RESEARCH ARTICLE

International Journal of Computer Communication and Informatics

DOI: 10.34256/ijcci2421

Crop Prediction using Supervised Learning

K. Kavitha [a,*], R. Senthil Kumar [b], S.R. Nishanthini [a]

[a] Department of Computer Applications, Dr.N.G.P. Arts and Science College, Coimbatore Tamil Nadu, India

[b] Department of Computer Science with Cognitive Science, Dr.N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India

* Corresponding Author: kavitha.k@drngpasc.ac.in

Received: 27-06-2024, Revised: 03-11-2024, Accepted: 15-11-2024, Published: 26-11-2024

Abstract: Agriculture is a key driver of a nation's economy, supplying raw materials, employment, and essential food. In India, the world's second most populated country, a significant portion of the population depends on agriculture for their livelihood. However, farmers face numerous challenges, including crop diseases, poor soil quality, unpredictable weather, and water scarcity. Additionally, repeated cultivation of the same crops and the indiscriminate use of fertilizers deplete soil nutrients, further reducing crop yields. Modern technology adoption may minimize these issues and improve the quality and production of agriculture. In agriculture, machine learning (ML), a branch of artificial intelligence (AI), makes automation, classification, and prediction easier. It supports well-informed decision-making for food security and effective crop management by assisting in the optimization of crop selection, fertilization, and irrigation. Using the Kaggle crop recommendation dataset, this paper presents a strong machine learning architecture for crop prediction. Important input characteristics including soil pH, temperature, humidity, and nutrient levels are included in the dataset. Using classification techniques like Decision Tree (DT) and Support Vector Machine (SVM), the system identifies the most appropriate crop for a specific type of soil based on its weather and soil data. It also offers information on the amount of fertilizer, seeds, and soil nutrients needed for production. By using this system, farmers can explore new crop varieties, increase agricultural productivity, enhance profit margins, and reduce soil pollution.

Keywords: Rainfall prediction, Crop recommendation, Decision tree, Support Vector Machine, Machine learning, Crop prediction

1. Introduction

Many Asian nations rely heavily on agriculture as their main source of income, especially India, where a sizable section of the populace makes their living from farming. Despite its importance, the sector faces numerous challenges, including inadequate quality control, low productivity, and financial instability among farmers. These issues often lead to dire consequences, such as overwhelming debt and, in extreme cases, farmer suicides [1-3]. The situation is further worsened by the unpredictability of changing environmental conditions. Additionally, farmers frequently cultivate the same crops repeatedly, avoiding new crop varieties, and apply fertilizers in arbitrary quantities without understanding soil nutrient deficiencies. This practice not only reduces crop yields but also causes soil pH reduction and damages the topsoil. All these factors drive towards the pressing need for innovative methods to improve agricultural efficiency and sustainability.

Advanced tools, such as statistical and mathematical models, offer promising approaches to address some of the critical challenges faced by agriculture. With the use of these data-driven insights, farmers are able to choose crops and farming methods with confidence. By forecasting crop growth, identifying plant illnesses, and automating crucial farming tasks like fertilization and irrigation, machine learning, enhances these efforts [4, 5]. However, challenges such as infrastructure limitations, limited data accessibility, and social acceptance continue to hinder the widespread adoption of these technologies in agriculture. Traditional methods often fail to address the complex and dynamic requirements of diverse agricultural ecosystems like those in India.

To tackle these challenges, we have developed a machine learning-based system to support farmers in making decisions that are more informed. Our methodology uses meteorological and soil characteristics to determine which crop is best suited for a certain plot of land. Additionally, it provides recommendations on the appropriate type and quantity of fertilizers, as well as the seeds required for cultivation. By leveraging this system, farmers can experiment with new crop varieties, potentially increase profit margins, and mitigate soil pollution, thereby improving both productivity and sustainability. The system is designed to be accessible and practical for farmers and policymakers alike.

The algorithm uses data from Kaggle to select crops depending on parameters including temperature, humidity, rainfall, and soil pH [6]. Farmers or sensors immediately enter temperature, humidity, and pH readings. To find patterns and provide predictions, this input data is processed using machine learning techniques like SVM and DT. Apart from proposing appropriate crops, the algorithm also suggests the amounts of nutrients required for the chosen crops. Additional capabilities include determining the necessary seed quantity (in kilos per acre), estimating yield (in quintals per acre), and delivering the crop's market price.

2. Literature Review

In order to predict agricultural yields, [6] employed Support Vector Regression (SVR), showcasing SVR's flexibility in dealing with non-linear correlations between variables. In order to raise farmer profitability and improve the quality of agricultural production, the author of paper [7] describes methods for predicting crop yields and suggests ideal crops. They have used big data, including soil and weather data, along with an agro algorithm in Hadoop platform to forecast agricultural yields. By using the data from the repository to determine the right crop for a particular condition, the crop quality also improved.

In Article [8], crop productivity and rainfall were estimated using a machine learning technique. A range of machine learning approaches were employed by the authors of this study to estimate rainfall and agricultural productivity. Additionally, they talked on the efficiency of certain ML algorithms, such as line regression, SVM, KNN, and decision trees. They conclude that, of all the algorithms they looked at, SVM has the highest rainfall prediction efficiency.

In the work [9], the author outlines the many machine-learning techniques that are employed to increase crop productivity. This article used a variety of AI techniques and big data analysis for precision agriculture. ML algorithms like neural networks, ensemble-based models, KNN, and other techniques were used to explain crop recommendations.

The paper [10] compares the effectiveness of SVM and RVM in forecasting rainfall. Both models demonstrate strong capabilities in handling complex weather data, with RVM offering additional advantages like sparse solutions and probabilistic outputs. While the study provides valuable insights into applying machine learning for meteorology, more details on data preprocessing and feature selection could enhance its utility. This work highlights the potential of AI-driven approaches for accurate weather prediction.

Article [11] used climate and remote sensing data to forecast agricultural yield using Decision Tree-based regression models. Their research emphasizes the importance of combining different data sources to enhance forecast accuracy.

Paper [12] focused on pest control which is essential for maximizing crop health and output which relates to our study. In their investigation of deep learning methods for predicting wheat production, showed how well neural networks comprehend intricate relationships seen in agricultural data.

Convolutional Neural Networks (CNNs) were used by [13] to estimate rice yield, highlighting the need of adding image-based data to yield prediction models.

Additionally, [14] explored the application of Long Short-Term Memory (LSTM) networks for spatial yield forecasting. The study contributes to advancements in agricultural production forecasting by emphasizing the effectiveness of recurrent neural networks in processing sequential data.

Understanding the variables influencing crop output is crucial for ensuring the efficacy of prediction models. New studies have shed important light on these variables. For example, [15] carried out an extensive investigation into the ways in which crop output is influenced by soil conditions, which is essential for choosing pertinent features for machine learning models.

In their investigation of how climatic factors affect agricultural production, [16] emphasized the need of weather information for precise yield forecasts. The integration of machine learning into agriculture has significantly enhanced precision farming, which focuses on optimizing agricultural inputs to achieve maximum productivity and sustainability.

Additionally, a number of research have suggested useful frameworks for machine learning-based intelligent crop management. Reference [17] developed a machine learning-based crop selection and optimization of yield system especially for small-scale farmers. A precision agriculture decision support system that combines crop, weather, and soil data to provide real-time suggestions was presented by [18].

There are still difficulties in using machine learning in agriculture, despite these developments. Reference [19] talked about the shortcomings of the machine learning models that are currently in use, especially with regard to data variety and model interpretability. Reference [20] discussed the infrastructural issues that farmers in poor countries confront and offered fixes to increase accessibility to machine learning technology

3. Proposed Methodology

The suggested method determines the most suitable crop for a specific soil based on soil composition and weather parameters like temperature, humidity, and soil pH. Additionally, it has the capability to predict rainfall. As illustrated in Figure (1), the architecture of the proposed system consists of multiple components.

3.1. Data Collection

Data is gathered from the 2023 rainfall dataset and crop recommendation datasets on Kaggle. Crop name, temperature, humidity, pH, rainfall, and NPK levels are among the attributes included in the crop suggestion dataset. Crop forecast takes these factors into account. We gather rainfall data from prior years in order to anticipate the annual rainfall. District, Taluk, Station, and Rainfall (mm) are among its features.

3.2. Data Preprocessing

Once the dataset is collected, it undergoes preprocessing before model training. Data preprocessing involves multiple stages, starting with reading the dataset followed by data cleaning. During the cleaning process, redundant attributes that are irrelevant to crop prediction are removed. Additionally, missing values in the dataset are either dropped or filled with appropriate values to improve accuracy. After cleaning, the target variable for the model is defined. Finally, the dataset is partitioned into training and test datasets using the sklearn library, preparing it for model development.

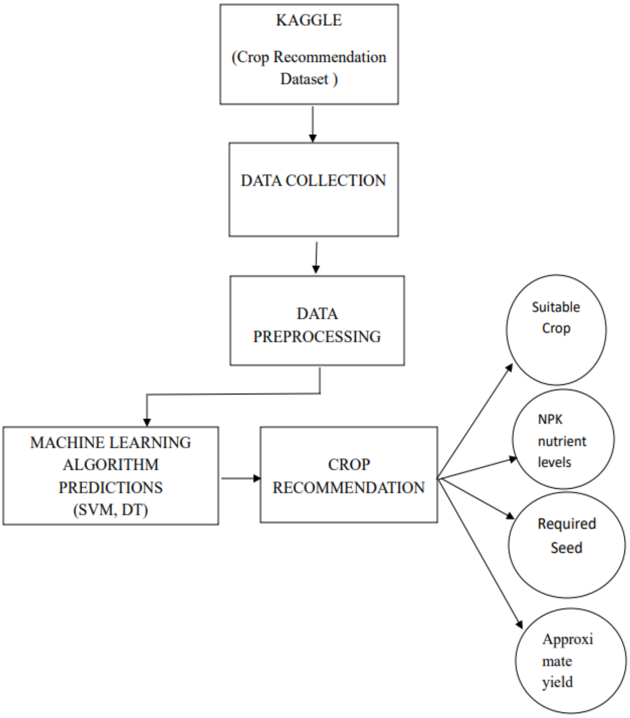

Figure 1. Crop Prediction System's Architecture

3.3. Prediction Using Machine Learning Algorithms

Prediction Algorithms of ML are models that forecast future events based on historical data. These techniques are classified under the supervised learning category, in which input-output pairs from labeled datasets are used to train the model. Finding patterns and links in the data to forecast a particular target variable is the main objective of predictive algorithms. The goal is to offer accurate predictions of future outcomes rather than merely analyzing past events.We employed supervised machine learning algorithms with classification and regression as subcategories in our system. Our system will work best using a classification method. The SVM algorithm's Rainfall Prediction uses a hyperplane to divide classes. The Decision Tree algorithm, which divides data according to feature values, is used for crop prediction.

3.3.1 Rainfall Prediction

Rainfall prediction flow chart is displayed in Figure 2. A supervised machine learning technique named Vector Machine is used for both regression and classification problems. It divides data into discrete classes by determining the best decision boundary, or hyperplane. The flowchart describes the step-by-step process for predicting rainfall for the year 2025 using the Kaggle's rainfall dataset.

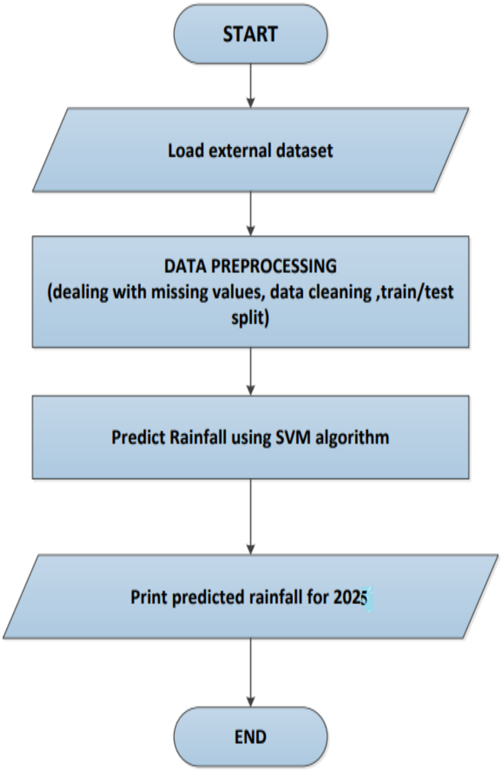

Figure 2. Rainfall Prediction using SVM

An external dataset with historical rainfall data is loaded at the start of the procedure. Then data preprocessing is then applied to the loaded dataset. The SVM algorithmic classifier with Radial Basis Function kernel is used to train the model after pre-processing. Equation (1) gives the Radial Basis Function's mathematical representation.

$$(p_1, p_2) = exponent\ (-\gamma||p_1 - p_2||\ 2\ ) \quad (1)$$

where, $||p_1\text{-}p_2||$ = Eucliden distance between p_1 & p_2, γ = Gamma.

The model predicts annual rainfall after fitting and testing, which serves as a key input for the prediction of crops.

3.3.2 Crop Prediction

Crop prediction begins by loading the crop recommendation datasets from Kaggle. After reading the dataset, data cleaning will be carried out as described in the Data Pre-processing section. Following this, use a decision tree classifier to train the models on the training set.

Step1: **Load the dataset**

- Store the data (N, P, K, and CROP values) in a DataFrame for easy handling.

Step2: **Convert Crop Names to Numbers**

- Replace each crop name with a unique number for easier processing by the algorithm.

Step3: **Separate Features and Target**

- N, P, K are the features.
- CROP is the target to predict.

Step4: **Train the Decision Tree**

- Use the DecisionTreeClassifier to train the model.

Step5: **Visualize the output**

- Plot the tree structure with feature names (N, P, K) and crop labels.

Step6: **Display the predicted crop**

Figure 3. Decision Tree algorithm for crop prediction

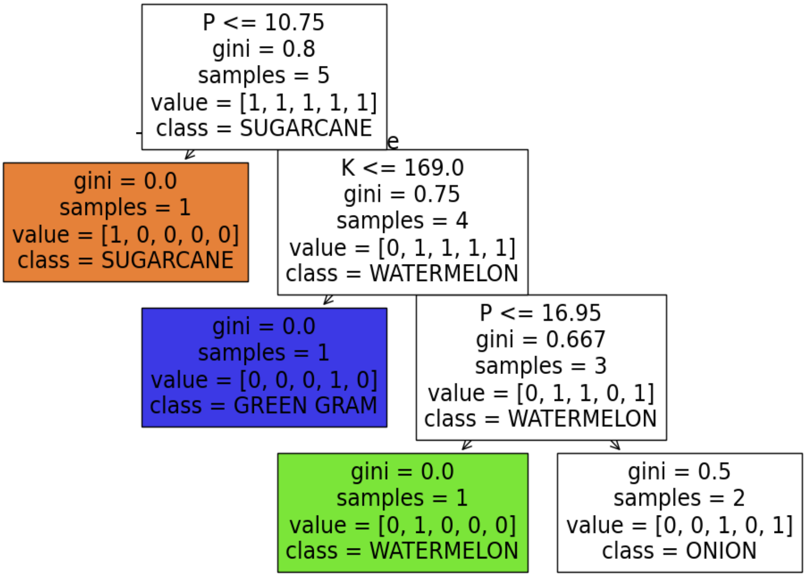

Figure 4. Decision Tree classification output

We take a number of factors into account when predicting the crop, including temperature, humidity, soil pH, and anticipated rainfall. A list including the input parameter values and predicted rainfall will be appended. The crop will be predicted by the decision tree algorithm using list data. Figure (3) illustrates the prediction algorithm's step-by-step execution, while Figure (4) displays the classification results.

3.4. Crop Recommendation

Depending on soil composition, determined rainfall and weather inputs, the system recommends the best crop for plantation. It also provides detailed information on the essential amount of fertilizers (ie.,NPK in kilograms per hectare) and seeds (in kilograms per acre) actually needed for the predicted crop. Additionally, the system estimates the approximate yield (in quintals per acre) of the suggested crop. These insights collectively assist farmers in selecting the most profitable crop.

4. Experimental Results

The proposed system is implemented in python language. It uses SVM and DT algorithms for crop classification and prediction. This system is trained and tested using data from Kaggle datasets.

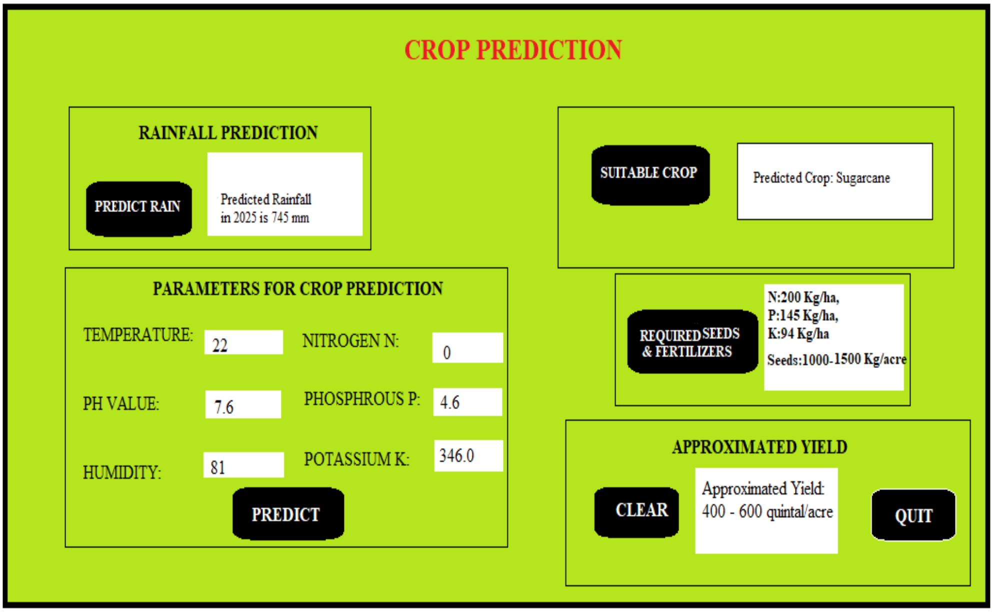

Figure 5. Crop Prediction with Recommendation

This system predicts the most suitable crop for a particular type of land by taking into consideration, parameters like yearly rainfall, soil pH, NPK values, humidity and temperature.

Out of these, the system first predicts annual rainfall using SVM algorithm and historical data from Kaggle rainfall dataset. Then it gets user input for soil parameters like temperature, PH value, humidity, Nitrogen level, Phosphorous level and Potassium level. These input values are passed to the Decision tree algorithm, which is already trained & tested with data from Kaggle's crop recommendation dataset. The proposed system displays the recommended crop, for given input values. Additionally, it also calculates the necessary NPK levels for the recommended crop, the required seed quantity per acre, and the estimated yield for the suggested crop. The results are displayed in a GUI form, as given below in Figure (5).

The output of the predictions made by Decision Tree algorithm is tabulated below in Table 1. The visualization of crop prediction is represented in Figure 6, and the recommendations for the crop, sugarcane is graphically shown in Figure 7.

Table 1. Crop Prediction Results

Parameters For Crop Prediction			Predicted Crop	User Entered Npk Values (Kg/Ha)			Reommended Npk Values (Kg/Ha)			Predicted Seeds For Plantation (Kg/Acre)	Approximated Yield (Quintal/Acre)
Ph	**Temp**	**Humidity**		**N**	**P**	**K**	**N**	**P**	**K**		
7.6	22	81	Sugarcane	0	4.6	346.0	200.0	145.0	94.0	1000 - 1500	400- 600
7.97	26	78	Watermelon	0	16.90	614.0	201.0	84.5	514.O	0.3	185 – 205
7.05	24	88	Onion	0	57.6	443.0	61.0	3.6	412.5	355	85 – 105
9	29	83	Green Gram	315.67	23.2	164	164	292.33	139	6 – 8	2 – 3
6.7	27	87	Groundnut	1	17.0	174.0	41.0	25.0	114.0	46	3.5 – 4.5

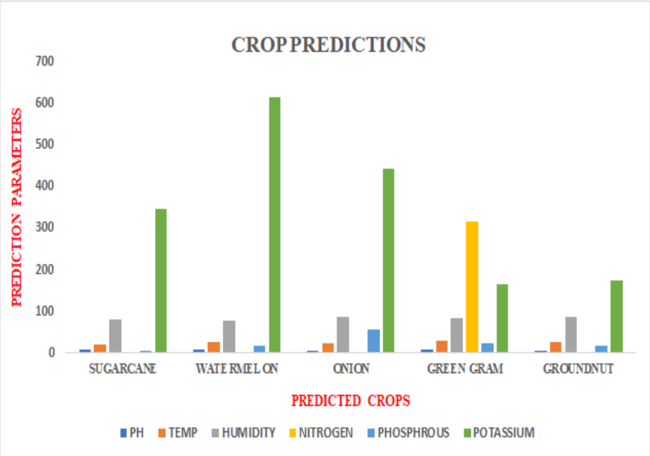

Figure 6. Crop Predictions

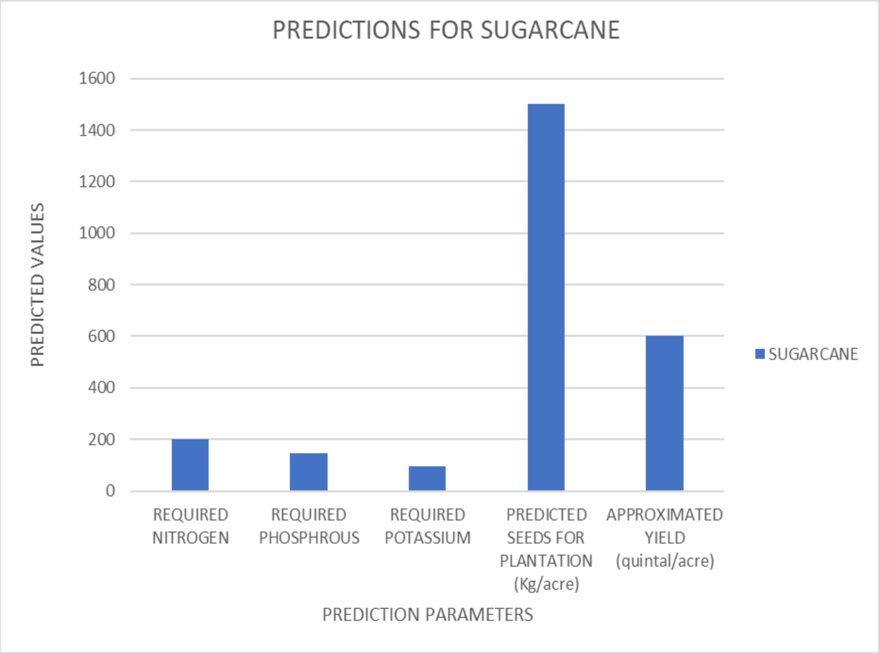

Figure 7. Recommendations for sugarcane

5. Conclusion

Currently, many farmers are unaware of technology usage in agriculture, which can lead to incorrect crop selection and a subsequent reduction in their production and income. To mitigate such losses, a GUI based farmer-friendly system is developed, which predicts the most suitable crop for a given type of land. This system also provides essential information on required nutrients, the appropriate quantity of seeds for cultivation, and the expected yield. Thus, this system empowers farmers to make informed decisions, enhancing growth and driving innovation in the agricultural sector. The future scope of this system is to enhance the system by collecting the necessary data through GPS locations of the land and integrating data from government-run rain forecasting systems. This will allow crop predictions based solely on the GPS location. Additionally, we plan to develop a model that can help address food supply imbalances, preventing both overproduction and shortages.

References

[1] A. Zakaria, A.Y.M. Shakaff, M.J. Masnan, F.S.A. Saad, A.H. Adom, M.N. Ahmad, M.N. Jaafar, A.H. Abdullah, L.M. Kamarudin, Improved maturity and ripeness classifications of Magnifera indica cv. Harumanis mangoes through sensor fusion of an electronic nose and acoustic sensor. Sensors, 12(5), (2012) 6023-6048. https://doi.org/10.3390/s120506023

[2] M. Garanayak, G. Sahu, S.N. Mohanty, A.K. Jagadev, Agricultural recommendation system for crops using different machine learning regression methods. International Journal of Agricultural and Environmental Information Systems, 12(1), (2021)1–20.

[3] D. Dighe, A. Pawar, R. Nagpure, and S. Kharat, Survey of crop recommendation systems. International Research Journal of Engineering and Technology (IRJET), 5, (2018) 476-481.

[4] R.V. Saraswathi, J. Sridharani, P.S. Chowdary, K. Nikhil, M.S. Harshitha, K. M. Sai, (2022) Smart farming: The IoT based future agriculture. 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), IEEE, India. https://doi.org/10.1109/ICSSIT53264.2022.9716331

[5] S. Pudumalar, E. Ramanujam, R.H. Rajashree, C. Kavya, T. Kiruthika, J. Nisha, (2017) Crop recommendation system for precision agriculture. Eighth International Conference on Advanced Computing (ICoAC), IEEE, India. https://doi.org/10.1109/ICoAC.2017.7951740

[6] M. Rajakumaran, G. Arulselvan, S. Subashree, R. Sindhuja, Crop yield prediction using multi-attribute weighted tree-based support vector machine, Measurement: Sensors, 31, (2024), 101002. https://doi.org/10.1016/j.measen.2023.101002

[7] A.K. Kushwaha, S. Bhattachrya, Crop yield prediction using Agro Algorithm in Hadoop, International Journal of Computer Science and Information Technology & Security, 5(2), (2015) 271-274.

[8] L. Girish, S. Gangadhar, T.R. Bharath, K.S. Balaji, K.T. Abhishek, Crop Yield and Rainfall Prediction in Tumakuru District using Machine Learning. International Journal for Research in Engineering Application & Management, (2018) 61-65.

[9] R. Katarya, A. Raturi, A. Mehndiratta, A. Thapper, (2020) Impact of machine learning techniques in precision agriculture. In 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), IEEE, India. https://doi.org/10.1109/ICETCE48199.2020.9091741

[10] P. Samui, V.R. Mandla, A. Krishna, T. Teja, Prediction of rainfall using support vector machine and relevance vector machine. Earth Science India, 4, (2011) 188-200. Available at: https://www.researchgate.net/publication/268439113_Prediction_of_Rainfall_Using_Support_Vector_Machine_and_Relevance_Vector_Machine

[11] X. Jiang, Y. Sun, F. Zhao, Decision tree-based regression models for predicting crop yield using remote sensing and climate data. Agricultural Systems, 211, (2023) 103277.

[12] N. Bali, A. Singla, Deep learning based wheat crop yield prediction model in Punjab region of North India. Applied Artificial Intelligence, 35(15), (2021) 1304-1328. https://doi.org/10.1080/08839514.2021.1976091

[13] P. Nevavuori, N. Narra, T. Lipping, Crop yield prediction with deep convolutional neural networks, Computers and Electronics in Agriculture, 163, (2019) 104859. https://doi.org/10.1016/j.compag.2019.104859

[14] S.M. Kuriakose, T.Singh. (2022) Indian Crop Yield Prediction Using LSTM Deep Learning Networks. 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, 2022, 1–5. https://doi.org/10.1109/ICCCNT54827.2022.9984407

[15] K. Juhos, S. Szabó, M. Ladányi, Influence of Soil Properties on Crop Yield: A Multivariate Statistical Approach. International Agrophysics, 29(4), (2015), 433–40. https://doi.org/10.1515/intag-2015-0049

[16] Q. Luo, L. Wen, The Role of Climatic Variables in Winter Cereal Yields: A Retrospective Analysis, International Journal of Biometeorology, 59(2), (2015), 181–92. https://doi.org/10.1007/s00484-014-0834-4

[17] T. Islam, T. A. Chisty A. Chakrabarty (2018) A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh. IEEE Region 10 Humanitarian Technology Conference (R10-HTC), IEEE, 2018, 1–6. https://doi.org/10.1109/R10-HTC.2018.8629828

[18] H. Singh, N. Sharma, Decision Support System for Precision Farming, International Journal of Computers & Technology, vol. 4(1), (2013), 76–81. https://doi.org/10.24297/ijct.v4i1b.3065

[19] A. Cravero, S. Pardo, S. Sepúlveda, L. Muñoz Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review, Agronomy, 12(3), (2022) 748. https://doi.org/10.3390/agronomy12030748

[20] Anusha S. Rai A, & R. Srinivasa Rao Kunte, Challenges in Implementing AI Technology Smart Farming in Agricultural Sector – A Literature Review, International Journal of Management, Technology, and Social Sciences, 9(2), (2024) 283-301. https://doi.org/10.5281/zenodo.12622780

Funding

No funding was received for conducting this study.

Conflict of interest

The Author's have no conflicts of interest to declare that they are relevant to the content of this article.

About The License