How Machine Learning Has Changed The Agriculture Sector? – Review

DANDU SAI SATYA VAMSI KRISHNA RAJU[1], EDIGA VINOD KUMAR[2], EDIGA MAHENDRA GOUD[3], SAPNA R[4]

[1, 2, 3] *Department of Computer Science and Engineering, Presidency University, Bengaluru, India*

[4] *Assistant Professor, Department of Computer Science and Engineering, Presidency University, Bangalore, India*

Abstract- With the development of big data technologies and high-performance computers, machine learning has opened up new possibilities for data-intensive science in the multidisciplinary field of agri-technology. This study provides a thorough analysis of applications-focused research of artificial intelligence in systems that produce food. The analysed works were divided into the following categories: (a) crop management, which includes applications on yield prediction, disease detection, and weed identification Crop quality and species identification; b) animal welfare and livestock production applications; c) water management; and d) soil management. The categorization and filtering of the articles shown here show how machine learning technologies will help agriculture. Farm management systems are developing into real-time artificial intelligence enabled programmes that offer detailed recommendations and insights for farmer decision support and action by using machine learning to sensor data.

I. INTRODUCTION

The global economy depends heavily on the agricultural sector. The system for farming will be under pressure. Increase in tandem with the ongoing growth of the population of humans. Precision farming and agri-technology have emerged as new data-driven scientific disciplines, and the latter is now referred to as digital agriculture measures that are aggressive in order to increase agricultural productivity while reducing its environmental impact. Modern agricultural operations generate data from a variety of sensors, allowing for a better understanding of both the activity itself (machinery data) and the operational atmosphere (an collaboration of dynamic crop, soil, and weather conditions). This enables more accurate and efficient decision-making.

Along with big data technology and high-performance computing, algorithms for learning (ML) has arisen to open up new possibilities for unravelling, quantifying, and comprehending data-intensive processes in agriculture operational environments. Among other things, machine learning (ML) is described as the branch of science that allows machines to gain knowledge without being explicitly programmed [1]. Every year, machine learning (ML) is used in an increasing number of scientific disciplines, such as bioinformatics [2,3], biochemistry, medicine, meteorology, economic sciences, robotic, aquaculture, food safety, and climatology.

We give a thorough analysis of machine learning's use in agriculture in this study. There are several pertinent publications that highlight important and distinctive characteristics of well-known ML models given. The vocabulary, definition, learning tasks, and analysis for the present work are presented in Section 2 along with the most often used learning models and algorithms. The established approach for gathering and categorization of the works offered is presented in Section 3. The benefits of applying ML to agri-technology are finally listed in Section 4, along with predictions for the future of the field. Tables 1-4 list the shorthand terms that appear in this work, categorised to machine learning (ML) algorithms, statistical measurements, and generic abbreviations, respectively. This is due to the huge amount of abbreviations used in the related scientific publications.

II. OVERVIEW OF MACHINE LEARNING

2.1 Terminology and Definitions of Machine Learning:

ML approaches often involve a method of learning with the goal of learning from "experience" (training data) in order to carry out a task. In ML, the input is a collection of examples. Typically, a set of properties, often referred to as features or variables, are used to characterise a specific example. A characteristic can be numerical (integer, real number, etc.), ordinal (such as A+ or B), binary (i.e., 0 or 1), nominal (enumeration), or any combination of these. A performance indicator that gets better with practise over time is used to assess how well the ML model performs in a particular activity. Several statistical and mathematical methods are used to determine how well ML models and algorithms function. After the learning process is complete, the model that was trained can be used to categorise, forecast, or cluster fresh instances (testing data) based on the knowledge gained. Figure 1 depicts a typical ML strategy.

According to the learning type (supervised/unsupervised), the models of learning (classification, regression, clustering, and dimensionality reduction), or the learning models used to carry out the chosen task, ML tasks are often divided into various broad groups.

2.2 Task of Learning

According to the learning signal provided by the learning system, both supervised and unsupervised educational activities are divided into two primary types. In supervised learning, examples of inputs and outputs are supplied together with data, and the goal is to develop a general rule which maps inputs to outputs. Some inputs may only be partially available, and some of the desired outcomes may not be present or may only be provided as response to the actions taken in an evolving setting (reinforcement learning). In the supervised environment, the trained model and gained knowledge are utilised to forecast the test data's missing outputs (labels). Although the data in unsupervised learning is unlabeled, there is no separation between the training and test sets. In order to find hidden patterns, the learner analyses the incoming data.

2.3 Analysis of Learning:

In order to preserve as much information from the original data as possible, dimensionality reduction (DR) is a technique that is used in both kinds of supervised and unsupervised learning types. Its goal is to provide a more compact, lower-dimensional representation of a dataset. In order to mitigate the effects of dimensionality, it is typically carried out before using a classification or regression model. The following represent a few of the most popular DR algorithms: Principal component analysis (PCA) [22], the partial least-squares regression (PLS) [23], and linear discriminant analysis (LDA) [24] are three examples.

III. REVIEW

On the most basic level, the examined publications have been divided into four general categories: agricultural management, animal administration, management of water, and soil management. The subcategories of ML applications in the crop area include predicting yields, disease detection, weed identification, quality of crops, and species recognition. The welfare of animals and livestock productivity were the two sub-categories into which the uses of machine learning in the livestock area were classified. Scopus, ScienceDirect, and PubMed were the search engines used. The chosen articles only discuss efforts that have been published as journal papers. The provided review does not cover climate prediction, despite the fact that it is crucial for agricultural productivity and that machine learning (ML) methods for climate prediction constitute a separate field in and of itself. Finally, all of the articles discussed here refer to the years 2004 until the present.

3.1. Crop Management

3.1.1. Yield Prediction

One of the most important aspects of precision agriculture is yield prediction, which is crucial for mapping and estimating yields, matching crop supply and demand, and managing crops to maximise production. A few instances of ML applications can be found in [74]'s publications; an effective, a low-cost, non-destructive approach that counted the coffee beans on a branch automatically. The approach divides the coffee fruits into three categories: those that can be harvested, those that cannot, and ones with ignored maturity stages. The approach also calculated the weight and maturity rate of the coffee fruits. The purpose of this effort was to advise coffee growers on how to maximise financial gains and organise their agricultural operations. The authors of [75], who created a system using machine vision towards automating shaking and capturing cherries during harvest, conducted another study that was used for yield prediction. Even inconspicuous cherry branches with full foliage, the algorithm segments and detects them. The system's primary goal was to cut down on the amount of labour needed for harvesting and handling tasks. A technique for early yield mapping was created in another investigation [76] to identify premature greenish citrus in a citrus orchard outdoors. The study's goal, as with many such research, was to give producers information on specific yields to help them optimise their grove for improved yield and profit.

In a different work, the authors [77] used artificial neural networks with multi temporal data from remote sensing to create an algorithm for the assessment of grasslands biomass (kg dry matter/ha/day). Another investigation into yield forecasting, and more particularly the forecasting of wheat yields, was published [78].

For a more precise prediction, the developed system used soil data, crop growth parameters, and satellite photos. In [79], the authors described a technique for the detection of tomatoes using remotely sensed RGB pictures and electromagnetic (EM) data from a unmanned aerial vehicle (UAV). Additionally, in the study of [80], authors created a system for the prediction of rice development stage based on SVM and fundamental geographic data gathered from meteorological stations in China. Finally, another study [81] proposed a generalised approach for predicting agricultural productivity. The approach depends upon an ENN application using agronomical data produced over an extended time period (1997–2014). The study's focus on regional forecasts (particularly for Taiwan) was on helping farmers prevent supply and demand imbalances in the market that could be sped up by poor harvest crop quality.

3.1.2. Disease Detection

The subcategories with the most publications included in this review are identifying diseases and yield prediction. The control of pests and diseases in outdoors (arable agriculture) and greenhouse settings is one of the biggest issues in agriculture. Spraying pesticides evenly across the crop field is the method of controlling pests and diseases that is most commonly utilised. Despite being useful, this practise has a tremendous cost to the environment and the economy. The repercussions on the environment can include residues in agricultural goods, consequences for groundwater contamination, affects on local fauna as well as eco-systems, and so forth. In precision agriculture management, where agrochemical application is selected in terms of time and place, ML is an integral component. A technique for the identification and differentiation between normal Silybum marianum, or marianum, plants and those affected by the smuts fungal Microbotyumsilybum in vegetative growth is given in the literature [82]. For the classification of parasites and the automatic identification of insect in strawberry greenhouse environments, authors of the study [83] created an innovative approach based on image processing procedures. This method allows for real-time control.
A method for diagnosing and screening Bakanae illness in rice seedlings was provided in the authos of [84].

The study's main objective was to precisely identify the pathogen Fusariumfujikuroi in two rice cultivars. When compared to a visual inspection, the automatic detection of sick plants boosted grain yield and took less time.

3.1.3 Weed Detection:

Another key issue in agriculture is the identification and management of weeds. Weeds are frequently cited by farmers as the biggest threat to crop productivity. Because weeds are challenging to identify and distinguish from crops, accurate weed identification is crucial for sustainable agriculture. Once more, ML algorithms in combination with sensors can result in accurate weed detection and discrimination at minimal cost and without any negative environmental effects. Herbicide use can be reduced by using robots and tools to eliminate weeds, thanks to machine learning for weed detection. There have been two research on the use of ML for agricultural weed detection problems. In the initial study [91], authors introduced a novel approach for determining the presence of Silybum marianum, a difficult-to-remove weed that significantly reduces agricultural output, based on counter propagating (CP)-ANN and multispectral photographs acquired by unmanned aerial systems (UAS). In the following research [92], the authors created a novel method for agricultural and weed species identification based on ML methods and hyperspectral imaging. More specifically, the authors developed an active learning system for the identification of Medicagolupulina, Ciriumarvense, Poaannua, Polygonumpersicaria, Urticadioica, and Ranunculusrepens as weed species. The proper identification and classification of these organisms for economic and environmental objectives was the main objective. In a different study, the authors [93] created an SVN-based weed detection approach for grassland cropping.

3.2 Livestock Management

The welfare of animals and livestock production are the two subcategories that make up the livestock category. Animal health and welfare are related, and machine learning is mostly used to track animal behaviour in order to identify diseases early. On the opposite hand, livestock production addresses problems with the production system, and the key application area for ML in this field is the precise estimation of farmers' economic balances using production line monitoring.

3.2.1 Animal Welfare

According to reports, a number of articles fall within the subcategory of animal welfare. A strategy for categorising cow behaviour based on ML models is provided in the first paper [98] using information gathered by collar sensors equipped with magnetic field metres and three-axis accelerometers. The study's objective was to identify dietary changes in cattle and forecast events like the oestrus. A technique for automatically recognising and categorising of chewing patterns in calves was given in the second article [99]. The scientists developed a machine learning (ML) system using information from chewed signals of nutritional supplements like ryegrass and hay along with information on behaviour like ruminating and inactivity. Optical FBG sensors were used to gather data. An automated surveillance system based on machine learning (ML) was described in another work [100] for tracking animal behaviour, including tracking the movements of animals by depth camera footage for monitoring various behaviours of an animal (standing, moving forward feeding, and drinking).

3.3 Soil Management

The last section of this paper discusses the use of ML for predicting and identifying agricultural soil characteristics, such as the assessment of soil temperature, moisture content, and dryness. The natural resource soil is heterogeneous and has difficult-to-understand complex processes and systems. Researchers can use soil parameters to comprehend how agriculture affects ecosystem dynamics. An precise assessment of the state of the soil can result in better soil management. For an accurate examination of a region's eco-environmental conditions and the effects of climate change, soil temperature alone is crucial. It is a crucial meteorological variable that regulates the interactions between the atmosphere and the earth. In addition, crop yield variability is significantly influenced by soil moisture. Yet, soil measurements are typically time-consuming and expensive, therefore using computer analysis based on ML approaches can result in a low cost and dependable solution for the precise assessment of soil. The research of [110] serves as the initial investigation for this final subcategory. This work provided a way for assessing the drying of soil for agricultural planning, to be more precise. Using data on evapo transpiration and precipitation, the approach provides an accurate assessment of soil drying in a region in Urbana, Illinois, in the United States.

The implementation of remote agriculture management decisions was the aim of this technique. The second study [111] was created to provide predictions about the state of the soil. Regarding the forecasting of organic matter in the soil (OC), moisture content (MC), and total nitrogen in the soil (TN), the study specifically offered an analysis of four regression models. More specifically, the scientists collected soil spectra from 140 wet, unprocessed samples of the top layer of Luvisol soil types using a visible-near infrared (VIS-NIR) spectrophotometer. The samples were taken in August 2013 from an agricultural area in Premslin, Germany, following the completion of wheat harvests. They came to the conclusion that effective soil property prediction could improve soil management. In an additional investigation [112], the authors created a new technique for estimating every day temperature of the soil at a total of six various depths of 5, 10, 20, 30, and 100 cm in two distinct climate regions of Iran, Bandar Abbas and Kerman. This technique is based on an autonomous adaptive evolutionary-extreme developing device and everyday weather data. The objective was to estimate soil temperature accurately for agricultural management. In the most recent study [113], a novel approach for estimating soil moisture was described. It was based on ANN models and used information collected by force sensors on a no-till chisel opener.

IV. DISCUSSION AND CONCLUSIONS

There were a total of forty papers included in this review. The remaining articles were published to the journals Sensors, Sustainability, Real-Time Imagining, Precision Agriculture, Earth Observations and Remote Sensing, Saudi Journal of Biological Sciences, Scientific Reports, and Computers in Industry. Of the presented articles, twenty-five (25) were published in the journal "Computer and Electronics in Agriculture," six (6) were published in the journal of "Biosystems Engineering," and the remaining articles were published to the journals of "Earth Observations and Remote Sensing," " Eight of the articles discuss the use of ML in managing livestock, four discuss the use of ML in managing water resources, four discuss the use of ML in managing soil resources, and the majority of the articles—24—discuss the use of ML in managing crops.

REFERENCES

[1] Samuel, A.L. Some Studies in Machine Learning Using the Game of Checkers. IBM J. Res. Dev. 1959, 44, 206–226. [CrossRef]

[2] Kong, L.; Zhang, Y.; Ye, Z.Q.; Liu, X.Q.; Zhao, S.Q.; Wei, L.; Gao, G. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007, 35, 345–349. [CrossRef] [PubMed]

[3] Mackowiak, S.D.; Zauber, H.; Bielow, C.; Thiel, D.; Kutz, K.; Calviello, L.; Mastrobuoni, G.; Rajewsky, N.; Kempa, S.; Selbach, M.; et al. Extensive identification and analysis of conserved small ORFs in animals. Genome Biol. 2015, 16, 179. [CrossRef] [PubMed]

[4] Richardson, A.; Signor, B.M.; Lidbury, B.A.; Badrick, T. Clinical chemistry in higher dimensions: Machine-learning and enhanced prediction from routine clinical chemistry data. Clin. Biochem. 2016, 49, 1213–1220. [CrossRef] [PubMed]

[5] Wildenhain, J.; Spitzer, M.; Dolma, S.; Jarvik, N.; White, R.; Roy, M.; Griffiths, E.; Bellows, D.S.; Wright, G.D.; Tyers, M. Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning. Cell Syst. 2015, 1, 383–395. [CrossRef] [PubMed]

[6] Kang, J.; Schwartz, R.; Flickinger, J.; Beriwal, S. Machine learning approaches for predicting radiation therapy outcomes: A clinician's perspective. Int. J. Radiat. Oncol. Biol. Phys. 2015, 93, 1127–1135. [CrossRef] [PubMed]

[7] Asadi, H.; Dowling, R.; Yan, B.; Mitchell, P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. PLoS ONE 2014, 9, e88225. [CrossRef] [PubMed]

[8] Zhang, B.; He, X.; Ouyang, F.; Gu, D.; Dong, Y.; Zhang, L.; Mo, X.; Huang, W.; Tian, J.; Zhang, S. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. Cancer Lett. 2017, 403, 21–27. [CrossRef] [PubMed]

[9] Cramer, S.; Kampouridis, M.; Freitas, A.A.; Alexandridis, A.K. An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. Expert Syst. Appl. 2017, 85, 169–181. [CrossRef]

[10] Rhee, J.; Im, J. Meteorological drought forecasting for ungauged areas based on

machine learning: Using long-range climate forecast and remote sensing data. Agric. For. Meteorol. 2017, 237–238, 105–122. [CrossRef]