iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60339

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

HarvestMax: A Predictive Model for Crop Yield and Fertilizer Optimization

S. Sreekanth[1], A. Sai Kumar[2], K. Pavan[3], K. Akash[4]
[1]*Associate Professor, Department of CSE (DS), Institute of Aeronautical Engineering, Hyderabad*
[2, 3, 4]*Department of CSE (DS), Institute of Aeronautical Engineering, Hyderabad*

Abstract: In the backdrop of India's agrarian-centric economy, the precision of crop yield prediction and the adoption of optimal farming practices emerge as critical components for sustainable agricultural development. This research paper introduces an innovative methodology that integrates data mining techniques, machine learning algorithms such as Support Vector Machines (SVM), and Regression algorithms, along with a comprehensive attribute analysis to establish a resilient crop yield prediction and recommendation system. The proposed system draws insights from a diverse range of attributes crucial for agriculture, including geographical location, soil pH for alkalinity assessment, nutrient percentages (Nitrogen, Phosphorous, and Potassium), real-time weather conditions sourced from third-party APIs, soil type, nutrient composition, and regional rainfall data. By amalgamating and scrutinizing these multifaceted attributes, our system aspires to furnish farmers with precise and dependable predictions regarding their crop yields.

Keywords: Support Vector Machines (SVM), Attribute Analysis, Nutrient Composition, Weather Conditions, Sustainable Agriculture.

I. INTRODUCTION

This Project represents a groundbreaking initiative aimed at revolutionizing modern agricultural practices through the integration of advanced machine learning models. The project addresses the complexities faced by farmers by offering a comprehensive solution that combines predictive analytics and recommendation systems. The project revolves around developing a sophisticated decision support system for agriculture, leveraging cutting-edge technologies to enhance the precision of crop yield predictions. By integrating data mining techniques and machine learning algorithms, the system aims to analyse a myriad of attributes crucial for agriculture, ranging from geographical location and soil pH to real-time weather conditions and nutrient compositions. Through this holistic approach, the system intends to empower farmers with actionable insights for informed decision-making in crop cultivation. This project involves various models that utilize both machine learning and deep learning techniques. The primary focus is on conducting a comparative analysis of these algorithms, evaluating their performance based on various metrics. The primary objectives of the project include: Crop Yield Prediction: Develop an accurate model for forecasting crop yields based on historical data, weather conditions, and soil characteristics. Crop Recommendation: Create a recommendation system to guide farmers in selecting the most suitable crops for cultivation, taking into account soil type, climate, and historical performance. Fertilizer Recommendation: Implement a model that optimizes fertilizer usage by analyzing soil nutrient levels, crop types, and historical fertilizer application data. Algorithm Comparative Analysis: Evaluate and compare the performance of various machine learning algorithms, including Support Vector Machines (SVM), regression, Gaussian processes, and neural networks, in the context of crop yield and fertilizer recommendation models.

II. LITERATURE SURVEY

S. Rai, J. Nandre and B. R. Kanawade [1] presented a study comparing different regression techniques for crop yield prediction, providing insights into their effectiveness.

R. J, V. K. G. Kalaiselvi, A. Sheela, D. S. D and J. G [2] explored the application of machine learning algorithms for crop yield prediction, emphasizing their potential in agricultural decision-making.

S. Vashisht, P. Kumar and M. C. Trivedi [3] proposed an enhanced version of the Extreme Learning Machine algorithm for crop yield prediction, aiming for improved accuracy.

Pradeep [4] investigated the use of gradient boosting techniques for crop yield prediction, focusing on enhancing agricultural outcomes.

Gul [5] introduced a novel approach using Spiking Neural Networks for crop yield prediction, exploring its potential in agricultural forecasting

Sharma [6] discussed the early prediction of crop yield in India using machine learning techniques, highlighting the importance of timely predictions for agricultural planning.

Thirumal and Latha [7] proposed an automated prediction model combining the Sine Cosine Algorithm and Weighted Regularized Extreme Learning Machine for rice crop yield prediction.

Rananavare and Chitnis [8] explored the utilization of Sentinel-2 satellite data for predicting monocot crop yields, leveraging remote sensing technology for agricultural forecasting.

A. k. Gajula, J. Singamsetty, V. C. Dodda and L. Kuruguntla [9] investigated machine learning techniques for crop and yield prediction in agriculture, providing insights into the applicability of these techniques in practical scenarios.

Bharathi [10] presented an experimental analysis of crop yield prediction employing a modified deep learning strategy, aiming to enhance prediction accuracy.

Saini and Nagpal [11] proposed a Deep Long Short-Term Memory (LSTM) model for wheat crop yield prediction in India, utilizing recurrent neural networks for accurate forecasting.

Thirumal and Latha [12] introduced an automated prediction model based on hyperparameter-tuned stacked autoencoders for rice crop yield prediction.

Patki and Wazurkar [13] explored the application of data sampling techniques for crop yield prediction using Stochastic Gradient Descent Neural Networks.

F. Shahrin, L. Zahin, R. Rahman, A. J. Hossain, A. H. Kaf and A. K. M. Abdul Malek Azad, [14] analyzed agricultural data and predicts crop yields in Habiganj using multispectral bands of satellite imagery and machine learning algorithms.

S. G L, N. V and S. U [15] provides a comprehensive review of crop yield prediction techniques using machine learning, summarizing the state-of-the-art methods and identifying future research directions.

Dhande and Malik [16] presented an empirical study on crop-disease detection and crop-yield analysis systems, offering a statistical perspective on agricultural analysis.

Suresh [17] investigated the application of the Random Forest algorithm for crop yield prediction, emphasizing its effectiveness in handling agricultural data.

S. M. G, S. Paudel, R. Nakarmi, P. Giri and S. B. Karki [18] proposed a prediction model based on soil moisture using various machine learning algorithms, aiming to improve crop yield forecasting accuracy.

R. ThangaSelvi and M. Sathish [19] introduced an optimal Bidirectional Gated Recurrent Neural Network (Bi-GRNN) model for crop yield prediction, leveraging its sequential modelling capabilities.

M. S. Teja, T. S. Preetham, L. Sujihelen, Christy, S. Jancy and M. P. Selvan [20] explored the utilization of Support Vector Machine (SVM) algorithm for crop recommendation and yield production, highlighting its potential in agricultural decision support systems.

K. Lohitha Reddy and A. P. Siva Kumar [21] investigated machine learning techniques for weather-based crop yield prediction, emphasizing the integration of weather data for more accurate forecasting.

This literature survey highlights the diverse approaches and methodologies employed in crop yield prediction using machine learning techniques, providing valuable insights for further research and development in agricultural forecasting and decision-making.

III. DATASET INTRO

The Crop Yield Prediction Model aims to forecast crop yields based on a combination of historical crop yield data, soil characteristics, and climate conditions.

This model is a crucial component of the Integrated Agricultural Decision Support System, providing farmers with insights into potential harvests and aiding in effective crop management. To train the Crop Yield Prediction Model effectively, crucial input features are considered. These include historical crop yield data, specifying crop types and yield per hectare, soil characteristics such as soil type, pH levels, nutrient content (N, P, K), and organic matter, and climate data encompassing temperature, precipitation, humidity, wind speed, and solar radiation.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

	Crop	Season	State	Area	Production	Annual_Rainfall	Fertilizer	Pesticide
0	0	4	2	73814.0	56708	2051.4	7024878.38	22882.34
1	1	1	2	6637.0	4685	2051.4	631643.29	2057.47
2	8	1	2	796.0	22	2051.4	75755.32	246.76
3	9	4	2	19656.0	126905000	2051.4	1870661.52	6093.36
4	11	1	2	1739.0	794	2051.4	165500.63	539.09
...	...	...	...	...	...	...	...	...
19684	44	1	19	4000.0	2000	1498.0	395200.00	1160.00
19685	53	2	19	1000.0	3000	1498.0	98800.00	290.00
19686	24	1	10	310883.0	440900	1356.2	29586735.11	96373.73
19687	40	1	10	275746.0	5488	1356.2	26242746.82	85481.26
19688	53	2	10	239344.0	392160	1356.2	22778368.48	74196.64

19689 rows × 8 columns

Fig :3.1 Crop yield Dataset

The Fertilizer Recommendation Model, a critical component of the Integrated Agricultural Decision Support System, is designed to optimize fertilizer usage by analyzing soil nutrient levels, crop types, and historical fertilizer application data. This model assists farmers in making informed decisions about the type and quantity of fertilizers to enhance crop productivity and sustainability. The Fertilizer Recommendation Model relies on specific input features for effective training. These include soil characteristics such as soil type, pH levels, and nutrient content (N, P, K), crop types, and historical fertilizer usage information detailing the types of fertilizers used, application rates, and timing.

	Temparature	Humidity	Moisture	Soil Type	Crop Type	Nitrogen	Potassium	Phosphorous	Fert N
0	26	52	38	Sandy	Maize	37	0	0	
1	29	52	45	Loamy	Sugarcane	12	0	36	
2	34	65	62	Black	Cotton	7	9	30	14-[illegible]
3	32	62	34	Red	Tobacco	22	0	20	[illegible]
4	28	54	46	Clayey	Paddy	35	0	0	

Fig:3.2 Fertilizer Recommendation Dataset

Exploratory Data Analysis (EDA) is a crucial step in the data analysis process. It involves the initial investigation and examination of data to understand its main characteristics, uncover patterns, identify trends, detect outliers, and test hypotheses. EDA is often performed before formal statistical modelling to gain insights into the data and inform subsequent analysis. The EDA has been performed on the data to make it ready for further process.

Label Encoding is a technique used in machine learning and data preprocessing to convert categorical data into numerical form. In Label Encoding, each unique category or label is assigned an integer value. This process is typically applied to categorical variables that have an ordinal relationship, meaning there is a meaningful order among the categories.

IV. METHODOLOGY

In this research paper, we present a comprehensive methodology for the implementation of a crop yield prediction model and fertilizer recommendation system. Our study focuses on leveraging machine learning techniques to predict crop yields accurately and recommend optimal fertilizer application strategies. The proposed methodology integrates various algorithms, including regression algorithms, support vector machines (SVM), and Gaussian process regression, to analyse and compare their performance in predicting crop yields and recommending the fertilizer for the soil.

The first step of our methodology involves collecting relevant data pertaining to crop characteristics, environmental factors, soil properties, and historical yield records. This dataset serves as the foundation for training and testing the prediction models. We then preprocess the data to handle missing values, normalize features, and address any outliers to ensure the robustness of the models.

Subsequently, we employ regression algorithms such as linear regression, polynomial regression, and decision tree regression to build baseline yield prediction models. These algorithms offer a traditional approach to modelling crop yield based on input features such as weather conditions, soil nutrients, and crop types.

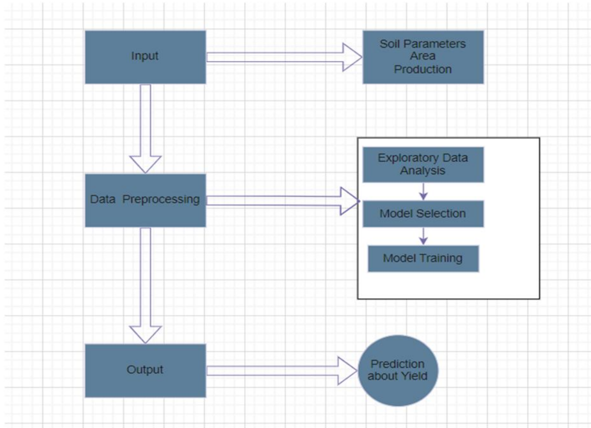

Fig:4.1 System Architecture

Next, we explore the efficacy of SVM, a supervised learning algorithm known for its ability to handle high-dimensional data and nonlinear relationships. By optimizing kernel functions and hyperparameters, SVM aims to capture complex patterns within the dataset to improve yield prediction accuracy.

Furthermore, we investigate Gaussian process regression, a probabilistic machine learning approach that models the uncertainty inherent in crop yield predictions. By estimating the distribution over possible functions, Gaussian process regression provides valuable insights into the variability of yield estimates and offers robustness against noise in the data.

The crop yield prediction model has been implemented using various regression algorithms, support vector machine and boosting algorithms. Similarly the fertilizer recommendation model has been implemented by using various algorithms such as regression, Support Vector Machine and KNN algorithm.

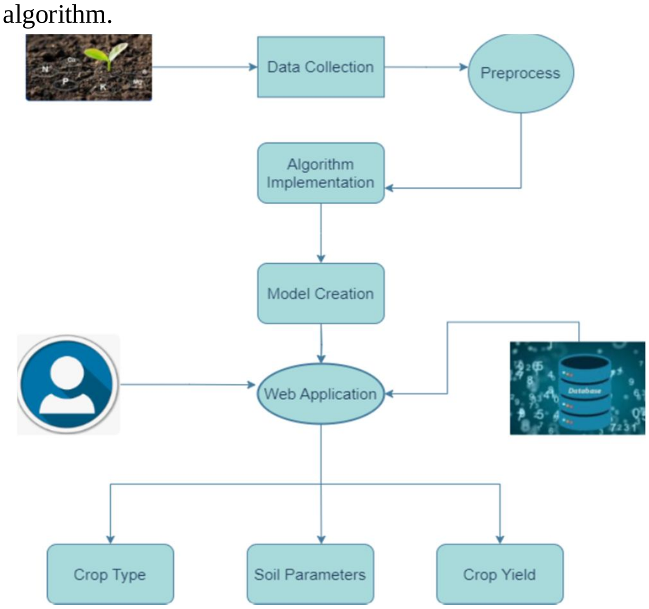

Fig:4.2 Data Flow Diagram

Finally, we conduct a comparative analysis of the performance of these algorithms based on key metrics such as mean absolute error, mean squared error, and R-squared. Through extensive experimentation and evaluation, we elucidate the strengths and weaknesses of each algorithm in accurately predicting crop yields and providing fertilizer recommendations.

Our methodology entails a rigorous comparative analysis of the aforementioned algorithms to discern their respective strengths and weaknesses in predicting crop yields and providing fertilizer recommendations. We assess the predictive accuracy, computational efficiency, and scalability of each algorithm to ascertain their suitability for real-world agricultural applications. By conducting this comparative analysis, we aim to provide valuable insights into the optimal choice of algorithms for crop yield prediction and fertilizer recommendation tasks, thereby facilitating informed decision-making for farmers and agricultural stakeholders.

V. RESULTS AND ANALYSIS

Crop Yield Prediction Model has implemented using following algorithms such as SVM ,Regression techniques such as Linear regression, Decision Tree Regressor ,Random Forest Regressor, Gradient Boosting Regressor, XGB Regressor, KNN Regressor and Gaussian Process. We used R2 score value as metric for evaluating the models and comparing them. The algorithm and its R2 score values are as follows: (only top 7 algorithms)

Algorithm	R2 Score
Gradient Boosting	0.91
Bagging Regressor	0.907
Random Forest	0.906
XGBoost	0.867
Decision Tree	0.867
KNN	0.83

Here the evaluation metric R2 Score measures the proportion of the variance in the dependent variable that is predictable from the independent variables.

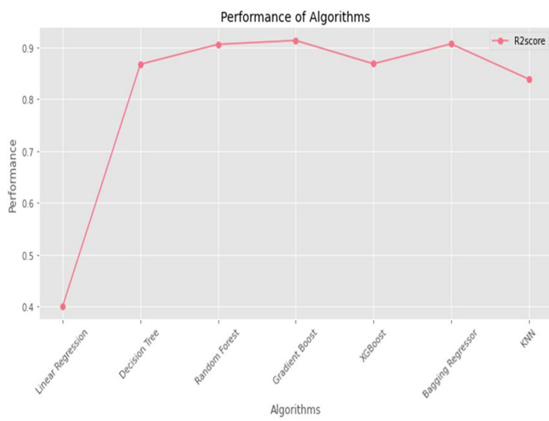

Fig:5.1 Comparison of Algorithms

Fertilizer Recommendation Model has been implemented by SVM, Gradient Boosting and Logistic Regression .Among those algorithms model performed well with SVM. The Boosting algorithm and Logistic Regression algorithm are not performed as the SVM has performed. Through our comparative analysis, we found that SVM effectively captured complex nonlinear relationships between soil attributes, environmental factors, and crop requirements, thereby generating more precise fertilizer recommendations. SVM's ability to identify intricate patterns within the dataset, coupled with its capacity to handle high-dimensional feature spaces, contributed to its enhanced performance in this context.

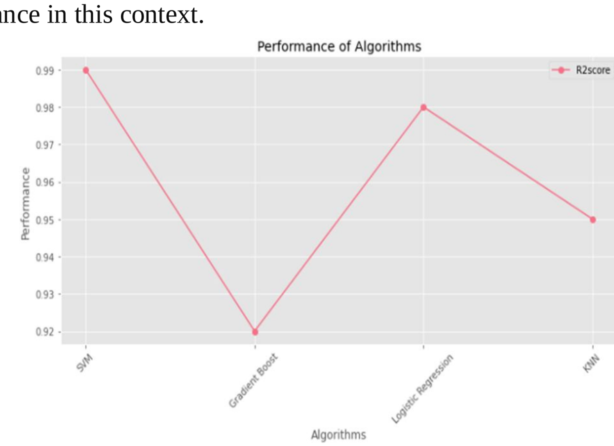

Fig :5.2 Comparing performance of Algorithms

A. Analysis

From this Project We can Understand that SVM has been performed well in the case of crop recommendation and fertilizer recommendation system .In the Case of Crop Yield Recommendation, Gradient Boosting algorithm has performed well . Using These algorithms ,we can predict the amount of yield that can obtained from the particular area based on some parameters. We can also recommend the crop for the land using various features and machine learning models. We can recommend required fertilizer for the land based on the crop and ither features. Some Machine Learning algorithms are performed well in some models. The Working Of the Algorithm for certain Model has been varied based on the type of the data and size of data Support Vector Machines (SVM) often perform well in scenarios where the data is characterized by clear margins of separation between different classes. SVM excels in high-dimensional spaces and is effective when the number of dimensions is greater than the number of samples. It is particularly powerful in situations where the data exhibits non-linear relationships and can be effectively employed for both classification and regression tasks. SVM's robustness and versatility make it suitable for various real-world applications with complex decision boundaries. Gradient Boosting algorithms, like XGBoost or Gradient Boosted Decision Trees, excel when modelling complex relationships in data. They perform well in situations with heterogeneous features and diverse patterns. The iterative boosting process, focusing on misclassified instances, enhances model accuracy over iterations. Gradient Boosting is effective for both regression and classification tasks, demonstrating high predictive performance and resilience against overfitting, making it a powerful choice for a broad range of machine learning applications.

VI. CONCLUSION AND FUTURE SCOPE

The project achieved significant milestones in the realm of precision agriculture by successfully implementing a comprehensive suite of machine learning models. These models encompassed crop recommendation, crop yield prediction, and fertilizer recommendation, collectively enhancing farming practices. Leveraging diverse datasets and robust preprocessing techniques, the system incorporated factors like soil type, climate, and historical crop performance. The user friendly interface facilitated seamless interaction for farmers, enabling them to access personalized recommendations effortlessly. The project's impact was further underscored through collaboration with agricultural experts and stakeholders, ensuring alignment with real-world farming practices. Rigorous performance evaluations showcased the reliability and accuracy of the implemented solutions, measured through metrics such as accuracy, precision, and recall. Notably, the system's scalability was prioritized, allowing for potential expansion and integration of additional features. Overall, the project has made a positive impact on agriculture, offering improved crop selection, increased yield predictions, and sustainable fertilizer use for the benefit of the farming community.

The proposed improvements for the crop recommendation, crop yield prediction, and fertilizer recommendation system encompass a multifaceted approach aimed at refining various facets of the project. To enhance data robustness, strategies for data augmentation and advanced data cleaning techniques will be implemented, ensuring high-quality datasets. Exploring advanced machine learning architectures, such as deep learning models, is underway to capture intricate patterns in agricultural data and improve prediction accuracy.

REFERENCES

[1] S. Rai, J. Nandre and B. R. Kanawade, "A Comparative Analysis of Crop Yield Prediction using Regression," 2022 2nd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2022, pp. 1-4.

[2] R. J, V. K. G. Kalaiselvi, A. Sheela, D. S. D and J. G, "Crop Yield Prediction Using Machine Learning Algorithm," 2021 4th International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2021, pp. 611-616.

[3] S. Vashisht, P. Kumar and M. C. Trivedi, "Improvised Extreme Learning Machine for Crop Yield Prediction," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 754-757

[4] G. Pradeep, T. D. V. Rayen, A. Pushpalatha and P. K. Rani, "Effective Crop Yield Prediction Using Gradient Boosting To Improve Agricultural Outcomes," 2023 International Conference on Networking and Communications (ICNWC), Chennai, India, 2023, pp. 1-6.

[5] M. U. Gul, K. John Pratheep, M. Junaid and A. Paul, "Spiking Neural Network (SNN) for Crop Yield Prediction," 2021 9th International Conference on Orange Technology (ICOT), Tainan, Taiwan, 2021, pp. 1-4.

[6] A. Sharma, A. Tamrakar, S. Dewasi and N. S. Naik, "Early Prediction of Crop Yield in India using Machine Learning," 2022 IEEE Region 10 Symposium (TENSYMP), Mumbai, India, 2022, pp. 1-6

[7] S. Thirumal and R. Latha, "Automated Rice Crop Yield Prediction using Sine Cosine Algorithm with Weighted Regularized Extreme Learning Machine," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 35-40, doi: 10.1109/ICICCS56967.2023.10142403.

[8] L. B. Rananavare and S. Chitnis, "Monocot Crop Yield Prediction using Sentinal-2 Satellite Data," 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2022, pp. 1-6, doi: 10.1109/GCAT55367.2022.9971820.

[9] A. k. Gajula, J. Singamsetty, V. C. Dodda and L. Kuruguntla, "Prediction of crop and yield in agriculture using machine learning technique," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-5, doi: 10.1109/ICCCNT51525.2021.9579843.

[10] P. S. Bharathi, V. Amudha, G. Ramkumar, T. J. Nagalakshmi, N. Nalini and P. Jagadeesh, "An Experimental Analysis of Crop Yield Prediction using Modified Deep Learning Strategy," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2022, pp. 1-6, doi: 10.1109/ACCAI53970.2022.9752492.

[11] P. Saini and B. Nagpal, "Deep-LSTM Model for Wheat Crop Yield Prediction in India," 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2022, pp. 73-78, doi: 10.1109/CCiCT56684.2022.00025

[12] S. Thirumal and R. Latha, "Automated Hyperparameter Tuned Stacked Autoencoder based Rice Crop Yield Prediction Model," 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2023, pp. 14-18, doi: 10.1109/ICOEI56765.2023.10126037

[13] V. Patki and P. Wazurkar, "Application of Data Sampling for Crop Yield Prediction using Stochastic Gradient Descent Neural Networks," 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2021, pp. 685-689, doi: 10.1109/CSNT51715.2021.9509682

[14] F. Shahrin, L. Zahin, R. Rahman, A. J. Hossain, A. H. Kaf and A. K. M. Abdul Malek Azad, "Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning," 2020 11th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 2020, pp. 21-24, doi: 10.1109/ICECE51571.2020.9393066

[15] S. G L, N. V and S. U, "A Review on Prediction of Crop Yield using Machine Learning Techniques," 2022 IEEE Region 10 Symposium (TENSYMP), Mumbai, India, 2022, pp. 1-5, doi: 10.1109/TENSYMP54529.2022.9864482

[16] A. Dhande and R. Malik, "Empirical Study of Crop-disease Detection and Crop-yield Analysis Systems: A Statistical View," 2022 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2022, pp. 1-4, doi: 10.1109/ESCI53509.2022.9758284.

[17] N. Suresh et al., "Crop Yield Prediction Using Random Forest Algorithm," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 279-282, doi: 10.1109/ICACCS51430.2021.9441871

[18] S. M. G, S. Paudel, R. Nakarmi, P. Giri and S. B. Karki, "Prediction of Crop Yield Based-on Soil Moisture using Machine Learning Algorithms," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 491-495, doi: 10.1109/ICTACS56270.2022.9988186

[19] R. ThangaSelvi and M. Sathish, "An Optimal Bidirectional Gated Recurrent Neural Network Model for Crop Yield Prediction," 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 1302-1305, doi: 10.1109/ICSSIT55814.2023.10060890

[20] M. S. Teja, T. S. Preetham, L. Sujihelen, Christy, S. Jancy and M. P. Selvan, "Crop Recommendation and Yield Production using SVM Algorithm," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2022, pp. 1768-1771, doi: 10.1109/ICICCS53718.2022.9788274.

[21] K. Lohitha Reddy and A. P. Siva Kumar, "Machine Learning Techniques for Weather based Crop Yield Prediction," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 1263-1268, doi: 10.1109/ICAIS56108.2023.10073740.