Integrating Soil Sensing and Rainfall Modelling for Precision Crop Recommendation Systems

T.Vasudeva Reddy
Dept. of ECE
B V Raju Institute of Technology
Narsapur, Medak (dist), Telangana
vasu.tatiparthi@bvrit.ac.in

K. Saikumar
Dept. of ECE
Koneru Lakshmaiah Education Foundation
Greenfields, Vaddeswaram, Andhra Pradesh
saikumarkayam4@gmail.com

R. Anirudh Reddy
Dept. of ECE
B V Raju Institute of Technology
Narsapur, Medak (dist), Telangana
anirudhreddy.r@bvrit.ac.in

V. Bhanuprasad Chary
Dept. of ECE
B V Raju Institute of Technology
Narsapur, Medak (dist), Telangana
20211a04m9@bvrit.ac.in

Ponnam Srinija
Dept. of ECE
B V Raju Institute of Technology
Narsapur, Medak (dist), Telangana
21215a0425@bvrit.ac.in

T. Harsha Vardhan Reddy
Dept. of ECE
B V Raju Institute of Technology
Narsapur, Medak (dist), Telangana
20211a04m2@bvrit.ac.in

2024 IEEE 9th International Conference for Convergence in Technology (I2CT) | 979-8-3503-9447-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/I2CT61223.2024.10543793

***Abstract*— This research article offers a sophisticated machine learning-based method for altitude, temperature, pressure, sunshine, rainfall, nutrients, and soil moisture data-driven soil state, and crop type analysis. The dataset is rigorously preprocessed, with seaborn used to visualize and handle null values. The process behind the article explains the four soil states of soil according to the encoding of categorical output classes: dry, very dry, wet, and very wet. Several crop types are to be predicted depending on the parameters taken. Random Forest Classifier, Decision tree, KNN, and Naive Bayes models were trained with accuracies of respectively, following data normalization and splitting into training and testing sets. This work demonstrates how machine learning can effectively classify soil conditions, which opens up new possibilities for agricultural and environmental management applications.**

Keywords-- Machine Learning, Rainfall forecast, Weather data, Soil state Data Visualization, K- Nearest Neighbors, Random Forest, Linear regression, decision Tree.

I. INTRODUCTION

For many uses in land management, environmental science, and agriculture, examining and describing soil conditions are essential. A thorough understanding of soil characteristics, including rainfall, nutrients like Nitrogen, phosphorus, and Potassium, temperature, pressure, altitude, and moisture content, is essential for forecasting environmental changes, planning land uses, and improving agricultural methods [1].

Conventional soil analysis techniques frequently rely on empirical evaluations and manual observation, which can be laborious, arbitrary, and prone to human error. Machine learning techniques have become attractive tools for improving and automating soil state assessments with the rise of data-driven approaches and improved technologies. This study aims to provide a novel framework for precisely classifying soil conditions into four distinct groups: dry, very dry, wet, and very wet. It also predicts which crop is suitable for that particular soil. These can predict about nearly 21 crops. It accomplishes this by applying state-of-the-art machine learning methods. [2]. This study's main dataset, which includes multi-dimensional data gathered from several geographic locations, offers a thorough depiction of soil properties.

To comprehend soil parameter correlations, the study carefully pre-processed the information by eliminating null values and used Python's Seaborn package for data visualization. One Hot Encoding gave the categorical output classes a numerical representation for machine learning. To evaluate the model, the dataset was divided into training and testing sets, and then it was standardized. Training Random Forest Classifier, Decision Tree, KNN, and Linear Regression yielded accuracy values. This project aims to revolutionize soil analysis and enhance land management, agriculture, and environmental monitoring by automating precise classification of soil states. The methodology, experimental design, findings, and discussions of the implications and subsequent actions are all provided in depth in the section that follows. [3].

II. LITERATURE SURVEY

The ability of machine learning (ML) techniques to predict soil characteristics, crop development, and soil fertility with high accuracy has attracted a lot of interest recently. The ML technique can be used to investigate a variety of soil parameters, including texture, organic matter, pH, nutrient content, soil moisture, and soil structure. Because machine learning approaches can process enormous volumes of complex data and reveal hidden patterns, they are superior to classical statistical methods. Developing methods for using machine learning to forecast soil characteristics [4,5,6], crop growth [7,8,9], and soil fertility [10,11] has been the subject of several studies.

A comprehensive assessment of the literature that was recently published in [12] identifies research needs in specific deep-learning applications and assesses the impact of environmental conditions and vegetation indicators on agricultural productivity. The previous research from 2012 to 2022 was reviewed by the writers using several databases. The paper focuses on the many variables that affect crop yield prediction, the benefits of estimating agricultural output using deep learning, and the best remote sensing technology depending on data gathering requirements. It has been demonstrated that machine learning algorithms work well for predicting crop yields, soil fertility, and soil properties. However, the volume and quality of the data used for training, along with the settings and methods employed, have a significant impact on how accurate machine learning models are.. To further understand how to build and improve machine learning models that predict soil parameters and assess how well these models work in various soil and environmental conditions, more study is required. Recommendations from machine learning can be advantageous for plant breeders, farmers, legislators, and other agricultural professionals.

In this project, an overall analysis of the soil is being conducted based on these respective factors.

For crop forecasts based on season, a crop selection method has been devised. Consequently, crops that are sown at the end of the monsoon or the beginning of the winter season—such as those planted between September and October—are referred to as Rabi and Kharif, respectively (these crops are called monsoon crops), and Zaid (the short season between Kharif and Rabi in March to July), the seasonal crops will be predicted.

III. EXISTING DESIGNS, MODELS AND THEIR APPROACH

A. *Internet of Things Technologies:*

The integration of satellite data with IoT sensors on the ground provides a comprehensive viewpoint, augmenting insights through an extensive dataset. Wireless network sensors enable the seamless collection of real-time data, promoting a synergistic approach that, by using the advantages of both technologies, empowers a variety of applications, from smart cities to environmental monitoring. [2].

B. *Cloud-based Platforms:*

Cloud-based systems that provide extensive capabilities for the easy management and analysis of Internet of Things (IoT) data include AWS IoT, Azure IoT, and Google Cloud IoT. By utilising these platforms, companies may leverage cloud computing's capacity to improve the scalability, efficiency, and insights of their IoT deployments. [2].

C. *Physically Based Models:*

Physically Based Models: These include hydraulic and water balance models. They carefully take into account input parameters in order to predict results. By taking into account physical principles, these models ensure a thorough knowledge of complicated systems and imitate real-world processes. They improve prediction accuracy of hydraulic behaviours and water-related phenomena by taking into account a variety of variables.

D. *Open-source models:*

Open-source models enable the creation of IoT applications and visual programming via readily available tools such as node-red and open IoT. In addition to promoting a dynamic ecosystem for the Internet of Things and widespread acceptance and technological improvement, this collaborative approach stimulates creativity by empowering developers to produce adaptable and customisable solutions.

E. *Automation:*

Farming techniques are revolutionised in the agricultural sector by the combination of automation and IoT devices. Through this synergy, real-time data-driven modifications are made possible, enabling precision agriculture. Enabled by the Internet of Things, automated farming maximises the use of resources, improves crop management, and guarantees effective, sustainable methods for a productive and technologically sophisticated agricultural environment. [2].

Limitations: Dependency on Rainfall forecasts, Limited Data quality, Calibration changes, Limited understanding soil type, data security, Cost constraints, and Complexity of soil-water dynamics.

IV. PROPOSED SYSTEM

With limited resources, the proposed system seeks to assist farmers in managing their farms sensibly. This system creates a smart irrigation system by combining ML and IoT features. Making informed irrigation decisions that maximizes water use could be greatly benefited by it. The machine learning model is to be trained by the system using reports on soil moisture, rainfall, and crop predictions. The optimal irrigation choice is subsequently determined by applying the machine learning model [7]. Five sections comprise the work: an IoT-based motor controller, an online Rainfall data collection platform, a soil prediction model, a web interface for crop prediction, and a data acquisition model.

TABLE I. DATASET FOR CROP PREDICTION [16]

N	P	K	temperatu	humidity	ph	rainfall	label
90	42	43	20.8797437	82.002744	6.502985	202.9355	rice
85	58	41	21.7704617	80.319644	7.038096	226.6555	rice
60	55	44	23.0044592	82.320763	7.840207	263.9642	rice
74	35	40	26.4910964	80.158363	6.980401	242.864	rice
78	42	42	20.1301748	81.604873	7.628473	262.7173	rice
69	37	42	23.0580487	83.370118	7.073454	251.055	rice
69	55	38	22.708838	82.639414	5.700806	271.3249	rice
94	53	40	20.2777436	82.894086	5.718627	241.9742	rice
89	54	38	24.5158807	83.535216	6.685346	230.4462	rice
68	58	38	23.2239739	83.033227	6.336254	221.2092	rice
91	53	40	26.5272351	81.417538	5.386168	264.6149	rice
90	46	42	23.9789822	81.450616	7.502834	250.0832	rice
78	58	44	26.800796	80.886848	5.108682	284.4365	rice
93	56	36	24.0149762	82.056872	6.984354	185.2773	rice

TABLE II. DATASET FOR SOIL MOISTURE PREDICTION[16]

Te	Press	Altit	Su		Ra	Sc	S
29.1	9984.5	-12.21	6.3	68	0	377	0
29.08	9984.4	-12.22	9.7	80	3.6	379	0
29.06	9984.6	-12.2	3.3	82	3.6	376	0
29.05	9984.4	-12.22	9.1	62	39.8	377	0
29.03	9984.4	-12.21	10.6	68	2.8	379	0
29.02	9984.6	-12.2	8.2	70	0	376	0
29	9984.4	-12.21	8.4	63	0.2	380	0
28.99	9984.3	-12.23	4.6	65	0	380	0
28.97	9984.1	-12.24	4.1	70	0	380	0
28.96	9984.1	-12.24	7.7	82	16.2	379	0
28.95	9984.3	-12.23	11.9	74	0	379	0
28.94	9984.1	-12.24	12.5	54	0.2	378	0
28.92	9984	-12.25	13	62	0	379	0
28.91	9984.2	-12.24	12.4	67	0	382	0

V. METHODOLOGY

A more sophisticated grasp of the distribution of data was made possible by the insights from Seaborn visualizations, which were crucial in directing feature selection. Seaborn visualizations provided valuable insight into feature selection, a crucial stage that took into account the model's impact and relevancy. The dataset was improved by additional pre-processing steps like variance reduction and standardization. Normalization guaranteed that each feature contributed uniformly, which is important for algorithms that are sensitive to scale, and variance reduction streamlined features by eliminating those that had little bearing. Regression and classification methods together demonstrated adaptability depending on the objective variable in model training. Your thorough explanation of model training and data feature extraction demonstrates a deep comprehension of these procedures, which is essential for interpreting the results.

Flow Chart:

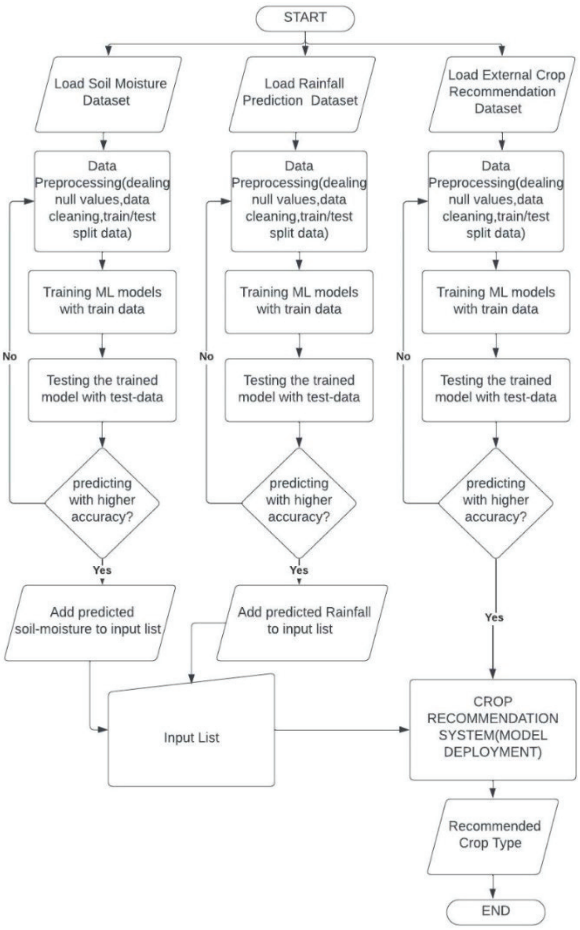

Fig. 1. Flow chart

A. Removing Null Values& Addressing Missing Values:

The first phase of data preprocessing was a thorough handling of missing values, with a special emphasis on null value removal and instances of missing data. Eliminating null values was essential to guaranteeing a reliable dataset for analysis. Several techniques were used to deal with missing values. To deal with null entries, imputation techniques such as mean, median, or mode imputation and constant value replacement were applied. Advanced techniques were also taken into consideration, like imputing the most frequent value, using random values, or using algorithms to forecast and fill in missing data. The distribution of the data and its effect on maintaining the integrity of the dataset for upcoming studies determined the technique to be used [4].

B. Characteristic Illustration:

To maximize the classification of ground states, One Hot Encoding was essential. This method assigned distinct binary values to each category to translate the categorical output into a format that was machine-readable. More specifically, model comprehension was streamlined by representing Dry, Very Dry, Wet, and Very Wet as 0, 1, 2, and 3, respectively. The smooth integration of categorical data made possible by this technique allowed machine learning models to classify and comprehend soil states according to the numerical values that were assigned [12].

C. Exploratory Data Analysis:

Understanding the complex correlations between soil properties required the application of exploratory data analysis or EDA. The comprehensive visualization made possible by Python's Seaborn package allowed for perceptive insights into the relationships between temperature, pressure, altitude, and soil moisture. EDA provided a thorough understanding of these variables' effects on soil state classification by revealing patterns, distributions, and possible linkages. Histograms, correlation matrices, and scatter plots were used to reveal important details about the properties of the dataset. Finding patterns, nonlinear relationships, and outliers was made easier with the help of these representations [4]. EDA functioned as a fundamental stage, offering crucial perspectives that

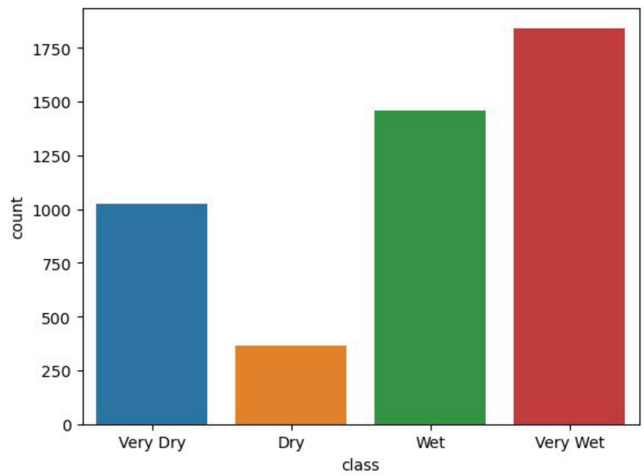

Fig. 2. Visualization of Soil Moisture

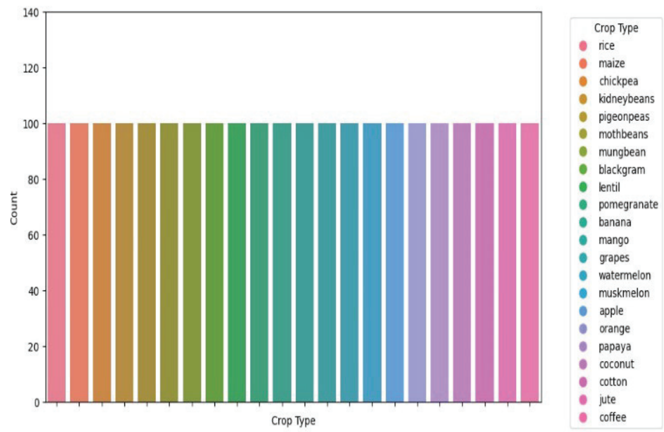

Fig. 3. Visualization of Crop Data

influenced later preprocessing stages and the feature engineering procedure, augmenting the efficacy of the machine learning models in precisely classifying soil conditions according to the multidimensional input parameters [9].

D. *Correlation of Columns:*

Investigating the relationships between columns was essential to comprehending how soil characteristics are interdependent. The correlations between rainfall, temperature, pressure, altitude, and soil moisture were explained using correlation matrices and statistical metrics. The degree and direction of linear correlations were revealed by this study, which helped to uncover important parameters for the classification of soil states and crop types. Comprehending these relationships played a pivotal role in the process of feature selection and model construction, guaranteeing the incorporation of relevant variables to enable precise soil state classification by the machine learning models [3].

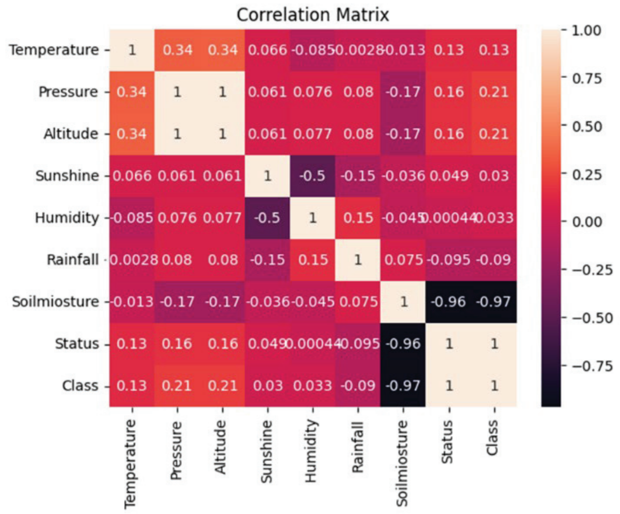

Fig. 4. Correlation Matrix of Soil Moisture

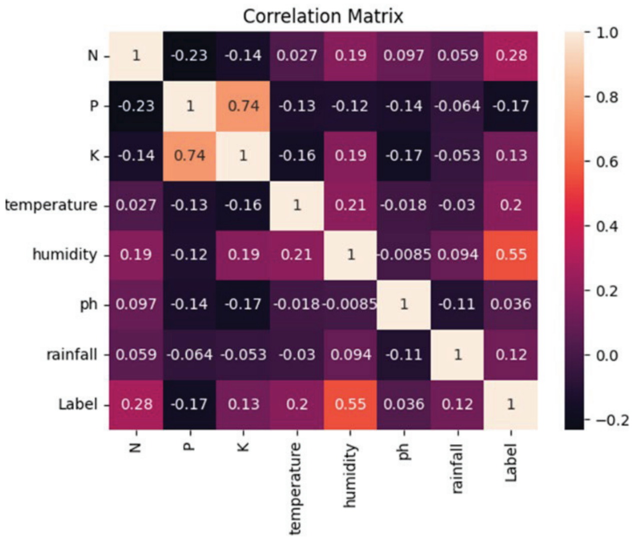

Fig. 5. Correlation Matrix of Crop Prediction

E. *Data Splitting and Standardization:*

The dataset was separated into subgroups for training and testing to facilitate the evaluation of the model. The division guaranteed unbiased model validation and assessment. Concurrently, data standardization methods were utilized to scale numerical attributes, guaranteeing consistency and reducing the influence of disparate scales across variables [11]. By making characteristics more comparable and avoiding some traits from predominating over others, standardization improved model performance. By ensuring that the machine learning models could efficiently learn from the training data and generalize well to new, untested data, this critical testing stage improved the process of classifying soil states. [11-13].

VI. RESULTS AND DISCUSSIONS

A. *Model Selection:*

Four distinct models were selected for training and evaluation:

1) *Linear Regression:*

A basic predictive modeling method correlated soil conditions with factors like temperature, pressure, height, and moisture to establish linear relationships. Using multidimensional data, the model forecasted and classified soil conditions based on these correlations.[1].

2) *Decision Tree:*

A decision tree predictive model connects features to inferences about a target variable by recursively segmenting data into subgroups based on key characteristics. Each node represents decision-making based on features, with branches indicating possible outcomes. [6].

3) *Random Forest algorithm:*

During training, a random forest ensemble technique constructs numerous decision trees, amalgamating them for a more accurate and reliable forecast. Each tree is trained on a random data subset, utilizing random features at each node. [13].

4) *K- Nearest Neighbours (KNN):*

K Nearest Neighbours (KNN) is a versatile machine learning method for regression and classification tasks. It relies on the principle that similar data points yield similar outcomes. By majority vote of an object's k-nearest neighbours, KNN determines its classification, a user-defined parameter. It's non-parametric, flexible, and applied in various domains like recommendation systems, image, and speech recognition.

B. *Prediction:*

1) *Rainfall Prediction:*

During the prediction stage, initially, we collected historical data from 1901 to 2022 of Telangana. We trained the models with three different algorithms, out of which the Linear Regression Algorithm gave us the best results [14].

2) *Soil Moisture and Crop Prediction:*

During the prediction stage, the three different models—Decision tree, KNN, and Random Forest—were each given the test data independently. By using the patterns, they had discovered during the training phase on the given test data, these models were able to forecast the soil conditions. [10]. Following the creation of the predictions, the accuracy, precision, recall, and F1-score of each model were carefully evaluated using metrics derived from the confusion matrix.

This evaluation offered information on how well, the models classified soil states, allowing for a comparative study to identify the most trustworthy model for soil state classification [9].

C. Model Evaluation:

1) Rain Fall Prediction:

TABLE III. ACCURACY OF DIFFERENT MODELS FOR RAINFALL PREDICTION.

MODELS APPLIED	R2 Score (%)
KNN Regression	92
Random Forest Regressor	97
Linear Regression	98

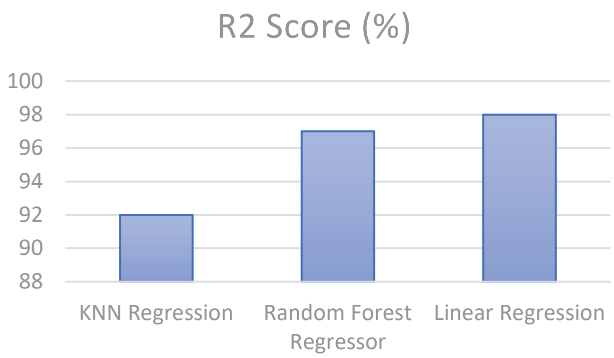

Fig. 6. Accuracy of Algorithms for Rainfall Prediction

2) Soil Moisture Prediction:

TABLE IV. ACCURACY OF DIFFERENT MODELS FOR SOIL MOISTURE

MODELS APPLIED	ACCURACY (%)
Naïve Bayes Classifier	94
Logistic Regression	97
Random Forest Classifier	99

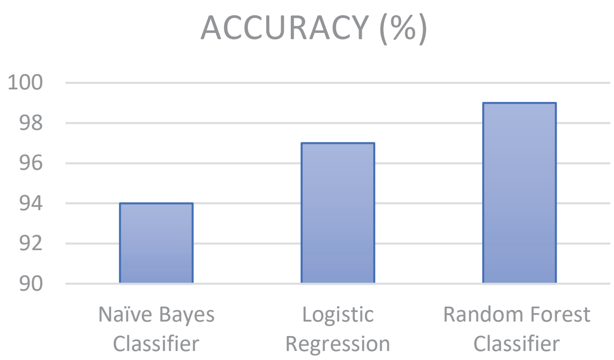

Fig. 7. Accuracy of Algorithms for Soil Moisture

3) Crop Prediction (Recommendation):

TABLE V. ACCURACY OF DIFFERENT MODELS FOR CROP PREDICTION

MODELS APPLIED	ACCURACY (%)
Decision Tree Classifier	93.08
KNN Classifier	94
Random Forest Classifier	99.4

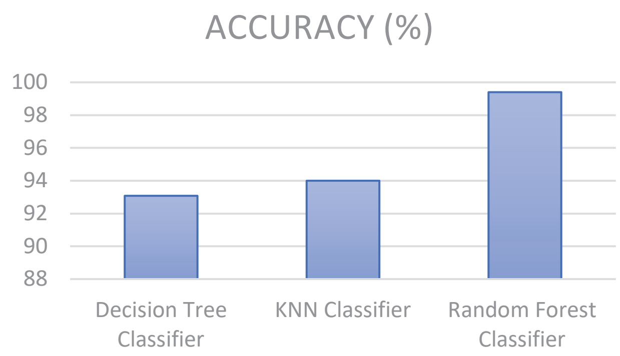

Fig. 8. Accuracy of Algorithms for Crop Prediction

Below Image represents the confidence values for some crops with which the model predicts.

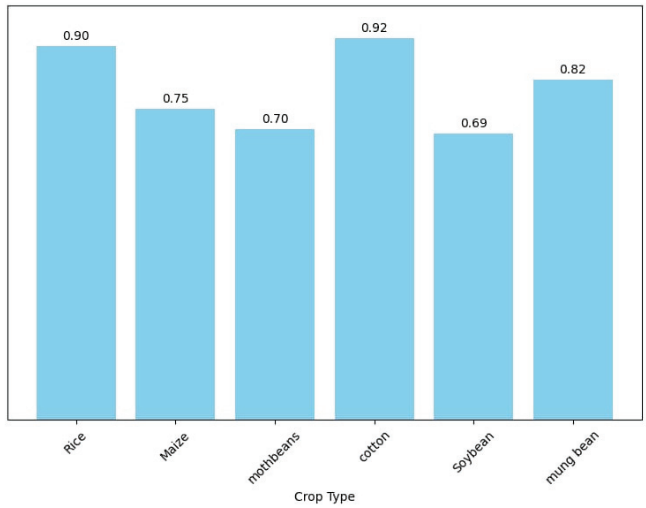

Fig. 9. Confidence values for selected crops

VII. CONCLUSION

This research article examines the soil and crop type based on rainfall, nutrients, temperature, humidity, sunshine, pressure, altitude, and soil moisture using machine learning techniques. When it comes to retrieving real-time data on rainfall, nutrients, temperature, humidity, sunshine, pressure, altitude, and soil moisture, the project is highly accurate and efficient. For Rain Fall Prediction Linear Regression gave us best results with accuracy 98%, For Soil Moisture Prediction and Crop Recommendation Random Forest Classifier gave us best results with accuracies 99% and 99.4% respectively. The project will help farmers increase agricultural yield and effectively manage food production because of the continuous assistance in providing them with accurate real-time feeds of environmental factors like rainfall and soil moisture. The research provides a reasonable agricultural model in the agricultural industry, which has historically valued machine learning. Due to the highly localized nature of agriculture and the unique conditions involved, this is an extremely difficult task. It is anticipated that having access to full real-time and historical environmental data will facilitate effective resource management and utilization.

REFERENCES

[1] Ren, Ye, Le Zhang, and Ponnuthurai N. Suganthan. "Ensemble classification and regression-recent developments, applications and

future directions." IEEE Computational intelligence magazine 11.1 (2016): 41-53.

[2] Shekhar, Yuthika, et al. "Intelligent IoT based automated irrigation system." International Journal of Applied Engineering Research 12.18 (2017): 7306-7320.

[3] Tantithamthavorn, Chakkrit, et al. "An empirical comparison of model validation techniques for defect prediction models." IEEE Transactions on Software Engineering 43.1 (2016): 1-18.

[4] Martinelli, Gaëtan, and Marc-Olivier Gasser. "Machine learning models for predicting soil particle size fractions from routine soil analyses in Quebec." Soil Science Society of America Journal 86.6 (2022): 1509-1522.

[5] Payen, Florian Thomas, et al. "Predicting the abatement rates of soil organic carbon sequestration management in Western European vineyards using random forest regression." Cleaner Environmental Systems 2 (2021): 100024.

[6] Liu, Dunwen, et al. "A GA-BP neural network regression model for predicting soil moisture in slope ecological protection." Sustainability 14.3 (2022): 1386.

[7] Han, Jichong, et al. "Prediction of winter wheat yield based on multi-source data and machine learning in China." Remote Sensing 12.2 (2020): 236.

[8] Kuwata, Kentaro, and Ryosuke Shibasaki. "Estimating crop yields with deep learning and remotely sensed data." 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2015.

[9] Maimaitijiang, Maitiniyazi, et al. "Soybean yield prediction from UAV using multimodal data fusion and deep learning." Remote sensing of environment 237 (2020): 111599.

[10] Keerthan Kumar, T. G., C. and Shubha, and S. A. Sushma. "Random forest algorithm for soil fertility prediction and grading using machine learning." Int J Innov Technol Explor Eng 9.1 (2019): 1301-1304.

[11] Benedet, Lucas, et al. "Rapid soil fertility prediction using X-ray fluorescence data and machine learning algorithms." Catena 197 (2021): 105003.

[12] Muruganantham, Priyanga, et al. "A systematic literature review on crop yield prediction with deep learning and remote sensing." Remote Sensing 14.9 (2022): 1990.

[13] Reddy, T. Vasudeva, et al. "Classification and detection of skin diseases based on CNN-powered image segmentation." 2023 3rd International Conference on Intelligent Technologies (CONIT). IEEE, 2023.

[14] Vasu Deva Reddy, T., et al. "COVID-19 Recognition in X-RAY and CTA Images Using Collaborative Learning." Artificial Intelligence for Smart Healthcare. Cham: Springer International Publishing, 2023. 431-445.

[15] Reddy, T. Vasudeva, et al. "Check for updates Design and Analysis of CNN-Based Skin Disease Detection System with Preliminary Diagnosis." Proceedings of the Second International Conference on Emerging Trends in Engineering (ICETE 2023). Vol. 223. Springer Nature, 2023.

[16] Data sets for soil moisture, rainfall, and Crop prediction:

https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset, https://www.kaggle.com/datasets/rajanand/rainfall-in-india,

https://www.kaggle.com/datasets/uttamkumar15802/soil-moisture-prediction.