Review Article

Leveraging Machine Learning Techniques in Agriculture: Applications, Challenges, and Opportunities

ABSTRACT

Machine learning (ML) is rapidly transforming agriculture by introducing data-driven insights and automation into farming practices, enabling precision, efficiency, and sustainability. This paper explores the foundational concepts of ML, distinguishing it from conventional programming approaches through its ability to learn patterns and make predictions from data without explicit instructions. The discussion delves into the applications of ML in agriculture, including crop management, water resource optimization, soil quality assessment, and livestock health monitoring, highlighting its potential to address complex challenges. Furthermore, the paper outlines various ML algorithms, such as decision trees, support vector machines, neural networks, and ensemble methods, emphasizing their suitability for specific agricultural tasks. Despite its promising potential, the adoption of ML in agriculture faces several challenges, including data scarcity, model interpretability, high implementation costs, and limited technical expertise among farmers. This study aims to provide a comprehensive overview of the transformative role of ML in agriculture while critically analyzing the barriers to its widespread adoption. By addressing these challenges, ML can become a cornerstone of sustainable and innovative agricultural practices.

Key words: Agriculture,Big-Data, Crop Management, Machine Learning, Sustainability

1. Introduction

1.1 Background and significance of machine learning in agriculture

The concept of the "Digital Agricultural Revolution," also known as "Agriculture 4.0," signifies a profound transformation within the agricultural sector, propelled by the integration of advanced technologies. This revolution encompasses a wide array of innovative tools such as the Internet of Things (IoT), cloud computing, artificial intelligence (AI), big data analytics, sophisticated sensing technologies, autonomous robotics, and decision support systems (DSS) (Mostafa *et al.*, 2019). The essence of this shift lies in the deployment of sensors and robotics to acquire crucial field data, which is then transmitted via IoT networks to either local or cloud-based servers for further processing, analysis, and storage (Liakos*et al.*, 2018). Through the application of big data methodologies and AI-driven analytics, this raw data is transformed into actionable insights. Decision support systems play a pivotal role by equipping stakeholders with the necessary analytical capabilities to optimize agricultural processes, thereby enhancing user engagement and fostering data-driven decision-making (Anagnostis*et al.*, 2020).A pivotal aspect of AI within Agriculture 4.0 is machine learning (ML), a subset of AI that has exhibited remarkable promise in augmenting various dimensions of this field. Machine learning refers to a computational system or algorithm that can autonomously improve its performance on specific tasks by analysing patterns within data, without requiring explicit programming for each task (Benos*et al.*, 2021). It is an iterative learning process, where the computer's decisions evolve based on diverse data inputs. In this context, the term "data" encompasses a multitude of instances, with supervised learning algorithms typically employing labelled datasets, while unsupervised learning relies on unlabelled data for discovering hidden patterns(Myttenaere*et al.*, 2016). The integration of these machine learning techniques within Agriculture 4.0 enables more sophisticated predictive modelling and decision-making processes that drive efficiency and innovation across the sector [45,46].

Machine learning (ML) achieves significant precision in task execution, largely due to the availability of extensive datasets. However, in the agricultural domain, acquiring vast and diverse data can often be a challenging endeavour, though it is critical for the evolution of robust machine learning models. IoT sensors are fundamental in the collection of a wide array of agricultural data. Strategically deployed across fields, these sensors capture vital information on various aspects such as crop health, livestock performance, soil properties, and climatic conditions (Meshram*et al.*,2021). The proliferation of IoT technology facilitates the continuous and real-time acquisition of data, allowing for the systematic accumulation of large datasets over time.Despite the convenience of IoT for data collection, ensuring the quality and representativeness of the data is paramount (Hosseini *et al.*, 2019). For instance, in crop management, it is essential to monitor the different stages of crop growth meticulously to generate realistic and actionable models for real-world agricultural applications. Building such comprehensive and representative datasets is a time-intensive process, but it is a necessary investment to ensure the reliability and effectiveness of ML models in agriculture (Benos*et al.*, 2020).Collaborative efforts are also instrumental in accelerating this data collection process. Partnerships between farmers, research organizations, and agricultural institutions can foster the sharing of valuable data resources, thereby enriching the pool of information available for machine learning applications. These cooperative endeavours can bridge data gaps and enhance the development of more accurate and versatile models, ultimately pushing the boundaries of what can be achieved with machine learning in agriculture (Chen *et al.*, 2019).

1.2 Objectives

This review encompasses a wide range of topics related to machine learning in agriculture. It begins by outlining the fundamental definitions and concepts of ML, offering readers a clear understanding of the technology's core principles. The scope extends to a classification of the different types of ML, including supervised, unsupervised, and reinforcement learning, and their relevance to agricultural systems. This review also delves into the practical applications of ML in agriculture, ranging from crop and livestock monitoring to resource management and automation. In addition, it explores the most commonly used ML algorithms in agriculture, such as decision trees, neural networks, and support vector machines, providing a critical evaluation of their strengths and limitations. Finally, the review addresses the various challenges faced in the adoption and implementation of ML in agriculture, including issues related to data collection, model accuracy, scalability, and the integration of technology in traditional farming systems. This review aims to serve as a foundational resource for researchers, practitioners, and stakeholders seeking to understand and leverage ML technologies in the agricultural domain.

2. Definition and basic concepts of machine learning

Machine learning (ML) is a subfield of artificial intelligence (AI) and computer science that uses data and algorithms to simulate human learning in order to gradually improve AI's accuracy. Similar to human learning, this field focuses on giving computers the ability to think, act, and eventually improve their performance on their own (Veeragandham and Santhi, 2020). It is accomplished by giving machines information and understanding via human interactions and real-world observations, enabling them to learn on their own.

2.1 How it works?

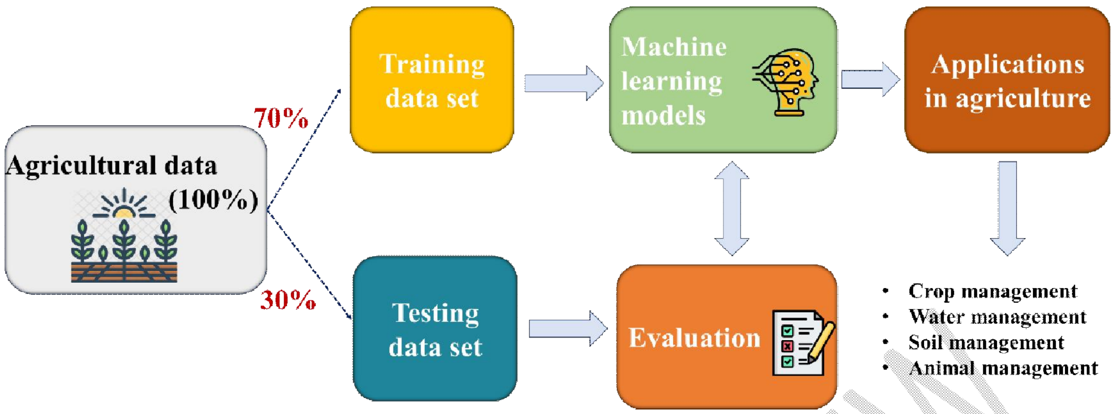

Fig: 1 General flow for the creation of Machine Learning models and their application in agriculture

Figure 1 illustrates the general process of developing machine learning (ML) models tailored for agricultural applications. The entire workflow begins with the collection of agricultural data from a variety of sources, such as satellite imagery, sensors, climate records, soil samples, and even data from farming machinery (Cai *et al.*, 2019). This diverse set of data is the foundation upon which the machine learning algorithms are built.

Step 1: Data Collection

The first crucial step involves gathering large quantities of data from agricultural fields, which can include sensor data on soil moisture, crop growth stages, weather conditions, water usage, and animal health. These data inputs form the building blocks that feed into the machine learning process. The richness and variety of the data are essential for producing accurate predictions and recommendations, as they allow the model to consider the complexities of agricultural environments.

Step 2: Data Preparation (Splitting into Training and Testing Sets)

Once the data is collected, it is divided into two key subsets: training and testing datasets. Typically, 70% of the total data is allocated to the training set, and 30% is reserved for testing purposes. This division is a standard practice in machine learning to ensure the model can learn from one set of data while being evaluated on another independent set to avoid overfitting.

Training Data: The training dataset, which constitutes 70% of the total data, is used to build and train the machine learning model. The algorithm analyses this dataset to detect patterns, relationships, and trends that will enable it to make decisions in agricultural contexts. During this phase, the model is fine-tuned through repeated iterations until it learns the complex dynamics of the agricultural data.

Testing Data: The remaining 30% of the data is set aside as the testing dataset, which acts as a mechanism to evaluate the performance of the model. By testing the model on unseen data, researchers can assess its ability to generalize its learning and make accurate predictions or classifications. This evaluation process ensures that the model is reliable and ready for real-world agricultural applications.

Step 3: Model Validation and Performance Evaluation

After the model is trained, it is validated using the testing dataset to gauge its effectiveness. Key performance metrics, such as accuracy, precision, recall, and error rates, are used to determine whether the model is performing as expected. This testing phase is critical because it reveals how well the model will perform when faced with new, unseen data in practical agricultural scenarios. Adjustments can be made during this phase to further refine the model, ensuring that it is robust and capable of delivering reliable outcomes.

Step 4: Model Deployment and Application in Agriculture

Once the machine learning model is trained, tested, and validated, it is ready to be deployed in various agricultural sectors.By doing so, machine learning brings precision and efficiency to agriculture, enabling farmers to make data-driven decisions that improve yields, reduce resource waste, and enhance sustainability. The process ensures that the model is not only accurate but also dependable across various agricultural domains such as crop, water, soil, and livestock management, thereby transforming modern farming practices (Sharma *et al.*, 2020).

2.2 What distinguishes it from conventional programming?

In traditional programming, a developer manually creates a set of rules or a program that dictates how input data is processed to produce the desired output. This approach follows a clear sequence of steps defined by the programmer: input data is fed into a predefined, meticulously tested program, and the machine executes these instructions to generate output.

The logic for solving a problem is explicitly coded based on the developer's understanding of the task, meaning that the computer follows these fixed rules exactly as written.

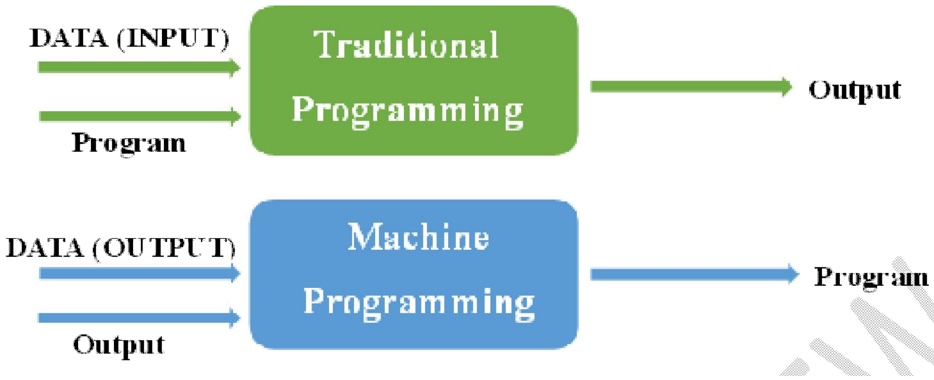

Fig 2: Traditional programming vs Machine programming

Machine learning (ML), on the other hand, flips this paradigm. Instead of explicitly programming rules and logic, in ML, both the input data and the corresponding output are provided to the machine during the training or learning phase. The goal is for the machine to identify patterns and relationships within the data and learn how to map inputs to outputs on its own. This process essentially allows the machine to "learn" the program or the underlying logic, rather than having it explicitly written by a human. The machine adjusts its internal parameters (e.g., weights in a neural network) to model the input-output relationship based on the data it has been trained on.

Once the learning phase is complete, the machine can make predictions or generate output for new, unseen inputs by applying the learned model. The advantage of ML is that it can handle complex problems where the rules are not easily identifiable or too intricate for manual coding, such as image recognition, natural language processing, or autonomous decision-making in dynamic environments like agriculture.

3. Types of Machine Learning

3.1. Supervised machine learning: Supervised machine learning, often referred to as supervised learning, is distinguished by the employment of annotated datasets to instruct algorithms in generating precise predictions or categorizing data. The model iteratively adjusts its internal parameters when exposed to input data, refining its predictive accuracy until an optimal alignment is reached. This calibration occurs during the cross-validation phase, which ensures that the model maintains a balance, neither excessively conforming to the training data (overfitting) nor inadequately capturing its patterns (underfitting). A prevalent application of supervised learning is the automated segregation of spam emails into designated folders, exemplifying its capacity to address complex, real-world challenges at scale. Techniques such as neural networks, naive bayes classifiers, linear regression, logistic regression, random forests, and support vector machines (SVM) are among the sophisticated methodologies utilized in this paradigm (Ouf, 2018).

3.2. Unsupervised machine learning: The practice of utilizing machine learning algorithms to analyse and classify unlabelled datasets (sometimes referred to as clusters) is known as unsupervised learning. Without the need for human intervention, these algorithms uncover hidden links or patterns in the data. Because it can find patterns and similarities in data, this method is ideal for consumer segmentation, cross-selling strategies, exploratory data analysis, and pattern and picture recognition. It can also be used to reduce the dimensionality of a model, which reduces its feature count. Principal component analysis (PCA) and singular value decomposition (SVD) are two widely used techniques for this. Probabilistic clustering approaches are used in unsupervised learning in addition to neural networks and k-means clustering (Attri*et al.,* 2024).

3.3. Reinforcement machine learning:Reinforcement learning, while analogous to supervised learning in its pursuit of optimized outcomes, diverges significantly in its approach by not relying on pre-labelled sample data for training. Instead, the algorithm learns dynamically through a process of trial and error, continuously interacting with its environment to discover the best course of action. This method enables the model to iteratively improve its performance by receiving feedback in the form of rewards or penalties based on the effectiveness of its decisions. Over time, the accumulation of successful outcomes strengthens the model's understanding, reinforcing the actions that yield favourable results. The ultimate goal of reinforcement learning is to devise an optimal strategy or policy for navigating a particular problem space, whereby each decision is informed by the model's evolving grasp of the environment. Such a framework proves particularly valuable in complex, real-world scenarios where explicit instruction is not feasible, allowing the algorithm to autonomously refine its behaviour (Storm *et al.,* 2020).

4. Applications of machine learning in agriculture

The area of machine learning is expanding and has a wide range of possible uses in agriculture. Using machine learning research to forecast pests and diseases, minimise water usage, and increase crop yields is a topic of investigation for farmers and agricultural experts. In the future, machine learning could assist farmers in producing food sustainably and making better use of their resources (Filippi*et al.*, 2019).In general, there are four broad applications of machine learning in agriculture as listed below:

4.1 Crop management: In order to regulate the biological, chemical, and physical crop environment and meet both quantitative and qualitative goals, a variety of farming techniques were combined to create the category known as crop management (Koul, 2021). Utilising cutting-edge crop management techniques, such as yield prediction, disease diagnosis, weed identification, crop recognition, and crop quality, helps to raise production and, in turn, revenue. Key objectives of precision agriculture include the aforementioned elements.

a) Yield Prediction: Yield prediction is often one of the most significant as well as challenging subjects in contemporary agriculture. For example, farm owners can make well-informed management decisions on what crops to cultivate to fit the crop to the demands of the current market with the use of an accurate model (Abrougui*et al.*, 2019). Numerous elements, including the crop's genotype and phenotypic traits, management techniques, and environment, might influence yield forecast. Therefore, it requires a basic understanding of how these interaction elements relate to yield. Therefore, finding these kinds of relationships requires extensive datasets in addition to potent technologies like machine learning approaches.

b) Disease Detection: A major threat to agricultural production systems is the development of crop diseases, which reduce output quantity and quality throughout production, storage, and transportation. Reports of yield losses on farms as a result of plant diseases are not uncommon. Moreover, crop diseases represent serious worldwide threats to food security (Ashapure*et al.*, 2020). An essential component of effective management is the prompt detection of plant diseases. Numerous types of bacteria, fungus, pests, viruses, and other agents can cause plant diseases. Disease symptoms, or the outward signs of pathogen presence and phenotypic changes in plants, might include leaf and fruit spots, wilting and colour changes, leaf curling, and more. In the past, field surveying was used to identify diseases by skilled agronomists. But this is a laborious procedure that relies only on visual examination. Sensing devices that are commercially accessible can now identify unhealthy plants before symptoms appear due to recent technology advancements. Moreover, maps showing the zones on the farm where the infection has spread can be made to show the spatial distribution of the plant disease (Chouhan *et al.*, 2018).

c) Weed Detection: Weeds typically develop and spread invasively across vast portions of the field very quickly due to their prolific seed production and extended lifespan. As a result, they compete with crops for resources like space, sunlight, nutrients, and water availability. Furthermore, without having to contend with natural enemies, weeds often emerge earlier than crops, a circumstance that negatively impacts crop growth (Dhingra *et al.,* 2019). Weed control is a crucial management obligation that can be achieved through mechanical treatment or herbicide application to prevent crop yield degradation. Herbicide spraying and mechanical treatment take a lot of time and money (Witten *et al.,* 1993). On the basis of smart agriculture, significant progress has been made in recent years in distinguishing weeds from crops. Instead of spraying the entire field and designing the shortest weeding path, ML algorithms combined with imaging technology or non-imaging spectroscopy can allow for real-time distinction and localization of target weeds, enabling accurate administration of herbicides to specified zones (Ozguven*et al.,* 2019).

d) Crop Recognition: Crop recognition software has drawn a lot of interest from scientists in a variety of disciplines, including plant taxonomy, botany, and the discovery of new species. Plant species can be identified and categorised by examining their leaves, stems, fruits, flowers, roots, and seeds, among other organs. The most popular method appears to be leaf-based plant recognition, which looks at the colour, shape, and texture of individual leaves. Remote sensing of crop attributes has made it easier to classify crops, and this has made it more common to utilise satellites and aerial vehicles for this purpose. Similar to the aforementioned subcategories, the automatic identification and classification of crops is a result of advances in computer software and image processing equipment when paired with machine learning.

e) Crop quality: The market is greatly impacted by crop quality, which is generally influenced by a number of factors including crop features, cultivation techniques, soil and climate conditions, and more. Better-quality agricultural products usually fetch higher prices on the market, which increases farmers' profits. For example, the most common maturity indices used for harvesting are skin colour, soluble solids concentration, flesh hardness, and fruit quality. The qualitative attributes of the harvested products in both arable and high-value crops (tree crops, grapes, vegetables, herbs, etc.) are significantly impacted by the date of harvesting. Thus, creating decision support systems can help farmers make the right management choices for higher-quality output (Geetharamani ad Arun, 2019). As an illustration, one management technique that may significantly improve quality is selective harvesting. Crop quality is also directly related to food waste, which is another issue facing modern agriculture because a crop that doesn't meet specifications for form, colour, or size may be thrown away. Like the previous subsection, using ML algorithms in conjunction with imaging technology can yield promising outcomes.

4.2 Water management: Since plant development is heavily dependent on the availability of water, the agricultural industry is the primary global consumer of fresh water. More efficient water management is required to better save water in order to achieve sustainable crop production, given the high rate of depletion of many aquifers with minimal recharge. Along with lowering pollution and health hazards, efficient water management can also result in better water quality.Variable rate irrigation presents a desirable option for achieving water reductions, according to a recent study on precision agriculture. This can be achieved by applying irrigation at rates that, rather than utilising a constant rate throughout the field, vary based on field variability and the unique water requirements of distinct management zones. In order to achieve both water savings and yield optimisation, agronomic parameters, such as topography, soil characteristics, and their effect on soil water, determine the viability and efficacy of the variable rate irrigation technique.Programming irrigation and effective water management can be aided by closely monitoring the soil water status, crop growth conditions, and temporal and spatial patterns in conjunction with weather monitoring and forecasting. Remote sensing is one of the ICTs that is used to produce images with geographical and temporal variability related to crop growth metrics and soil moisture status for precise water management (Crane, 2018). It's interesting to note that managing water in dry regions where irrigation relies on groundwater supplies is difficult enough, with precipitation meeting only a portion of the crop's evapotranspiration (ET) needs (Habib *et al.*, 2020).

4.3 Soil management: As a diverse natural resource, soil involves a great deal of intricate systems and processes. Accurate data on soil at the regional level is essential for improved soil management that aligns with land potential and supports sustainable agriculture as a whole. Issues like land degradation (loss of biological productivity), soil-nutrient imbalance (from excessive fertiliser use), and soil erosion (from overcutting vegetation, unbalanced crop rotations, livestock overgrazing, and unsustainable fallow periods) make better soil management even more important. Texture, organic matter, and nutrient content are a few examples of useful soil characteristics. Conventional techniques for evaluating soil consist of soil sampling and laboratory analysis, both of which are typically costly and require a significant amount of time and work (Ramesh *et al.,* 2020). However, inexpensive and simple solutions for the investigation of soil spatial variability can be found with remote sensing and soil mapping sensors. When using typical data analysis methodologies, there may be significant limitations related to data fusion and processing of such heterogeneous "big data." ML approaches can be a reliable and affordable way to accomplish this kind of work.

4.4 Livestock management: It is commonly acknowledged that techniques for producing livestock have become more intensive when considering the productivity of each animal. This intensification encompasses social concerns centred on human and animal health that might affect consumers' perceptions of food security, safety, and sustainability. To improve production processes, it is very important to assess animal welfare and total productivity. The aforementioned domains operate within the context of precision livestock farming, with the goal of utilising engineering methods to identify warning signs and monitor animal health in real time, as well as enhancing early-stage productivity. Non-invasive sensors including cameras, gyroscopes, accelerometers, radio-frequency identification systems, pedometers, and optical and temperature sensors are essential to precision livestock production. Variable physical quantities, or VPQs, are used by IoT sensors to sense various physical parameters, such as humidity, sound, and temperature. IoT sensors, for example, can provide important details about specific animals in real-time if a VPQ deviates from normal bounds. Modern animal's husbandry now includes ML approaches as a necessary component to leverage the vast volumes of data. It is possible to create models that can describe how a biological system functions by utilising causal links and taking advantage of this biological knowledge to produce predictions and recommendations (Abdulridha*et al.,* 2020).

a) Animal welfare: Animal welfare is a persistent problem because animal health is closely linked to product quality, which in turn is closely linked to consumer health and, secondly, to increased economic efficiency. Physiological stress and behavioural markers are two of the many indices available for assessing animal welfare. Animal behaviour, which can be influenced by illnesses, feelings, and living circumstances and may reveal physiological problems, is the most widely utilised indication (Cravero and Sepulveda, 2021). Commonly used sensors to identify behavioural shifts in animals (such as altered food or water intake or decreased activity), they include cameras, accelerometers, microphone systems, and more.

b) Livestock production: Combined with cutting-edge machine learning algorithms, sensor technology can boost cattle productivity. Because animal management practices have an impact on productive factors, livestock owners are becoming more cautious with their assets. It becomes increasingly challenging to properly care for each and every animal in a larger livestock holding, though. From this angle, the previously indicated assistance to farmers through precision livestock farming is a positive development for factors related to financial viability and the creation of environmentally friendly jobs. In animal production, a variety of models have often been employed, with the main goals being to raise and feed animals as efficiently as possible. However, again, ML techniques are required because to the massive volumes of data involved.

5. Machine learning algorithms used in agriculture

- Linear regression: One of the most widely used simple machine learning methods for predictive analysis is linear regression. Predictive analysis is used here to describe predictions about anything, while linear regression forecasts continuous variables like age and pay. It illustrates how the dependent variable (y) varies in response to the independent variable (x) and the linear relationship between the dependent and independent variables. The regression line is the best fit line those results from trying to find the best possible relationship between the dependent and independent variables(Akhter and Sofi, 2022).
- Logistic regression: The supervised learning approach used to predict discrete values or categorical variables is called logistic regression. The logistic regression algorithm can be applied to machine learning classification issues, and its output can take the form of Yes or NO, 0 or 1, Red or Blue, etc. With a few exceptions, logistic regression and linear regression are comparable in that logistic regression is used to solve classification problems and predict discrete values, whereas linear regression is used to solve regression problems and predict continuous values (Badage, 2018). It creates an S-shaped curve between 0 and 1 in place of fitting the best fit line. Another name for the S-shaped curve is a logistic function that makes use of the threshold idea. Every value will gravitate towards 1 if it is above the threshold and towards 0 if it is below it.
- Decision tree: A supervised learning approach called a decision tree is mostly used to tackle classification problems, while it can also be used to handle regression problems. It is applicable to both continuous and categorical variables. It displays a node-and branch-

filled structure like a tree, with the root node at the base expanding to the leaf node via additional branches. The dataset's features are represented by the internal node, decision rules by the branches, and problem outcomes by the leaf nodes(Uddin and Bansal, 2022).

- Support vector machine: An approach for supervised learning that can be applied to regression and classification issues is the support vector machine, or SVM. It is mostly applied to categorization difficulties, though. Creating a decision boundary or hyperplane that may divide datasets into distinct classes is the aim of support vector machines (SVM) (Swain *et al.,* 2020). The approach is called the support vector machine algorithm because the data points that aid in defining the hyperplane are referred to as support vectors. Face identification, picture categorization, drug discovery, and other real-world uses for SVM.
- Naviebayes: A supervised learning system called the Naive Bayes classifier is used to forecast things based on how likely they are to occur. Because it is predicated on the Bayes theorem and operates on the naive assumption that variables are independent of one another, the algorithm is known as Naive Bayes. One of the best classifiers for a given problem that yields good results is the Naive Bayes classifier. A naïve Bayesian model is simple to construct and works well with large datasets. Text classification is its primary use.
- K Nearest Neighbour (KNN): K-Nearest Neighbour can be applied to regression and classification issues. The way this algorithm operates is by presuming that the new and existing data points are identical. The new data points are grouped into the most similar groups based on these commonalities. Because it keeps all of the available datasets and uses K-neighbors to classify each new example, it is also known as the lazy learner algorithm. Any distance function calculates the separation between the data points, and the new case is allocated to the closest class with the greatest similarities (Vasilyevich, 2018). Depending on the needs, the distance function can be manhattan, euclidean, minkowski, or hamming distance.
- K- means clustering: One of the most basic unsupervised learning strategies for clustering problems is K-means clustering. Based on similarities and differences, the datasets are divided into K distinct clusters; that is, the datasets with the greatest degree of commonality stay in a single cluster while the datasets with the least amount of commonality or none at all remain in separate clusters. K-means denotes the number of clusters, and means denotes the dataset's average used to determine the centroid. Every cluster in the method is linked to a centroid, which is based on centroid theory. The goal of this technique is to shorten the distances between data points and cluster centroids. This

approach begins with a set of randomly chosen centroids that initially form the clusters. It then iteratively refines the placements of these centroids. It can be applied to the identification of bogus news, spam detection and filtering, and other tasks (Habib *et al.,* 2020).

- Random forest: Random forest is useful for machine learning tasks involving both regression and classification. Through the combination of several classifiers, this ensemble learning technique improves the model's performance and generates predictions. In order to increase the model's forecast accuracy, it includes several decision trees for different dataset subsets and calculates their averages. 64–128 trees are the ideal number for a random forest. An increased number of trees increases the algorithm's accuracy. When classifying a new dataset or object, the algorithm predicts the eventual result based on the majority vote, which is given by each tree. The quick random forest algorithm effectively handles inaccurate and missing data.
- Apriori algorithm: The unsupervised learning algorithm known as the apriori algorithm is used to resolve association difficulties. It is intended to operate on databases that contain transactions and generates association rules using frequent itemset. It establishes the strength or weakness of the relationship between two objects with the aid of these association rules. This approach computes the itemset in an efficient manner by using a Hash Tree and a breadth-first search. The method searches the huge dataset iteratively for the frequently occurring itemset. In 1994, R. Agrawal and Srikant presented the apriori algorithm. It aids in understanding the products that can be purchased together and is primarily used for market basket analysis. It is also applicable to the medical field.
- Principle Component Analysis: One method for reducing dimensionality in unsupervised learning is Principle Component Analysis (PCA). It facilitates the reduction of the dataset's dimensionality, which is made up of numerous features that are interrelated. It is a statistical procedure that uses orthogonal transformation to turn a set of correlated feature observations into a set of linearly uncorrelated features (Ang and Seng, 2021). It is a widely used tool for both predictive modelling and exploratory data analysis. In order to minimise dimensionality, PCA takes into account each attribute's variance. A large variance indicates a successful class split. Image processing, movie recommendation systems, and power allocation optimisation across several communication channels are a few examples of real-world PCA uses.

6.List -1 Challengesof machine learning in agriculture

Challenges	Explanation	Solution	References
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.	Charan and Anand, 2020
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.	Dhokane and Kulkarni, 2020
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.	Habib *et al.*, 2020
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.	Karunamurthy*et al.*, 2022
Data consistency	Consistent data ensures that ML models are reliable and reproducible. For example, inconsistent labelling 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.	Liakos*et al.*, 2018
Data context	ML models need to be trained on data that are	Investigating techniques for adapting ML models based on	Olaimat*et al.*, 2020

	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.	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.	Sahoo *et al.*, 2020
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.	Benos*et al.*, 2020
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.	Zecca*et al.*, 2019
Interpretability and explainability	ML-based systems pose a significant challenge in gaining the trust and acceptance of farmers,	Ensuring that ML models are transparent and that their inner workings are accessible. This means providing	Sorensen*et al.*, 2019

	stakeholders, and the agricultural industry. It is important to understand how models achieve their outputs.	information on the features, variables, and algorithms that contribute to a model's results. Techniques such as SHAP values or LIME 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.	Wang *et al.*, 2019
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.	Salina *et al.*, 2020

7. Conclusion

A significant improvement in agricultural production efficiency credited to machine learning (ML)-based precision farming methods, primarily in terms of developments in resource allocation plans that guarantee the prudent use of inputs like water, fertiliser, and pesticides. As a result, the application of ML technologies has improved the environmental impact of agricultural activities. Farmers can adopt sustainable methods that minimise resource usage and limit environmental impacts by using data-driven decision making. For instance, using real-time soil moisture data, ML-powered precision irrigation systems can adaptively control water use, encouraging water conservation and preserving ideal soil conditions. Furthermore, integrating ML has significantly improved the agricultural sector's capacity to control weeds, pests, and illnesses. Plant disease early detection and categorization using machine learning algorithms has proven to be very accurate, enabling prompt intervention and mitigation strategies. This minimises possible yield losses while simultaneously protecting crop health. All things considered, the use of ML in agriculture marks a paradigm change that is advancing the industry towards a future that is more technologically advanced, efficient, and sustainable. The advantages extend beyond increased output; they include a comprehensive adjustment of farming methods to conform to modern standards for food safety and environmental conservation. This is encouraging for the agricultural sector's adaptation and resilience to changing global problems. Identifying and examining some of the difficulties and offering solutions for mitigation are crucial. These include guaranteeing flexible machine learning models, maximising data accessibility, and preserving data consistency, accuracy, and completeness. It's also critical to contextualise data consumption, handle privacy and security issues, and guarantee timely data. Furthermore, it is imperative to focus on fostering human-machine collaboration, improving interpretability, and addressing the issue of low digital literacy among traditional farmers. Designing machine learning applications with intuitive user interfaces that need little technical knowledge is essential. Simple dashboards and easily understood visualisations can improve accessibility for less literate farmers. Furthermore, outreach and training programmes that are specifically designed to meet the requirements of agricultural communities can be put into place. Farmers can be equipped with the information and abilities needed to successfully integrate ML-based technology into their daily operations through workshops, demonstrations, and instructional initiatives. Furthermore, creating lightweight models and investigating distributed computing techniques are essential first steps towards successfully integrating machine learning into agriculture in contexts with limited resources. By addressing these issues, ML technologies will be used more widely and more effectively in the agriculture industry. To summarise, the application of machine learning (ML) in agriculture yields significant advantages, such as enhanced yield and efficient use of resources, enhanced identification of illnesses and pests, and diminished ecological consequences. These developments open the door to an agricultural industry that is more flexible and sustainable and prepared to fulfil the demands of the future. With the continued success of data-driven techniques, the agricultural landscape is poised for a more dynamic and sophisticated future where tradition and technology coexist peacefully for the benefit of global agriculture.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

8. References

1. Abdulridha, J., Ampatzidis, Y., Qureshi, J., and Roberts, P. (2020). Laboratory and UAV-based identification and classification of tomato yellow leaf curl, bacterial spot, and target spot diseases in tomato utilizing hyperspectral imaging and machine learning. *Remote sensing of environment*,12, 2732.
2. Abrougui, K., Gabsi, K., Mercatoris, B., Khemis, C., Amami, R., and Chehaibi, S. (2019). Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). *Soil and Tillage Research*,190, 202–208.
3. Akhter, R., and Sofi, S.A. (2022). Precision agriculture using IoT data analytics and machine learning. *Journal of King Saud University-Computer and Information Sciences*, 34(8), 5602-5618.
4. Anagnostis, A., Asiminari, G., Papageorgiou, E., and Bochtis, D. (2020). A Convolutional Neural Networks Based Method for Anthracnose Infected Walnut Tree Leaves Identification. *Applied Sciences*,10, 469.
5. Ang, K. L. M., and Seng, J. K. P. (2021). Big data and machine learning with hyperspectral information in agriculture. *Institute of Electrical and Electronics Engineers*, 9, 36699-36718.
6. Ashapure, A., Jung, J., Chang, A., Oh, S., Yeom, J., Maeda, M., Maeda, A., Dube, N., Landivar, J., and Hague, S. (2020). Developing a machine learning based cotton yield estimation framework using multi-temporal UAS data.*Remote sensing of environment*,169, 180–194.

7. Attri, I., Awasthi, L. K., and Sharma, T. P. (2024). Machine learning in agriculture: a review of crop management applications. *Multimedia Tools and Applications,* 83(5), 12875-12915.
8. Badage, A. (2018). Crop disease detection using machine learning: Indian agriculture.*International Research Journal of Engineering and Technology,*5(9), 866-869.
9. Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., and Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors,* 21(11), 3758.
10. Benos, L., Bechar, A., andBochtis, D. (2020). Safety and ergonomics in human-robot interactive agricultural operations. *Biosystems engineering,*200, 55–72.
11. Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., and You, L. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agricultural and Forest Meteorology,*274, 144–159.
12. Charan, M. P., and Anand, G. (2020). Application of artificial intelligence in agriculture: A review. *International Journal of Computer Sciences and Engineering,* 8(2), 222-227.
13. Chen, Y.Y., Lin, Y.H., Kung, C.C., Chung, M.H., and Yen, I.H. (2019).Design and Implementation of Cloud Analytics-Assisted Smart Power Meters Considering Advanced Artificial Intelligence as Edge Analytics in Demand-Side Management for Smart Homes. *Sensors,* 19, 2047.
14. Chouhan, S.S., Kaul, A., Singh, U.P. and Jain, S. (2018). Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards plant pathology. *Institute of Electrical and Electronics Engineers,*6, 8852–8863
15. Crane, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters,* 13(11), 114003.
16. Cravero, A., and Sepulveda, S. (2021). Use and adaptations of machine learning in big data applications in real cases in agriculture. *Electronics,* 10(5), 552.
17. Dhingra, G., Kumar, V., and Joshi, H.D. (2019). A novel computer vision based neutrosophic approach for leaf disease identification and classification. *Measurement: Journal of the International Measurement Confederation*,135, 782–794.

18. Dhokane, S., and Kulkarni, A. (2020). Smart farming: Role of AI and IoT in agriculture. *International Journal of Engineering Research & Technology,* 9(1), 32-36.
19. Filippi, P., Jones, E.J., Wimalathunge, N.S., Somarathna, P.D.S.N., Pozza, L.E., Ugbaje, S.U., Jephcott, T.G., Paterson, S.E., Whelan, B.M., and Bishop, T.F.A. (2019). An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precision Agriculture,* 20, 1015–1029.
20. Geetharamani, G., and Arun Pandian. (2019). Identification of plant leaf diseases using a nine-layer deep convolutional neural network.*Journal of Computational Engineering,*76, 323–338
21. Habib, M.T., Majumder, A., Jakaria, A.Z.M., Akter, M., Uddin, M.S., and Ahmed, F. (2020). Machine vision based papaya disease recognition. *Journal of Computing and Information Science in Engineering,* 32, 300–309.
22. Hosseini, S., Ivanov, D., and Dolgui, A. (2019). Review of quantitative methods for supply chain resilience analysis. *Transportation Research Part E: Logistics and Transportation Review,*125, 285–307.
23. Karunamurthy, A., Kulunthan, K., Dhivya, P., and Vickson, A. V. S.(2022). A Knowledge Discovery Based System Predicting Modelling for Heart Disease with Machine Learning.*International Journal of Innovative Research in Science and Engineering,* 1(1), 14-22. https://doi.org/10.54368/qijirse.1.1.0005
24. Koul, S. (2021).Smart Agriculture: Emerging Pedagogies of Deep Learning, Machine Learning and Internet of Things. *Applied Sciences,* 3(1), 1-19.
25. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., and Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors,* 18(8), 2674.
26. Meshram, V., Patil, K., Meshram, V., Hanchate, D., andRamkteke, S. D. (2021). Machine learning in agriculture domain: A state-of-art survey. *Artificial Intelligence in the Life Sciences,* 1, 100010.
27. Mostafa, S.S.Mendonca, F., Ravelo-Garcia, A.G., and Morgado-Dias, F. (2019). A Systematic Review of Detecting Sleep Apnea Using Deep Learning. *Sensors,* 19, 4934.
28. Myttenaere, A., Golden, B., Le Grand, B., and Rossi, F. (2016). Mean Absolute Percentage Error for regression models. *Neurocomputing,* 192, 38–48.

29. Olaimat, A. N., Holley, R. A., and Al-Holy, M. A. (2020). Artificial intelligence in the food industry: A comprehensive review. *Comprehensive Reviews in Food Science and Food Safety*, 19(6), 2635-2666
30. Ouf, N. S. (2018). A review on the relevant applications of machine learning in agriculture. *Sensors*, 6(8).
31. Ozguven, M.M., and Adem, K. (2019). Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithms. *Physica A: Statistical Mechanics and its Applications*,535, 122537.
32. Ramesh, S., and Vydeki, D. (2020). Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm. *Information Processing in Agriculture*, 7, 249–260.
33. Sahoo, J., Kandpal, L. M., and Singh, A. (2020). Applications of artificial intelligence in agriculture: A comprehensive review. *Journal of Ambient Intelligence and Humanized Computing*, 11(7), 2961-2989.
34. Salina, A.B., Hassan, L., Saharee, A.A., Jajere, S.M., Stevenson, M.A., and Ghazali, K. (2020). Assessment of knowledge, attitude, and practice on livestock traceability among cattle farmers and cattle traders in peninsular Malaysia and its impact on disease control. *Tropical Animal Health and Production*, 53, 15.
35. Sharma, A., Jain, A., Gupta, P., and Chowdary, V. (2020). Machine learning applications for precision agriculture: A comprehensive review. *Institute of Electrical and Electronics Engineers*, 9, 4843-4873.
36. Sorensen, C.A.G., Kateris, D., and Bochtis, D. (2019). ICT Innovations and Smart Farming. *Institute of Electrical and Electronics Engineers*,953,1–19.
37. Storm, H., Baylis, K., andHeckelei, T. (2020). Machine learning in agricultural and applied economics. *European Review of Agricultural Economics*, 47(3), 849-892.
38. Swain, M., Singh, R., Thakur, A. K., andGehlot, A. (2020). A machine learning approach of data mining in agriculture 4.0. *International Journal of Emerging Technologies in Learning*,101(1).
39. Uddin, M. S., and Bansal, J. C. (2022). Computer Vision and Machine Learning in Agriculture. *Institute of Electrical and Electronics Engineers*,2, 234-242.
40. Vasilyevich, K. A. (2018). Machine learning methods in digital agriculture: Algorithms and cases. *International Journal of Advanced Studies*, 8(1), 11-26.
41. Veeragandham, S., andSanthi, H. (2020). A review on the role of machine learning in agriculture. *Scalable Computing: Practice and Experience*, 21(4), 583-589.

42. Wang, A., Zhang, W., and Wei, X. (2019). A review on weed detection using ground-based machine vision and image processing techniques. *Computers and Electronics in Agriculture*, 158, 226–240.
43. Witten, I. H., Holmes, G., McQueen, R. J., Smith, L. A., and Cunningham, S. J. (1993). Practical machine learning and its application to problems in agriculture.*International Journal of Advanced Studies*, 4, 126–140.
44. Zecca, F. (2019). The Use of Internet of Things for the Sustainability of the Agricultural Sector: The Case of Climate Smart Agriculture. *International journal of civil engineering*, 10, 494–501.
45. Swarnalatha, P., V. Srinivasa Rao, G. Raghunadha Reddy, Santosha Rathod, D. Ramesh, and K. Uma Devi. 2024. “Application of Machine Learning Techniques Models for Forecasting of Redgram Prices of Andhra Pradesh, India”. Journal of Scientific Research and Reports 30 (7):252-71. https://doi.org/10.9734/jsrr/2024/v30i72142.
46. Prajapati, Harshad A., D. M. Kadam, Shivkanya S. Aitwar, PrathameshDilip Jagtap, Debesh Singh, NirjharneeNandeha, and Deepanshu Mukherjee. 2023. “Application of Robotics, Artificial Intelligence and Deep Learning in Modern Agriculture Technology: A Review”. International Journal of Plant & Soil Science 35 (23):106-16. https://doi.org/10.9734/ijpss/2023/v35i234222.