*Review*

# A Review of Machine Learning Techniques in Agroclimatic Studies

Dania Tamayo-Vera [1,2,3], Xiuquan Wang [3,4,*] and Morteza Mesbah [2]

1 School of Mathematical and Computational Sciences, University of Prince Edward Island, Charlottetown, PE C1A 4P3, Canada; dtamayovera@upei.ca
2 Charlottetown Research and Development Centre, Agriculture and Agri-Food Canada, Charlottetown, PE C1A 4N6, Canada; morteza.mesbah@agr.gc.ca
3 Canadian Center for Climate Change and Adaptation, University of Prince Edward Island, St. Peters Bay, PE C0A 2A0, Canada
4 School of Climate Change and Adaptation, University of Prince Edward Island, Charlottetown, PE C1A 4P3, Canada
* Correspondence: xxwang@upei.ca

**Abstract:** The interplay of machine learning (ML) and deep learning (DL) within the agroclimatic domain is pivotal for addressing the multifaceted challenges posed by climate change on agriculture. This paper embarks on a systematic review to dissect the current utilization of ML and DL in agricultural research, with a pronounced emphasis on agroclimatic impacts and adaptation strategies. Our investigation reveals a dominant reliance on conventional ML models and uncovers a critical gap in the documentation of methodologies. This constrains the replicability, scalability, and adaptability of these technologies in agroclimatic research. In response to these challenges, we advocate for a strategic pivot toward Automated Machine Learning (AutoML) frameworks. AutoML not only simplifies and standardizes the model development process but also democratizes ML expertise, thereby catalyzing the advancement in agroclimatic research. The incorporation of AutoML stands to significantly enhance research scalability, adaptability, and overall performance, ushering in a new era of innovation in agricultural practices tailored to mitigate and adapt to climate change. This paper underscores the untapped potential of AutoML in revolutionizing agroclimatic research, propelling forward the development of sustainable and efficient agricultural solutions that are responsive to the evolving climate dynamics.

**Keywords:** machine learning; agricultural research; deep learning; agricultural data; data processing; AutoML; crop management; pet diseases; smart farming; soil assessment

## 1. Introduction

Climate change presents a global challenge with diverse regional manifestations, impacting various facets of human life and the environment. Its effects are particularly acute in agriculture, necessitating in-depth studies to understand and mitigate these impacts [1]. In the realm of agriculture, climate change poses a significant threat to food security, due to the susceptibility of production systems to fluctuating environmental conditions. The effects of climate change on agriculture are both direct and indirect, ranging from altered precipitation patterns and increased temperatures to more frequent extreme weather events and shifts in pest and disease distributions [2,3]. These changes threaten agricultural productivity and call for innovative solutions to adapt and mitigate their impacts. Moreover, they have profound implications for nations worldwide, underscoring the need for innovative and adaptive solutions [4].

Numerous studies have demonstrated that machine learning (ML) and deep learning (DL) hold significant applications in agriculture [5–7]. ML and DL can enhance agricultural practices through precision farming, disease detection, yield prediction, and climate

impact modeling, thereby improving resilience to climate variability [7,8]. For instance, ML-enabled precision agriculture allows for the optimization of inputs, such as water and fertilizers, which is critical under the constraints of changing climatic conditions [9].

Although ML and DL have shown remarkable potential in revolutionizing agricultural practices, their application across the field exhibits notable inconsistency [10]. The complexities inherent in designing and implementing these models require deep technical knowledge, and the absence of a unified approach across studies highlights a significant barrier in environmental science research [10]. This diversity leads to challenges in data-processing standardization, model selection, hyperparameter tuning, and evaluation metrics, thereby questioning the replicability and scalability of such innovations. These challenges underscore the necessity for a more streamlined approach in employing ML and DL in agriculture.

Automated Machine Learning (AutoML) has emerged as a pivotal solution to these challenges. By automating critical aspects of model development, including selection and optimization, AutoML democratizes access to sophisticated data-analysis techniques. Automation is crucial for researchers and practitioners who lack extensive ML expertise, enabling them to leverage advanced computational tools more effectively [11,12]. Despite its transformative potential, the adoption of AutoML in agricultural climate science has remained limited [13]. This underutilization represents a significant missed opportunity to advance adaptive strategies that could mitigate the impacts of climate change on agriculture. Addressing this gap can unlock new avenues for enhancing the resilience and sustainability of agricultural practices through more accessible and efficient technological solutions.

This study aims to delve into the application of ML, DL, and AutoML within the agricultural sector, focusing on their role in addressing the impacts of climate change. We pose several research questions:

- What are the fundamental ML and AutoML methods used in assessing climate change impacts in agriculture?
- What performance metrics and evaluation methods are utilized to gauge the effectiveness of ML models in climate adaptation and mitigation within agriculture?
- What are the limitations and challenges in applying ML to climate change studies in agriculture?
- How prevalent are ML techniques compared to AutoML approaches in current climate science research?

The subsequent sections detail our methodology and explore these questions, aiming to shed light on the role of AutoML in transforming agricultural practices in the face of climate change.

## 2. Advancing Agriculture through Machine Learning

Climate change presents formidable challenges to agriculture, necessitating innovative approaches to sustainable and efficient farming practices. Machine learning has emerged as a pivotal technology in this domain, offering advanced solutions for planning, forecasting, and optimizing resource use to mitigate environmental impacts.

### 2.1. ML's Techniques in Agricultural Practices

Machine learning techniques are revolutionizing agricultural practices by enabling crop yield forecasts, optimizing farming conditions based on historical data and future projections, and facilitating precise predictions and planning [14]. These methodologies not only enhance the sustainability of agricultural outputs but also optimize resource use, such as water and fertilizers, contributing to environmental stewardship [15]. For instance, ML algorithms have been utilized to optimize fertilizer application rates, striking a balance between maximizing crop productivity and minimizing environmental impacts [16]. Additionally, ML is crucial in soil health assessment, evaluating critical soil properties to inform crop rotation and soil management strategies, thus preserving soil fertility and promoting robust crop yields [17].

The successful application of these techniques has marked the advent of a new era in precision farming, significantly improving crop management and yield. Studies such as [18,19] highlight the accuracy of ML models in forecasting crop yields, facilitating informed decision making among farmers and stakeholders. The integration of ML in disease management, demonstrated by [20,21], showcases the potential of convolutional neural networks (CNNs) in early plant disease identification, thus mitigating losses and reducing chemical pesticide dependency. Furthermore, advancements in soil analyses through ML, as discussed in [22–24], have enabled more precise soil health assessments, leading to optimized nutrient management and soil conservation practices. The development of smart agriculture applications, from plant classification to soil erosion modeling, illustrates the transformative role of these technologies [23,25].

These examples underscore the diverse applications of these technologies in agriculture, from disease detection and yield prediction to the automation of harvesting and navigation in orchards. By employing advanced algorithms, the agricultural sector can significantly improve productivity, sustainability, and resilience against climate change challenges. Table 1 illustrates the range of techniques currently applied across various agricultural practices, summarizing key methods alongside their targeted applications. From traditional models like Decision Trees (DT) and Random Forests (RF) to complex architectures such as CNNs and Long Short-Term Memory (LSTM) networks, the table below highlights how each method addresses specific agricultural challenges. This overview not only reaffirms the transformative potential of these technologies in enhancing agricultural outcomes but also emphasizes the importance of ongoing research and development to fully exploit their capabilities in addressing climate change and food security issues.

**Table 1.** Summary of Machine Learning and Deep Learning Techniques in Agricultural Applications.

| Reference | ML Technique | Agricultural Application |
|---|---|---|
| [18,26] | Decision Tree | Crop Yield Prediction, Disease Detection, Soil Assessment |
| [18–20] | Random Forest | Crop Yield Prediction, Disease Detection, Soil Assessment |
| [18,27] | Extreme Gradient Boosting | Crop Yield Prediction, Soil Assessment |
| [18,20] | Naive Bayes | Crop Yield Prediction, Disease Detection |
| [18,21] | K-Nearest Neighbors | Crop Yield Prediction, Disease Detection |
| [28] | Ensemble Traditional ML Models | Crop Yield Prediction |
| [26] | Multi-Linear Regressor | Crop Yield Prediction |
| [29] | RNN | Crop Yield Prediction |
| [29] | LSTM | Crop Yield Prediction |
| [29] | Support Vector Regression | Crop Yield Prediction |
| [23,24,30,31] | CNN | Crop Yield Prediction, Disease Detection |
| [30] | GNN | Crop Yield Prediction |
| [30] | U-Net | Crop Yield Prediction |
| [23,25,32] | ANN | Crop Yield Prediction, Disease Detection |
| [25] | DBSCAN | Crop Yield Prediction |
| [23,25] | Support Vector Machine | Crop Yield Prediction, Disease Detection, Smart Farming |
| [33] | Vision Transformers | Disease Detection |
| [22] | VGG-RNN Hybrid | Soil Assessment |
| [23,24] | MLP | Soil Assessment |

### 2.2. Distinguishing between ML and DL in Agricultural Applications

Machine learning encompasses a broad spectrum of algorithms that range from traditional methods to deep learning. Traditional methods, like Support Vector Machines (SVMs) and Random Forests, find extensive application in agriculture for tasks such as crop type classification and pest detection. These approaches are particularly effective for scenarios where the relationships between input variables and the target predictions are less complex and can be modeled with fewer data points [18,34].

DL, utilizing neural networks with multiple layers, excels at discerning complex patterns and predictions from large, unstructured datasets. It proves to be especially potent in processing vast image data, such as satellite imagery, for tasks including land-use classification and crop monitoring. DL has been pivotal in creating computer vision systems

for automated weed detection, fostering precision agriculture techniques that markedly reduce herbicide usage [31,35,36].

Agricultural ML tasks typically categorize into supervised or unsupervised learning [19,32,37]. Supervised learning, employing methods like Gradient Boosting (GB) and neural networks (NNs), suits predictive modeling tasks, such as estimating crop yields. Conversely, unsupervised learning, through methods such as K-Means clustering and a Principal Component Analysis (PCA), is instrumental in revealing hidden patterns within agricultural data [38]. Figure 1 delineates the key distinctions between supervised and unsupervised learning in analyzing agricultural data.
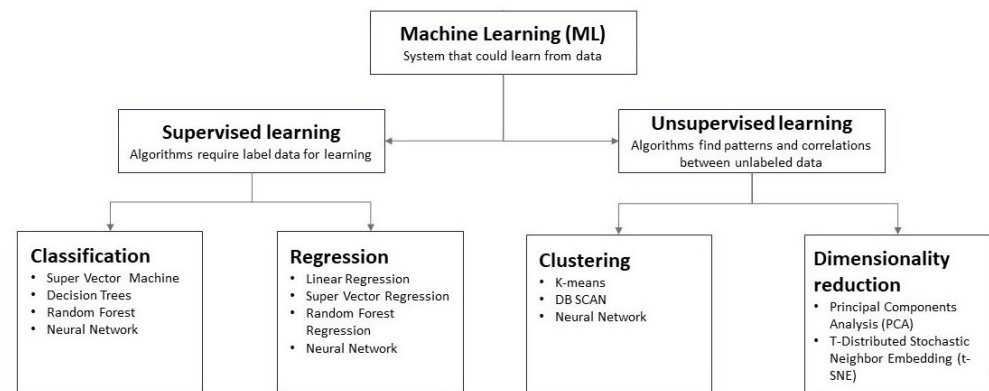
**Machine Learning (ML)**
System that could learn from data

**Supervised learning**
Algorithms require label data for learning

**Unsupervised learning**
Algorithms find patterns and correlations between unlabeled data

**Classification**
• Super Vector Machine
• Decision Trees
• Random Forest
• Neural Network

**Regression**
• Linear Regression
• Super Vector Regression
• Random Forest Regression
• Neural Network

**Clustering**
• K-means
• DB SCAN
• Neural Network

**Dimensionality reduction**
• Principal Components Analysis (PCA)
• T-Distributed Stochastic Neighbor Embedding (t-SNE)

**Figure 1.** Comparison of supervised and unsupervised learning in agriculture.

Moreover, semi-supervised learning, which uses both labeled and unlabeled data, has become an effective approach in situations where data labeling is too costly or impractical. This learning paradigm is notably beneficial for agricultural monitoring, especially in classifying crop types. By combining a small amount of labeled satellite imagery with larger volumes of unlabeled data, semi-supervised learning algorithms significantly improve classification accuracy and efficiency. Recent advancements have furthered these applications, showcasing semi-supervised learning's efficacy in analyzing complex agricultural structures via remote sensing images. Techniques such as the semi-supervised Extreme Learning Machine (SS-ELM) have been employed for enhanced precision and computational efficiency in classifying crop types and land uses [39,40]. These developments highlight the essential role of semi-supervised learning in leveraging remote sensing data for agricultural and environmental applications, providing innovative solutions to the enduring issues of data scarcity and labeling constraints.

*2.3. Enhancing ML Accessibility in Agriculture with AutoML*

The transformative potential of ML in agriculture is vividly illustrated through applications like crop yield forecasting, where ML models analyze historical yield data, weather patterns, and satellite imagery to predict future yields. This application not only aids in planning and resource allocation for farmers but also helps mitigate risks associated with climate variability. However, applying ML in agriculture faces hurdles due to the sector's complexity, characterized by nonlinear interactions among various factors. Advanced ML algorithms, capable of parsing through extensive datasets, are vital for modeling these complex dynamics. Yet, the need for consistent methodologies and variable selection remains challenging, highlighting the importance of systematic approaches in agricultural ML modeling [10].

AutoML simplifies ML applications by automating the selection, tuning algorithms within ML pipelines, and covering all steps from data preparation to model training [12]. It addresses the complexity of choosing and configuring the best algorithms for specific tasks, including supervised, unsupervised, and reinforcement learning, streamlining the model development process. This innovation democratizes ML, enabling practitioners across fields to employ cutting-edge solutions without requiring deep expertise. The necessity of

AutoML is highlighted by theoretical foundations like the no-free-lunch theorem, which argues that no single algorithm excels at every problem, advocating for a tailored approach to algorithm selection and configuration [41].

AutoML bridges the gap between complex models and practical agricultural applications, offering scalable solutions to food security and sustainability challenges. By automating the ML workflow, including algorithm selection and hyperparameter tuning, AutoML simplifies ML deployment in agriculture [11,42]. This extends from data preprocessing to model evaluation, highlighting its potential to democratize ML applications in the sector. Figure 2 showcases the AutoML framework construction process, demonstrating its role in making ML applications more accessible in agriculture.
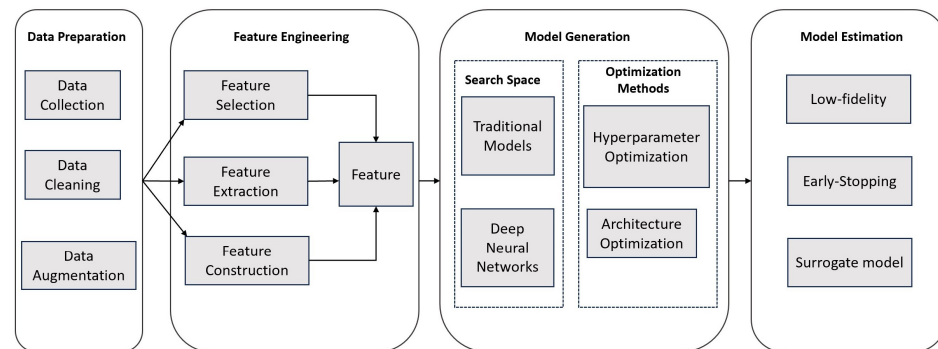


**Figure 2.** Overview of the AutoML workflow.

Many agricultural applications of ML and DL have focused on model development and improvement. For example, an advanced spatiotemporal convolutional neural network (STCNN) model was developed for detecting pineapples in complex field conditions, achieving high detection accuracy by leveraging the shifted window transformer architecture to outperform traditional methods under occlusion and varying light conditions [43]. Despite this success, AutoML could further enhance model efficiency and accuracy, potentially reducing inference time while maintaining high detection rates. AutoML-driven data augmentation techniques could improve model robustness against diverse environmental conditions. Automatic feature selection and engineering through AutoML could uncover new insights and enhance model performance, particularly in managing occlusion and fruit overlap scenarios. Moreover, AutoML can enhances the evaluation of diverse CNN architectures, facilitating the identification of optimal strategies for pineapple detection across varied scenarios. Additionally, leveraging Neural Architecture Search (NAS) within AutoML frameworks can further refine the selection process, tailoring architecture choices to specific environmental contexts.

The rise of ML and DL applications in agriculture brings an influx of novel metaheuristic approaches [44], signaling advanced model innovations. However, the challenge lies in effectively leveraging these methodologies due to the manual selection process from myriad potential combinations, hindering the full exploitation of these advanced techniques in agriculture. AutoML streamlines the optimization process, identifying novel approaches or configurations that elevate model performance in agricultural settings. It allows researchers and practitioners to focus on addressing agricultural challenges rather than on the intricacies of model building. This is particularly significant in climate change research, where AutoML can deepen insights into its impacts on agriculture and inform mitigation strategies. The shift from traditional techniques to AutoML marks a transformative advancement in agricultural research and practices, enabling the sector to better navigate the complexities of climate change and foster more resilient and productive farming systems.

## 3. Applications of ML and DL in Agriculture

Agriculture's pivotal role in global economic growth, contributing to approximately 4% of global GDP and over a quarter in the least-developed countries, underscores its

importance in poverty alleviation and food security, especially in the face of climate change challenges [45,46]. ML and DL technologies have emerged as key players in addressing these challenges, offering innovative solutions across various agricultural domains.

Crop Modeling and Yield Prediction: The quest to accurately predict crop yields has led to adopting ML and DL techniques, which analyze the intricate interplay between climate conditions and soil characteristics. While traditional ML methods such as Linear Regression (LR) and Random Forest (RF) lay the groundwork [34], DL approaches, including neural networks and convolutional neural networks (CNNs), are increasingly favored for their ability to process large datasets, thereby improving predictive accuracy [47]. The integration of ML and DL with remote sensing technologies has further revolutionized crop yield predictions. By harnessing the detailed, large-scale observational data provided by satellite and aerial imagery, these technologies enhance the analysis of climate conditions and soil characteristics. DL approaches, particularly CNNs, are adept at processing the complex, high-dimensional data from remote sensing, offering unprecedented accuracy in predictive analytics. This synergy enables more precise and informed agricultural planning, leveraging the global coverage and temporal frequency of remote sensing to monitor crop health and predict yields with greater accuracy [48].

Pest and Disease Management: The redistribution of pests and diseases due to climate change is a formidable challenge in agriculture, where ML demonstrates remarkable capabilities in early disease detection and efficient disease classification using advanced algorithms. The integration of technologies such as drones for data acquisition has been complemented by the adoption of hyperspectral imaging and Internet of Things (IoT) based sensors, enhancing the precision of ML algorithms and facilitating proactive pest management strategies. These technologies enable the detailed monitoring of crop health in real-time, significantly improving the management of pests and diseases [20,36,49].

Soil Health Assessment: Recognizing soil health as a critical component of agricultural productivity, ML has been pivotal in predicting soil properties, integrating high-resolution spatial data with climate dynamics to anticipate soil and crop yield outcomes. This convergence with crop modeling underscores the holistic approach essential in agricultural research, applying predictive models to soil fertility and crop yields through algorithms like RF, SVM, and GB. The inclusion of satellite imagery and IoT-based soil sensors enhances this process, offering deeper insights into soil conditions, thereby refining predictions and fostering sustainable farming practices [50,51].

Toward Smart Farming: The union of ML and DL methodologies in agriculture signifies a shift toward smart farming. This approach leverages data and ML to make farming more efficient and sustainable. Smart farming employs a comprehensive perspective by integrating soil, crop, and remotely sensed data, ensuring informed decision making. This paradigm enhances soil health assessment and pest management and extends to precision agriculture, where technology-driven solutions optimize farming practices to the specific conditions of each plot, thereby maximizing yields and minimizing environmental impact [52].

The integration of ML, DL, and other advanced technologies into agriculture signifies a pivotal shift toward precision farming. These technologies enhance crop yield predictions, pest management, and soil health assessments through sophisticated data analyses. Utilizing ML/DL in conjunction with remote sensing and IoT-based sensors provides deep insights, facilitates improved decision making, and enables real-time monitoring, promoting sustainable and efficient agricultural practices. This innovative approach not only confronts the challenges posed by climate change but also establishes a new benchmark for smart farming innovation, underscoring the vital role of technology in agriculture's future.

Tackling agricultural research complexities begins with defining the problem and assessing ML methods' potential over traditional models [10]. From data collection to preprocessing and feature engineering, the process requires substantial datasets for pattern recognition, given the diversity and noise in agricultural data [17,53,54]. Transforming raw data into actionable insights demands meticulous hyperparameter tuning and model opti-

mization, a process fraught with challenges requiring significant expertise [55,56]. Model explanation, often overlooked, is crucial for evaluating a model's practical applicability [57].

Innovations such as Auto-sklearn [58], Tree-based Pipeline Optimization Tool (TPOT) [59], and H2O AutoML [60,61] represent groundbreaking advancements in automating this process, particularly through hyperparameter optimization and NAS. These tools can be specialized in refining ML and DL applications in agriculture by streamlining the selection and tuning of models to enhance performance. Auto-sklearn leverages ensemble methods and Bayesian optimization, tailoring model selection and preprocessing steps to specific agricultural tasks, demonstrating the power of automated, data-driven approaches [58]. TPOT advances this automation using genetic programming, iteratively evolving machine learning pipelines for optimal performance, thereby simplifying the development of robust models for complex challenges [59,62]. H2O AutoML further democratizes ML application, offering an accessible interface for exploring diverse models and employing model stacking for superior predictive accuracy, invaluable in rapid model deployment scenarios [60,61]. In addition, these approaches significantly reduce the barrier to entry for applying sophisticated ML models, enabling more accurate and informed decisions without requiring users to have in-depth algorithmic understanding [63]. By abstracting the complexity of model selection and hyperparameter optimization, these technologies empower practitioners across disciplines to harness advanced ML models, facilitating informed decision making and innovation.

The forthcoming sections will examine the applications of ML, DL, and AutoML in existing literature, highlighting both the limitations and potential opportunities within the field.

## 4. Search, Screening, and Review Process

This study adopts a systematic literature review approach to comprehensively synthesize existing research on the application of ML, DL, and AutoML in the context of agriculture, particularly focusing on climate change impacts and adaptation strategies. Our methodology is designed to ensure rigor and transparency, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

To begin, we crafted a search strategy employing a set of keywords that encapsulate the core themes of our research. These keywords include "climate change impacts", "machine learning", "AutoML", "deep learning", "smart farming", "soil assessment", "crop modeling", "crop yield", "disease detection", "feature selection", "model selection", and "climate change adaptation and mitigation". Utilizing these terms, our initial query across multiple databases yielded a total of 1006 papers, setting the stage for our multi-layered screening process.

Our screening and selection process unfolds in three distinct stages, each designed to progressively refine the pool of relevant literature based on specific criteria aligned with our study's objectives.

**Stage 1:** Abstract Screening—The first layer of our screening process involved evaluating the abstracts of all retrieved papers. The goal at this stage was to filter out studies based on their direct relevance to the intersection of ML/AutoML and agricultural practices under the lens of climate change. This preliminary screening resulted in the retention of 564 papers, deemed potentially relevant for an in-depth analysis.

**Stage 2:** Full-Text Review—Subsequently, the narrowed selection underwent a rigorous full-text review, wherein each paper was assessed for its specific contributions to the domains of ML, DL, and their applications within agriculture. This stage was critical in identifying studies that offered substantial insights into modeling approaches, challenges, and solutions pertinent to agricultural climate research. The outcome of this review further reduced our pool to 232 papers.

**Stage 3:** In the final stage, we proceed with a detailed evaluation focusing on the specificity of modeling problems addressed, the robustness of parameter selection, the validity of model validation techniques, and the appropriateness of metrics, including considerations

for transfer learning. Adherence to these criteria led to the exclusion of papers that lacked comprehensive methodological details, culminating in a final selection of 66 papers that form the basis of our review.

The subsequent section, Section 5, delves into the critical advancements, challenges, and opportunities uncovered in our review.

## 5. Results and Discussion

### 5.1. Algorithms and Metrics Used in Agriculture Applications

The most commonly used algorithms in agriculture-related ML and DL studies are Artificial Neural Networks (ANNs), Random Forest, and Support Vector Machine (SVM) [16]. These techniques are applied in various agro-meteorological applications like maximizing crop yield and minimizing water use [64]. Deep learning approaches like AlexNet and GoogleNet have shown superior performance in plant classification over traditional methods like SVM. Multi-layer perceptron (MLP) neural networks and Random Forest Regression models are commonly used in agriculture for tasks like yield prediction, disease detection, and environmental monitoring. Ensemble methods and weight optimization techniques like ensemble modeling, where multiple models are combined, and weight optimization in classifiers like Support Vector Machine (SVM) are also employed. These methods aim to enhance the accuracy and reliability of predictions in agricultural applications [65].

Figure 3 provides a visual representation of the prevalence of various machine learning and deep learning algorithms across different agricultural applications. The data for this graph were derived from a comprehensive literature review of recent scientific papers that focus on the use of ML and DL in agriculture. Key algorithms such as Artificial Neural Networks, Random Forest, Support Vector Machines, AlexNet, GoogleNet, Convolutional Neural Networks, Long Short-Term Memory networks (LSTM), eXtreme Gradient Boosting (XGBoost), and Generative Adversarial Networks (GANs) are included. Each bar in the graph represents a specific algorithm, and the height of the bar indicates the frequency with which a particular algorithm is mentioned or utilized in that context. The applications include crop yield and water use, plant classification, disease detection, pest detection, and soil and water management.
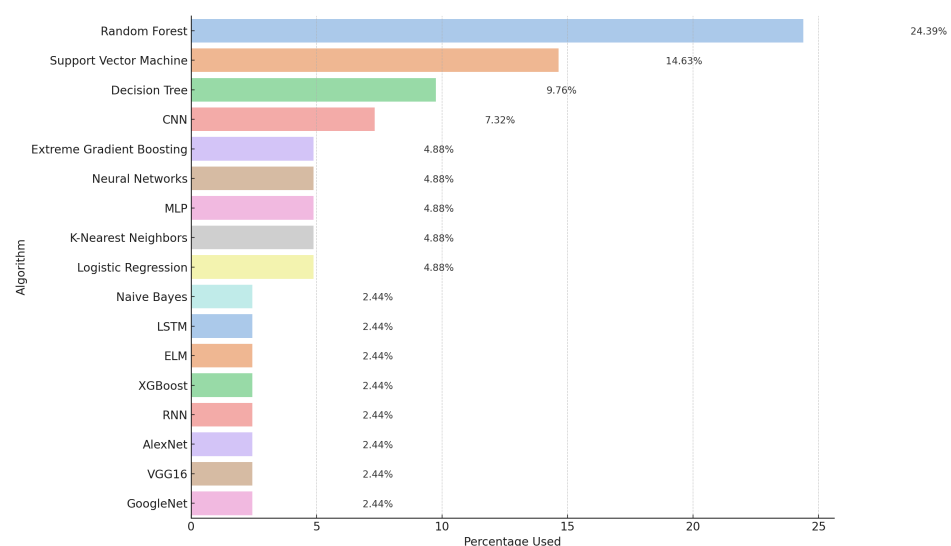


**Figure 3.** The graph represents the prevalence of machine learning and deep learning algorithms across various agricultural applications.

The performance of these models is typically evaluated using metrics like Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE) [19]. These metrics are crucial for evaluating models in agriculture, particularly for regression tasks. They measure the average magnitude of the errors in a set of predictions without

considering their direction. In classification tasks, the most used metrics are accuracy, Area Under the Curve (AUC), Sensitivity, specificity, False Negative Rate (FNR), and False Positive Rate (FPR). Accuracy measures the model's overall correctness, while AUC measures the model's ability to distinguish between classes. Sensitivity and specificity are used to evaluate the model's ability to identify positive and negative instances correctly. Other metrics used in the literature are Precision, Recall, and F1-score [18]. These metrics are particularly important in scenarios where the balance between false positives and false negatives is crucial, such as in disease detection or pest identification in crops [66]. Figure 4 shows the most frequently employed metrics in classification and regression tasks as identified in the reviewed studies.
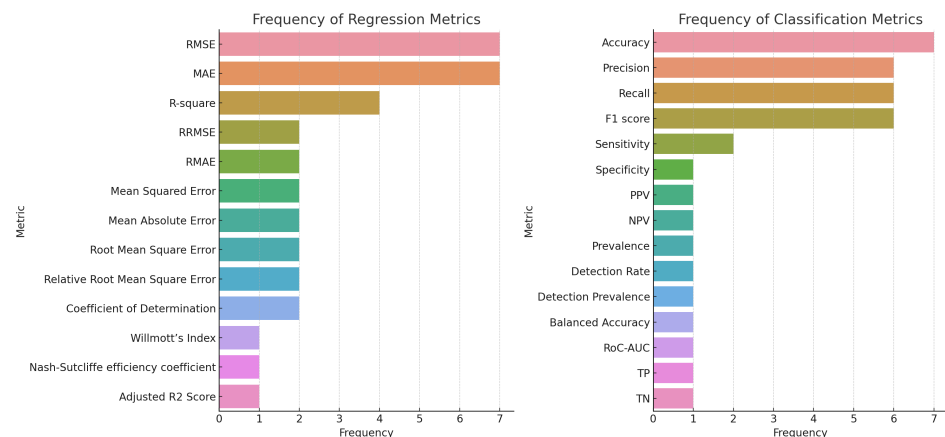


**Figure 4.** Metrics commonly utilized in both classification and regression tasks across reviewed studies.

Some studies compare various machine learning and deep learning models using a combination of the above metrics to determine the best-performing model for specific tasks, such as drought prediction or disease identification [67]. The choice of metrics and methods for model evaluation in agriculture depends on the specific task, whether it is classification or regression, and the nature of the agricultural problem being addressed. The key is to select metrics that accurately reflect the model's performance in real-world agricultural scenarios.

### 5.2. Challenges and Best Practices in Applying ML to Agriculture

Most applications in agriculture tend to use supervised learning, especially for classification and prediction tasks, as seen in studies like leaf disease detection [49,68] and corn plant disease classification [67]. Deep learning applications in agriculture face several challenges, including the need for large labeled datasets, high computational costs, and the complexity of interpreting DL models. Issues like overfitting, data imbalance, and variability in environmental conditions also pose significant challenges. There is a growing interest in integrating ML and DL with other technologies like IoT for precision agriculture [69,70] and using deep learning for more nuanced tasks like pest detection [71] and yield prediction [28]. However, there are some challenges and limitations, hence the approach in this area. Often, the datasets used for training DL and ML models in crop modeling need to be increased in size and diversity. This inadequacy can lead to models that do not generalize well to real-world conditions [21]. Accuracy issues arise due to environmental variations like nutrient levels, soil dampness, and temperature fluctuations. Farmers often need help in performing the optical observation of plant leaves for disease diagnoses due to limited resources [72]. Challenges in model interpretability and the need for significant computational resources to train and deploy models are notable issues [14]. In other applications, accurately classifying various rice leaf diseases and achieving high validation accuracy in models is difficult [68]. Inadequate pre-processing steps, lack of accurate feature identification, and suboptimal classification algorithms hinder accurate

disease grade measurement. Challenges related to similarity in disease symptoms and the extraction of irrelevant features are common.

### 5.3. Transparency Gaps in Data Processing for Agricultural ML

A critical observation in reviewing recent literature on ML and DL applications in agriculture is the frequent omission of detailed descriptions regarding data processing architectures and specifics of model construction [18,68]. This gap presents significant challenges in replicating studies, understanding the nuances of model performance, and applying these models in varied agricultural contexts.

One of the primary challenges in agricultural ML applications is data acquisition and quality [73]. The effectiveness of ML models heavily relies on the volume and quality of training data [10]. In agricultural contexts, obtaining a large, well-annotated dataset is often challenging, affecting the robustness and reliability of the models [74]. Few-shot learning methods and transfer learning have been utilized to address data scarcity. While these methods show promise, they introduce complexity and may only sometimes be directly applicable to diverse agricultural datasets [33,67]. Data processing is crucial in ML and DL applications, directly impacting the model's performance [10]. However, many studies need more information on how data is cleaned, transformed, or augmented. Key aspects such as handling missing values, normalization techniques, and data augmentation strategies should be reported and included more. This lack of transparency hinders the ability of other researchers to understand and replicate the research fully.

### 5.4. Challenges in Model Architecture and Training Transparency

Selecting appropriate model architectures and training methodologies is critical. Agricultural datasets often exhibit high variability, necessitating careful consideration in model selection to ensure adaptability to specific characteristics of agricultural data [73]. Additionally, the performance evaluation of these models is complicated by the diversity of datasets used in different studies.

Several ML and DL studies tend to focus on the outcomes rather than the journey of reaching those outcomes [19–21,37]. For instance, studies often highlight the accuracy and effectiveness of their models without providing a comprehensive breakdown of the model architecture, parameter settings, or training processes [18,20,28]. Information about the layers in a neural network, the activation functions used, or the specifics of the optimization algorithms is vital for a thorough understanding of the model's performance and applicability.

Figure 5 outlines the methodologies employed in the analyzed studies, focusing on feature engineering, model selection, hyperparameter optimization, and validation strategies. Of the 66 papers examined, merely 12 provided comprehensive discussions on these methodologies, supported by examples and citations [27,67,75]. Notably, 39% (26 studies) omitted documentation on feature selection, and a significant 61% (40 studies) acknowledged the features used without detailing the selection process. The practice of feature scaling was largely overlooked, with 92% (61 studies) not detailing the method and only 7% (5 papers) mentioning normalization. Regarding model validation and testing, 44% (29 studies) mentioned cross-validation without specifying the k-fold value, and 26 papers outlined a data split for training and testing without justifying the chosen split ratio.

Moreover, for some papers that did provide certain details, the information concerning the model was often insufficient for result reproduction [19,26,30,32,68]. A substantial inconsistency was observed regarding metric selection, with many papers comparing their results to randomly chosen algorithms. While they claim improved performance, particularly with DL techniques, they frequently omitted architecture search details or failed to explain how they determined their architecture as superior to other possible combinations. This issue persisted even when parameters and hyperparameters were reported to have been tuned, suggesting a need for more transparency in the tuning process.
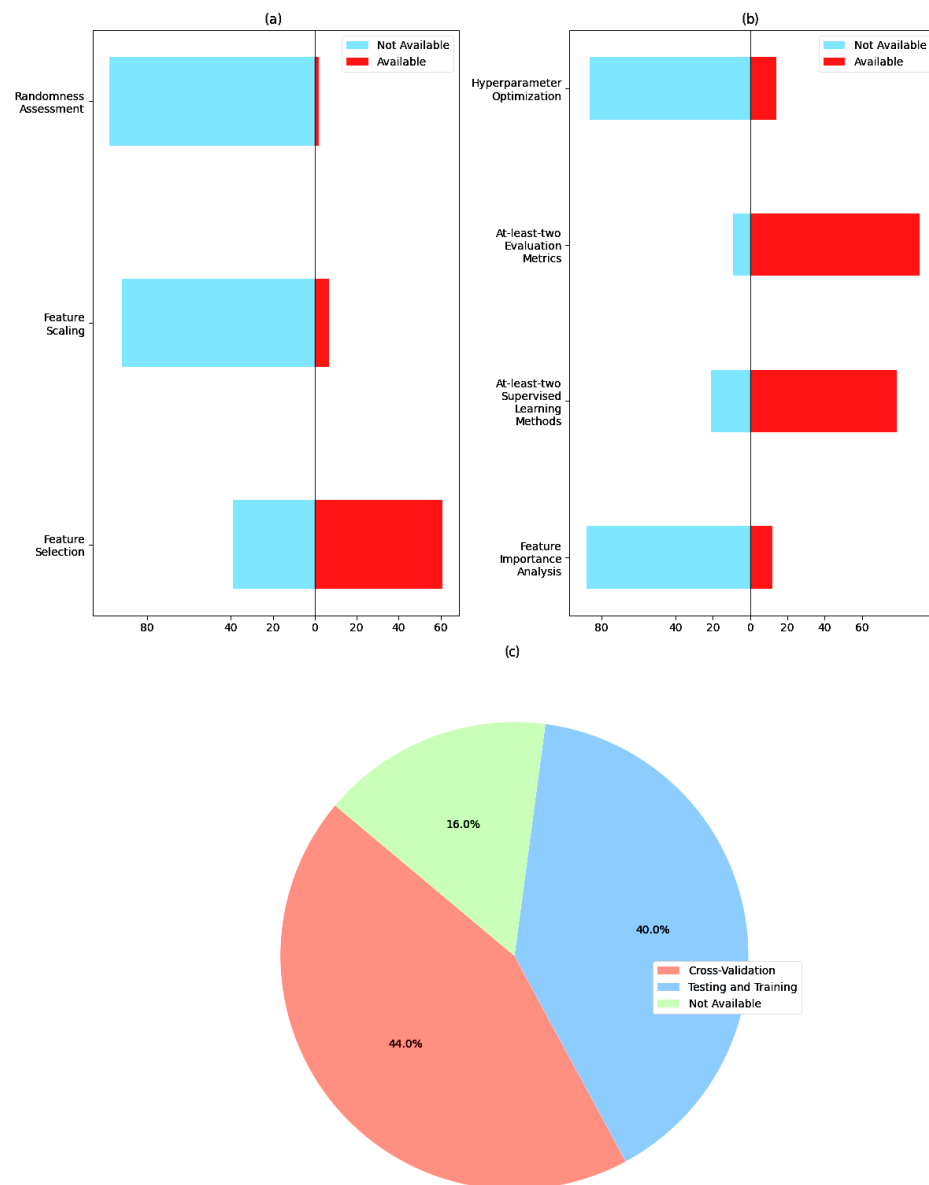
**Figure 5.** Quantitative assessment of data preprocessing, model selection, fine-tuning of parameters, feature selection, and model validation techniques. (**a**) The bar graph shows the types of preprocessing techniques used in different studies to process the dataset. It indicates that randomness assessment and feature scaling are widely selected, whereas feature selection is less frequently present. (**b**) This graph presents a similar analysis, revealing that hyperparameter optimization is commonly addressed in the studies, while the use of multiple evaluation metrics and multiple ML algorithms is less common, and feature importance analysis is notably scarce. (**c**) The pie chart illustrates the distribution of dataset split methodologies, with cross-validation being the most prevalent, followed by separate testing and training sets, and a significant portion of studies not mentioning the protocol used for evaluation.

### 5.5. Enhancing Replicability and Scalability in Agriculture through AutoML

The absence of comprehensive descriptions regarding data processing and model construction in agricultural ML and DL applications can pose significant challenges in terms of replicability, scalability, and adaptability of these models. Agricultural datasets often exhibit intricate patterns and domain-specific nuances, making it imperative to address these challenges. Traditional ML models may need help to effectively capture the inherent complexities of agricultural data, leading to suboptimal results. Moreover,

developing custom ML models for agricultural applications demands considerable time and expertise [73]. This may impede research progress and limit the accessibility of advanced ML techniques to a broader audience. Additionally, these customized models may need more scalability and reproducibility to apply research findings across diverse agricultural scenarios, hindering their broader utility.

On the contrary, AutoML tools offer a promising solution by streamlining critical aspects of the modeling process [12]. They automate tasks such as model selection, hyperparameter tuning, and feature engineering, resulting in substantial time and resource savings for researchers. Furthermore, AutoML democratizes ML by making it accessible to researchers with varying levels of expertise, potentially fostering a wider adoption of ML techniques within the agricultural community [11]. AutoML pipelines provide standardized workflows that enhance experiment reproducibility across various agricultural studies and settings. These tools also adapt to evolving agricultural data and research questions, offering flexibility and agility in addressing complex challenges [76].

However, it is essential to consider these types of frameworks' potential benefits and inherent limitations. AutoML, by automating ML pipelines, it can significantly reduce the time and computational resources required for model development [77]. This automation, however, confronts the combinatorial optimization problem, where there is a trade-off between computational resources and the time necessary for finding optimal solutions [78]. Implementing heuristic and metaheuristic algorithms within AutoML frameworks can mitigate this issue by providing efficient solutions within a reasonable timeframe. Additionally, AutoML's ease of use might lead to a superficial grasp of machine learning concepts among users, potentially causing misinterpretations or difficulties in troubleshooting suboptimal model performance [77]. The opaque nature of AutoML systems further complicates transparency and accountability, particularly critical in precision agriculture where model trustworthiness is crucial. The risk lies in the potential for AutoML to exacerbate existing biases within agricultural datasets, potentially leading to skewed results. However, it is important to recognize that these limitations are not unique to AutoML, but rather extend to the broader field of machine learning [79].

Mitigating risks associated with AutoML in agriculture goes beyond simply using the tool. Rigorous validation and testing protocols are crucial to guard against overfitting and bias. This involves splitting data into training, validation, and testing sets, employing cross-validation techniques, and selecting relevant evaluation metrics specific to the agricultural task. Additionally, emphasizing model interpretability is essential [79]. A feature importance analysis and the use of interpretable models like decision trees can help growers understand the factors influencing model outputs. Furthermore, Explainable AI (XAI) techniques can shed light on the model's internal workings, fostering trust and ensuring responsible application [80]. By combining these practices, AutoML frameworks can be leveraged to democratize access to advanced analytics in agriculture. Standardizing validation and testing procedures within the framework ensures a consistent level of rigor across users. Similarly, promoting interpretable models and XAI techniques empowers growers to understand the model's reasoning and assess its alignment with established agricultural practices.

The success of AutoML in agriculture hinges on a symbiotic relationship between automated tools and the intrinsic understanding of machine learning principles by the researchers [81]. Such knowledge is crucial for selecting suitable AutoML tools, interpreting results accurately, diagnosing issues, and critically evaluating model outputs within the agricultural context. This foundational understanding is indispensable for identifying and correcting errors, as the sophistication of a model does not exempt it from generating erroneous outputs if the input data or underlying assumptions are flawed [82]. Future responsible deployment of AutoML in agriculture relies on the integration of automated efficiency and the invaluable depth of human expertise, establishing a balance where automation enhances analytical accessibility without compromising the need for critical, domain-specific insight. In this manner, AutoML can serve not just as a tool

for efficiency but as a collaborative partner in the nuanced field of agricultural research, blending the strengths of technology with the irreplaceable value of human knowledge and experience [82].

The no-free-lunch theorem reminds us that no single algorithm excels at solving every problem [41]. Despite this, AutoML frameworks democratize access to machine learning expertise, particularly in the agricultural context [11]. They shift the focus toward more knowledge-intensive tasks, such as identifying meaningful features for inclusion in datasets, thus emphasizing a data-centric approach over model development [83]. This shift can enhance the quality of agricultural research outcomes. Even slight improvements in model precision, which might seem marginal in statistical terms, can have significant practical implications in the field. For instance, a small increase in model precision can be critical in disease detection within agriculture, as timely detection is crucial. Though numerically modest, such improvements could substantially impact farmers' decision making and operational efficiency [8].

An illustrative example of how the Tree-based Pipeline Optimization Tool (TPOT) can be applied in agricultural applications is showcased in [59]. This research highlights TPOT's utility in addressing the challenges of high-dimensional datasets prevalent in vegetation mapping and analyses. Multi-date Sentinel-2 images, rich in phenological and canopy structural information, present a complex, high-dimensional dataset crucial for enhancing mapping accuracy. The study focuses on optimizing classification accuracies for landscapes infested by the invasive parthenium weed (Parthenium hysterophorus), a significant agricultural pest. By employing TPOT, which utilizes a genetic algorithm to automatically generate and optimize machine learning pipelines, the researchers were able to navigate the complexity of the multi-date image data effectively. The TPOT's application facilitated the identification and selection of the most relevant features from the dataset, optimizing the classification process despite the data's high dimensionality and the presence of redundant variables. A novel approach, named "ReliefF-Svmb-EXT-TPOT", combining feature selection with TPOT's optimization capabilities, was also tested to enhance the efficiency of the classification. This integrated method aimed to reduce computational costs while maintaining high accuracy levels. The study reported overall accuracies of 91.9% with the TPOT model and an improved 92.6% using the ReliefF-Svmb-EXT-TPOT approach, demonstrating the potential of TPOT, particularly when combined with feature selection techniques, to handle complex agricultural datasets effectively. This example underscores the promise of TPOT in agricultural applications, particularly in precision agriculture tasks such as invasive species mapping, by leveraging its sophisticated algorithmic capabilities to process and analyze high-dimensional remote sensing data, thus offering a powerful tool for enhancing agricultural decision making and environmental management.

Figure 6 showcases the application of ML in agricultural research, highlighting the deployment of traditional ML and DL methods. While adopting AutoML in agriculture is currently limited, its benefits are increasingly evident. As AutoML tools continue to evolve and cater to the specific requirements of agricultural research, we anticipate a shift toward more widespread adoption. This transition holds the potential to overcome the limitations associated with traditional ML approaches and accelerate innovation and problem solving within the agriculture sector. To facilitate this shift, it is recommended that future research places a greater emphasis on documenting and sharing comprehensive details of data processing techniques and model architectures. This approach enhances research transparency, contributes to the collective knowledge base, and paves the way for more robust and adaptable ML and DL solutions in agriculture.
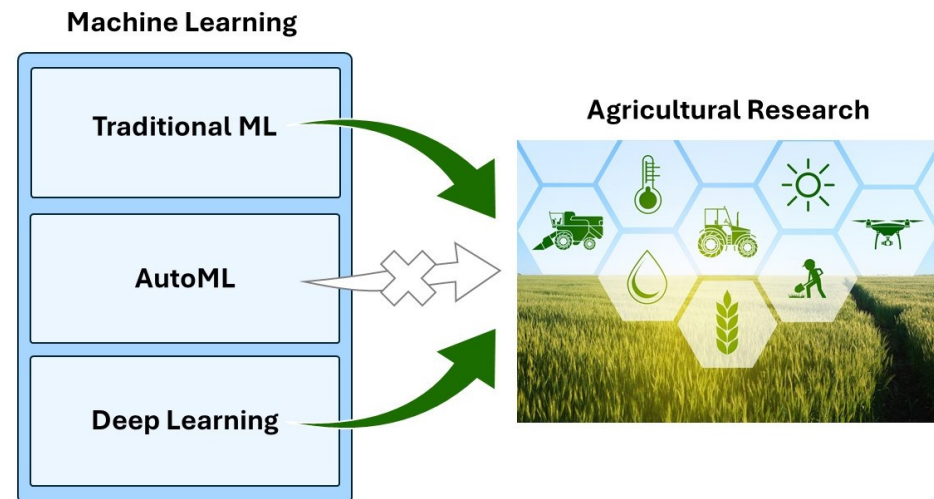
**Machine Learning**

**Agricultural Research**

Traditional ML

AutoML

Deep Learning

**Figure 6.** The integration of machine learning techniques in agricultural research, depicting the use of traditional ML and deep learning, with a noted absence of AutoML applications.

*5.6. Future Research Directions*

Addressing the challenges and future directions in agricultural machine learning and deep learning research requires a focused approach that acknowledges current successes and recognizes improvement areas. A key concern in the field is the frequent omission of detailed reporting on model architectures, training processes, and data handling, as exemplified in Section 5. This gap in the literature hampers the replication of studies and the broader application of successful models. To enhance the transparency and replicability of ML and DL applications in agriculture, future studies must adopt more rigorous standards for documenting and sharing models and datasets.

Despite these issues, the availability of diverse datasets from satellite imagery has been instrumental in advancing the field. These datasets provide invaluable insights into crop health, growth patterns, and environmental impacts. However, integrating such complex and high-dimensional data into ML models introduces additional complexity, necessitating sophisticated modeling and analysis techniques. AutoML emerges as a promising solution to these challenges, offering a pathway to simplify the development and application of ML models [12]. By automating critical aspects of the ML pipeline, such as model selection, hyperparameter optimization, and feature engineering, AutoML can make advanced ML techniques more accessible and applicable to a broader range of agricultural problems. Consider the application of satellite data for crop yield prediction. Traditional approaches might involve manually developing and tuning CNN models to analyze satellite images and extract relevant features for yield prediction. This process requires significant expertise and can be time-consuming and prone to error. In contrast, AutoML tools can automate the selection of the optimal model architecture and preprocessing steps, streamlining the development process. For example, an AutoML framework could automatically test various CNN architectures, adjust preprocessing techniques for satellite imagery, and tune hyperparameters based on performance metrics, all without extensive manual intervention [84]. This AutoML-driven approach simplifies the modeling process and opens up new possibilities for integrating diverse data sources into predictive models. By reducing the barriers to advanced ML applications, AutoML can accelerate innovation in agricultural research, leading to more accurate, efficient, and scalable solutions for crop yield prediction and other critical challenges.

To fully realize the potential of AutoML in agriculture, future research must focus on developing AutoML tools tailored to the unique characteristics of agricultural data and challenges. This includes improving data preprocessing techniques, incorporating domain-specific knowledge into AutoML algorithms, and enhancing model interpretability. By addressing these areas, the agricultural research community can leverage AutoML to

advance the state of the art in agricultural ML and DL, ultimately contributing to more sustainable and productive farming practices worldwide.

## 6. Conclusions

The exploration of AutoML systems within the agricultural sector uncovers a transformative potential to redefine traditional machine learning methodologies. This paper has highlighted critical challenges facing agricultural ML applications, including the need for detailed documentation on model architectures, training processes, and data handling. Such gaps hinder the replicability of studies and the broad application of successful models, ultimately limiting the field's progress.

AutoML stands out as a promising solution to these challenges. By automating the selection of models, tuning hyperparameters, and streamlining data preprocessing, AutoML has the potential to democratize ML technologies, making them accessible to a broader range of researchers and practitioners. This accessibility is crucial for tackling complex agricultural challenges, where integrating advanced ML techniques can significantly enhance decision-making processes, improve yield predictions, and contribute to sustainable farming practices. Furthermore, this paper underscores the necessity of integrating domain-specific knowledge and enhancing data diversity in ML models. Emphasizing the importance of transparency in ML applications, it advocates for rigorous documentation standards that can facilitate the replication of research findings. Looking ahead, the incorporation of AutoML into agricultural research represents a paradigm shift toward more efficient, scalable, and adaptable ML applications. This shift promises to advance the sector's technological capabilities and foster a more collaborative and innovative research environment. To realize these benefits, it is imperative for the research community to prioritize the comprehensive documentation of ML workflows, from data acquisition to model deployment. Such efforts will contribute to building a transparent, reproducible, and robust foundation for future ML and DL solutions in agriculture.

In conclusion, the paper advocates for a concerted effort to embrace AutoML technologies, alongside a commitment to enhancing research transparency and collaboration. By addressing the identified challenges and harnessing the potential of AutoML, we can unlock transformative opportunities for progress in agricultural research, driving forward the global agenda for food security and sustainability.

**Author Contributions:** Conceptualization, X.W.; methodology, D.T.-V.; validation, D.T.-V.; formal analysis, D.T.-V.; investigation, D.T.-V.; resources, X.W.; data curation, D.T.-V.; writing—original draft preparation, D.T.-V.; writing—review and editing, M.M. and X.W.; visualization, D.T.-V.; supervision, X.W. and M.M.; project administration, X.W.; funding acquisition, X.W. All authors have read and agreed to the published version of the manuscript.

## References

1. Calvin, K.; Fisher-Vanden, K. Quantifying the indirect impacts of climate on agriculture: An inter-method comparison. *Environ. Res. Lett.* **2017**, *12*, 115004. [CrossRef]
2. Wheeler, T.; Von Braun, J. Climate change impacts on global food security. *Science* **2013**, *341*, 508–513. [CrossRef] [PubMed]
3. Lesk, C.; Rowhani, P.; Ramankutty, N. Influence of extreme weather disasters on global crop production. *Nature* **2016**, *529*, 84–87. [CrossRef] [PubMed]
4. Futia, G.; Vetrò, A. On the Integration of Knowledge Graphs into Deep Learning Models for a More Comprehensible AI—Three Challenges for Future Research. *Information* **2020**, *11*, 122. [CrossRef]

5.  Domingues, T.; Brandão, T.; Ferreira, J.C. Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey. *Agriculture* **2022**, *12*, 1350. [CrossRef]
6.  Sharma, A.; Jain, A.; Gupta, P.; Chowdary, V. Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access* **2021**, *9*, 4843–4873. [CrossRef]
7.  Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674. [CrossRef]
8.  Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
9.  Weersink, A.; Fraser, E.; Pannell, D.; Duncan, E.; Rotz, S. Opportunities and challenges for big data in agricultural and environmental analysis. *Annu. Rev. Resour. Econ.* **2018**, *10*, 19–37. [CrossRef]
10. Zhu, J.J.; Yang, M.; Ren, Z.J. Machine Learning in Environmental Research: Common Pitfalls and Best Practices. *Environ. Sci. Technol.* **2023**, *57*, 17671–17689. [CrossRef]
11. He, X.; Zhao, K.; Chu, X. AutoML: A survey of the state-of-the-art. *Knowl.-Based Syst.* **2021**, *212*, 106622. [CrossRef]
12. Salehin, I.; Islam, M.S.; Saha, P.; Noman, S.; Tuni, A.; Hasan, M.M.; Baten, M.A. AutoML: A systematic review on automated machine learning with neural architecture search. *J. Inf. Intell.* **2024**, *2*, 52–81. [CrossRef]
13. Li, K.Y.; Burnside, N.G.; de Lima, R.S.; Peciña, M.V.; Sepp, K.; Cabral Pinheiro, V.H.; de Lima, B.R.C.A.; Yang, M.D.; Vain, A.; Sepp, K. An automated machine learning framework in unmanned aircraft systems: New insights into agricultural management practices recognition approaches. *Remote Sens.* **2021**, *13*, 319. [CrossRef]
14. Sharma, R.; Kamble, S.S.; Gunasekaran, A.; Kumar, V.; Kumar, A. A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Comput. Oper. Res.* **2020**, *119*, 104926. [CrossRef]
15. Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors* **2021**, *21*, 3758. [CrossRef] [PubMed]
16. Peng, W.; Karimi Sadaghiani, O. A review on the applications of machine learning and deep learning in agriculture section for the production of crop biomass raw materials. *Energy Sources Part A Recover. Util. Environ. Eff.* **2023**, *45*, 9178–9201. [CrossRef]
17. Mohamed, S.A.; Metwaly, M.M.; Metwalli, M.R.; AbdelRahman, M.A.E.; Badreldin, N. Integrating Active and Passive Remote Sensing Data for Mapping Soil Salinity Using Machine Learning and Feature Selection Approaches in Arid Regions. *Remote Sens.* **2023**, *15*, 1751. [CrossRef]
18. Maheswari, M.U.; Ramani, R. A Comparative Study of Agricultural Crop Yield Prediction Using Machine Learning Techniques. In Proceedings of the 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 17–18 March 2023; IEEE: Piscataway, NJ, USA, 2023; Volume 1, pp. 1428–1433.
19. Kuradusenge, M.; Hitimana, E.; Hanyurwimfura, D.; Rukundo, P.; Mtonga, K.; Mukasine, A.; Uwitonze, C.; Ngabonziza, J.; Uwamahoro, A. Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. *Agriculture* **2023**, *13*, 225. [CrossRef]
20. Prem, G.; Hema, M.; Basava, L.; Mathur, A. Plant Disease Prediction using Machine Learning Algorithms. *IJCA* **2018**, *182*, 0975-8887. [CrossRef]
21. Sai, P.M.; SushmaSri, V.; Sailu, V.H.; Pradeepthi, U.; Kavitha, M.; Kavitha, S. Detection Of Leaf Diseases In Modern Agriculture Using Deep Learning Techniques. In Proceedings of the 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 23–25 January 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [CrossRef]
22. Pawar, D.; Rais Allauddin Mulla, D.S.H.; Shikalgar, S.; Jethva, H.B.; Patel, G.A. A Novel Hybrid AI Federated ML/DL Models for Classification of Soil Components. *Int. J. Recent Innov. Trends Comput. Commun.* **2022**, *10*, 190–199. [CrossRef]
23. Kumar, R.; Chug, A.; Singh, A.P.; Singh, D. A Systematic Analysis of Machine Learning and Deep Learning Based Approaches for Plant Leaf Disease Classification: A Review. *J. Sens.* **2022**, *2022*, e3287561. [CrossRef]
24. Sirsat, M.; Cernadas, E.; Fernández-Delgado, M.; Barro, S. Automatic prediction of village-wise soil fertility for several nutrients in India using a wide range of regression methods. *Comput. Electron. Agric.* **2018**, *154*, 120–133. [CrossRef]
25. Elavarasan, D.; Vincent, D.R.; Sharma, V.; Zomaya, A.Y.; Srinivasan, K. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Comput. Electron. Agric.* **2018**, *155*, 257–282. [CrossRef]
26. Kouadio, L.; Deo, R.C.; Byrareddy, V.; Adamowski, J.F.; Mushtaq, S. Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. *Comput. Electron. Agric.* **2018**, *155*, 324–338. [CrossRef]
27. Chala, A.T.; Ray, R.P. Machine Learning Techniques for Soil Characterization Using Cone Penetration Test Data. *Appl. Sci.* **2023**, *13*, 8286. [CrossRef]
28. Keerthana, M.; Meghana, K.; Pravallika, S.; Kavitha, M. An ensemble algorithm for crop yield prediction. In Proceedings of the 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 4–6 February 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 963–970.
29. Shook, J.; Gangopadhyay, T.; Wu, L.; Ganapathysubramanian, B.; Sarkar, S.; Singh, A.K. Crop yield prediction integrating genotype and weather variables using deep learning. *PLoS ONE* **2021**, *16*, e0252402. [CrossRef] [PubMed]
30. Singh, G.; Singh, S.; Sethi, G.; Sood, V. Deep learning in the mapping of agricultural land use using Sentinel-2 satellite data. *Geographies* **2022**, *2*, 691–700. [CrossRef]

31. Wolanin, A.; Mateo-García, G.; Camps-Valls, G.; Gómez-Chova, L.; Meroni, M.; Duveiller, G.; Liangzhi, Y.; Guanter, L. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. *Environ. Res. Lett.* **2020**, *15*, 024019. [CrossRef]

32. Fenu, G.; Malloci, F.M. An Application of Machine Learning Technique in Forecasting Crop Disease. In Proceedings of the 2019 3rd International Conference on Big Data Research, Cergy-Pontoise, France, 20–22 November 2019; pp. 76–82. [CrossRef]

33. Zhuang, L. Deep-Learning-Based Diagnosis of Cassava Leaf Diseases Using Vision Transformer. In Proceedings of the 2021 4th Artificial Intelligence and Cloud Computing Conference, AICCC '21, New York, NY, USA, 17–19 December 2022; pp. 74–79. [CrossRef]

34. Cravero, A.; Sepúlveda, S. Use and Adaptations of Machine Learning in Big Data—Applications in Real Cases in Agriculture. *Electronics* **2021**, *10*, 552. [CrossRef]

35. Elnashar, H.S. Deep Learning: Potato, Sweet Potato Protection and Leafs Diseases Detections. In *Integrated Emerging Methods of Artificial Intelligence & Cloud Computing*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 529–539.

36. Kumar, M.; Kumar, A.; Palaparthy, V.S. Soil Sensors-Based Prediction System for Plant Diseases Using Exploratory Data Analysis and Machine Learning. *IEEE Sens. J.* **2021**, *21*, 17455–17468. [CrossRef]

37. Fernández, D.; Adermann, E.; Pizzolato, M.; Pechenkin, R.; Rodríguez, C.G.; Taravat, A. Comparative Analysis of Machine Learning Algorithms for Soil Erosion Modelling Based on Remotely Sensed Data. *Remote Sens.* **2023**, *15*, 482. [CrossRef]

38. Xu, H.; Croot, P.; Zhang, C. Discovering hidden spatial patterns and their associations with controlling factors for potentially toxic elements in topsoil using hot spot analysis and K-means clustering analysis. *Environ. Int.* **2021**, *151*, 106456. [CrossRef]

39. Feng, Z.; Huang, G.; Chi, D. Classification of the complex agricultural planting structure with a semi-supervised extreme learning machine framework. *Remote Sens.* **2020**, *12*, 3708. [CrossRef]

40. Khan, S.; Tufail, M.; Khan, M.T.; Khan, Z.A.; Iqbal, J.; Alam, M. A novel semi-supervised framework for UAV based crop/weed classification. *PLoS ONE* **2021**, *16*, e0251008. [CrossRef]

41. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [CrossRef]

42. Barbudo, R.; Ventura, S.; Romero, J.R. Eight years of AutoML: Categorisation, review and trends. *Knowl. Inf. Syst.* **2023**, *65*, 5097–5149. [CrossRef]

43. Meng, F.; Li, J.; Zhang, Y.; Qi, S.; Tang, Y. Transforming unmanned pineapple picking with spatio-temporal convolutional neural networks. *Comput. Electron. Agric.* **2023**, *214*, 108298. [CrossRef]

44. Sörensen, K. Metaheuristics—The metaphor exposed. *Int. Trans. Oper. Res.* **2015**, *22*, 3–18. [CrossRef]

45. Loizou, E.; Karelakis, C.; Galanopoulos, K.; Mattas, K. The role of agriculture as a development tool for a regional economy. *Agric. Syst.* **2019**, *173*, 482–490. [CrossRef]

46. Gitz, V.; Meybeck, A.; Lipper, L.; Young, C.D.; Braatz, S. Climate change and food security: Risks and responses. *Food Agric. Organ. United Nations (FAO) Rep.* **2016**, *110*, 3–36.

47. Zhang, Q.; Liu, Y.; Gong, C.; Chen, Y.; Yu, H. Applications of Deep Learning for Dense Scenes Analysis in Agriculture: A Review. *Sensors* **2020**, *20*, 1520. [CrossRef]

48. Joshi, A.; Pradhan, B.; Gite, S.; Chakraborty, S. Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review. *Remote Sens.* **2023**, *15*, 2014. [CrossRef]

49. Kumar, S.; Prasad, K.; Srilekha, A.; Suman, T.; Rao, B.P.; Vamshi Krishna, J.N. Leaf Disease Detection and Classification based on Machine Learning. In Proceedings of the 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 9–10 October 2020; pp. 361–365. [CrossRef]

50. Khanal, S.B.; Lynch, J.P. The opening of Pandora's Box: Climate change impacts on soil fertility and crop nutrition in developing countries. *Plant Soil* **2010**, *335*, 101–115. [CrossRef]

51. Khanal, S.; Fulton, J.; Klopfenstein, A.; Douridas, N.; Shearer, S. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Comput. Electron. Agric.* **2018**, *153*, 213–225. [CrossRef]

52. Prabha, C.; Pathak, A. Enabling Technologies in Smart Agriculture: A Way Forward Towards Future Fields. In Proceedings of the 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 5–6 May 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 821–826.

53. Blair, G.S.; Henrys, P.; Leeson, A.; Watkins, J.; Eastoe, E.; Jarvis, S.; Young, P.J. Data Science of the Natural Environment: A Research Roadmap. *Front. Environ. Sci.* **2019**, *7*, 121. [CrossRef]

54. Mafarja, M.M.; Mirjalili, S. Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing* **2017**, *260*, 302–312. [CrossRef]

55. Raschka, S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv* **2018**, arXiv:1811.12808.

56. Liaw, R.; Liang, E.; Nishihara, R.; Moritz, P.; Gonzalez, J.E.; Stoica, I. Tune: A research platform for distributed model selection and training. *arXiv* **2018**, arXiv:1807.05118.

57. Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. A survey of methods for explaining black box models. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–42. [CrossRef]

58. Feurer, M.; Eggensperger, K.; Falkner, S.; Lindauer, M.; Hutter, F. Auto-sklearn 2.0: Hands-free automl via meta-learning. *J. Mach. Learn. Res.* **2022**, *23*, 11936–11996.

59. Kiala, Z.; Odindi, J.; Mutanga, O. Determining the capability of the tree-based pipeline optimization tool (tpot) in mapping parthenium weed using multi-date sentinel-2 image data. *Remote Sens.* **2022**, *14*, 1687. [CrossRef]

60. LeDell, E.; Poirier, S. H$_2$O automl: Scalable automatic machine learning. In Proceedings of the AutoML Workshop at ICML, Online, 17–18 July 2020; Volume 2020.

61. Lee, S.; Kim, J.; Bae, J.H.; Lee, G.; Yang, D.; Hong, J.; Lim, K.J. Development of Multi-Inflow Prediction Ensemble Model Based on Auto-Sklearn Using Combined Approach: Case Study of Soyang River Dam. *Hydrology* **2023**, *10*, 90. [CrossRef]

62. Laadan, D.; Vainshtein, R.; Curiel, Y.; Katz, G.; Rokach, L. MetaTPOT: Enhancing a tree-based pipeline optimization tool using meta-learning. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Online, 19–23 October 2020; pp. 2097–2100.

63. Azevedo, K.; Quaranta, L.; Calefato, F.; Kalinowski, M. A Multivocal Literature Review on the Benefits and Limitations of Automated Machine Learning Tools. *arXiv* **2024**, arXiv:2401.11366. [CrossRef]

64. Jala, P.K.; Meenal, R.; Nagabushanam, P.; Selvakumar, A.I.; Jude Hemanth, D.; Rajasekaran, E. Machine Learning, Deep Learning Models for Agro-Meteorology Applications. In Proceedings of the 2023 4th International Conference on Signal Processing and Communication (ICSPC), Coimbatore, India, 23–24 March 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 196–200.

65. Peppes, N.; Daskalakis, E.; Alexakis, T.; Adamopoulou, E.; Demestichas, K. Performance of machine learning-based multi-model voting ensemble methods for network threat detection in agriculture 4.0. *Sensors* **2021**, *21*, 7475. [CrossRef]

66. Shoaib, M.; Shah, B.; EI-Sappagh, S.; Ali, A.; Ullah, A.; Alenezi, F.; Gechev, T.; Hussain, T.; Ali, F. An advanced deep learning models-based plant disease detection: A review of recent research. *Front. Plant Sci.* **2023**, *14*, 1158933. [CrossRef]

67. Rahul Kumar, V.; Shrishti, V.; Sridhar, P.A. Corn Plant Disease Classification using a combination of Machine Learning and Deep Learning. In Proceedings of the 2022 International Conference on Futuristic Technologies (INCOFT), Belgaum, India, 24–26 November 2022; pp. 1–4. [CrossRef]

68. Aggarwal, M.; Khullar, V.; Goyal, N. Exploring Classification of Rice Leaf Diseases using Machine Learning and Deep Learning. In Proceedings of the 2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM), Uttar Pradesh, India, 22–24 February 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.

69. Memon, K.; Umrani, F.A.; Baqai, A.; Syed, Z.S. A Review Based On Comparative Analysis of Techniques Used in Precision Agriculture. In Proceedings of the 2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 17–18 March 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–7.

70. Sharma, A.; Sharma, A.; Tselykh, A.; Bozhenyuk, A.; Choudhury, T.; Alomar, M.A.; Sánchez-Chero, M. Artificial intelligence and internet of things oriented sustainable precision farming: Towards modern agriculture. *Open Life Sci.* **2023**, *18*, 20220713. [CrossRef]

71. Golatkar, N.; Hemalatha, N. Applications of deep learning in agriculture (pest-detection). *Redshine Arch.* **2023**, *1*. [CrossRef]

72. Ahmed, A.A.; Reddy, G.H. A mobile-based system for detecting plant leaf diseases using deep learning. *AgriEngineering* **2021**, *3*, 478–493. [CrossRef]

73. 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. [CrossRef]

74. Jones, J.W.; Antle, J.M.; Basso, B.; Boote, K.J.; Conant, R.T.; Foster, I.; Godfray, H.C.J.; Herrero, M.; Howitt, R.E.; Janssen, S.; et al. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agric. Syst.* **2017**, *155*, 269–288. [CrossRef] [PubMed]

75. Crane-Droesch, A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ. Res. Lett.* **2018**, *13*, 114003. [CrossRef]

76. Karmaker, S.K.; Hassan, M.M.; Smith, M.J.; Xu, L.; Zhai, C.; Veeramachaneni, K. Automl to date and beyond: Challenges and opportunities. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–36. [CrossRef]

77. Hutter, F.; Kotthoff, L.; Vanschoren, J. *Automated Machine Learning: Methods, Systems, Challenges*; Springer Nature: Berlin/Heidelberg, Germany, 2019.

78. Feurer, M.; Hutter, F. Hyperparameter optimization. In *Automated Machine Learning: Methods, Systems, Challenges*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 3–33.

79. Gardner, S.; Golovidov, O.; Griffin, J.; Koch, P.; Shi, R.; Wujek, B.; Xu, Y. Fair AutoML Through Multi-objective Optimization. In Proceedings of the ESEC/FSE 2023: 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering Fix Fairness, Don't Ruin Accuracy: Performance Aware Fairness Repair Using AutoML, San Francisco, CA, USA, 3–9 December 2023; pp. 502–514.

80. Ryo, M. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artif. Intell. Agric.* **2022**, *6*, 257–265.

81. Khuat, T.T.; Kedziora, D.J.; Gabrys, B. The roles and modes of human interactions with automated machine learning systems. *arXiv* **2022**, arXiv:2205.04139.

82. Lee, D.J.L.; Macke, S. A Human-in-the-loop Perspective on AutoML: Milestones and the Road Ahead. *IEEE Data Eng. Bull.* **2020**, *42*, 59–70.

83.  Liashchynskyi, P.; Liashchynskyi, P. Grid search, random search, genetic algorithm: A big comparison for NAS. *arXiv* **2019**, arXiv:1912.06059.
84.  Stamoulis, D.; Ding, R.; Wang, D.; Lymberopoulos, D.; Priyantha, B.; Liu, J.; Marculescu, D. Single-path mobile automl: Efficient convnet design and nas hyperparameter optimization. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 609–622. [CrossRef]