A Review of IoT Systems and Machine Learning Techniques in Crop Yield Prediction

Nitin Zachariah
Department of CSE
CMR Institute of Technology
Bengaluru, India
nitinzachariah8@gmail.com

Chaitra Ankesh
Department of CSE
CMR Institute of Technology
Bengaluru, India
chan16cs@cmrit.ac.in

Aliva Mohanty R
Department of CSE
CMR Institute of Technology
Bengaluru, India
almo18cs@cmrit.ac.in

Bhuvaneshwari R D
Department of CSE
CMR Institute of Technology
Bengaluru, India
bhrd18cs@cmrit.ac.in

Preethi Sheba Hepsiba
Department of CSE
CMR Institute of Technology
Bengaluru, India
preethisheba.h@cmrit.ac.in

Abstract— The agricultural sector across the world has always faced challenges in ensuring adequate production amidst changing seasons, weather conditions, arable land, and unpredictable natural disasters. Internet of Things (IoT) and data analytics play an important role in monitoring various parameters affecting the crop yield and using state-of-the-art machine learning techniques to predict the yield based on dependent variables. The IoT systems related to agriculture and the most commonly used components of the system are presented. The trends of various machine learning techniques and the emphasis on certain dependent parameters are identified. The only way forward is to ensure that an affordable and robust solution for crop yield prediction is reaching the farmers at all levels in the agricultural sector.

Keywords— IoT analytics, IoT system, Machine Learning, Multi-Layer Perception, Random Forest

I. INTRODUCTION

In India, there are more than a hundred crops planted around the whole country. As reported by Kapil [17], agricultural share in GDP hit 20% after 17 years in 2020-2021. India's rice production is at a record 102.36 million tons in the Kharif season of 2020-2021 [18]. However, there exists a high risk of vulnerability as reported by [19] on the yield because of climactic factors. Another major problem is pointed out by Puja Mondal [20] because of small and fragmented land holding due to the inheritance laws of our nation.

Even though IoT Analytics and Decision Support Systems (DSS) have greatly improved the prediction of crop yield, the research challenges remain. The benefits of IoT systems in agriculture are manifold.

The contributions of this review work are as follows:

- Examine existing IoT systems for crop yield prediction
- Determine dependent parameters that affect crop yield prediction from literature.
- Identify techniques that provide the best results for crop yield prediction.

In the following section, a review of IoT systems and data analytics for agriculture is examined.

II. IOT SYSTEMS FOR CROP YIELD PREDICTION

A typical IoT system for agriculture requires sensors for sensing various parameters affecting crop yield, a microcontroller to control the actions of the sensors, a cloud to store the data, and a data analytics system to process the information and predict or offer decision support.

Sushanth et al [21] proposed an IoT system using Arduino Uno R3 microcontroller board with LM35 temperature sensor and also sensors for humidity and moisture. The IoT gateway then transmits the data to the cloud for processing. Prathiba et al [22] use CC3200, a single chip with an integrated microcontroller, network processor, and Wi-Fi with temperature sensor TMP007 and humidity sensor, HCD1010 to set up a system for monitoring farmland.

Lee et al [23] use an IoT gateway with temperature, humidity, soil electrical conductivity (EC), and pH sensors attached. Their system runs an analytics decision support at the back end to help to make decisions and they report an increase in the yield of apple and tomato yield.

Garcia et al [24] report that

- YL69 is the most popular sensor to monitor soil moisture.
    - Its output voltage goes from 3.3V to 5V.
    - Outputs 0-300 for dry soil and 300-700 for humid soil.
- LDR is the most used luminosity sensor
- DHT11 is the most popular sensor for humidity and temperature.
    - DHT11 has a range of measuring 20% to 95% humidity and a temperature of 0°C to 50°C
- Arduino Uno is the most popular node in IoT systems for agriculture.
- WiFi is the most utilized communication system
- ESP82666 is the most utilized communication module
- The database and cloud are mostly varied but MySQL and ThingSpeak are most popular.

The various components highlighted by Garcia et al [24] are shown in Table I.

III. MACHINE LEARNING TECHNIQUES

Machine learning techniques use the dependent variables that affect crop yield. Using various clustering approaches or models, a prediction model can predict the crop yield with a certain accuracy. The various techniques used include Linear Regression, Logistic regression, Random Forest and decision tree, K Nearest Neighbour (KNN), Artificial Neural Network

(ANN), and Multilayer Perceptron (MLP) neural network. Autoregression models are also used. The following discussion reports the effectiveness of various clustering and learning models.

TABLE I. COMPONENTS OF THE PHYSICAL AN IOT SYSTEM

Type of Sensor	***Working Voltage***	***Specifications***
Soil Moisture Sensor YL69 	3.3V-5V	Analog and digital output
Luminosity Sensor LDR 	The maximum voltage at 0 lux: 200V	Peak wavelength: 600nm
Humidity and Temperature Sensor DHT11 	3.3V-5V	Temperature: 0°C to 50°C Humidity: 20% to 95%
Microcontroller Arduino Uno 		14 digital I/O pins 6 analog inputs, 16MHz quartz crystal USB connection power jack ICSP header reset button
Communication Module ESP82666 		Micro USB for power, programming, and debugging 15-pin header

Neural Network-based techniques

Kadir et al [1] predict wheat yield prediction using Multi-Layer Perceptron (MLP) backpropagation based-feedforward artificial neural networks (ANN). Many parameters such as temperature, frost, and rain are passed through neurons. The model predicts with an accuracy of 98%. Gandhi et al [2] use precipitation, minimum temperature, maximum temperature, area, production, and yield for the Kharif season (June to November) for rice production in various districts in Maharashtra. A Multilayer Perceptron Neural Network (MLP) performs with an accuracy of 97.5%. Another work by Gandhi et al [3] report that when MLP was compared with another classifier, Sequential Minimal Optimization (SMO), it performed much better. Ahamad et al [8] have considered rainfall, temperature, and soil attributes such as pH, salinity, and area of production to predict rice yield in Bangladesh. They have used Linear Regression, KNN, and Neural networks. No one method produces the best results for all crops. ANN performs better for some crops that have more missing values. Linear regression performs better for most crops. They also report that their training set was small.

Islam et al [6] used soil type, type of land, fertilizer, rainfall, temperature, and humidity information about various agriculture zones in Bangladesh to predict the crop yield. They concluded that deep neural network performs better than logistic regression and SVM and Random Forest with an accuracy of over 94% for all the crops considered.

Random Forest and Decision Tree

Champaneri et al [15] use a Random Forest classifier on crop yield prediction based on temperature, rainfall, and area. They report a prediction accuracy of over 75%. Kumar et al [10] use parameters such as pH, temperature, humidity, rainfall, and crop name to predict crop yield using Random Forest, and Decision Tree. Their results help farmers decide on which crops would give the greatest yield and hence the highest profit. Jeong et al [14] use Random Forest and Multiple Linear Regression (MLR) on global wheat grain yield and, US maize grain yield, North-eastern Seaboard region potato and maize yield. RF gave a better prediction than MLR in all instances. Kumar et al [10] use parameters such as pH, temperature, humidity, rainfall, and crop name to predict crop yield using Random Forest, and Decision Tree. Their results help farmers decide on which crops would give the greatest yield and hence the highest profit.

K-Means clustering techniques

Awan et al [4] use Weighted Kernel K-Means with Spatial Constraints (SWK-Means) for predicting oil-palm yield by analyzing rainfall values. The model can also be used for other parameters such as precipitation, temperature, pressure in different areas, and their effect on oil palm production. Ananthara et al [16] use region, pH, sunlight, soil type and use a beehive clustering algorithm to predict crop yield. The proposed algorithm overcomes shortcomings of K-Means clustering and performs with an accuracy of over 84% for all years for paddy and sugarcane yield.

Autoregression Models

Bang et al [5] use three methods, Auto Regressive Moving Average (ARMA), Seasonal Auto-Regressive Integrated Moving Average (SARIMA), and ARMA with exogenous variables (ARMAX). They take the temperature and rainfall data, use it to predict the temperature and rainfall of the Kharif and Rabi season and in turn use the crop yield dataset to predict the crop yield. They report that temperature is predicted accurately by the SARIMA model.

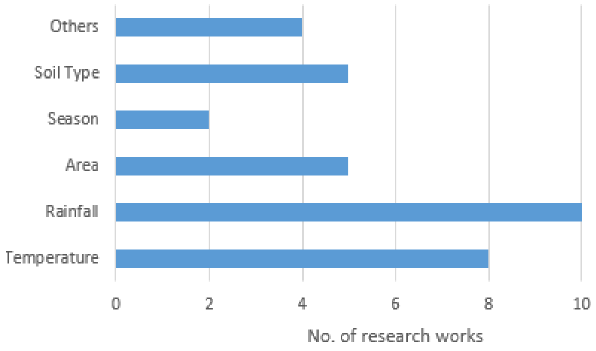
Tools used

Bhosale et al [7] use data visualization tool, Tableau. Paul et al [9] use RapidMiner 5.3 to run KNN and Naïve Bayes classifiers on a soil dataset that was taken from Madya Pradesh. Gandhi et al [11] used Netbeans IDE and Microsoft access database for developing a Decision Support System(DSS).

Fig. 1. Type of parameters used in various research works

IV. DISCUSSION

From the survey done, research done on data set from various parts of India [2,3,5,15,16], Bangladesh [6,8], Malaysia [1,4],



U.S. [14], and worldwide data [14] is considered. Some data sets were collected from labs in specific areas such as Madhya Pradesh [11], Allahabad [5] whereas others are those provided by government organizations [6,14].

TABLE II. COMPARISON OF VARIOUS PARAMETERS

Author	*Temperature*	*Rainfall*	*Area*	*Season*	*Soil Type*	*Others*
Kadir et al[1]		✓				
Gandhi et al [2]	✓	✓	✓	✓		
Gandhi et al [3]	✓	✓	✓	✓		
Awan et al [4]	✓	✓				pressure
Bang et al [5]	✓	✓				
Islam et al [6]	✓	✓			✓	humidity, fertilizer, type of land
Bhosale et al [7]		✓	✓		✓	
Ahamad et al [8]			✓			salinity
Paul et al [9]					✓	
Kumar et al [10]	✓	✓				
Jeong et al [14]	✓	✓			✓	Radiation
Champaneri et al [15]			✓			
Ananthara et al [16]	✓	✓			✓	

The crop yield prediction of major interest is rice [2,3,6,8,16] and wheat [1,5,6,14] and instance of sugarcane [7,16], jute [6], oil palm [4], potato [6,14] and maize [14].

The parameters that affect crop yield that is used for prediction are predominantly temperature and rainfall [1-6,10,16] as can be seen in Table II and also Figure. 1. In India, the Kharif and Rabi seasons are of particular interest due to the rice yield predictions in those seasons [2,3].

The soil type is used by some researchers to predict yield. Ananthara et al [16] classify it as black soil, clay soil, and alluvial soil. Paul et al [9] examine the nutrients and micronutrients of the soil such as organic carbon, nitrogen, potassium, sulfur, zinc, and iron. In certain regions, the land type is of importance such as in [6] where land is classified as highlands, medium highland, medium low land, low land, very low land, and miscellaneous land. Jeong et al [14] use additional parameters such as seasonal radiation, irrigation, clay content, and others to predict. Pressure and other climactic factors are used by [4] and classified.

The popular classifiers (Table III) are Naïve Bayes and KNN [7-9] are used to classify a set of climatic and agronomic parameters to predict yield or the best crop to grow in such circumstances to get maximum profit. LR or RF [8,10,14,15] to obtain a good prediction of a particular crop type given a set of dependent variables. ANN gives the best prediction with 98% prediction accuracy for the wheat [1]. The visual representation of the summary of various classifiers is summarized in Figure 2.

TABLE III. COMPARISON OF VARIOUS TECHNIQUES

Author	*Naïve Bayes*	*KNN*	*ANN*	*LR/RF*	*SVM*	*Others*
Kadir et al[1]			✓			
Gandhi et al [2]			✓			
Gandhi et al [3]			✓		✓	SMO
Awan et al [4]						SWK-Means
Bang et al [5]						ARMA, SARIMA, ARMAX
Islam et al [6]					✓	Logistic Regression
Bhosale et al [7]	✓	✓				
Ahamad et al [8]		✓	✓	✓		
Paul et al [9]	✓	✓				
Kumar et al [10]				✓		Decision Tree
Jeong et al [14]				✓		MLR
Champaneri et al [15]				✓		
Ananthara et al [16]						Bee Hive Clustering Algorithm

V. RESEARCH CHALLENGES AND RESEARCH DIRECTIONS

IoT Systems and ML techniques for DSS in crop yield have a tremendous positive impact in the agriculture sector. Elijah et al [24] indicate the impact of IoT systems on operational efficiency wherein the data can be put to use to optimize and plan out agricultural interventions. Although the benefits are manifold, research challenges remain. The following are some research challenges and possible future research directions stemming from the review of IoT systems and ML techniques.

Scientists struggle to channel the research findings and convey them in a timely, relevant, and user-friendly manner to the stakeholders. Some research has reached the farmers in the form of a mobile app [11,13,26,27]. Developing **user-friendly apps** and the maintenance thereof still poses a formidable challenge.

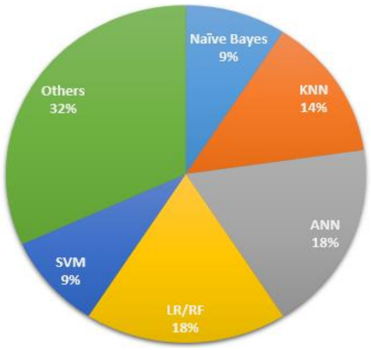


Fig. 2 Summary of various classifiers

A challenge in machine learning techniques is that no single classifier works best for all types of crops. Even though MLP provides the best results, there is a need to use bio-inspired approaches or **develop other meta-heuristics** to effectively predict the crop yield of various crops. Gangde [12] point out that more effective classifiers are required to predict yield with greater accuracy.

Developing **cost-effective** IoT solutions for farmers with fragmented landholding is another research challenge. The cost factor is not delved into in most research work. The cost involves the cost for deployment and maintenance of equipment and usage of the cloud.

There is a need to develop IoT systems and DSS for **post-drought or flood scenarios.** In most research work, the parameters considered for predicting crop yield are stable climactic factors or the soil type depending on the region. More research is required to consider adverse soil and climate conditions following a natural disaster.

There are very few research works that take into account the **season as a dependent variable** in predicting crop yield. Farmers understand seasons and it will be valuable to develop a solution as that proposed by Incant et al [26] to give solutions to farmers that they can use and understand easily.

VI. Conclusion

The IoT system and the various prediction models that have been used both nationwide and worldwide for predicting crop yield in agriculture are presented in this paper. Though the technological innovations in IoT and machine learning algorithms have helped us make strides in giving the agricultural sector a competitive edge, the technical, as well as the social-economic factors, have limited the reach of such technology to medium or small-scale farmers.
The awareness of the existence of such systems has increased in recent years but recommendations from such systems are not reaching the farmers at an affordable cost. It is expected that researchers will continue working on cost-effective solutions and develop user-friendly apps to support the agricultural sector.

References

[1] K. Muhd Khairulzaman Abdul, M. Zaki Ayob, and Nadaraj Miniappan, "Wheat yield prediction: Artificial neural network based approach," *in 4th International Conference on Engineering Technology and Technopreneuship (ICE2T). IEEE*, 2014.

[2] G. Niketa, P. Owaiz, and A. Leisa J.. "Rice crop yield prediction using artificial neural networks," *IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*,2016.

[3] G. Niketa, A. Leisa J., P. Owaiz, and T. Amiya Kumar, "Rice crop yield prediction in India using support vector machines," in *13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 1-5., 2016.

[4] A. A. Majid, and S. Mohd Noor Md, "An intelligent system based on kernel methods for crop yield prediction," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining,* pp. 841-846. Springer, Berlin, Heidelberg, 2006.

[5] B. Shivam, Rajat Bishnoi, C. Ankit Singh, D. Akshay Kumar, and C. Indu, "Fuzzy logic based crop yield prediction using temperature and rainfall parameters predicted through ARMA, SARIMA, and ARMAX models," in *2019 Twelfth International Conference on Contemporary Computing (IC3),* pp. 1-6. IEEE, 2019.

[6] I. Tanhim, C. Tanjir Alam, and C. Amitabha, "A deep neural network approach for crop selection and yield prediction in Bangladesh," in *IEEE Region 10 Humanitarian Technology Conference (R10-HTC),* pp. 1-6. IEEE, 2018.

[7] Bhosale, Shreya V., Ruchita A. Thombare, Prasanna G. Dhemey, and Anagha N. Chaudhari. "Crop yield prediction using data analytics and hybrid approach," in *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),* pp. 1-5. IEEE, 2018.

[8] Ahamed, AT M. Shakil, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, and Rashedur M. Rahman, "Applying data mining techniques to predict the annual yield of major crops and recommend planting different crops in different districts in Bangladesh," in *IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD),* pp. 1-6. IEEE, 2015.

[9] P. Monali, V. Santosh K. Vishwakarma, and V. Ashok, "Analysis of soil behaviour and prediction of crop yield using data mining approach," in *International Conference on Computational Intelligence and Communication Networks (CICN),* pp. 766-771. IEEE, 2015.

[10] Y. Kumar, Y. Jeevan Nagendra, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi. "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector, in *5th International Conference on Communication and Electronics Systems (ICCES),* pp. 736-741. IEEE, 2020.

[11] G. Niketa, A. Leisa J., and P. Owaiz, "Proposed decision support system (DSS) for Indian rice crop yield prediction," in *IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR),* pp. 13-18. IEEE, 2016.

[12] G. Yogesh. "A study on various data mining techniques for crop yield prediction," in *International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT),* pp. 420-423. IEEE, 2017.

[13] A. Manjula and G. Narsimha. "XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction," in *IEEE 9th International Conference on Intelligent Systems and Control (ISCO),* pp. 1-5. IEEE, 2015.

[14] J. Jig Han, R. Jonathan P., M. Nathaniel D., F. David H., Y. Kyungdahm, B. Ethan E., and T. Dennis J, "Random forests for global and regional crop yield predictions." *PLoS One 11*, no. 6 (2016): e0156571.

[15] C. Mayank, & C. Chaitanya., C. Darpan, R. Mansing, "Crop Yield Prediction using Machine Learning," *International Journal of Science and Research (IJSR),* April 2020.

[16] M. Ananthara, Gunasundari, T. Arunkumar, and R. Hemavathy. "CRY—an improved crop yield prediction model using bee hive

clustering approach for agricultural data sets," in *Proc. of International Conference on Pattern Recognition, Informatics and Mobile Engineering*, pp. 473-478. 2013.

[17] S. Kapil, (2021, January), Agri Share in GDP hit 20% after 17 years: Economic Survey, https://www.downtoearth.org.in/news/agriculture/agri-share-in-gdp-hit-20-after-17-years-economic-survey-75271, Accessed on 4.03.2021

[18] Rice Production Estimated At Record 102.36 Million Tonnes In 2020-21 Kharif Season: Government, https://www.bloombergquint.com/business/rice-production-estimated-at-record-102-36-mln-tonnes-in-2020-21-kharif-season-govt, Accessed On 3.03.2021

[19] S. Tarul, H. Vittal, K. Subhankar, and G. Subimal Ghosh, "Increasing agricultural risk to hydro-climatic extremes in India," *Environmental Research Letters* 15, no. 3, 2020.

[20] M. Puja., 10 Major Agricultural problems of India and their possible solutions. https://www.yourarticlelibrary.com/agriculture/10-major-agricultural-problems-of-india-and-their-possible-solutions/20988 Accessed on 2.03.2021

[21] G. Sushanth, and S. Sujatha, "IoT based smart agriculture system," in *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 1-4. IEEE, 2018.

[22] S. Prathiba, H. Anupama Mongol, and M. P. Jyothi. "IoT based monitoring system in smart agriculture," in *International conference on recent advances in electronics and communication technology (ICRAECT)*, pp. 81-84. IEEE, 2017.

[23] L. Meonghun, H. Jeonghwan, and Y. Hyun, "Agricultural production system based on IoT," in *IEEE 16Th international conference on computational science and engineering*, pp. 833-837. IEEE, 2013.

[24] G. Laura, P. Lorena, J. Jose M., L. Jaime, and L. Pascal, "IoT-based smart irrigation systems: An overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture," *Sensors* 20, no. 4, 1042, 2020.

[25] O. Elijah, R. Tharek Abdul, O. Igbafe, L. Chee Yen, and MHD Nour Hindia. "An overview of Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges," *IEEE Internet of Things Journal* 5, no. 5, 3758-3773, 2018.

[26] N. Potnuru Sai, V. Pinapa Sai, B. Lakshmi Avinash, and B. Jabber. "Crop Yield Prediction based on Indian Agriculture using Machine Learning," in *International Conference for Emerging Technology (INCET)*, pp. 1-4. IEEE, 2020.

[27] S. Geetha, P. Deepalakshmi and S. Pande, "Managing Crop for Indian Farming Using IoT," in *IEEE International Conference on Clean Energy and Energy Efficient Electronics Circuit for Sustainable Development (INCCES)*, 1-5, 2019.