IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 8, August 2022

Features of Machine Learning in Agriculture

Divya Prakash Singh, Assistant Professor
College of Agriculture Sciences, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India
Email id- d.p.singhg@gmail.com

ABSTRACT: ***Machine learning has evolved with big data technologies and high-performance computers to provide new opportunities for data intensive research in the multi-disciplinary agri-technology industry. In this article, we offer a comprehensive review of research dedicated to applications of machine learning in agricultural production systems. The works analyzed were categorized in (a) crop management, including applications on yield prediction, disease detection, weed detection crop quality, and species recognition; (b) livestock management, including applications on animal welfare and livestock production; (c) water management; and (d) soil management. The filtering and classification of the supplied articles demonstrate how agriculture will benefit from machine learning technologies. By integrating machine learning to sensor data, farm management systems are evolving into real time artificial intelligence driven apps that provide rich recommendations and insights for farmer decision support and action.***

KEYWORDS: ***Agricultural, Data, Machine Learning, Models, Prediction, Soil Management, Water Management.***

1. INTRODUCTION

Agriculture plays a crucial role in the global economy. Pressure on the agriculture sector will increase with the continuing growth of the human population. Agri-technology and precision farming, nowadays often termed digital agriculture, have developed as new scientific fields that use data intensive techniques to improve agricultural output while minimizing its environmental impact. The data generated in modern agricultural operations is provided by a variety of different sensors that enable a better understanding of the operational environment (an interaction of dynamic crop, soil, and weather conditions) and the operation itself (machinery data), leading to more accurate and faster decision making [1].

Machine learning (ML) has evolved together with big data technologies and high-performance computers to provide new opportunities to unravel, measure, and understand data heavy processes in agricultural operational settings. Among other definitions, ML is defined as the scientific field that gives computers the ability to learn without being strictly programmed. Year by year, ML applies in more and more scientific fields including, for example, bioinformatics, biochemistry medicine, meteorology, economic sciences, robotics, aquaculture, and food security, and climatology [2].

In this post, we offer a comprehensive review of the application of ML in agriculture. A number of significant papers are provided that highlight crucial and unique features of popular ML models. The structure of the present work is as follows: the ML terminology, definition, learning tasks, and analysis are initially given, along with the most popular learning models and algorithms. This paper offers the applicable method for the collection and categorization of the provided works. Finally, in this paper, the advantages obtained from the use of ML in agri-technology are highlighted, as well as the future aspirations in the field [3].

An Overview on Machine Learning

Machine Learning Terminology and Definitions:

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 8, August 2022

Typically, ML techniques involve a learning process with the aim to learn from "experience" (training data) to perform a task. Data in ML consists of a collection of instances. Usually, each unique example is described by a set of characteristics, commonly called as features or variables. A feature may be nominal (enumeration), binary (i.e., 0 or 1), ordinal (e.g., A+ or B-), or quantitative (integer, real quantity, etc.). The performance of the ML model in a given task is evaluated by a performance metric that is improved with experience over time. To calculate the performance of ML models and algorithms, various statistical and mathematical models are used. After the end of the learning phase, the trained model may be used to classify, predict, or cluster new examples (testing data) utilizing the knowledge acquired throughout the training process [4].

ML tasks are typically classified into several broad categories depending on the learning type (supervised/unsupervised), learning models (classification, regression, clustering, and dimensionality reduction), or the learning models used to perform the selected job.

2.1 Tasks of Learning:

ML tasks are classified into two main categories, which is, supervised and unsupervised learning, depending on the learning signal of the learning system. In supervised learning, data are provided with example inputs and the corresponding outputs, and the aim is to construct a general rule that maps inputs to outputs. In some cases, inputs may be only partially available with portions of the intended outputs missing or given only as feedback to the actions in a dynamic environment (reinforcement learning). In the supervised scenario, the acquired information (trained model) is used to predict the missing outputs (labels) for the test data. In unsupervised learning, however, there is no distinction between training and test sets with data being unlabeled. The learner examines incoming data with the goal of discovering hidden patterns [5].

2.2. Analysis of Learning:

Dimensionality reduction (DR) is an analysis that is done in both families of supervised and unsupervised learning types, with the aim of creating a more compact, lower-dimensional representation of a dataset to preserve as much information as possible from the original data. It is usually done prior to employing a classification or regression model in order to avoid the effects of dimensionality. Some of the most common DR techniques include the following: (i) principle component analysis, (ii) partial least squares regression, and (iii) linear discriminant analysis [6].

2. LITERATURE REVIEW

H. Asadi et al. presented in this article that stroke is a significant cause of mortality and disability. Accurately predicting stroke outcome from a collection of predictive factors may identify high-risk individuals and guide treatment methods, leading to reduced morbidity. Logistic regression models allow for the discovery and validation of predictive factors. However, sophisticated machine learning techniques provide an alternative, in particular, for large-scale multi-institutional data, with the benefit of readily integrating newly available data to enhance prediction accuracy. Our goal was to develop and evaluate various machine learning techniques, capable of predicting the outcome of endovascular intervention in acute anterior circulation ischemic stroke. Method: We performed a retrospective analysis of a prospectively collected database of acute ischemic stroke treated by endovascular intervention. Using SPSS, MATLAB, and Rapidminer classical statistics as well as artificial neural network and support vector techniques were used to build a supervised machine capable of categorizing these

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 8, August 2022

predictors into probable good and bad outcomes. These methods were trained, verified and tested using randomly split data. We included 107 consecutive acute anterior circulation ischemic stroke patients treated by endovascular method. Sixty-six were male and the mean age was 65.3. All the relevant demographic, procedural and clinical variables were incorporated into the models. The final confusion matrix of the neural network, showed an overall congruency of 80 percent between the target and output classes, with good receiving operational properties. However, following optimization, the support vector machine showed a significantly higher performance, with a root mean squared error of 2.064 (SD: ±0.408). We demonstrated good accuracy of outcome prediction, utilizing supervised machine learning methods, with potential for integration of bigger multicenter datasets, likely significantly enhancing prediction. Finally, we suggest that a strong machine learning system may possibly improve the selection process for endovascular vs medicinal therapy in the management of acute stroke [7].

S. Cramer et al. presented in this article that challenging research challenges in the field of machine learning, and more generally intelligent systems, when the predictions of certain target variables are important to a given application. Rainfall is a great example, since it displays unique features of high volatility and chaotic patterns that do not present in other time series data. This work's primary effect is to demonstrate the advantage machine learning methods, and more generally intelligent systems have over the present state-of-the-art approaches for rainfall prediction inside rainfall derivatives. We apply and compare the predictive performance of the current state-of-the-art (Markov chain extended with rainfall prediction) and six other popular machine learning algorithms, namely: Genetic Programming, Support Vector Regression, Radial Basis Neural Networks, M5 Rules, M5 Model trees, and k-Nearest Neighbors. To help in the comprehensive assessment, we conduct experiments utilizing the rainfall time series across data sets for 42 cities, with highly different climatic characteristics. This comprehensive study demonstrates that the machine learning techniques are able to surpass the present state-of-the-art. Another aspect of this study is to identify connections between various climates and prediction accuracy. Thus, these findings demonstrate the beneficial impact that machine learning-based intelligent systems have for forecasting rainfall based on predictive accuracy and with low correlations occurring across climates [8].

J. Rhee et al. presented in this article that a high-resolution drought prediction model for ungauged regions was created in this research. The Standardized Precipitation Index (SPI) and Standardized Precipitation Evapotranspiration Index (SPEI) with 3-, 6-, 9-, and 12-month time scales were predicted with 1–6-month lead periods at 0.05 × 0.05° resolution. The use of long-range climate prediction data was contrasted to the use of climatological data during periods with no observation data. Machine learning models using drought-related variables based on remote sensing data were compared to the spatial interpolation of Kriging. Two performance measures were used; one is producer's drought accuracy, defined as the number of correctly classified samples in extreme, severe, and moderate drought classes over the total number of samples in those classes, and the other is user's drought accuracy, defined as the number of correctly classified samples in drought classes over the total number of samples classified to those classes. One of the machine learning models, highly randomized trees, performed the best in most instances in terms of producer's accuracy reaching up to 64 percent, while spatial interpolation performed better in terms of user's accuracy up to 44 percent. The contribution of long-range climate prediction data was not substantial under the circumstances employed in this research, but additional improvement is anticipated if forecast ability is increased or a more complex downscaling technique is applied. Simulated reductions of forecast error in

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 8, August 2022

precipitation and mean temperature were tested: the simulated decrease of forecast error in precipitation improves drought prediction while the decrease of forecast error in mean temperature does not help much. Although there is still some space for development, the created model may be utilized for drought-related decision making in ungauged regions [9].

A. Aybar-Ruiz et al. presented in this article a new method for global solar radiation prediction, based on a hybrid neural-genetic system. Specifically a Grouping Genetic Algorithm (GGA) and an Extreme Learning Machine (ELM) algorithm have been combined in a single algorithm, in such a manner that the GGA solves the optimum selection of features, and the ELM carries out the prediction. The suggested method is particularly unique since it utilizes as input of the system the output of a numerical weather meso-scale model (WRF), i.e., atmospherically variables predicted by the WRF at various nodes. We examine then various issues connected with this broad algorithmic framework: Initially, we assess the capability of the GGA-ELM for carrying out a statistical downscaling of the WRF to a particular location of interest (where a measure of solar radiation is available), i.e., we only take into consideration predictive factors from the WRF and the objective variable at the same time tag. In a second assessment method, we attempt to forecast the solar radiation at the location of interest at various time tags t+x, utilizing predictive factors from the WRF. Finally, we handle the full prediction issue by incorporating past values of observed solar radiation in the forecast. The suggested method and its efficiency for choosing the optimal set of features from the WRF are examined in this article, and we also present various operators and dynamics for the GGA. Finally, we evaluate the performance of the system with these different characteristics in a real problem of solar radiation prediction at Toledo's radiometric observatory (Spain), where the proposed system has shown an excellent performance in all the sub problems considered, in terms of different error metrics [10].

Water Management

Water management in agricultural production requires significant efforts and plays a vital role in hydrological, climatological, and agronomical balance. This section consists of four studies that were primarily intended for the estimation of daily, weekly, or monthly evapotranspiration. The accurate estimation of evapotranspiration is a complex procedure that is of a significant importance for resource management in crop production, as well as for the design and the operation management of irrigation systems. In another study, the authors developed a computer method for the estimation of monthly mean evapotranspiration for arid and semi-arid regions. It used monthly mean climatic data of 44 meteorological locations over the period 1951–2010. In another study dedicated to ML applications on agricultural water management, two scenarios were provided for the estimation of the daily evapotranspiration utilizing temperature data collected from six meteorological stations of a region over the long period (i.e., 1961–2014). Finally, in another study, scientists developed a method based on ELM model supplied with temperature data for the weekly estimation of evapotranspiration for two meteorological weather stations. The goal was the accurate estimation of weekly evapotranspiration in arid regions of India based on inadequate data situation for agricultural water management.

Daily dew point temperature, on the other hand, is a significant element for the identification of expected meteorological phenomena, as well as for the estimation of evapotranspiration and evaporation. In another article, a model is provided for the prediction of daily dew point temperature, based on ML. The weather data were collected from two different weather stations.

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 8, August 2022

Soil Management

The final topic of this study includes ML application on prediction-identification of agricultural soil parameters, such as the estimation of soil dryness, condition, temperature, and moisture content. Soil is a diversified natural resource, with intricate processes and systems that are difficult to understand. Soil characteristics allow academics to comprehend the dynamics of ecosystems and the effect on agriculture. The proper evaluation of soil conditions may lead to improved soil management. Soil temperature alone plays a significant role for the proper understanding of the climate change effects of a region and eco-environmental factors. It is a fundamental meteorological feature controlling the interaction processes between ground and atmosphere. In addition, soil moisture has a major impact on crop output unpredictability. However, soil measurements are typically time-consuming and expensive, thus a low cost and reliable solution for the accurate estimation of soil may be achieved with the use of computer analysis based on ML techniques. More specifically, this study offered a method for the evaluation of soil drying for agricultural planning. The method properly evaluates the soil drying, using evapotranspiration and precipitation data, in a region located in Urbana, IL of the United States.

The objective of this method was the provision of remote agricultural management options. The second study was intended for the prediction of soil condition. In particular, the study offered the comparison of four regression models for the prediction of soil organic carbon (OC), moisture content (MC), and total nitrogen (TN). More specifically, the scientists used a visible-near infrared (VIS-NIR) spectrophotometer to collect soil spectra from 140 raw and wet samples of the top layer of Luvisol soil types. The samples were collected from an agricultural area near Premslin, Germany in August 2013, after the harvest of wheat harvests. They discovered that the accurate prediction of soil properties may enhance soil management. In a third study, the authors developed a new method based on a self-adaptive evolutionary-extreme learning machine (SaE-ELM) model and daily weather data for the estimation of daily soil temperature at six different depths of 5, 10, 20, 30, 50, and 100 cm in two different in climate conditions regions of Iran; Bandar Abbas and Kerman. The aim was the accurate estimation of soil temperature for agricultural management. The final study offered a novel method for the evaluation of soil moisture, based on ANN models using data from force sensors on a no-till chisel opener.

3. DISCUSSION

It is demonstrated that ML models have been used in various applications for crop management (61 percent); mainly yield prediction (20 percent) and disease detection (22 percent). This trend in the applications distribution represents the data intensive applications inside crop and high usage of pictures (spectral, hyperspectral, NIR, etc.). Data analysis, as an established scientific discipline, offers the foundation for the creation of many applications linked to crop management since, in most instances, ML-based predictions can be derived without the requirement for fusion of data from other resources. In contrast, when data records are involved, sometimes at the level of large data, the implementations of ML are fewer in number, primarily due of the additional efforts needed for the data analysis job and not for the ML models per se. This finding partly explains the nearly equal distribution of ML applications in livestock management (19 percent), water management (10 percent), and soil management (10 percent). It is also apparent from the analysis that most of the research utilized ANN and SVM ML models. More precisely, ANNs were utilized mainly for applications in agricultural, water, and soil management, whereas SVMs were used mostly for animal management.

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 8, August 2022

4. CONCLUSION

By integrating machine learning to sensor data, farm management systems are evolving into true artificial intelligence systems, providing richer recommendations and insights for the following decisions and actions with the ultimate scope of productivity improvement. For this reason, in the future, it is expected that the usage of ML models will be increasingly more common, allowing for the possibility of integrated and appropriate tools. At the moment, all of the techniques involve individual approaches and solutions and are not adequately connected with the decision-making process, as seen in other application areas. This combination of automated data recording, data analysis, ML implementation, and decision-making or support will provide practical tolls that fall in line with the so-called knowledge-based agriculture for increasing production levels and bio-products quality.

REFERENCES

[1] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM J. Res. Dev.*, 2000, doi: 10.1147/rd.441.0206.

[2] L. Kong *et al.*, "CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine," *Nucleic Acids Res.*, 2007, doi: 10.1093/nar/gkm391.

[3] A. Richardson, B. M. Signor, B. A. Lidbury, and T. Badrick, "Clinical chemistry in higher dimensions: Machine-learning and enhanced prediction from routine clinical chemistry data," *Clinical Biochemistry*. 2016, doi: 10.1016/j.clinbiochem.2016.07.013.

[4] J. Wildenhain *et al.*, "Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning," *Cell Syst.*, 2015, doi: 10.1016/j.cels.2015.12.003.

[5] J. Kang, R. Schwartz, J. Flickinger, and S. Beriwal, "Machine learning approaches for predicting radiation therapy outcomes: A clinician's perspective," *International Journal of Radiation Oncology Biology Physics*. 2015, doi: 10.1016/j.ijrobp.2015.07.2286.

[6] B. Zhang *et al.*, "Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma," *Cancer Lett.*, 2017, doi: 10.1016/j.canlet.2017.06.004.

[7] H. Asadi, R. Dowling, B. Yan, and P. Mitchell, "Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy," *PLoS One*, 2014, doi: 10.1371/journal.pone.0088225.

[8] S. Cramer, M. Kampouridis, A. A. Freitas, and A. K. Alexandridis, "An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives," *Expert Syst. Appl.*, 2017, doi: 10.1016/j.eswa.2017.05.029.

[9] J. Rhee and J. Im, "Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data," *Agric. For. Meteorol.*, 2017, doi: 10.1016/j.agrformet.2017.02.011.

[10] A. Aybar-Ruiz *et al.*, "A novel Grouping Genetic Algorithm-Extreme Learning Machine approach for global solar radiation prediction from numerical weather models inputs," *Sol. Energy*, 2016, doi: 10.1016/j.solener.2016.03.015.