JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Review on Machine Learning Algorithm in the Field of Agriculture for Prediction

Mohammed Arif Khan, *Sunil K. Sahu, Abha Mahalwar and N. Kumar Swamy

ISBM University Nawapara (Kosmi), Block & Tehsil- Chhura, Gariyaband, Chhattisgarh- 493996, India.

Abstract:

The integration of machine learning (ML) techniques into the field of agriculture has led to significant advancements in predicting various agricultural outcomes. This review aims to provide an overview of recent developments and applications of machine learning for prediction in agriculture. The review begins by highlighting the importance of accurate predictions in agriculture, emphasizing how traditional methods often fall short due to the complexities of biological, environmental, and climatic factors. Machine learning offers a data-driven approach that can capture intricate relationships within these multifaceted systems, thus enabling more precise predictions. Various machine learning techniques employed in agriculture are explored, ranging from classical algorithms such as decision trees and support vector machines to more advanced approaches like random forests, gradient boosting, neural networks, and deep learning. The strengths and weaknesses of each technique in the context of agricultural prediction are discussed. The application areas covered in this review include crop yield prediction, disease and pest outbreak forecasting, soil health assessment, weather and climate impact analysis, and precision agriculture. For each application, representative studies are presented to illustrate the efficacy of machine learning in improving prediction accuracy and informing decision-making processes for farmers and stakeholders. Challenges associated with implementing machine learning in agriculture are also addressed. These challenges encompass data quality and availability, model interpretability, scalability, and the need for domain expertise. Potential solutions and ongoing research directions are discussed to mitigate these challenges and promote the widespread adoption of machine learning technologies in agriculture.

Keywords: Crop yield prediction, Machine learning algorithms, agriculture and Classification techniques.

1. Introduction

The agricultural sector is a vital part of the global economy, as it faces increasing pressure due to the growth of the population. The development of precision farming and agri-technology has revolutionized the way we think about farming. These new scientific fields are designed to help improve the efficiency and sustainability of the agricultural sector. Modern agriculture relies on sensors and software to collect and analyze data, which allows them to improve the efficiency and effectiveness of their operations. Through the use of machine learning, these tools can help them identify and analyze the various factors that affect the farm's operations. Big data and high-performance computing are being used to create new opportunities in the field of agricultural research. Machine learning is a scientific field that allows machines to learn on their own without requiring specific instructions. [1]. ML frameworks typically involve a learning process that aims to help users perform a task by learning from experience. Data in ML is typically described by a set of examples that are attributes, such as features or variables. These are typically described as values

ranging from nominal to binary. The performance of an ML model is measured by a metric that can be improved over time. To determine the efficiency of an algorithm or model, various mathematical and statistical models are used. Following the training phase, the trained model may be used to classify or predict new data. The classification of ML tasks into two main categories is shown in Figure 1. The first is supervised learning, which involves presenting data with associated outputs and examples. The goal is to develop a rule that maps the outputs to the inputs. In certain cases, the outputs may only be partially available with the help of a dynamic environment or with feedback only. In supervised learning, a trained model is used to identify the missing labels. Unsupervised learning doesn't differentiate between test sets and training. Instead, it focuses on uncovering hidden patterns in the input data. A dimensionality reduction analysis is a process that is commonly used in both unsupervised and supervised learning frameworks. It aims to provide a more compact representation of the collected information. This process is usually performed before a regression or classification model is applied to the collected data.

1.1 Motivation of research

Conventional methods of guess are often based on assumptions and do not provide a suitable response to issues. For instance, in assessing plant health, it is important to consider the various factors that affect its development. The use of machine learning techniques in agricultural science can improve the accuracy of this prediction. In agriculture, machine learning can be used for various tasks such as soil salinity, pH level, temperature update, crop yield prediction, and humidity update. It can perform automatic decision-making without requiring explicit programming in diverse scenarios.

2. Machine Learning Models

Regression

A supervised learning model that aims to predict an output variable based on the input variables is known as regression. Most commonly, it uses logistic regression and linear regression. More complex models are also being developed, such as multi-linear regression, scatterplot smoothing, and adaptive linear regression [2].

Clustering

In unsupervised learning models, clustering is a widely used technique to find clusters of data. There are various well-established techniques used in clustering, such as the k-means method, the hierarchical approach, and the expectation maximization method [3].

Bayesian Model

A type of graphical model known as a Bayesian model is commonly used in the analysis of complex problems such as classification and regression. It can be used for solving both the classification and regression problems. Some of the most popular algorithms used in this type of model include the Naive Bayes, the gaussian naive model, and the multinomial model [4].

Decision tree

A decision tree is a type of model that's formulated in a tree structure. It divides the data into smaller sub-populations and produces an associated tree graph. The internal nodes of the tree represent the various features of the classification model and their outcomes. The leaf nodes are the final steps in a decision or prediction that's been taken following a

path from one point to another. Some of the most popular learning algorithms used in this category include the regression and classification trees [5].

Artificial neural networks

There are two types of artificial neural networks: Deep ANNs and Traditional ANNs. These are modeled after the functions of the human brain, which include learning, decision making, and pattern generation. The billions of neurons in the brain are responsible for communicating and processing information. An ANN is a simplified representation of the structure of the neural network. It consists of interconnected units that are arranged in a specific topology. The system's input layer is where the data is collected and fed into it, while the learning process takes place. Finally, the output layer is where the prediction or decision is made. Artificial neural networks (ANNs) are commonly used for classification and regression problems. They are equipped with various learning algorithms, such as radial basis functions, perceptron systems, and resilient back-pagation. Numerous ANN-based learning systems have been presented and published. Some of these include autoencoder, Kohonen networks, adaptive-neuro fuzzy systems, and XY-Fusion. They are also equipped with various other learning algorithms such as multi-level perceptron systems, extreme learning machines, and self-organising maps. The term deep learning or deep neural networks is often used to refer to these types of systems. They are relatively new areas of research in the field of machine learning. With the help of several processing layers, a computational model can learn complex data sets. One of the most important advantages of DL is that it can perform the extraction step directly [6]. The applications of deep learning models in various industries, such as agriculture, have greatly improved. These are ANNs with multiple hidden layers that can be supervised or unsupervised. One of the most common models is the CNN, which is a convolutional neural network. This type of network is used to extract feature maps from images.A comprehensive overview of CNNs is provided in the literature. Other commonly used deep learning architectures are the deep Boltzmann machine, the deep belief network, and auto-encoders.

Support vector machine

The first support vector machine was introduced in the field of statistical learning. It is a binary classification system that is designed to classify data instances. It can be enhanced by using a kernel trick, which is a feature space transformation. SVMs are commonly used in the fields of classification, clustering, and regression. They are designed to solve overfitting problems in high-dimensional spaces, which are attractive in various applications. Most of the time, these algorithms are used for support vector regression, followed by the least squares and successive projection[7].

3. Agriculture data collection and prediction

Machine learning can help predict the broad-scale crop yield. This technology can be used in three different areas: remote sensing, crop modeling, and field surveys. The first method uses random sampling to forecast the crop yield in most countries. This method is costly and time-consuming, and it has a weak prediction accuracy. In addition, the survey field suffers from the lack of real data. Since the last decade, the field of agriculture has been greatly benefited by the use of remote sensing. This technology allows the gathering and storing of precise data related to various factors that affect the production of crops. It plays a vital role in identifying the plant's weaknesses and strengths during the growth period. One of the most common techniques utilized for forecasting the climate variability is process-based modeling. This method is commonly used to predict the crop yield. It involves analyzing the physiological growth of the plants and applying statistical modeling techniques to improve the accuracy of the prediction.

Yield Prediction

One of the most critical issues in precision agriculture is the prediction of yield. This is done through the use of various methods such as mapping, forecasting, and crop management. There are also various ML applications that can help improve the efficiency of this process. One example of an efficient method is the automated counting of coffee fruits[8]. This method is designed to calculate the various types of coffee fruits, such as harvestable, non-harvestable, and disregarded maturation. It also took into account the weight and percentage of the fruits that have already been matured. The objective of this project was to provide coffee farmers with the necessary information to plan their agricultural activities[8]. A study conducted on the use of machine vision technology for predicting the yield of cherries revealed that the system could automate the shaking and catching of cherries. The researchers noted that the system could help reduce the labor required in manual operations[9].In another study, the researchers developed a mapping system that could identify the early stages of green citrus trees in a citrus grove[10]. The goal of the study was to provide a comprehensive analysis of the various factors that affect the development of this tree. This method could help improve the efficiency of the citrus grove[11]. A study presented a method that can estimate the biomass of a grassland using remote sensing data and ANNs[12]. Another study focused on the prediction of wheat yield. The researchers used satellite imagery and soil data to improve the accuracy of their method[13]. The authors of a study presented a method that can detect the presence of tomatoes using remote sensing data and images taken by an unmanned aerial vehicle. In another study, the researchers used data collected by China's weather stations to predict the rice development stage.A new method for predicting agricultural yields was presented in a study[14]. The method is based on the data collected from long-term agricultural records. It was designed to provide a reliable prediction of the future crop production. The study focused on the Taiwan region.

4. Summary of related work

The paper compares the two methods for predicting the rice yield at the district level in West Bengal. The first is the weather indices-based model, while the other is the ML-ANN method. The inputs are the temperature, relative humidity, rainfall, minimum and maximum temperature, and time variable t. The outputs are the output and time variables of the model. According to the study, the ANN method is more accurate than the conventional regression approach when it comes to predicting the rice production. The error percentage in the prediction method is lower than 5% in the MLP ANN framework except in one district[15]. The objective of this project is to analyze the relationship between the climatic parameters and the paddy yield in different areas of Sri Lanka, such as Batticaloa, Hambantota, Ampara, Badulla, Kurunegala, and Puttalam. Through a combination of training algorithms, such as the Levenberg-Marquardt, Bayesian Regularization, and SCG, the researchers can then identify the best one. The performance indicators of the ANN models were evaluated using the Correlation coefficient and the Mean Squared Error. The researchers found that the LM training algorithm performed better than the other two in determining the paddy yield and climatic factors relationship [16]. The authors of this study utilized neural networks and machine learning to predict the yield of crops in Kazakhstan. The researchers trained various models using the collected data, including multi-layer regression, support vector machines, and polynomial regression models. They found that the multi-layered perceptron models performed the best, with an accuracy of 85% [17].

The authors of this study used data collected from various sources, such as climate and agricultural data, to predict rice yield in eastern China. They then trained various machine learning models using the collected information. The researchers found that the FFBN models performed the best when it came to predictive accuracy, with an R2 of 0.843 [18]. The goal of this study was to analyze the role that individual nutrients play in achieving a crop's growth and yield. The AI-based model was able to predict the maximum performance of various growth parameters, such as the plant height, leaf area index, and tiller number, at different doses.The study also analyzed the effects of individual nutrients on a plant's growth and yield. This information could help improve the management of nutrients and minimize environmental pollution [19] . A pair of statistical models were used to analyze the

predictability of rice yield prediction using long-term weather and crop data in 11 districts of Karnataka. They were validated using the AI neural network model in 2019 and 2020. The results of the analysis revealed that the ANN model was more accurate than the SMLR in predicting rice yield, with the observed deviations being smaller at around 4%. On the other hand, the district-wise forecasting model indicated that the yield prediction was underestimation by overestimations in some districts, such as Mysuru and Uttara Kannada. In 2019, the predicted rice yield in Shivamogga, Davanagere, Hassan, Ballari, Belagavi, and Kodagu was underestimated by around 0.6%, 0.1%, and 0.7%, respectively. In the other districts, the yield was overestimated by around 4.8%.[20]

Table 1 Description of the data set used generally in Machine leaning for crop yield prediction

Feature ID	Feature type	Data types	Features Category	Description
Area	Predictor	Integer	Continuous	Total land area used for Rice cultivation in hectare
Rainfall	Predictor	Float	Continuous	Average rainfall for the year in mm
Temperature	Predictor	Float	Continuous	Average temperature for the year in Celsius
Humidity	Predictor	Float	Continuous	Average humidity for the year in %
Wind Speed	Predictor	Float	Continuous	Average wind speed for the year in km/h
Production	Predictor	Float	Continuous	Total Rice production of the year in tones
Yields	Target	Float	Continuous	Total amount of crop grown per unit of land in tones/hectare

5. Conclusion:

Predictive machine learning in agriculture has emerged as a powerful tool for optimizing various aspects of farming, enabling more efficient resource allocation, improved crop yields, and better risk management. This review explores the current state of machine learning applications in agriculture, focusing on prediction tasks, and draws insightful conclusions about its impact.In conclusion, predictive machine learning holds immense promise in revolutionizing agriculture. By harnessing the power of data and advanced algorithms, it empowers farmers to optimize their practices, increase productivity, and contribute to sustainable and efficient agricultural systems. While challenges remain, the synergy between technology and agriculture has the potential to address global food security challenges and promote responsible land management.

References

1. Samuel, A.L. Some Studies in Machine Learning Using the Game of Checkers. IBM J. Res. Dev. 1959, 44, 206–226.
2. Cleveland, W.S. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 1979, 74, 829–836
3. Tryon, R.C. Communality of a variable: Formulation by cluster analysis. Psychometrika 1957, 22, 241–260.

4. Russell, S.J.; Norvig, P. Artificial Intelligence: A Modern Approach; Prentice Hall: Upper Saddle River, NJ, USA, 1995; Volume 9, ISBN 9780131038059.
5. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. Classification and Regression Trees; Routledge: Abingdon, UK, 1984; Volume 19, ISBN 0412048418.
6. Riedmiller, M.; Braun, H. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In Proceedings of the IEEE International Conference on Neural Networks, San Francisco, CA, USA, 28 March–1 April 1993; pp. 586–591.
7. Suykens, J.A.K.; Vandewalle, J. Least Squares Support Vector Machine Classifiers. Neural Process. Lett. 1999, 9, 293–300.
8. Ramos, P.J.; Prieto, F.A.; Montoya, E.C.; Oliveros, C.E. Automatic fruit count on coffee branches using computer vision. Comput. Electron. Agric. 2017, 137, 9–22.
9. Amatya, S.; Karkee, M.; Gongal, A.; Zhang, Q.; Whiting, M.D. Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting. Biosyst. Eng. 2015, 146, 3–15.
10. Sengupta, S.; Lee, W.S. Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions. Biosyst. Eng. 2014, 117, 51–61.
11. Ali, I.; Cawkwell, F.; Dwyer, E.; Green, S. Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data—A Machine Learning Approach. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2016, 10, 3254–3264.
12. Pantazi, X.-E.; Moshou, D.; Alexandridis, T.K.; Whetton, R.L.; Mouazen, A.M. Wheat yield prediction using machine learning and advanced sensing techniques. Comput. Electron. Agric. 2016, 121, 57–65.
13. Senthilnath, J.; Dokania, A.; Kandukuri, M.; Ramesh, K.N.; Anand, G.; Omkar, S.N. Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. Biosyst. Eng. 2016, 146, 16–32.
14. Su, Y.; Xu, H.; Yan, L. Support vector machine-based open crop model (SBOCM): Case of rice production in China. Saudi J. Biol. Sci. 2017, 24, 537–547.
15. Gupta, K. Sarkar, D. Dhakre, and D. Bhattacharya, "Weather based crop yield prediction using artificial neural networks: A comparative study with other approaches," *MAUSAM*, vol. 74, no. 3, pp. 825–832, Jul. 2023, doi: 10.54302/mausam.v74i3.174.
16. V. Amaratunga, L. Wickramasinghe, A. Perera, J. Jayasinghe, U. Rathnayake, and J. G. Zhou, "Artificial Neural Network to Estimate the Paddy Yield Prediction Using Climatic Data," *Math Probl Eng*, vol. 2020, 2020, doi: 10.1155/2020/8627824.
17. M. Sadenova, N. Beisekenov, P. S. Varbanov, and T. Pan, "Application of Machine Learning and Neural Networks to Predict the Yield of Cereals, Legumes, Oilseeds and Forage Crops in Kazakhstan," *Agriculture*, vol. 13, no. 6, p. 1195, Jun. 2023, doi: 10.3390/agriculture13061195.
18. Y. Guo, H. Xiang, Z. Li, F. Ma, and C. Du, "Prediction of rice yield in east China based on climate and agronomic traits data using artificial neural networks and partial least squares regression," *Agronomy*, vol. 11, no. 2, Feb. 2021, doi: 10.3390/agronomy11020282.
19. T. Shankar *et al.*, "Prediction of the Effect of Nutrients on Plant Parameters of Rice by Artificial Neural Network," *Agronomy*, vol. 12, no. 9, Sep. 2022, doi: 10.3390/agronomy12092123.
20. M. N. Thimmegowda *et al.*, "Weather-Based Statistical and Neural Network Tools for Forecasting Rice Yields in Major Growing Districts of Karnataka," *Agronomy*, vol. 13, no. 3, Mar. 2023, doi: 10.3390/agronomy13030704.