S.; Torreggiani, D.; Tassinari, P. Random forest modelling of milk yield of dairy cows under heat stress conditions. *Animals* **2021**, *11*, 1305. [[CrossRef](#)] [[PubMed](#)]
93. Xu, J.; Zhou, S.; Xu, A.; Ye, J.; Zhao, A. Automatic scoring of postures in grouped pigs using depth image and CNN-SVM. *Comput. Electron. Agric.* **2022**, *194*, 106746. [[CrossRef](#)]
94. Nasir, A.; Ullah, M.O.; Yousaf, M.H. Ai in apiculture: A novel framework for recognition of invasive insects under unconstrained flying conditions for smart beehives. *Eng. Appl. Artif. Intell.* **2023**, *119*, 105784. [[CrossRef](#)]
95. Ranjan, R.; Tsukuda, S.; Good, C. Effects of image data quality on a convolutional neural network trained in-tank fish detection model for recirculating aquaculture systems. *Comput. Electron. Agric.* **2023**, *205*, 107644. [[CrossRef](#)]
96. Mei, W.; Yang, X.; Zhao, Y.; Wang, X.; Dai, X.; Wang, K. Identification of aflatoxin-poisoned broilers based on accelerometer and machine learning. *Biosyst. Eng.* **2023**, *227*, 107–116. [[CrossRef](#)]
97. SHAP. Welcome to the SHAP Documentation. Available online: <https://shap.readthedocs.io/en/latest/> (accessed on 19 September 2023).
98. Data Imaginist. LIME. Available online: <https://lime.data-imaginist.com/> (accessed on 19 September 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.