
AI ASSISTED FARMING FOR CROP RECOMMENDATION AND FARM YIELD PREDICTION

Dr.V.Vignesh¹, Mr.S.Hanuman Shabarish², Ms.B.Sasi Priya³, Mr.A.P.Sooriakand⁴, Ms.D.Sujitha⁵

¹ Associate Professor, Computer Science Engineering, KPR INSTITUTE OF ENGINEERING AND TECHNOLOGY, Coimbatore, Tamilnadu

² UG- Computer Science Engineering, KPR INSTITUTE OF ENGINEERING AND TECHNOLOGY, Coimbatore, Tamilnadu

³ UG- Computer Science Engineering, KPR INSTITUTE OF ENGINEERING AND TECHNOLOGY, Coimbatore, Tamilnadu

⁴ UG- Computer Science Engineering, KPR INSTITUTE OF ENGINEERING AND TECHNOLOGY, Coimbatore, Tamilnadu

⁵ UG- Computer Science Engineering, KPR INSTITUTE OF ENGINEERING AND TECHNOLOGY, Coimbatore, Tamilnadu

ABSTRACT

Machine learning is a crucial decision-support tool for forecasting agricultural yields, enabling judgments about which crops to cultivate and what to do when in the growing season. For this study, we performed a Systematic Literature Review (SLR) to find and combine the methods and components that are employed in agricultural prediction research. Using inclusion and exclusion criteria from six internet databases, we chose 50 publications out of a total of 567 that met our search criteria for relevancy. We thoroughly examined the chosen publications, applied, and offered recommendations for additional studies.

Our data show that temperature, rainfall, and soil type are the most often used characteristics in these models, and artificial neural networks are the most frequently used methodology. This observation was based on an examination of 50 publications, and we next looked for studies employing deep learning in additional electronic databases. We gathered the deep learning algorithms from 30 of these publications that we discovered. Convolution Neural Networks (CNN), Long-Short Term Memory (LSTM), and Deep Neural Networks are the three deep learning algorithms that are used in these investigations, according to this additional analysis (DNN).

INTRODUCTION

Many industries employ machine learning (ML) techniques, from estimating customer phone usage to examining customer behavior in supermarkets. Agriculture has long used machine learning. Crop production prediction is one of the difficult issues with precision farming. This far, several models have been proposed and confirmed. Since agricultural production depends on many parameters, including climate, weather, soil, fertilizer use and seed type, this difficulty necessitates the usage of several datasets. This demonstrates that forecasting agricultural production is a complex process that requires a number of difficult steps. The actual yield can currently be roughly predicted using crop yield prediction models, but improved yield prediction performance is still desired. Using a range of features, machine learning, a branch of artificial intelligence (AI) that focuses on learning, is a practical technique that can estimate yields more precisely. By identifying patterns, correlations, and relationships, machine learning (ML) may extract knowledge from datasets. The datasets must contain results that are modelled based on prior information in order to train the models. The predictive model is constructed from multiple elements, and as a result, the model's parameters are determined using previous data collected during the practice period. Previously collected data from the training phase is used in part during the testing phase to assess performance. An ML model can be either descriptive or predictive, depending on the research problem and goals. Descriptive models are used to learn from the data gathered and explain what has happened, while predictive models are used to forecast the future. ML research encounters a number of challenges while seeking to build a

high-performance predictive model. The right algorithms must be selected in order to solve the current issue, and both the algorithms and the supporting platforms must be capable of handling the volume of data. To get a general overview of the work that has been done on the application of ML in crop production prediction, we carried out a comprehensive literature review (SLR). A systematic literature review (SLR) identifies possible research holes in a particular area of study and offers advice to academics and practitioners who want to do new research in that area. Utilizing an SLR technique, all relevant studies are retrieved from electronic sources, summarized, and presented to answer the research questions outlined in the research. An SLR investigation offers fresh perspectives and supports knowing the state of the art through academia.

CROP YIELD PREDICTION

Crop production prediction is one of the tough issues in agriculture. It is essential to decision-making on a global, regional, and local scale. When estimating crop yield, crop, soil, climatic, environmental, and other aspects are taken into consideration. The three factors that are most usually employed are temperature, precipitation, and soil type. The most popular machine learning algorithm is neural networks. The most widely used deep learning algorithm is CNN. For the globe to produce enough food, it is essential to be able to predict crop yield. Policymakers must decide when to import and export food in a timely manner based on accurate projections in order to boost national food security.

DECISION SUPPORT SYSTEM

An organization or business employs a decision support system (DSS), which is computer software, to support decisions, assessments, and courses of action. Massive amounts of data are sorted and examined by a DSS, which builds up in-depth knowledge that may be used for problem-solving and decision-making. A decision support system (DSS) is a computer programmer that aids in decision-making for enterprises. There are huge amounts of data. The best options are then provided to an organization after being assessed.

MACHINE LEARNING

A subfield of computer science and artificial intelligence (AI) called "machine learning" aims to mimic human learning by using data and algorithms to gradually increase a system's accuracy. Many commercial and professional procedures as well as our daily lives have substantially benefited from contemporary advances like machine learning. This branch of artificial intelligence (AI) focuses on employing statistical methods to create intelligent computer systems that can gain knowledge from pre-existing databases.

DEEP LEARNING

A neural network is a network having three or more layers, and deep learning is a branch of machine learning. These neural networks attempt to emulate human brain activity and allow it to "learn" from enormous volumes of data, despite the fact that they are unable to match the human brain's powers in any way. A machine learning technique known as deep learning teaches computers to learn by imitating human behavior. One of the key technologies in autonomous vehicles enables them to recognize a stop sign or distinguish a pedestrian from a lamppost.

RELATED WORK

Using data mining techniques, the tropical wet and dry climate zones of India's rice crop yield can be predicted.

Gandhi et al. have made a nice proposal in their study. Data mining is the process of uncovering hidden patterns in massive, complicated data sets. It could be crucial in the process of making choices regarding complex agricultural challenges. Data visualization is also important to understand the broad trends of the impact of the many factors affecting crop output. The productivity of the rice crop is investigated in this study using data visualization tools to identify correlations between meteorological factors. The study also employs data mining techniques to extract knowledge from the historical agriculture data set in order to anticipate rice crop output for the Khari season in India's Tropical Wet and Dry climate zone. Microsoft Office Excel has been used to illustrate the data set using scatter plots. The categorization algorithms have been tested using WEKA, a free and open

source data mining tool. The experimental data contains information on sensitivity, specificity, accuracy, F1 score, Mathew's correlation coefficient, mean absolute error, root mean squared error, relative absolute error, and root relative squared error. The data visualization shows broad trends showing higher minimum, average, or maximum temperature for the season as well as lower precipitation in the selected climate zone increase rice crop productivity. The experimental results showed that for the current data set, J48 and LAD Tree had the best accuracy, sensitivity, and specificity. The accuracy, sensitivity, and specificity of the LWL classifier's classification results were the lowest.

Utilizing the Hadoop Framework and the Random Forest Method

In this study, ShriyaSahu et al., proposes With the development of information technology, big data has become a trendy topic. Since agriculture is essential to human survival, it must significantly advance crop data analysis. This study demonstrates how to employ a big data technique to derive insights from reliable agricultural data. In crop analysis where data is collected remotely, successfully gathering the valuable data pushes a framework toward severe computing challenges. In order to store a sizable amount of crop data for the storage purpose of massive data availability in agriculture, we intend to employ the Hadoop architecture for our work. This method enables farmers to predict which crops to plant in their fields in order to boost output more precisely. The Map Reduce programming model and the Random Forest technique are both used in the Hadoop architecture. The world's population, which is projected to grow daily during the next 35 years, will probably surpass 10 billion people. Significant improvements in agricultural productivity and disaster management are now needed to feed the world's population. Making projections requires gathering data from a variety of agricultural sources. Weather conditions have an impact on agricultural management and productivity. Future agriculture will depend heavily on weather forecasts.

Utilizing Data Mining to Predict Crop Yield

In this research, Shute Mishra et al. make the argument that the agriculture sector has a particularly large impact on India's economy. 50% of India's population works in it, and it contributes 18% of the nation's GDP. The people of India have long engaged in agriculture, but the outcomes have never been satisfactory due to a variety of factors that affect agricultural output. To meet the needs of approximately 1.2 billion people, a sufficient crop output is essential. Variables including soil type, precipitation, seed quality, a lack of technical capabilities, etc. have a direct impact on agricultural production. As a result, farmers must make informed decisions when selecting new technologies rather than depending on easy fixes in order to fulfil the growing demand. This project analyses agricultural datasets and builds a crop yield forecast system using data mining techniques. Several classifiers, including J48, LWL, LAD Tree, and IBK, are used for prediction. Then, the effectiveness of each is contrasted using the WEKA tool. One of the qualities used to assess performance is accuracy. The values of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Relative Absolute Error are used to further assess the classifiers' accuracy (RAE). The algorithm will operate with more accuracy as the error value decreases. The basis for the result is a comparison of the classifiers. The process of analyzing, extracting, and anticipating pertinent information from enormous amounts of data in order to find patterns is known as data mining. Using this technique, businesses can turn client data from its raw form into information that is useful. After choosing the data, data mining preprocesses, transforms, and searches for patterns that can be applied to forecast pertinent insights. The pre-processing step includes the detection of outliers and missing values, while the transformation step includes the discovery of object correlation.

Predicting Annual Yield of Major Crops Using Data Mining Techniques and Suggestions for Planting Different Crops in Different Bangladeshi Districts

The development of agricultural products is controlled by a number of factors, including geography, economy, biology, and climate, according to A.T.M. ShakilAhamed et al. in this research. With the appropriate statistical approaches, the effects of many variables on agriculture can be quantified. By using such procedures and approaches to historical agricultural yields, it is possible to obtain

information or knowledge that can help farmers and government organizations establish better decisions and policies that enhance productivity. The main focus of this article is the use of data mining techniques to estimate crop production for major cereal crops in important districts of Bangladesh. An important field of research that supports the security of the world's food supply is crop yield forecasting. Bangladesh is one of the top rice producers in the world, and the country has good soil for growing rice. It was the sixth-largest crop producer in the world in 2012, producing 33,889,632 metric tons of rice overall [9]. To take full advantage of the subtropical climate and soil of Bangladesh, farmers must be properly informed on the best time to plant crop seeds. The yield from the annual crop also contributes to the national economy. It is essential to consider environmental factors particular to each of these locations because different districts in Bangladesh have distinct climates. The best areas for various types of agricultural cultivation can be identified using this data. Additionally, there are regional variations in rainfall, which has a big impact on farming. Because the ideal crop output depends on the proper amount of rain, whereas crops might be damaged by either too little or too much rain. Since the amount of rain varies by district, so does the humidity that comes with it. A area with the ideal annual rainfall average and humidity are required. Humidity controls how much water the atmosphere can absorb before keeping crops too wet or too dry, preventing them from developing properly and yielding.

Global and Regional Crop Yield Predictions Using Random Forests

Accurate crop output estimates are crucial for the creation of effective agricultural and food policy at the regional and global levels, according to Jig Han Jeonget. In order to assess Random Forests' (RF) ability to predict crop production responses to meteorological and biophysical variables at both the global and regional levels in wheat, maize, and potato, we compared it to multiple linear regressions (MLR), the industry standard. We employed crop yield data from numerous sources and regions, including gridded global wheat grain production, maize grain yield from US counties spanning thirty years, potato tuber and maize silage output from the northeastern seaboard, for the training and testing of our model. RF was found to be highly good at projecting crop yields and beat MLR benchmarks in all performance data that were examined. For example, in every test case, the root mean square errors (RMSE) for RF models ranged from 6 to 14% of the average observed yield, whereas the same values for MLR models ranged from 14% to 49%. Our results show that RF is an effective and adaptable machine-learning method for crop production estimates at regional and global sizes because to its high accuracy and precision, ease of use, and value in data analysis. RF may result in a loss of accuracy when expecting the extreme ends or answers outside the parameters of the training data.

EXISTING SYSTEM

The harvest's yield is affected by the occasional climate. India's weather conditions are continually moving. During dry seasons, ranchers deal with difficult issues. This prompted the utilization of a few AI calculations to assist ranchers with choosing the harvest that would offer the best yield. They utilize a scope of information from earlier years to figure future information. SMO classifiers in WEKA were utilized to classifications the outcomes. The essential elements considered are the typical temperature, least temperature, greatest temperature, and information on the harvests and yields from the earlier year. Utilizing the SMO device, the former information was parted into two classes, high return and unfortunate yield. The harvest yield conjecture result acquired utilizing the SMO classifier is less exact when contrasted with innocent Bayes, multi-facet insights, and Bayesian organizations. Every one of the temperatures saw, normal, least, and most extreme have an influence. They likewise added a pristine variable called crop evapotranspiration. The yield's evapotranspiration is affected by both the climate and the phase of plant development. This characteristic is considered to pursue an educated choice with respect to the yield regarding the gatherings. They gathered the informational collection with these qualities, took care of it into the Bayesian organization, and afterward partitioned it into the valid and bogus classes. The exactness not set in stone by looking at the noticed orders in the model to the anticipated groupings in the model using a chaos organization. They in the long run arrived at the goal that using Blameless Bayes to measure crop yields Besides,

a Bayesian association has more prominent exactness than a SMO classifier, making it more valuable to foresee crop yield under fluctuated climatic and farming conditions. Utilizing information mining procedures and verifiable horticultural result and climatic information, a few forecasts that increment crop yield are delivered. It is important to set up a choice emotionally supportive network to help ranchers in making taught decisions with respect to the yield and soil to be developed. They gathered the information, which incorporated the harvest season, region, and result in hectares, and used WEKA to dissect it utilizing various calculations. The exactness of four unmistakable information investigation strategies was assessed and differentiated. The four techniques utilized in WEKA are J48, IBK, Fellow trees, and LWL. Contingent upon the kind of dataset and its qualities, they arrived at the resolution that the IBK had achieved more exactness than the others.

PROPOSED FRAMEWORK

Before directing the deliberate survey, an audit technique is laid out. The assessment was led as per Kitchen Ham's popular survey systems. The initial step is to characterize the exploration questions. When the exploration questions are ready, data sets are utilized to choose the applicable investigations. The data sets utilized in this examination were Science Direct, Scopus, Web of Science, Springer Connection, Wiley, and Google Researcher. After the relevant examinations had been chosen, they were sifted and assessed utilizing a bunch of prohibition and quality standards. From the chose studies, all pertinent information is extricated, and the joined information with the removed information is utilized to respond to the exploration questions. The three components of the methodology we utilized are plan audit, direct survey, and report survey.

The underlying stage is to design the survey. Right now, research questions are recognized, a convention is made, and ultimately the interaction is tried to check whether the arrangement will work. Alongside the review subjects, the distributing settings, beginning inquiry terms, and distribution determination measures not entirely settled. The convention is indeed different to check whether it really epitomizes a satisfactory audit strategy after this information has been all characterized.

TABLE: 1

Key	Evaluation parameter	# of times used
RMSE	Root mean square error	29
R2	R-squared	19
MAE	Mean absolute error	8
MSE	Mean square error	5
MAPE	Mean absolute percentage error	3
RSAE	Reduced simple average	3
LCCC	ensemble	1
MFE	Lin's concordance correlation coefficient	1
SAE	Multi factored evaluation	1
Rcv	Simple average ensemble	1
MCC	Reference change values	1
	Matthew's correlation coefficient	1

ARCHITECTURE DIAGRAM:

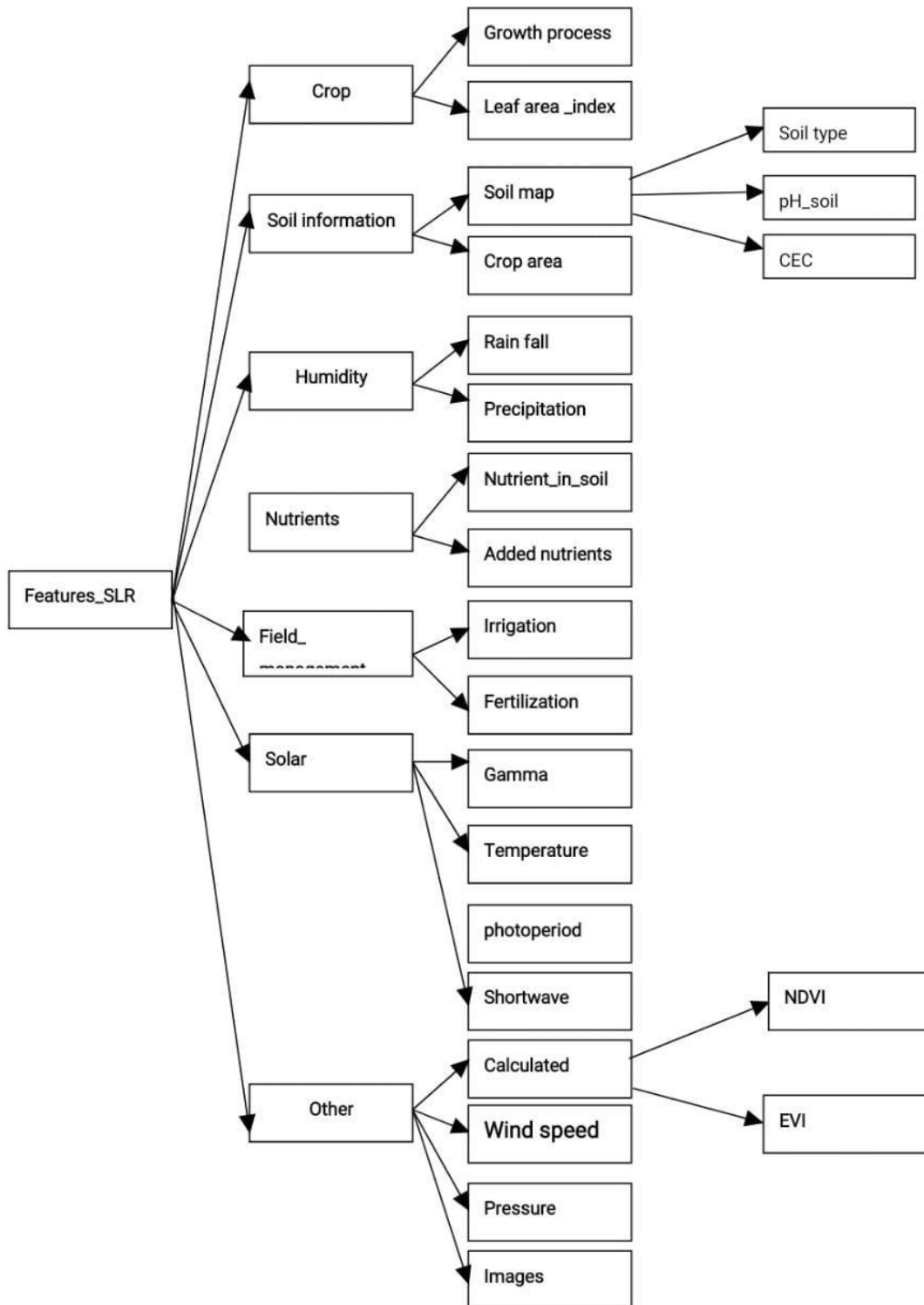


TABLE: 2

Most used machine learning algorithms.

Most used machine learning algorithms	# of times used
Neural Networks	27
Linear Regression	14
Random Forest	12
Support Vector Machine	10
Gradient Boosting Tree	4

CONCLUSION

Artificial intelligence in the agricultural field not only helped farmers to automate farming but also shifts to precise cultivation for higher crop yield and better quality using fewer resources. Companies involved in improving machine learning or artificial intelligence based products or services will get technological advancement in the future and will provide more useful applications to this sector helping the world deal with food production issues for the growing population.

REFERENCES

- [1] International Conference on Computer Applications Advances, "Rice Crop Yield Forecasting in Tropical Wet and Dry Climatic Zone of India Using Data Mining Techniques," by Nicety Gandhi et al. (ICACA) , 2016.
- [2] An Effective Analysis of Crop Yield Prediction Using Hadoop Framework Based on Random Forest Technique by ShriyaSahu.
- [3] According to IEEE Xplore, the article "Use of Data Mining in Crop Yield Prediction" by Shruti Mishear, Priyanka Paygude, Snehal Chaudhary, and SonaliDate has the part and Dnumber. CFP18J06-ART and CFP18J06DVD, respectively.
- [4] "Using Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh," copyright 2015 IEEE SNPD 2015, June 1-3, 978-1-4799-8676-7/15/\$31.00. Mohammad TanzirKabir, Kallal Das, A.T.M. ShakilAhamed, NavidTanzeem Mahmood, Nazmul Hoss.
- [5] Jig Han Jeong et al., Random Forests for Global and Regional Crop Yield Predictions, PLOS-ONE.
- [6] G. Banerjee, U. Sarkar, S. Das, I. Ghosh. Artificial Intelligence in Agriculture: A Literature Survey[J]. International Journal of Scientific Research in Computer Science Applications and Management Studies, 2018, 7(3):1-6.
- [7] "2050: A Third More Mouths to Feed." Food and Agriculture Organization of the United Nations, Food and Agriculture Organization of the United Nations, 2020, www.fao.org/news/story/en/item/35571/icode/.
- [8] Popa, Cosmin. Adoption of Artificial Intelligence in Agriculture[J]. Bulletin of the University of Agricultural Sciences & Veterinary, 2011.



- 9] Mckinion J M, Lemmon H. E. Expert systems for agriculture[J]. Computers & Electronics in Agriculture, 1985, 1(1):31-40.
- 10] Gertsis AC, Galanopoulou-Sendouca S, Papathanasiou G, Symeonakis A. USE OF GOSSYM-A COTTON GROWTH SIMULATION MODEL-TO MANAGE A LOW INPUT COTTON PRODUCTION SYSTEM IN GREECE. In First European Conference for Information Technology in Agriculture, 199.
- 11] P. P. Ray, "Internet of things for smart agriculture: Technologies, practices and future direction," Journal of Ambient Intelligence and Smart Environments, vol. 9, no. 4, pp. 395-420, 2017.
- 12] M. S. Mekala and P. Viswanathan, "A Survey: Smart agriculture IoT with cloud computing," in 2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS), 2017, pp. 1- 7: IEEE.
- 13] Singh M V. (2008). Micronutrient Deficiencies in Crops and Soils in India. In Micronutrient,Deficiencies in Global Crop Production.
- 14] MoA. (2013). State of Indian agriculture, 2012-13, Department of Agriculture and Cooperation, Ministry of Agriculture.
- 15] Kanan, E., & Sundaram, S. (2011). Analysis of trends in India's Agricultural Growth, Working