International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Crop Recommendation using XG Boost Algorithm for Sustainable Agrarian Application

Durai Arumugam S. S. L.[1], Praveen Kumar R.[2], Mahadevan B.[3], Akash V.[4]

Submitted: 27/01/2024 **Revised**: 05/03/2024 **Accepted**: 13/03/2024

Abstract: India is primarily an agrarian nation, where agriculture holds significant sway over both the country's economy and the livelihoods of its population. Most of the Indian farmers face a common challenge of not selecting the most appropriate crop for their land in accordance with the environmental requirements. As a result, they will see a major decline in their total level of productivity. So to overcome this problem crops are suggested based on soil and crop pattern, weather information which includes humidity, rainfall, and temperature which plays a vital role in attaining quantity and quality of crops massively. So our paper's aim is to analyse all these factors and match it with the required parameters of each crops and provide the best possible option to the farmers. This leads to a decrease in crop selection errors and an increase in yield. To achieve our aim we will use a recommendation system using an Extreme Gradient (XG) Boost and Random Forest, Decision tree and CNN algorithms to Suggesting a suitable crop based on the specific site parameters involves considering the unique conditions and characteristics of the area.

Keywords: *CNN, Crop pattern, Decision tree, Extreme Gradient (XG) Boost, Random Forest, Recommendation System, Weather information.*

1. Introduction

In the modern world, farmers are facing a major dilemma on how much to invest on certain specific crops on specific fields. Our paper provides a unique solution and an insight and resolution for that problem by creating a recommendation model using machine learning**[26]** which can predict the yield of specific crop and give good ideas and suggestions to the end users who are the farmers. This initiative aims to enhance the efficiency of the agricultural output. By offering insights into the cultivation prospects of specific crops, this application empowers farmers and stakeholders to make informed decisions about their agricultural practices. This endeavor represents a significant step towards improving overall agricultural productivity and sustainability.

Our model will predict the output or result based on the input data which can be extracted from the past years. The Core aim of the paper is to design a model with machine learning which can be used to get the information about the weather, soil, other geographical factors and so on of the selected Field under study which will be suitable for the cultivation of the individual crop. The envisioned model is set to be constructed employing XG Boost and Random Forest machine learning algorithms. These sophisticated algorithms will enable the model to generate highly accurate predictions by meticulously analyzing input data.

The predictive outcomes will be seamlessly presented to users through a web interface, implemented using the Python Django web framework. This integration of machine learning and web development ensures a robust and user-friendly platform, where individuals can readily access and interpret the predicted results for informed decision-making in agriculture**[29]**.

2. Literature Survey

1. Gupta and Sharma presented a model for collecting and analyzing environmental and agricultural data to enhance crop yield for farmers. The process involves data preprocessing in Python, utilizing the MapReduce framework for large-scale analysis, and applying k-means clustering for accuracy. The study explores correlations between crops, weather information and soil data in two regions using bar graphs and scatter plots. A self-designed recommender system predicts crops, displayed on a Flask-based Graphic User Interface. The scalable system can extend its crop recommendations to other states in the future.[1]

2. Dhivya Elavarasan and P. M. Durairaj Vincent proposed research on crop yield prediction utilizes deep learning, facing challenges like indirect data-to-yield mapping and featurequality dependence. The proposed system achieves superior accuracy (93.7%) by

[1] *Assistant Professor, Department of Information Technology, Easwari Engineering College, Ramapuram, Chennai,, Tamil Nadu. Mail ID: duraiarumugam36@gmail.com*
[2] *Assistant Professor, Department of Electronics and Communication Engineerng , Easwari Engineering College, Ramapuram, Chennai,, Tamil Nadu. Mail ID: rpjcspraveen@gmail.com*
[3] *Student, Department of Information Technology, Easwari Engineering College, Ramapuram, Chennai,, Tamil Nadu. Mail ID: haricool0808@gmail.com*
[4] *Student, Department of Information Technology, Easwari Engineering College, Ramapuram, Chennai, Tamil Nadu. Mail ID: akash65190@gmail.com*

preserving data distribution.[2]

3. Reddy, Dadore and Watekar proposed a model that considers soil pattern and yield data of a crop to suggest suitable crops, enhancing overall productivity. This method offers benefits such as optimal resource utilization and improved decision-making in farming. This system ensures high accuracy and efficiency in suggesting suitable crops based on soil pattern. [14]
4. Charvat, Reznik and Lukas presented a model that allows farmers to grapple with diverse data—local weather, GNSS, orthophotos—essential for quality and yield. Securing Big Data is crucial, but competing storage and visualization needs pose challenges. This paper stresses trust in data security as farmers nagivate complexities in handling (Big) data.[9]
5. Nishant, Venkat and Avinash proposed a model that leverages basic data such as location and season allowing users to forecasting crop yield for a particular year. It employs regression techniques prediction. [15]
6. Shah, Hiremath and Chaudhary presented a data management system leveraging Apache Spark for efficient and scalable information processing, aiming to bridge the scientific divide between agrarian and Knowledge reserves. This model is designed to streamline data management, fostering more accessible and user-friendly interactions with agricultural information, thereby contributing to a more inclusive and efficient agricultural technology landscape.[10]
7. Rajeswari, Scholar, and Rajakumar proposed a system in which an IOT device is deployed for sensing agricultural data, transmitting it to a Cloud based database. Cloud-powered Big Data insight is employed to examine various aspects such as fertilizer needs, crop analysis, market trends, and stock requirements. Predictions are generated using data mining techniques, and the insights are conveyed to farmers through a mobile app, facilitating informed decision-making in agriculture.[8]
8. Doshi, Nadkarni, and Shah presents an Agro Consultant, an learning system designed to help Indian agrarian in decision-making regarding crop selection. This system takes into account crucial factors like sowing season, land details, soil pattern, and ecological elements. By synthesizing this information, Agro Consultant aims to guide farmers towards optimal crop choices, contributing to improved agricultural outcomes.[13]

3. Existing System

Before starting the steps for the proposed system, we just analyze two existing systems which more or less gives the similar kind of output. The first system is **[1]** WB-CPI: Weather based crop prediction in India using big data analytics, which uses weather API where it gets various inputs regarding the weather such as temperature, humidity and rainfall in the particular area along with the crop selection and then based upon these inputs it just determines whether the particular crop which has been selected can be cultivated in that particular area for that particular time or season. The second existing system is **[2]** Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications which uses land pattern. It gets various parameters from the land which are the nutrients such as potassium level, sodium level and other nutrients available in the sand and it also gets the historical agricultural data of the land, the past crops being grown in that particular land. Based upon these inputs and the selected crop, it just determines whether the particular crop can be grown on the particular land for the given time frame. Based on these two existing systems, we have come up with a new proposed system which gets all the inputs together and gives the required output.

4. Limitations in Existing System

Taking into consideration the two existing systems which we already discussed the first one the weather API which completely gets all the inputs regarding the weather and the crop selection to give the required output which is the yield efficiency. The second existing system which is our land pattern model gets all the inputs regarding the land the sand and its nutrients and the selected crop to give the required output. However, there may be a scenario like in the first model it does not take the sand into consideration which will reduce the efficiency of the output. Similar way in the second system which is the land pattern model it won't take the atmospheric conditions into consideration decreasing the efficiency of the output. In order to overcome these two problems or in order to overcome the problems in these two existing systems we are proposing a new system which takes all the inputs into consideration with respect to giving a proper output thus increasing the efficiency of our yield output which would be useful for the end users.

5. Problem Statement

In the modern world, due to global warming and various other environmental factors, major problem which the farmers face is to select a particular crop for a particular field to get the desired output which is the profit. So this is one of the major concerns why most of the farmers in our country are not earning properly due to incorrect crop selection for a particular field. So our paper provides a valid solution to this modern day problem by providing an efficient solution in giving the proper crop for the land selected thus maximizing the yield of cultivation

6. Proposed System: An Overview

The system that is being proposed is built keeping in mind the disadvantages of the existing systems that are already available in the market. Our system requires various inputs such as the atmospheric, natural parameters such as the temperature, humidity, rainfall in the particular area and the area under question and the historical cultivation data of the particular land. So if there is any previous crop production in the land and in the is already done in the land under question then our system predicts that and gives the most suitable crop that can be cultivated on that land after the previous crop such that we can get maximum profit. If there is no historical cultivation in the land and the land is a new one and just based upon the atmospheric natural parameters and the soil parameter the system provides the best possible selection of crop for the particular land for the particular season to acquire maximum profit.

7. Methodology

Let us now begin with the working principle of our proposed system. The system which is being proposed get mainly two major inputs, one the atmospheric parameters, the other the land parameters. The atmospheric parameter includes the temperature, the humidity, average rainfall in the given area.

The land parameter consists of various nutrients level available in the land, the soil fertility and also the historical cultivation data for the given selected land.

Once all the inputs that are being collected are pre-processed and analyzed, the system then processes all the inputs for the given crop and for the given land. The system provides the output in the form of desired crop that can be grown on the selected field taking into account the various atmospheric factors and soil pattern as the end output for the user.

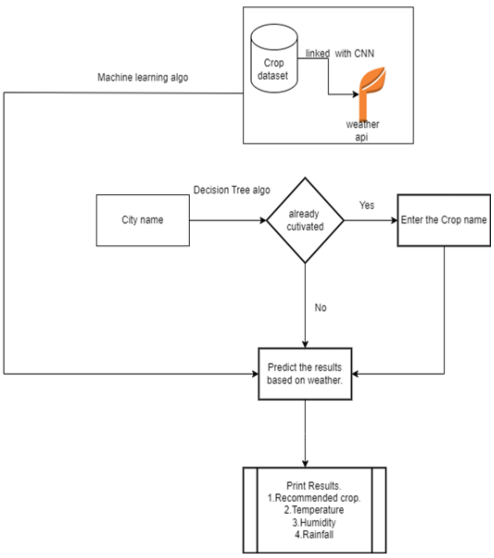

Fig 1. Overall Architecture of the Model

7.1. Data Collection and Pre-Processing

Identify relevant data sources (soil type, climate, temperature, rainfall, historical crop yield). Cleanse data by handling missing values and outliers. Normalize or standardize numerical features. Encode categorical variables for machine learning compatibility**[28]**.

7.1.1. Identify Relevant Data Sources:

i. Soil data:

Once the field is selected Soil API fetches all the relevant details of the land. These details involve the soil nutrient levels such as potassium, sodium, urea, ph level and so on.

ii. Weather data:

The Weather API incorporated in our system collects various input parameters such as Temperature, Humidity level, average Rainfall once

the field is selected for cultivation analysis. These inputs plays a major role in determining the crop to be cultivated on that particular field.

iii. Historical Crop Yield:

The Historical crop cultivation data for the selected field is fed as an input by the user to the system. This input plays the major role in helping the soil API to fetch soil nutrient level with accuracy.

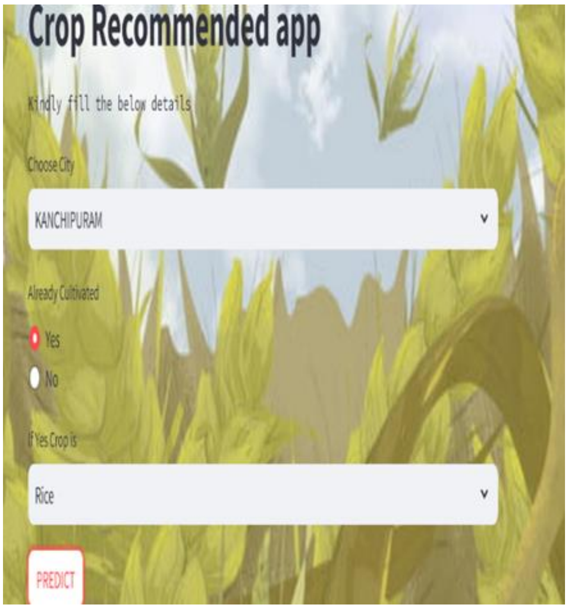

Fig 2. shows the user interface screen where the end user can enter inputs for Processing.

7.1.2. Data Cleaning:

i. Handling Missing Values:

Identify and assess missing values in each dataset.

ii. Depending on the amount of missing data:

Remove rows with missing values. Impute missing values using methods like mean, median, or machine learning-based imputation.

iii. Handling Outliers:

Detect and analyze outliers using statistical methods or visualization tools. Decide whether to remove outliers or transform them based on domain knowledge.

7.2. GIS (Geographical Information System)

MAPPING: Utilize GIS to map the geographical features of the farm, including topography and elevation.

SPATIAL ANALYSIS: Analyze spatial data to understand how geographical factors affect crop suitability.

7.3. Weather Integration

Real-time Weather Updates: Integrate APIs or services to provide real-time weather updates which includes Temperature, humidity and average rainfall for accurate decision-making.

7.4. XG Boost Model Training and Hyperparameter Tuning

i. Split the dataset into testing sets and training sets.

ii. Choose XG Boost algorithm for its advantages in handling complex relationships.

iii. Fine-tune hyperparameters (learning rate, tree depth, regularization) through cross-validation.

$$f(x) \approx y(t-1) \longrightarrow \quad (1)$$

This represents the prediction from the previous iteration (t-1), denoted as y(t-1).

$$Z(q) \approx \Sigma\ \chi(x)^2 * (\ y(t-1) - f(x))^2 + \lambda * T * sum\ (\textstyle\sum \chi^2(x))^2 \longrightarrow \quad (2)$$

This formula represents the objective function that XG boost aims to minimize during training.

Using (**1**) and (**2**)

$$g(x) \approx y(t-1) + T * sum\ (\Sigma\ \chi(x)^2 * (\ y(t-1) - f(x))\ )^2)$$

This formula represents the update to the prediction for the current iteration.

T is a hyperparameter that controls the step size of the update.

These formulas describe the prediction, regularization, and update steps in the XG Boost algorithm. The algorithm seeks to find the optimal combination of weak learners by minimizing the objective function, which includes terms for prediction accuracy and regularization**[27]**.

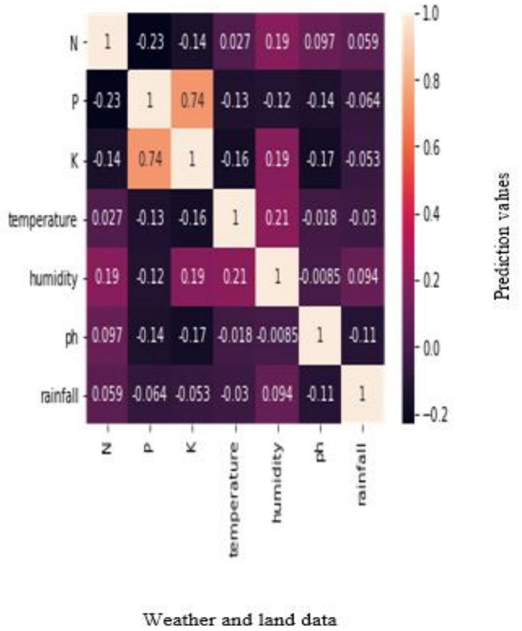

Fig 3. Shows how XG Boost model is trained with the collected data that is Temperature, Rainfall, Humidity and land details.

8. Result and Discussion

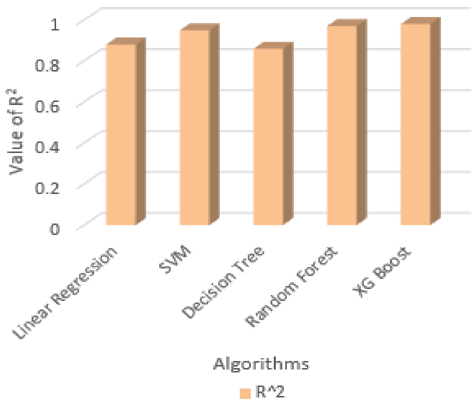

Fig 4. Shows the comparison of R^2 for XG Boost algorithm with other algorithms.

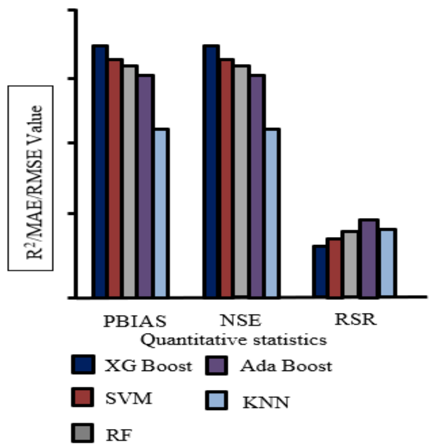

Fig 5. Shows the comparison of three quantitative statistics for XG Boost algorithm with other algorithms.

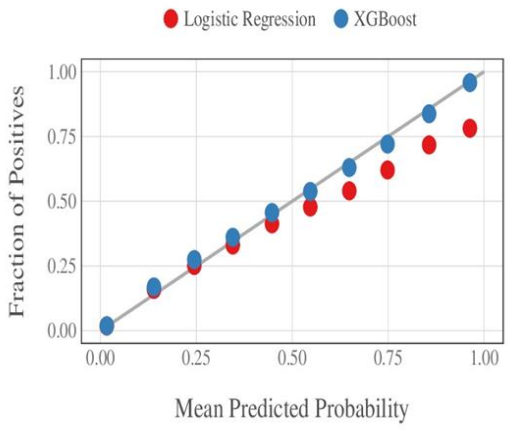

Fig 6. Shows the graph between Fraction of Positives and Mean Predicted Probability for XG Boost and Logistic Regression algorithms.

Fig 7. Shows the user interface showing the final output to the end user (Farmer).

9. Conclusion

This paper provides a commendable endeavor aimed at addressing the challenge of cost loss through the application of a machine learning recommendation model. The efficiency of this model hinges on the quality of the data it processes, with crucial inputs derived from historical crop production information. This dataset encompasses key factors such as the production area of the crop and weather information, playing a pivotal role in ensuring the accuracy of the predictive outcomes. To optimize the dataset, the removal of irrelevant data and the creation of a CSV file are essential steps in the computation process. While the current paper focuses on specific parameters for productivity evaluation, there is potential for future enhancements and expansions. In conclusion, the importance of initiatives like these in supporting agriculture, the backbone of our country's economy, cannot be overstated, and the encouragement of more such projects is crucial for sustained growth.

One of the major future enhancements that can be made to the proposed system is weather forecasting. To be more precise, the crop that is being recommended, we can set a specific cultivation time for the crop, such that the forecasted weather for the given duration can be predicted This also contributes as a major input for the crop selection. This might be pretty difficult given the current technology we have today, but in future this can also be added to the proposed system to make the output of the system much more reliable and efficient.

References

[1] Gupta, R., Sharma, A. K., Garg, O., Modi, K., Kasim, S., Baharum, Z., ... & Mostafa, S. A. (2021) “WB-CPI: Weather based crop prediction in India using big data analytics,” in IEEE access, 9, 137869-137885.

[2] Dhivya Elavarasan and P. M. Durairaj Vincent (2020) “Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications,” in IEEE access, 8, 86886-86901.

[3] R. Priya, D. Ramesh, and E. Khosla, ‘‘Crop prediction on the region belts of india: A Naïve Bayes MapReduce precision agricultural model,’’ in Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI), Sep. 2018, pp. 99–104.

[4] W. Fan, C. Chong, G. Xiaoling, Y. Hua, and W. Juyun, ‘‘Prediction of crop yield using big data,’’ in Proc. 8th Int. Symp. Comput. Intell. Design (ISCID), Dec. 2015, pp. 255–260, doi: 10.1109/ISCID.2015.191.

[5] M. Ramya, C. Balaji, and L. Girish, ‘‘Environment change prediction to adapt climate-smart agriculture using big data,’’ Int. J. Adv. Res. Comput. Eng. Technol., vol. 4, no. 5, pp. 1995–2000, 2015.

[6] P. C. Reddy and A. S. D. Babu, ‘‘Survey on weather prediction using big data analystics,’’ in Proc. 2nd IEEE Int. Conf. Electr., Comput. Commun. Technol., Feb. 2017, pp. 1–6.

[7] P. S. Vijayabaskar, R. Sreemathi, and E. Keertanaa, ‘‘Crop prediction using predictive analytics,’’ in Proc. Int. Conf. Comput. Power, Energy Inf. Commun. (ICCPEIC), Mar. 2017, pp. 370–373.

[8] S. Rajeswari, R. Scholar, K. Suthendran, and K. Rajakumar, ‘‘A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics,’’ in Proc. Int. Conf. Intell. Comput. Control, 2017, pp. 1–5.

[9] K. Charvat, K. C. Junior, T. Reznik, V. Lukas, K. Jedlicka, R. Palma, and R. Berzins, ‘‘Advanced visualisation of big data for agriculture as part of databio development,’’ in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Jul. 2018, pp. 415–418.

[10] P. Shah, D. Hiremath, and S. Chaudhary, ‘‘Towards development of spark based agricultural information system including geo-spatial data,’’ in Proc. IEEE Int.

Conf. Big Data (Big Data), Dec. 2017, pp. 3476–3481.

[11] M. P. Darji, V. K. Dabhi, and H. B. Prajapati, ‘‘Rainfall forecasting using neural network: A survey,’’ in Proc. Int. Conf. Adv. Comput. Eng. Appl., Mar. 2015, pp. 706–713, doi: 10.1109/ICACEA.2015.7164782.

[12] M. Motlhabi, P. Pantsi, and R. Netshiya, ‘‘A machine learning deepdive analysis into network logs,’’ in Proc. ICCWS, 2021, p. 213, doi: 10.34190/IWS.21.019.

[13] Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018, August). AgroConsultant: intelligent crop recommendation system using machine learning algorithms. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-6). IEEE

[14] Reddy, D.A.; Dadore, B.; Watekar, A. Crop Recommendation System to Maximize Crop Yield in Ramtek region using MachineLearning. Int. J. Sci. Res. Sci. Technol. 2019,6, 485–489.

[15] P. S. Nishant, P. Sai Venkat, B. L. Avinash, and B. Jabber, ”Crop Yield Prediction based on Indian Agriculture using Machine Learning,” 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-4, DOI: 10.1109/INCET49848.2020.9154036.

[16] G. E. O. Ogutu, W. H. P. Franssen, I. Supit, P. Omondi, and R. W. Hutjes, ‘‘Probabilistic maize yield prediction over East Africa using dynamic ensemble seasonal climate forecasts,’’ Agricult. Forest Meteorol., vols. 250–251, pp. 243–261, Mar. 2018.

[17] K. Ryan, P. Agrawal, and S. Franklin, ‘‘The pattern theory of self in artificial general intelligence: A theoretical framework for modeling self in biologically inspired cognitive architectures,’’ Cognit. Syst. Res., to be published.

[18] M. E. Holzman, F. Carmona, R. Rivas, and R. Niclòs, ‘‘Early assessment of crop yield from remotely sensed water stress and solar radiation data,’’ ISPRS J. Photogramm. Remote Sens., vol. 145, pp. 297–308, Nov. 2018.

[19] M. M. Rahman, D. Hagare, and B. Maheshwari, ‘‘Bayesian belief network analysis of soil salinity in a peri-urban agricultural field irrigated with recycled water,’’ Agricult. Water Manage., vol. 176, pp. 280–296, Oct. 2016.

[20] Y. Dash, S. K. Mishra, and B. K. Panigrahi, ‘‘Rainfall prediction for the Kerala state of India using artificial intelligence approaches,’’ Comput. Elect. Eng., vol. 70, pp. 66–73, Aug. 2018.

[21] P. S. Dutta and H. Tahbilder, ‘‘Prediction of rainfall using data mining technique over Assam,’’ Indian J. Comput. Sci. Eng., vol. 5, no. 2, pp. 85–90, Apr./May 2014

[22] W. Wieder, S. Shoop, L. Barna, T. Franz, and C. Finkenbiner, ‘‘Comparison of soil strength measurements of agricultural soils in Nebraska,’’ J. Terramech., vol. 77, pp. 31–48, Jun. 2018.

[23] Y. Cai, K. Guan, J. Peng, S. Wang, C. Seifert, B. Wardlow, and Z. Li, ‘‘A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach,’’ Remote Sens. Environ., vol. 210, pp. 35–47, Jun. 2018.

[24] A. Kaya, A. S. Keceli, C. Catal, H. Y. Yalic, H. Temucin, and B. Tekinerdogan, ‘‘Analysis of transfer learning for deep neural network based plant classification models,’’ Comput. Electron. Agricult., vol. 158, pp. 20–29, Mar. 2019.

[25] K. Abrougui, K. Gabsi, B. Mercatoris, C. Khemis, R. Amami, and S. Chehaibi, ‘‘Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR),’’ Soil Tillage Res., vol. 190, pp. 202–208, Jul. 2019.

[26] Praveenkumar R, Kirthika, Durai Arumugam and Dinesh (2023), “Hybridization of Machine Learning Techniques for WSN Optimal Cluster Head Selection”. IJEER 11(2), 426-433.

[27] Praveen Kumar, R, Jennifer S Raj & Smys, S 2022, “Performance Analysis of Hybrid Optimization Algorithm for Virtual Head Selection in Wireless Sensor Networks”, Wireless Personal Communications.

[28] Praveen Kumar, R, Jennifer S Raj & Smys, S 2021, “Analysis of dynamic topology wireless sensor network for the internet of things”, International Journal of Communication Systems, vol. 34, no. 17.

[29] SSL, D. A., Praveenkumar, R., & Balaji, V. (2023). An Intelligent Crop Recommendation System using Deep Learning. International Journal of Intelligent Systems and Applications in Engineering, 11(10s), 423-428.