IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Comprehensive Analysis Of Machine Learning Methods In Agricultural Crop Cultivation

Patan Rasool[1], Pidugu Deepthi[2], Sk.Sahera Bhanu[3]

PG Students[123], Dept.of MCA[123], Sree Venkateswara College of engineering[123], Nellore, Ap.

Abstract

In an attempt to reevaluate the research studies on the applicability of machine learning techniques in the field of agricultural crop production, this paper was prepared. This technique is a novel approach to agricultural crop management production. The directorate of economics and statistics issues forecasts of crop production that are both accurate and timely, which are essential for making important policy decisions like import-export, pricing, marketing distribution, etc. But it's important to realize that these earlier estimates weren't objective since they needed a lot of descriptive analysis based on numerous qualitative variables. Therefore, it is necessary to develop objective crop production predictions that are statistically sound. A significant amount of data has been made available by the advancements in computing and information storage. The challenge has been extracting complex knowledge from this raw data, which has prompted the creation of new methods and strategies like machine learning that can be applied to combine data knowledge with crop yield assessment. The goal of this study is to assess these novel approaches so that important relationships can be discovered by applying them to the different variables included in the database.

Keywords: Artificial Neural Network, Decision Tree, Machine Learning, Regression Analysis, Time Series Analysis

I. INTRODUCTION

Civilization began with agriculture. India is predominantly an agrarian nation, with a crop-based economy. Thus, in India, agriculture is the foundation of all business. India currently holds the second-place position in the world for agricultural output. In 2015, agriculture and related industries like fishing and forestry accounted for 14.5% of the country's GDP and roughly 50% of its labor force. Though agriculture's contribution to India's GDP is steadily decreasing, it still constitutes the country's broadest economic sector and is vital to the country's diverse socioeconomic framework. Climate is just one of the many variables that impact Indian agriculture. Other factors include topography, history, geography, biology, politics, institutions, and socioeconomics. Over time, there have been changes.

Climate is just one of the many variables that impact Indian agriculture. Other factors include topography, history, geography, biology, politics, institutions, and socioeconomics. As natural factors and technological advancements evolved over time, policies also had to adapt. Thus, the performance of agricultural production also varies dramatically and with significant gaps in various national geographic locations. Agriculture is impacted by a variety of independent factors. So this arise risk and the consistent output of food also affected. Environmental factors have a major impact on agricultural production. Large intra-seasonal yield variability results from weather influences on crop growth and development. Furthermore, the interaction between soil property variability and weather patterns results in variations in spatial yield. Weather-related yield losses can be compensated for through crop agronomic management practices such as planting, fertilizer application, irrigation, tillage, etc. Therefore, yield forecasting is a crucial tool for assessing crop-area insurance contracts and optimizing crop yield.

The population, solid waste management, surface and ground water hydrology, pollution, and other factors all contribute to periodic changes in the climate, as does the impact of these changes on developing nations, as outlined by G. Yamuna. That gives us the notion that we also need to examine production in light of climate change1. In [2], they discussed the analysis of rainfall runoff using artificial neural networks and came to the conclusion that the model of artificial neural networks could accurately predict the data. Using a neural network, In forecasted the reservoir water level stage.

Giving knowledge to a machine is known as machine learning. Supervised and unsupervised learning are two examples of the different kinds of machine learning approaches. Under supervised learning, a single supervisor keeps an eye on the process by which the program learns from training examples and then applies that knowledge to accurately infer conclusions for fresh data. Supervised learning includes the use of artificial neural networks, Bayesian networks, decision trees, support vector machines, ID3, k-nearest neighbor, hidden markov models, and more. With unsupervised machine learning, a large amount of data is fed to the system, and it uses this data to identify patterns and relationships. Consequently, unsupervised learning can be used to find hidden patterns in the data. Some examples of unsupervised learning algorithms are k-nearest neighbor, self organizing map, and partial based clustering, hierarchical clustering, k-means clustering etc.

Machine learning, when combined with statistics and computer science, increases prediction accuracy. Data scientists, data analysts, and those wishing to use raw data to forecast or identify trends in data are the main users of it. Machine learning techniques can be applied to agriculture and agricultural production to find accurate crop production predictions because there is a growing amount of data in these fields.

II.APPLICATION OF MACHINE LEARNING TECHNIQUES IN AGRICULTURE

Crop prediction can be done through a variety of machine learning algorithms, including statistical and mathematical methods. Here are some of the techniques that have already been researched.

a).Time Series Analysis

Time series analysis is a technique for examining time in parametrimetric series data in order to derive useful statistics and additional features. A model to forecast future values based on previously observed values is called a time series forecasting model. Time series analysis techniques come in different varieties, including frequency domain and time domain, parametric and non-parametric, linear and nonlinear, univariate and multivariate. Spectral and wavelet analysis are examples of frequency domain analysis; auto-correlation and cross-correlation are examples of time domain analysis; parametric approaches include autoregressive or moving average models; non-parametric approaches include covariance or the process spectrum. Crop forecasting relies heavily on time series analysis. We can determine the relationship between yield and time by considering the dependent variable yield as a function of time.

outlined the novel idea of crop yield under average climatic conditions in [11] and established a forecasting model using time series analysis of historical yield data. They used data on grain per unit yield, which was obtained by adding various crop per unit yields from 1949 to 2005, and tested the model in the province of Liaoning. They started by using the moving average method, then they used the regression equation to find the difference and determine how the climate affected yield. They therefore come to the conclusion that the moving average model is thought to be a good model for yield forecasting. The strong point is it needed a relatively small amount of data.

b). Markov Chain Model

A stochastic process that can be represented mathematically in a probabilistic way is the Markov chain. When dealing with a markov chain, an experiment's result is solely dependent on the results of its predecessors. Put another way, we could say that the system's future state is dependent upon its current state. Markov chains bear the name of the Russian mathematician who founded the field of stochastic process theory in 1856–1922. The Markov chain approach [12] is used to forecast cotton yield from pre-harvest crop data. They have looked into how useful the markov chain method is for forecasting crop yields. They examine separately the data on cotton yields for the four-year period spanning 1981–1984, pertaining to two important producing states: Texas and California. They use a Markov chain to estimate the probability distribution. They have employed two regression models, such as multiple linear regressions and multiple rank regressions, to help choose the important variables within each period for the baseline 1981–1983 data. Four transition matrices were computed from the states of Texas and California.

Next, the consecutive transition matrices are multiplied to determine the expected yield distribution. They have utilized the means of these anticipated yield distributions to forecast yield. In 1946, the second order Markov chain model was utilized to forecast sugarcane yield, allowing for the simultaneous use of data from two stages. They discovered that this model outperforms the current models, which include the regression model and the first-order Markov chain model. Consequently, they draw the conclusion that, when it comes to crop yield forecasting, the second order Markov chain model is more useful than both the regression model and the first order Markov chain model.

A machine learning technique is suggested to commendable crop rotations which can predict with good ability, at the outset of agricultural season, the crops all most achievable to be presented in a given field by implementing the cropping system of coming 3-5 years. The access capable to analyze from data and acquire professional awareness as mentioned by first order logic rules.

c). Artificial Neural Network

The network of artificial neurons is known as an artificial neural network, or ANN. It is predicated on the biological functions of the human brain. It is among the instances of learning that is supervised. After a neural network has been trained once, similar patterns in subsequent data can be predicted, for example, and meaningful solutions to problems can be generated even in cases where the input data is inaccurate or lacking.The addition of more and more data causes the accuracy of ANN to continue rising. Furthermore, even in the absence of knowledge of the underlying principles, ANNs are able to adopt their complexity. Any process can use an ANN to determine the relationship between input and output.

Artificial neural networks (ANN) were utilized to forecast input energy-based potato production in Iran by designing output energy and greenhouse gas emissions (GHG). They used a face-to-face approach to gather data from 260 farmers. As a result, different ANN were developed, and their forecasting efficiency was evaluated using the quality factor. The average total input and output energy for potato production, according to the results, was 83,723 and 83,059 MJ ha-1, respectively.

When it came to energy utilization, electricity, chemical fertilizer, and seed were the most important factors. Measures were taken of energy productivity and efficiency. The optimal ANN design for estimating or assessing the output energy and overall greenhouse gas emissions in potatoes was the one with the 12-8-2 structure. For total greenhouse gas emissions and potato output energy, the optimal topology's coefficient of determination (R2) was 0.99 and 0.98, respectively.

In order to validate the symptoms of the tomato crop, In5 used artificial intelligence, machine learning algorithms—specifically, ID3 algorithms and a few optimization algorithms—to build a web-based expert system with Java serving as the front end and SQL as the back end. The most significant and widely grown vegetable crop worldwide is the tomato. It is grown all over the world in market gardens, farm gardens, and kitchen gardens for both processing and raw or fresh use. This tomato crop expert conductive system was created collaboratively by a team of skilled computer engineers, programmers, and designers, as well as eminent agricultural scientists and experts in the field of tomato cultivation.

The expert system and the tomato knowledge systems are the two main parts of the professional setup. The applicant can get all the trustworthy guidance they require via the messaging system on a range of subjects, such as varieties, disease and pest symptoms, cultural practices, and the mosaic of tomato fruits and plants. The client and the system interact online when using an expert system. When you ask a question of a client, they have to answer. Once the expert system has the response from the client, it will advise and show its control measure for each risk. Expert systems with knowledge linked to identifying different pests, diseases, and varieties typically arise in tomato crops.

Sixth-generation maize breeding techniques utilizing machine learning (ML) have demonstrated that these ML algorithms show promise and can be applied to statistical techniques in maize, such as the more recently popularized linear mixed models. One new area of technology that is being used to speed up the release of new genotypes is machine learning. The literature currently addresses several key areas, including the popular genome-wide selections, quantitative trait loci mapping heterotic group assignment, and several strategic uses of ML in maize breeding. Together with rice and wheat, corn is one of the most significant cereals in the world and the main source of calories for humans. Over time, genotypes have evolved to adapt to an increasingly harsh climate, particularly drought situations where crops must be grown in marginal laws and with shifting climate conditions.

d). Information Fuzzy Network

In order to predict crop yield using an aggressive neural network, remote sensing data and other parameters were analyzed. The flexible Neuro-fuzzy Inference system (ANFIS) was utilized by them. The inputs to ANFIS are the soil moisture content, ground biomass, and repository organ. It has only a single number is sought or otherwise one output node, i.e. yield. An additional issue with yield forecasting is the short latency of remote sensing data. Thus, in order to construct a model that can predict future values, any forecasting effort is forced to apply a very limited number of previous years. Discipline is applied to the arrangement by excluding one year and utilizing all other data.

Since it is acknowledged that climate change will have an impact on crop arrangement and high productivity in Malaysian agriculture, researchers in 20 studied how to group the aspects of incomplete information and employed fuzzy modeling advancement to overcome confusion during the development of an agriculture web-based decision support system. In order to administer transparent and more acceptable information during user communication, data synthesis in the progeny tests requires decision of combat issuing from the heterogeneity of data. They conclude decision algorithms classifier from incomplete knowledge. In order to achieve better results, the detailed data regarding the region's soil structure and rainfall pattern will be allocated, fully utilizing the available incomplete information.

Newly developed algorithms and an investigation into planting material classification are the improvements of this study. In order to implement decision-making policies that could lead to a high recovery in the planting of material breeders in agriculture management, the researcher addressed the experimental results.

e). Decision Tree

Concepts like nodes, branches, terminal values, strategy, payoff distribution, certain equivalent, and the rollback method are all included in decision tree models. There are two types of branches and three types of nodes. A decision must be made at the decision node, which is symbolized by a square. A decision node is the source of the decision branches. A terminal value, also known as a payoff value, outcome value, or endpoint value, is associated with each terminal node. Every terminal value represents a scenario's outcome or a decision-making sequence.

A decision tree algorithm is constructed in two steps: first, a large decision tree is grown; in the second step, the tree is pruned to reduce size and overfit the data. The classification tree described24 is the name given to the pruned decision tree that is used for classification purposes.
The prediction is affected by a number of factors. In [16] used agronomic variables, nitrogen application and weed control using the machine learning algorithm such as artificial neural network and Decision tree to develop the yield mapping and to forecast yield. They come to the conclusion that using ANNs yields high prediction accuracies.

Outlined the decision tree algorithm modeling of soybean productivity in In26. They had gathered information on the district of Bhopal's climate from 1984 to 2003. They took into account meteorological variables like rainfall, maximum temperature, maximum relative humidity, evaporation, and crop yield soybean. The Interactive Dichotomizer3 (ID3) algorithm, which is based on two assumptions and information, has been applied. The relative humidity was found to be a significant influencing factor on the yield of soybean crops using the induction tree analysis.

f).Regression Analysis

For consultants, producers, and other agriculture-related organizations, crop productivity is crucial. Crop forecasting can be accomplished by gathering data from various sources and utilizing various data types, including soil, remote sensing, agricultural statistics, metrological, agro-metrological, and soil data. Decisions about marketing, storage, and transportation need to be made with an accurate and timely forecast. In Dahikar MSS, Rode SV., research was conducted to develop a framework for a crop prediction model. The findings indicated that planting practices, especially the proper application of fertilizer, had a greater impact on corn yield than climate-related variables.

By Kumar AVTV, Rajini Kanth R., it was predicted how temperature would affect the yield of the Jowar crop. The experiment was conducted by taking the Jowar production and both maximum and minimum temperature of India during the period from 1950-2011. When using Pearson correlation coefficients, a 2-tailed significance test was used, and the results were significant at the 0.01 level. The relationship between crop yield and temperature was then analyzed using regression. It was concluded that while other factors affected the yield at a higher level, the yield of Jowar crops increased when the temperature was lowered. Jowar yields were found to be very less dependent on temperature.

By utilizing the yield data along with the fortnightly weather variables, such as average daily maximum and minimum temperature, relative humidity in the morning and evening, and total fortnightly rainfall, a model was created by 30 to forecast the sugarcane yield in the Coimbatore district. They used the 1981–2004 data and created a model based on that data for the years 1981–2001. Using data from (2002-2004), the model was validated. They conclude that two months prior to harvest, the sugarcane yield can be accurately predicted using the regression technique because their forecast model was able to account for 87% of the variation in the yield.

g).Clustering

The process of identifying objects that are similar to each other but distinct from individuals in other groups is known as cluster analysis or clustering. Data analysis is its primary application. Several domains, including agriculture, machine learning, pattern recognition, image analysis, and information retrieval, use clustering. While there are other clustering algorithms, such as k-means and k-medoid, k-means is the most widely used and significant clustering algorithm.

An assessment of the modified k-Means clustering algorithm for crop prediction is shown in In Glauston R, Liman TD, Stephany S. paper. The comparison between the modified k-Means clustering algorithm and both k-Means and Means++ clustering algorithms was demonstrated by their evaluation and results, which indicated that the modified k-Means achieved the highest number of high-quality clusters, accurate crop prediction, and maximum accuracy count.

The metrological data is classified by the weather forecast model by [18] based on the frequency of variables that are available. The patterns linked to severe convective activity have been identified. During the summer of 2007, their results indicated good classification performance for a few chosen mini-regions of Brazil. They reasoned that their metrological model, Eta, would help meteorologists spot patterns ahead of time.

h). Principal Component Analysis

In [18], a principal component analysis-based data mining process was used to determine the arrangement and familiarity of the data used for accurate forecasting. Rainfall is attainable, and the primary case variable in the monsoon evidence is that the amount of rainfall varies periodically depending on the location. For a catchment area where water storage or rainfall operations have been conducted, rainfall outlining is crucial, especially for flood-prone techniques. For the past ten years, there have been problems with the precise implementation of the statistically based approach currently used to predict the Indian monsoon rainfall. The economy of India may be severely impacted by the slight fluctuations in the country's recurring rainfall.

Weather and water asset analysts typically use the component derivation method to assess large-scale spatial datasets like total sea surface temperature and rainfall intervals. In order to determine the time-based evolution of India's monsoon rainfall, the principal component analysis is examined.

i). Bayesian Belief Network

Statistical models of this kind are known as Bayesian networks, Bayes networks, belief networks, Bayesian models, or probabilistic directed acyclic graph models.

[19] developed a belief network to evaluate how climate change may affect potato production. A belief network that takes into account the variability of current weather parameters like temperature, radiation, and rainfall, as well as knowledge about potato development, has been demonstrated. This network combines the uncertainty of future climate change. They believed that the policy makers in agriculture received support from their network. They test their model by using synthetic weather scenarios and then the results are com¬pared with the conventional mathematical model and conclude that the efficiency is more for the belief network.

A detailed about the study is shown in the **Table 1.**

Fusion type	**Application area**
Nonlinear Regression36	Forecasting Corn Yields
Markov Chain Approach45	Forecasting Cotton Yields from Surveys
Linear Regression34	Estimating Grain Yield of Maturing Rice
Belief Networks 43	Future Crop Production
Neuro-Fuzzy Modeling18	For Crop Yield Prediction
Secpnd Order Markov Chains46	Forecasting of Crop Yields
Polynomial Regression35	Factors Affecting the Yield of Winter Cereals in CropMargins
Deterministic and Probabilistic Prediction Approaches38	Season to Inter-annual Climate Forecasting
FINkNN: a fuzzy interval number k-nearest neighbour classifier19	for prediction of sugar production from populations of samples
Artificial neural networks13	corn and soybean yield prediction
Neural Network11	Rice Crop Monitoring
Artificial Neural Networks12	Forecasting Thailands Rice Export
Building a fuzzy logic information network and a decision-support system8	for olive cultivation in Andalusia
Regression30	sugarcane yield using climatic variables
Decision Tree Algorithms26	Soybean Productivity Modelling
A Fuzzy Modelling of Decision Support System20	for Crop Selection
Time Series Techniques44	Crop Yield Forecast
Datamining with climate variable29	Jowar Crop Yield in India
Fuzzy Cognitive Map learning approach21	Yield prediction in apples
Regression and Neural Networks Models14	for Prediction of Crop Production
Modified K-Means Clustering40	Crop Prediction
Artificial Neural Network Approach17	Agricultural Crop Yield Prediction
Markov logic model47	Crop rotations for early crop mapping

III.CONCLUSIONS

A growing number of machine learning applications in agriculture are needed these days, and in order to uncover hidden knowledge, a significant amount of data from various sources can be analyzed. This area of advanced research is anticipated to expand in the coming years. Crop forecasting is made easier by the combination of computer science and agriculture. Developing an objective methodology for pre-harvest crop forecasting is necessary. There are some advantages to developing a suitable model over the conventional forecasting approach.

IV.REFERENCES

1. Yamuna G. Impacts of climate change in the developing world. Indian Journal of Science and Technology. 2016 Feb; 9(6):1–1.
2. Chakravarti A, Joshi N, Panjiar H. Rainfall runoff analysis using artificial neural network. Indian Journal of Science and Technology. 2015 Jul; 8(14):1–7.
3. Ashaary N, Ishak WHW, Ku-Mahamud KR. Neural network application in the change of reservoir water level stage forecasting. Indian Journal of Science and Technology. 2015 Jul; 8(13):1–6.
4. Khoshnevisan B, Shahin Rafiee S, Omid M, Mousazadeh H, Rajaeifar MA. Application of artificial neural networks for prediction of output energy and GHG emissions in potato production in Iran, Elsevier. Agricultural Systems. 2014; 123:120–7.
5. Babu MSP, Ramana Murty NV, Narayana SVNL. A web based tomato crop expert information system based on artificial intelligence and machine learning algorithms. International Journal of Computer Science and Information Technologies. 2010; 1(3):1–5.
6. Ornella L, Cervigni G, Tapia E. Applications of Machine Learning for Maize Breeding. In: Venkateswarlu B, Shanker AK, Shanker C. Book chapter of Crop stress and its management: Perspectives and Strategies, Springer, New York, USA. 2012; 1–29.
7. Attonaty JM, Chatelin MH, Garcia F, Ndiaye N. Using extended machine learning and simulation technics to design crop management strategies. First European Information Technology in Agriculture, Copenhagen. 1997; 1–500.
8. Gorni G, Augusto A. The Application of Neural Networks in the Modeling of Plate Rolling Processes, JOM. The Minerals, Metals and Material Society. 2008; 49(4):1–4.
9. Drummond ST, Suddeth KA, Birrell SJ. Analysis and Correlation Methods for Spatial Data. ASAE. 1995; 1–6.
10. Baskar SS, Arockiam L, Arul Kumar V, Jeyasimman L. Brief survey of application of data mining techniques to agriculture. Agricultural Journal. 2010; 5(2):116–8.
11. Hong-Ying L, Yan-Lin H, Yong-Juan Y, Hui-Ming Z. Crop yield forecasted model based on time series techniques. Journal of Northeast Agricultural University (English Edition). 2012; 19(1):73–7.

12. Matis JH, Birkett T, Boudreaux D. An Application of the Markov Chain Approach to Forecasting Cotton Yields from Surveys. Agricultural Systems. 1989; 29(4):357–70. 13. Monisha Kaul M, Robert L, Hill H, Walthall C. Artificial neural networks for corn and Soybean yield prediction, Elsevier. Agricultural System. 2005; 85(1):1–18.

14. Prasad PR, Begum SA. Regression and neural networks models for prediction of crop production. International Journal of Scientific and Engineering Research. 2013 Sep; 4(9):98–108.

15. Agrawal R, Mehta SC. Weather Based Forecasting of Crop Yields, Pests and Diseases - IASRI Models. Journal of Ind Soc Agril Statistics. 2007; 62(2):1–12.

16. Uno Y. Artificial Neural Networks to Predict Corn Yield from Compact Airborne Spectographic Imager Data. Computers and Electronics in Agriculture. 2005; 47(2):149– 61. 18. Stathakis D, Savin I, Negre T. Neuro-fuzzy modeling for crop yield prediction. The International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences. 1994; 34:1–4.

17. Rossana MC, Leon D, Rex EL, Jalao J. A Prediction Model Framework for Crop Yield Prediction. Asia Pacific Industrial Engineering and Management Society Conference Proceedings Cebu, Phillipines. 2013. p. 185.

18. Brindha KP. Data mining based on principal component analysis for rainfall forecasting in India. International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE). 2013 Sep; 3(9):1–11.

19. Yiqun Gu Y, James W, McNicol M. An Application of Belief Networks to Future Crop Production. IEEE Conference on Artificial Intelligence for Applications, San Antonia, TX. 1994. p. 305–9. 20. Salleh MNM. A Fuzzy Modelling of Decision Support System for Crop Selection. IEEE Symposium on Industrial Electronics and Applications (ISIEA2012), Bandung, Indonesia. 2012; 17–22.

21. Papageorgiou EI, Aggelopoulou KD, Gemtos TA, Nanos GD. Yield prediction in apples using Fuzzy Cognitive Map learning approach. Computers and Electronics in Agriculture, Elsevier. 2013; 91:19–21.

22. Delgado G, Aranda V, Calero J, Sanchez-Maranon M, Serrano J M, Sanchez D, Vila MA. Building a fuzzy logic information network and a decision-support system for olive cultivation in Andalusia. Spanish Journal of Agricultural Research. 2008; 6(2):252–63.

23. Daryl H, Hepting H, Maciag T, Hill H. Web-Based Support of Crop Selection for Climate Adaptation. 45th Hawaii International Conference on System Sciences, IEEE, Maui, HI. 2012. p. 1227–605.

24. Vijaysinh A, Solanki S. Data mining techniques using WEKA classification for Sickle Cell Disease. (IJCSIT) International Journal of Computer Science and Information Technologies. 2014; 5(1):1–26.

25. Uno Y. Artificial Neural Networks to Predict Corn Yield from Compact Airborne Spectographic Imager Data. Computers and Electronics in Agriculture. 2005; 47(2):149– 61.

26. Veenadhari S, Mishra B, Singh CD. Soybean Productivity Modelling using Decision Tree Algorithms. International Journal of Computer Applications. 2011; 27(7):975–8887.