Machine Learning Applications in Agriculture

Subjects: Computer Science, Artificial Intelligence

Contributor: Sara Oleiro Araújo , Ricardo Silva Peres , José Cochicho Ramalho , Fernando Lidon , José Barata

Progress in agricultural productivity and sustainability hinges on strategic investments in technological research. Evolving technologies such as the Internet of Things, sensors, robotics, Artificial Intelligence, Machine Learning, Big Data, and Cloud Computing are propelling the agricultural sector towards the transformative Agriculture 4.0 paradigm.

Agriculture 4.0 machine learning Agriculture

1. Introduction

Agriculture 4.0 [1][2][3][4][5], also known as "Digital Agricultural Revolution" [6], represents a paradigm shift in agriculture, leveraging cutting-edge technologies to optimise various aspects of farming operations. These technologies encompass the Internet of Things (IoT), Artificial Intelligence (AI), Big Data, cloud computing, Decision Support System (DSS), advanced sensing technology, and autonomous robots [1][6][7]. Sensors and robotics play a crucial role in collecting essential field data, which is then transmitted to a local or cloud server via IoT technology for storage, processing, and analysis. Big data and AI-based techniques can be used to convert these data into valuable insights. To facilitate user interaction and informed decision making, a DSS equips users with the necessary tools to optimise the agricultural system and undertake appropriate actions.

Machine Learning (ML), a subset of AI, has shown great potential in enhancing various aspects of Agriculture 4.0. It can be defined as a computer program or system that can learn specific tasks without being explicitly programmed to do so [8][9][10]. It is a process that involves the use of a computer to make decisions based on multiple data inputs [8]. In this case, data mean a set of examples. Labeled data is often used for supervised learning tasks (where the model learns from labeled examples), and unlabeled data might be used for unsupervised learning tasks (where the model finds patterns and structures in the data) [9].

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

2. Machine Learning in Agriculture

2.1. Crop Management Domain

Crop management is associated to several agricultural practices that profoundly influence the growth and yield of cultivated crops. These practices encompass a wide range of activities, starting with the meticulous sowing process, extending to the vigilant maintenance of crops throughout their growth and development phases, and concluding with the phases of harvest [1]. The optimisation of crop management strategies is essential to increase agricultural productivity, thereby addressing the escalating global requisites for sustenance, textile fibers, energy sources, and fundamental raw materials [11]. The application of ML techniques in crop management has significantly revolutionised conventional farming practices, offering capabilities such as crop mapping and recognition, yield prediction, optimal irrigation scheduling, pest and weed management, and disease detection [1].

Crop Quality

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

Crop Mapping and Recognition

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

Crop Yield

Crop yield refers to the quantity of agricultural produce obtained from a specific area of land during a growing season. Ensuring high crop yields is of utmost importance for addressing global food challenges and meeting the demands of a growing population [11]. There has been a growing application of ML methods to estimate crop yield, aiming to facilitate farming planning, resource allocation (such as water, fertilisers, and pesticides), enhance storage management and marketing strategies, and tackle the pressing challenges of food security in the forthcoming years [1].

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

Crop Disease

Crop disease refers to the study and management of various diseases that affect agricultural crops, leading to reduced yields and economic losses for farmers and the agricultural industry as a whole. Several techniques are applied to discern disease patterns, anticipate outbreaks, and implement targeted interventions, thereby offering a promising avenue for detection, diagnosis, and control of crop diseases [1]. Through the fusion of ML models with diverse data sources, such as IoT-generated data and satellite and UAV imagery, these studies showcase the capacity to accurately categorise and identify diseases across various crops, enabling timely and effective responses to mitigate their impact.

Pest and Weed Detection

Instances of crop pest infestations, ranging from weeds, insects, pathogens, and rodents [12], have emerged as factors affecting global agricultural production. This sub-domain focuses on the utilisation of advanced technologies, such as sensors, imaging systems, and ML algorithms, to detect and mitigate the presence of unwanted organisms that can negatively impact crop growth and yield. It is possible to understand that ML techniques can help analyse complex data from various sources (such as satellites, UAV, or sensors) and identify patterns and anomalies associated with pest and weed presence that may not be easily recognisable to the human eye. ML-powered systems can detect pests and weeds at their early stages, enabling swift intervention before infestations become widespread [1].

2.2. Water Management Domain

As water resources become increasingly finite and their management more complex, the fusion of cutting-edge technology with robust data analytics holds great promise in promoting more sustainable water management practices. IoT technology, sensors and actuators networks, data analytics, and predictive models have enabled farmers to monitor water quality, soil moisture levels, weather forecasts, and Crop Evapotranspiration (ETc) rates [13].

Table 1 exemplifies the utilisation of an array of ML algorithms, coupled with remote and proximal sensing techniques as well as innovative IoT technologies, to address diverse water-related challenges encompassing irrigation management, water quality surveillance, and ETc prediction.

Table 1. Machine learning applications in water management domain.

Ref.	Crop Field	Models Used	Summary
[14]	Maize	Linear regression, RF, Cubist, PLS, PCA, GBT	Uses remote sensing data and regression algorithms for predicting ETa and soil water content to enable remote irrigation management. The study employs VIs for training and phenology observations. Cubist showed slightly better performance for predicting ETa and RF for soil water content.
[15]	Cranberry	RF, XGBoost	Forecasts water table depth using DT-based modeling approaches for optimised irrigation management. XGBoost demonstrated superior predictive ability, accurately simulating water table depth fluctuations for longer periods than RF. Despite limitations with extrapolation and extreme events, the models hold potential with broader dataset ranges for practical applications.
[16]	Not applicable	KNN	Portable smart sensing system based on IoT for detecting nitrate, phosphate, pH, and temperature in water. KNN algorithm is used to enhance the accuracy of the system's analysis. The proposed system offers early hazard detection and promotes regular contaminant level evaluation.
[17]	Not specified	PCA, SVM, GBT	Focuses on accurately predicting crop ETo for efficient water resource management and irrigation. The research employs PCA techniques to identify key factors influencing ETo that are then used as inputs for prediction models. PSO was used to optimise SVM and GBT models. The PSO-GBT model exhibits the highest accuracy.
[18]	Maize	DT, RF, SVM, ANN, PLS	Uses UAV multispectral data and ML for estimating water content indicators, including equivalent water thickness, fuel moisture content, and specific leaf area of maize crops in smallholder farms. RF and SVM outperform others in predicting water content indicators. This approach offers accurate insights into drought-related water stress on smallholder farms.
[19]	Banana plants	KNN, GBT, LSTM	Employs IoT components to gather data (soil moisture, temperature, and weather conditions) and ML to optimise irrigation requirements and reduce energy consumption. The hybrid model predicts real-time and time-series water needs based on various observations. The work is demonstrated using banana cultivation, achieving up to a 31.4% water optimisation for a single banana tree.
[20]	Grains, vegetables, fruits, flowers	RF, NN, SVM	Predicts phosphorus concentrations in shallow groundwater in intensive agricultural regions. SVM achieved the highest accuracy (R2 = 0.60). These findings support groundwater phosphorus

2.3. Soil Management Domain

Ref.	Crop Field	Models Used	Summary
		[1]	monitoring, early warning, and pollution management decision making in intensive agricultural regions.

cultivation ... real-time parameter measurements, AI-driven data analysis techniques, and DSS for informed decision making equips farmers with the tools to effectively oversee their fields in a manner that is both efficient and sustainable [1][21]. ML-based techniques can process vast amounts of soil-related data (such as soil composition, texture, and moisture measurements) and generate insights into optimal irrigation schedules, nutrient management strategies, and soil health assessments.

s is clear from **Table 2**, ML techniques possess the ability to predict soil properties and behaviours, empowering farmers to make well-informed choices pertaining to soil fertility, structure, moisture levels, and nutrient concentrations, all aimed at enhancing crop growth and yield. Additionally, by leveraging computer vision and the remote sensing data, ML simplifies the monitoring of both crops and soil conditions. This technological synergy allows for a comprehensive assessment of crop health, growth stages, and potential stressors. Beyond remote sensing, one particularly notable application of ML involves the utilisation of cell phone images, as demonstrated in the study by [22]. This innovative approach showcases the potential of ML to develop efficient proximal soil sensors capable of swiftly and accurately predicting crucial soil properties. By harnessing readily available technology, this advancement exemplifies the adaptability and practicality of ML solutions in modern soil management practices. This not only exemplifies the adaptability and practicality of ML solutions in modern soil management practices, but it also underscores the transformative impact that technology-driven approaches can have on agricultural sustainability.

Table 2. Machine learning applications in soil management domain.

Ref.	Crop Field	Models Used	Summary
[23]	Various soil samples	RF, SVM, Logistic Regression	Predicts disease occurrence with high accuracy by analysing soil macroecological patterns of Fusarium wilt, a destructive soil-borne plant disease. The research employs a ML approach using bacterial and fungal data sets from diseased and healthy soils across various countries and plant varieties. The results reveal distinct differences in bacterial and fungal communities between healthy and diseased soils.
[24]	Canola	RF	The research utilises a ML approach to determine key predictors of soil nitrous oxide (N2O) emissions, including soil temperature, moisture, and nitrate availability. The results highlight that N2O emissions were influenced by these factors, with emission factors being lower in high yield zones compared to low yield zones.
[22]	Maize, soybean	DT, RF, Cubist, Gaussian Process, SVM, ANN	Estimates soil organic matter (SOM) and soil moisture content (SMC) based on 22 color and texture features extracted from cell phone images. The study demonstrates the potential of using computer vision and ML to create an efficient proximal soil sensor for quick and accurate predictions of soil properties. Gaussian Process and Cubist models performed the best for SMC prediction, while ANN and Cubist showed satisfactory accuracy for SOM prediction.

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

and dairy products, but it also supplies other high-quality goods, such as wool and leather. Global demand for animal products is expected to increase further due to population growth [11], meaning that agrifood industries must optimise production practices by ensuring the welfare and safety of animals and increasing the capacity to prevent, detect, diagnose, and treat animal diseases. Considering this, there is a growing awareness that animal management can no longer be performed via traditional means and requires the adoption of new digital technologies.

Smart animal monitoring systems have been viewed with great interest in the academic community, agrifood industries, and markets. Sensor-based animal wearables, computer vision systems, and other detection devices can capture the status of animals and environment in real time, which can be analysed afterwards with the aid of AI-based mechanisms to control and predict animals' health, welfare, production, etc. Livestock monitoring includes information related to animals' behaviour, physiology, clinical status, and performance [29], while in aquaculture, the desired information is more focused on water quality (water temperature, pH, dissolved oxygen content, ammonia, salt, etc.) [30][31].

2.5. Summary

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

2.5.1. Crop Management

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

2.5.2. Water Management

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

2.5.3. Soil Management

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

2.5.4. Animal Management

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

References

1. Araújo, S.O.; Peres, R.S.; Barata, J.; Lidon, F.; Ramalho, J.C. Characterising the Agriculture 4.0 Landscape—Emerging Trends, Challenges and Opportunities. Agronomy 2021, 11, 667.
2. De Clercq, M.; Vats, A.; Biel, A. Agriculture 4.0: The future of farming technology. In Proceedings of the the World Government Summit, Dubai, United Arab Emirates, 11–13 February 2018; pp. 11–13.
3. Zambon, I.; Cecchini, M.; Egidi, G.; Saporito, M.G.; Colantoni, A. Revolution 4.0: Industry vs. agriculture in a future development for SMEs. Processes 2019, 7, 36.
4. Liu, Y.; Ma, X.; Shu, L.; Hancke, G.P.; Abu-Mahfouz, A.M. From Industry 4.0 to Agriculture 4.0: Current Status, Enabling Technologies, and Research Challenges. IEEE Trans. Ind. Inform. 2020, 17, 4322–4334.
5. Zhai, Z.; Martínez, J.F.; Beltran, V.; Martínez, N.L. Decision support systems for Agriculture 4.0: Survey and challenges. Comput. Electron. Agric. 2020, 170, 105256.
6. Trendov, N.M.; Varas, S.; Zeng, M. Digital Technologies in Agriculture and Rural Areas; Briefing paper; FAO: Rome, Italy, 2019.
7. Rose, D.C.; Chilvers, J. Agriculture 4.0: Broadening responsible innovation in an era of smart farming. Front. Sustain. Food Syst. 2018, 2, 87.
8. Ahmed, M.; Pathan, A.S.K. Data Analytics: Concepts, Techniques, and Applications; CRC Press: Boca Raton, FL, USA, 2018.
9. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. Sensors 2018, 18, 2674.
10. Mahesh, B. Machine learning algorithms-a review. Int. J. Sci. Res. (IJSR) 2020, 9, 381–386.
11. FAO. The Future of Food and Agriculture—Trends and Challenges; Food and Agriculture Organization of the United Nations: Rome, Italy, 2017.
12. Abbas, T.; Zahir, Z.A.; Naveed, M.; Kremer, R.J. Limitations of existing weed control practices necessitate development of alternative techniques based on biological approaches. Adv. Agron. 2018, 147, 239–280.
13. Pereira, L.S.; Perrier, A.; Allen, R.G.; Alves, I. Evapotranspiration: Concepts and future trends. J. Irrig. Drain. Eng. 1999, 125, 45–51.
14. Filgueiras, R.; Almeida, T.S.; Mantovani, E.C.; Dias, S.H.B.; Fernandes-Filho, E.I.; da Cunha, F.F.; Venancio, L.P. Soil water content and actual evapotranspiration predictions using regression algorithms and remote sensing data. Agric. Water Manag. 2020, 241, 106346.

15. Brédy, J.; Gallichand, J.; Celicourt, P.; Gumiere, S.J. Water table depth forecasting in cranberry fields using two decision-tree-modeling approaches. Agric. Water Manag. 2020, 233, 106090.

16. Akhter, F.; Siddiquei, H.R.; Alahi, M.E.E.; Jayasundera, K.P.; Mukhopadhyay, S.C. An IoT-enabled portable water quality monitoring system with MWCNT/PDMS multifunctional sensor for agricultural applications. IEEE Internet Things J. 2021, 9, 14307–14316.

17. Zhao, L.; Zhao, X.; Zhou, H.; Wang, X.; Xing, X. Prediction model for daily reference crop evapotranspiration based on hybrid algorithm and principal components analysis in Southwest China. Comput. Electron. Agric. 2021, 190, 106424.

18. Ndlovu, H.S.; Odindi, J.; Sibanda, M.; Mutanga, O.; Clulow, A.; Chimonyo, V.G.; Mabhaudhi, T. A comparative estimation of maize leaf water content using machine learning techniques and unmanned aerial vehicle (UAV)-based proximal and remotely sensed data. Remote Sens. 2021, 13, 4091.

19. Vianny, D.M.M.; John, A.; Mohan, S.K.; Sarlan, A.; Ahmadian, A. Water optimization technique for precision irrigation system using IoT and machine learning. Sustain. Energy Technol. Assessments 2022, 52, 102307.

20. Yang, H.; Wang, P.; Chen, A.; Ye, Y.; Chen, Q.; Cui, R.; Zhang, D. Prediction of phosphorus concentrations in shallow groundwater in intensive agricultural regions based on machine learning. Chemosphere 2023, 313, 137623.

21. Na, A.; Isaac, W.; Varshney, S.; Khan, E. An IoT based system for remote monitoring of soil characteristics. In Proceedings of the 2016 International Conference on Information Technology (InCITe)—The Next Generation IT Summit on the Theme—Internet of Things: Connect your Worlds, Noida, India, 6–7 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 316–320.

22. Taneja, P.; Vasava, H.K.; Daggupati, P.; Biswas, A. Multi-algorithm comparison to predict soil organic matter and soil moisture content from cell phone images. Geoderma 2021, 385, 114863.

23. Yuan, J.; Wen, T.; Zhang, H.; Zhao, M.; Penton, C.R.; Thomashow, L.S.; Shen, Q. Predicting disease occurrence with high accuracy based on soil macroecological patterns of Fusarium wilt. ISME J. 2020, 14, 2936–2950.

24. Glenn, A.J.; Moulin, A.P.; Roy, A.K.; Wilson, H.F. Soil nitrous oxide emissions from no-till canola production under variable rate nitrogen fertilizer management. Geoderma 2021, 385, 114857.

25. Fournier, B.; Steiner, M.; Brochet, X.; Degrune, F.; Mammeri, J.; Carvalho, D.L.; Siliceo, S.L.; Bacher, S.; Peña-Reyes, C.A.; Heger, T.J. Toward the use of protists as bioindicators of multiple stresses in agricultural soils: A case study in vineyard ecosystems. Ecol. Indic. 2022, 139, 108955.

26. Li, P.; Hao, H.; Mao, X.; Xu, J.; Lv, Y.; Chen, W.; Ge, D.; Zhang, Z. Convolutional neural network-based applied research on the enrichment of heavy metals in the soil–rice system in China.

Environ. Sci. Pollut. Res. 2022, 29, 53642–53655.

27. Zhao, J.; Zhang, C.; Min, L.; Guo, Z.; Li, N. Retrieval of farmland surface soil moisture based on feature optimization and machine learning. Remote Sens. 2022, 14, 5102.
28. Wan, H.; Qi, H.; Shang, S. Estimating soil water and salt contents from field measurements with time domain reflectometry using machine learning algorithms. Agric. Water Manag. 2023, 285, 108364.
29. Nasirahmadi, A.; Edwards, S.A.; Sturm, B. Implementation of machine vision for detecting behaviour of cattle and pigs. Livest. Sci. 2017, 202, 25–38.
30. Raju, K.R.S.R.; Varma, G.H.K. Knowledge based real time monitoring system for aquaculture using IoT. In Proceedings of the 2017 IEEE 7th International Advance Computing Conference (IACC), Hyderabad, India, 5–7 January 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 318–321.
31. Shi, X.; An, X.; Zhao, Q.; Liu, H.; Xia, L.; Sun, X.; Guo, Y. State-of-the-art internet of things in protected agriculture. Sensors 2019, 19, 1833.