**Data-Driven Agro-Informatic System for Sustainable Farming Solutions**

**Monika R. 1 , Dr. Nirmala Sugirtha Rajini S. 2**

1. **PG Student 2. Professor**
2. **monikaramesh2002@gmail.com 2.nirmalasugirtharajini.mca@drmgr.ac.in**

**Department of Computer Applications**

**Dr. M.G.R. Educational and Research Institute, Chennai – 95.**

**ABSTRACT**

Crop production needs to double by the year 2050 to meet the food requirements of the growing global population. Proper management of fertilization and timely disease control are essential in this endeavour. Sophisticated techniques for providing crop and fertilizer guidance, along with rapid detection of plant diseases, are pivotal. Machine learning technologies are poised to address these challenges by evaluating various elements, including geographical location, soil acidity, and the presence of essential nutrients such as nitrogen, phosphorus, and potassium. Our expectation is that, by incorporating critical inputs from users such as soil nutrient content and crop images, we can achieve the most accurate predictions. By uniting third-party apps that offer live weather data, and gaining insights into soil types, nutrient concentrations, precipitation patterns, and soil concentration, we are able to assemble a thorough dataset. This dataset affords the training of machine learning models with suitable algorithms, crafting systems that can precisely advise on the optimal crops and fertilizers and identify diseases with finer accuracy. The proposed system is designed for efficient crop management. Following an evaluation of accuracy, we have selected the Random Forest Algorithm, which demonstrates 99.03% accuracy in most instances within our model. These sophisticated systems are indispensable tools for farmers aiming to increase their crop yields and economic viability.

**Keywords :** Crop prediction, Fertilizer recommendation, Disease detection, Machine Learning, Random Forest, SVM, Decision Tree, Logistic regression.

**I. INTRODUCTION**

Anticipated to significantly influence Smart Farming and encompass the entire supply chain, especially in rice cultivation, are the domains of Big Data, Machine Learning, and the Internet of Things. The burgeoning volume and diversity of data harnessed by these innovative technologies through IoT grant new predictive and analytical capabilities to strategies in rice smart farming. The calibre of data gleaned from sensory devices plays a pivotal role in the efficacy of model development employing ML techniques [1].

Recent advancements, along with the effects of the Green Revolution occurring from the 1990s to the 2010s, have led to a boost in Bangladesh's agricultural output, which has been crucial for food security, despite persistent yield gaps. However, the agricultural sector in Bangladesh continues to rely heavily on intensive inputs, such as agrochemicals (including fertilizers and pesticides), the introduction of modern crop varieties, and expanded irrigation, as well as practices like inversion tillage. This traditional reliance on chemicals for small-scale farming risks harming the environment, causing soil degradation, increasing soil acidity, and reducing soil fertility over time [2].

Models for crop production have been crafted and are extensively utilized for hypothesizing, conducting simulated experiments, and executing assessments involving various scenarios and risks across diverse scales, thereby amplifying our scientific comprehension of the intricate interplay among soil, crops, the environment, and management strategies. Crafting fertilizer guidance that is customized to particular crop types, climatic variables, and soil fertility levels, along with taking into account the socio-economic circumstances of farmers, can enhance yields while mitigating climate-induced production uncertainties and lessening the environmental repercussions associated with fertilizer use [3].

Conventionally, the identification and diagnosis of plant diseases rely on visual inspections conducted by specialists. Such a process is time-consuming and dependent on expertise, rendering it impractical for the surveillance of expansive agricultural areas. Consequently, there is a necessity for automated mechanisms for the continuous observation and predictive analysis of crop health to surmount the constraints of manual detection [4].

Historically, yield estimation hinged on a farmer's accumulated knowledge, but with the potential for unpredictable weather variations, such intuitive forecasting is unreliable. Technological advancements can assist in determining whether a crop is viable by projecting yield outcomes. Machine learning models are capable of recognizing trends related to crop growth under various environmental scenarios and can therefore estimate the expected yield for a specific planting region [5].

**II. LITERATURE SURVEY**

Chakraboty et al. (2021) emphasizes a case study of farmers using traditional methods, struggle to choose the most suitable crops for their land, leading to severe consequences such as suicide, abandoning agriculture, or migrating to urban areas for livelihood. To address this, proposed system has been designed that aids farmers in crop selection by considering crucial factors like soil conditions, sowing season, and geographical location [6].

Pande et al. (2021) proposed user-friendly yield prediction system accessible through a mobile application. Using GPS for location identification, farmers input area and soil type, and machine learning algorithms, including Random Forest with 95% accuracy, predict the most profitable crops or estimate yields for selected crops. The system also recommends optimal fertilizer application timings to enhance crop yield [7].

Kalimuthu et al. (2020) built the model for crop prediction which can be carried out by using the machine learning, specifically the Naive Bayes algorithm, to assist novice farmers in choosing suitable crops based on collected seed data and essential parameters like temperature, humidity, and moisture content that is being developed using an Android mobile application offers a user friendly interface, allowing users to input parameters for automated location-based crop predictions [8].

Chohan et al. (2020) proposed a plant disease detection model, employing deep learning techniques and referred to as the "plant disease detector." The model demonstrates proficiency in identifying various plant diseases by analyzing leaf images. The methodology involves the development of a neural network-based plant disease detection model. Initially, dataset augmentation is applied to enhance the sample size. Subsequently, a Convolutional Neural Network (CNN) is implemented, incorporating multiple convolution and pooling layers. The training of the model utilizes the PlantVillage dataset. Following the training phase, rigorous testing is conducted to validate the accuracy and effectiveness of the model's disease detection capabilities [9].

Reddy et al. (2020) presents a major exporter of farm products, faces productivity challenges impacting farmers incomes, crop recommendation algorithms are implemented to the system that can enhance land cultivation by suggesting suitable crops based on environmental factors. This study explores how machine learning algorithms can forecast crops and prices, offering a solution to farmers’ confusion. Additionally, global soil deterioration poses a threat to food security due to urbanization, industry, habitat loss, and ineffective land management, emphasizing the crucial role of soil in sustaining the ecosystem [10].

**III. PROPOSED SYSTEM**

Data analysis technology is pivotal for monitoring and predicting crop yield changes. This system empowers farmers to select optimal crops based on diverse land conditions. Using Decision Tree, Support Vector Machine, Logistic Regression, and Random Forest algorithms, the system excels in accurate predictions. Trained on a vast dataset encompassing historical crop yield, land conditions, and weather patterns, these algorithms become valuable tools for forecasting future yields and fertilizer needs. We compare algorithm performance, selecting the most accurate one for effective forecasts. The developed image classification, utilizing ResNet, aids in disease detection. Trained intensively on a large dataset containing diverse plant leaf disease information, this technology offers a breakthrough in precise crop and fertilizer prediction and disease detection. Farmers can then take informed actions based on this detailed information.



**Figure 1. Flow diagram of the proposed system**

**INPUT PARAMETERS :**

The dataset comprises parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), soil pH value, Humidity, Temperature, and Rainfall. The data were obtained from the Kaggle website and consist of 2200 instances or data points extracted from historical records. This dataset encompasses different crops, including rice, maize, chickpea, kidneybeans, pigeonpeas, mothbeans, mungbean, blackgram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, coffee.

**MODULES :**

In this paper, we have proposed three modules.

**1. Crop prediction**

This model requires historical data on crop yields, weather patterns, soil conditions, such as Nitrogen, Phosphorus, Potassium, pH level, Rainfall, State and location. This data helps the model to entity patterns and trends helping farmers to suggest best crop.

**2. Fertilizer Recommendation :**

This model is used to recommend fertilizers to crop based on user inputs like soil nutrients (N, P, K) and the type of crop which the farmer wants to grow.

**3. Disease prediction :**

This model is designed to identify the type of the plant leaf disease by using ResNet image processing technique. Firstly, Users must upload images of infected plants. It processes the data set and then indicates the disease of the plant using ResNet image processing technique. ResNet or Residual Network is specifically designed for tasks like image classification which helps in mitigating the vanishing gradient problem and facilitates the training of very deep networks.

**ALGORITHMS USED :**

1. **Decision Tree**

The primary machine learning method employed in constructing our targeted crop system is the Decision Tree. This nomenclature is derived from its graphical representation and hierarchical structure. It encompasses two key node types: decision nodes, facilitating decision-making through branching, and leaf nodes, indicating final outcomes without branching. The accuracy of the model we tested stands at approximately 90.0%. While this accuracy is commendable, we opted against its selection due to suboptimal precision. Although the accuracy is satisfactory, the decision was influenced by the recognition that precision levels did not meet our desired standards.

#Decision Tree

from sklearn.tree import DecisionTreeClassifierDT =DecisionTreeClassifier(criterion="entropy",random\_state=2,max\_depth=5)DT.fit(Xtrain,Ytrain)

1. **SVM**

Support Vector Machine (SVM) is a machine learning method that establishes an optimal hyperplane or decision boundary to separate dimensional spaces into classes, aiding in categorizing new data in the future. Utilizing support vectors, this method creates a hyperplane, supported by two vectors on either side. These support vectors are lines drawn through two data points, closest to the hyperplane on each side. The model's accuracy stands at approximately 97.09%, surpassing the accuracy of the Random Tree algorithm.

#SVM

from sklearn.svm import SVCfrom sklearn.preprocessing import MinMaxScalernorm = MinMaxScaler().fit(Xtrain)X\_train\_norm = norm.transform(Xtrain)X\_test\_norm = norm.transform(Xtest)SVM = SVC(kernel='poly', degree=3, C=1)SVM.fit(X\_train\_norm,Ytrain)

1. **Logistic Regression**

Logistic Regression, a machine learning algorithm, aims to model the relationship between dependent and independent variables, particularly applied to solve categorical data problems. It presents a straightforward and efficient model for implementation.The accuracy of the logistic regression model is reported at 95.02%. While surpassing the accuracy of the Random Tree algorithm, it falls short compared to SVM. Consequently, this model is not considered for further use in our system.

#Logistic Regression

from sklearn.linear\_model import LogisticRegressionLogReg = LogisticRegression(random\_state=2)LogReg.fit(Xtrain,Ytrain)

1. **Random Forest**

Random Forest is a machine learning algorithm where, during the training phase, multiple decision trees are created, and the output is divided based on the number of classes for classification or prediction of class for regression. The accuracy of prediction is proportional to the number of trees. The dataset includes variables like rainfall, perception, temperature, and production, which are used for training. Only two-thirds of the dataset are considered, and the remaining portion is reserved for experimental purposes. The Random Forest algorithm has three parameters: n tree, indicating the number of trees to grow, m try – specifying how many variables to consider at a node split; and node size, suggesting the number of observations to consider in terminal nodes .

In our model, we have implemented Random Forest (RF) by:(i) Importing the RandomForestClassifier library from the sklearn.ensemble class.(ii) Creating an RF classifier object.(iii) Fitting our data in the last step.

# Random Forest

from sklearn.ensemble import RandomForestCl assifier

RF = RandomForestClassifier(n\_estimators=20, random\_state=0)

RF.fit(Xtrain,Ytrain)

 **ResNet**

Identifying crop diseases is essential for maintaining food safety, yet rapid detection continues to pose a global challenge due to insufficient infrastructure. As digital cameras proliferate worldwide and computer vision advances, automated disease detection techniques are increasingly sought after in precision agriculture, high-yield plant phenotype analysis, smart greenhouse applications, and more. Employing an open dataset comprising 15,200 crop leaf images, a Residual Network (ResNet-9) was trained for disease classification. The proposed ResNet-9 model achieved an impressive accuracy of 99.40% on the test set, highlighting its effectiveness. Overall, training ResNet models on open image datasets presents a robust approach to crop disease detection using automated networks on a global scale.

**IV. RESULTS AND DISCUSSIONS**

An exhaustive examination of the dataset was conducted, and diverse results were obtained during the testing phase to assess the accuracy of the model. This process aimed to formulate a system capable of recommending the optimal crop based on the gathered insights.

 **Evaluation Metrics**

|  |  |
| --- | --- |
| **ALGORITHM** | **ACCURACY (%)** |
| Decision Tree (DT) | 90.0 |
| Support Vector Machine (SVM) | 97.9 |
| Logistic Regression (LR) | 95.2 |
| Random Forest (RF) | 99.3 |

**Table 1: Algorithm vs. Accuracy Result in Percentage**

The above table shows the result of Accuracy between different algorithms in percentage. It is clear that Random Forest algorithm has the greatest accuracy of all.

**Experimental result**

****

**Figure 2: Accuracy Comparison**

Figure 2. demonstrates the machine learning model based on Random Forest demonstrated an impressive accuracy of approximately 99.03%.

**OUTPUT**

1. **CROP PREDICTION**

****

**Figure 3. Crop Recommendation**

Figure 3, demonstrates crop prediction that takes user input of soil data, such as nutrient values (N, P, K), pH level, rainfall, as well as location information (state and city) to predict the crop suitable to the land.

1. **FERTILIZER RECOMMENDATION**

****

**Figure 4. Fertilizer Recommendation**

In Figure 4, User gives Soil nutrient values like Nitrogen, Potassium, Phosphorus and the type of crop to be grown in the land to suggest the suitable fertilizer to be used for the crop.

1. **DISEASE DETECTION**

****

**Figure 5. Disease Detection**

In Figure 5 , Disease detection can be done by uploading images of the infected plant leaf. After uploading, the image classification technique called ResNet is used to detect what type of disease is the plant get infected.

**V. CONCLUSION**

Addressing challenges in India's crucial agricultural sector, the system provides personalized recommendations to optimize crop production, simultaneously increasing yield and reducing environmental impact. By empowering farmers with data-driven insights, the platform facilitates informed decisions on crop yields and estimates fertilizer costs. Objectives of this system include region-specific recommendations, with a focus on enhancing yield, reducing costs, and promoting sustainability. Thus with the help of this agro informatic system and integrating machine learning algorithms such as Random Forest, is poised to revolutionize precision agriculture. Through maximizing productivity, minimizing soil degradation, and optimizing fertilizer usage, this agro informatic system emerges as an indispensable tool, contributing significantly to the sustainable development of India's agriculture and economy.

 **REFERENCES**

[1] Alfred R, Obit J H, Chin C P Y, Haviluddin H and Lim Y (2021) , "Towards Paddy Rice Smart Farming: A Review on Big Data, Machine Learning, and Rice Production Tasks", IEEE , Vol. 9, pp. 50358-50380.

[2] Hossain M A and Siddique M N A (2020), “Online Fertilizer Recommendation System (OFRS): A Step Towards Precision Agriculture And Optimized Fertilizer Usage By Smallholder Farmers In Bangladesh: Online fertilizer recommendation”,  European Journal of Environment and Earth Sciences, Vol. 4(1).

[3] Jairos Rurinda, Shamie Zingore (2020), “Science-based decision support for formulating crop fertilizer recommendations in sub-Saharan Africa”, Agricultural Systems, Vol. 180, pp. 180-102790.

[4] Kartikeyan P and Shrivastava G (2021), “Review on emerging trends in detection of plant diseases using image processing with machine learning” , International Journal of Computer Application, Vol.975, pp.975-8887.

[5] Palaniraj A, Balamurugan A S, Prasad R D and Pradeep P (2021), “Crop and fertilizer recommendation system using machine learning”, IRJET, Vol 8.

[6] P. A, Chakraborty S, Kumar A and Pooniwala O R (2021), "Intelligent Crop Recommendation System using Machine Learning", 5th International Conference on Computing Methodologies and Communication (ICCMC)*,* pp. 843-848.

[7] Pande S M, Ramesh P K, Anmol A, Aishwarya B R, Rohilla K and Shaurya K (2021), "Crop Recommender System Using Machine Learning Approach",  5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, pp. 1066-1071.

[8] Kalimuthu M, Vaishnavi P and Kishore M (2020), "Crop Prediction using Machine Learning",  Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*,* pp. 926-932.

[9] Chohan M, Khan A, Chohan R, Katpar S H and Mahar M S (2020), ” Plant disease detection using deep learning”, International Journal of Recent Technology and Engineering, 9(1), pp.909-914.

[10] Reddy P S, Amarnath B and Sankari M (2023), "Study on Machine Learning and Back Propagation for Crop Recommendation System*",* International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*,* pp. 1533-1537.

11. For Data sets: <https://www.kaggle.com/code/nirmalgaud/crop-recommendation-system-using-machine-learning>.