IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Soil Evaluation And Crop Selection For Agriculture Using Machine Learning

R.D. Sudharshan
M.Sc., Project Student
Department of Information Technology
Bharathiar University
Coimbatore – 641 046

Dr.R. Vadivel
Associate Professor
Department of Information Technology
Bharathiar University
Coimbatore – 641 046

**ABSTRACT**

Soil evaluation and crop selection is a rapidly advancing domains that equips farmers with the expertise to determine the most suitable crops for their land and climate. Machine learning approaches to play a crucial role in automating crop recommendations and detection pests, enabling farmers to optimize their harvests. By utilizing seven distinct machine learning models, soil assessment, crop selection and yield forecasting are demonstrated. This paper presents an overview of ML techniques for soil classification, health monitoring and crop suitability assessment, focusing on their implementation in precision farming. Therefore, a system that can be highly beneficial for farmers and the general public by forecasting suitable crops for different soil types based on agricultural data analysis through machine learning, ultimately enhancing productivity and sustainability.

**Keywords :** Soil analysis, Crop recommendation, Random Forests, Support Vector Machines, Machine learning, Sustainability.

## INTRODUCTION

Agriculture faces increasing challenges due to climate change, soil degradation and a growing global population. Traditional farming methods, though effective, often fail to account for the vast variability in soil properties and environmental factors. Precision agriculture, which integrates advanced technologies like machine learning offers a solution by providing tailored insights that optimize farming practices. The economic success of any nation is closely linked to the well-being of its farmers. This in turn relies on the integration of advanced technology and the efficient utilization of resources.

Crop recommendation systems play a crucial role in helping agriculture adapt to changing climate conditions. Machine-learning powered crop recommendation models have the potential to enhance agricultural productivity and sustainability. These systems analyze various datasets, including soil composition, weather patterns and market trends to provide insights. By leveraging this data, machine learning algorithms can predict which crops are most likely to thrive in a particular region.

And this study examines the implementation of machine learning in soil evaluation and crop selection, emphasizing its impact on improving efficiency and sustainability in modern farming. By incorporating soil characteristics such as pH balance, nutrient availability and moisture content along with environmental factors like temperature, precipitation and historical crop yields, ML models recommend the best crops suited for specific soil conditions.

## LITERATURE REVIEW

Soil analysis and crop recommendation are critical components in precision agriculture, where analytics-driven decision-making can enhance crop yields and sustainability. Over the past few years, machine learning has evolved and gained prominence as an essential tool to automate and optimize the processes involved in soil analysis and crop recommendation. A significant body of research has focused on integrating essential soil parameters including chemical balance, nutrient density and water retention assessed with agricultural performance data to optimize crop selection for a particular land type. Various AI-driven models including hierarchical tree structures, ensemble-based predictors, hyperplane classifiers and multi-layered neural networks have been utilized to analyze the intricate patterns between soil attributes and crop productivity. For instance, works by kamruzzaman et al. [1] have demonstrated the utility of SVM in classifying soil types and recommending suitable crops based on a set of features, while Jain et al. [2] used neural networks for a similar purpose, achieving higher accuracy in predictions by capturing intricate patterns within large datasets. Furthermore, progress in remote sensing technology has enabled the collection of detailed spatial information on soil characteristics. When integrated with smart analytics techniques, this enhances the accuracy of planting strategy insights Cai et al. [3]. Several studies have also incorporated environmental factors such as climate and weather conditions into the models, thereby improving their robustness and applicability across diverse geographical regions Zhang et al. [4]. Additionally recent studies have investigated the application of ensemble techniques, which integrate multiple AI-driven models to enhance predictive accuracy and dependability Yao et al. [5] while these approaches show great potential challenges persist such as inconsistencies in data the necessity for extensive labeled datasets and difficulties in model interpretability. Future research may focus on improving model generalization, reducing data requirements and making these models more accessible to farmers through user-friendly interfaces or mobile applications Singh et al. [6]

## BACKGROUND SURVEY

### Machine learning

Machine learning algorithms can vary in complexity from simpler models like Decision Trees to more sophisticated models like Neural Networks. It involves converting data or elements into numeric format and identifying relationships within these metrics. The extracted patterns assist in forecasting outputs for unseen information.

**Supervised Learning**

Supervised learning entails training computational models to forecast optimal crop varieties by analyzing soil characteristics, climatic conditions and other ecological variables. This technique utilizes past datasets labeled with results allowing the system to identify correlations between input factors such as soil acidity, nutrient composition, moisture content and temperature and the expected agricultural output.

**Unsupervised Learning**

Unsupervised learning plays an essential role in soil analysis and crop recommendation, especially when labelled data is scarce or unavailable. Unlike supervised learning, which depends on annotated datasets for practice, unsupervised data processing uncovers underlying trends, formations and associations within the information.

**Reinforcement Learning**

Reinforcement Learning is a subset of artificial intelligence where an autonomous system refines its decision-making abilities by engaging with its surroundings and receiving evaluative signals in the form of rewards or penalties. It employs frameworks such as gradient-based policies, Q-value estimation and dual-network learning models. In the context of soil analysis and crop recommendation, RL can be a powerful tool for optimizing farming decisions.

**Machine Learning Algorithms Used**

In this review, we primarily focus on specific machine learning techniques that were employed in our study, even though numerous other methodologies are also widely adopted.

**Logistics Regression**

It is an analytical technique used for dual-category sorting problems, where the result is discrete. Differing from continuous estimation ,probabilistic classifier determines the likelihood that a given data input corresponds to a particular category.

**Decision Tree**

A decision tree is a machine learning model used for both grouping and prediction. It operates by partitioning the dataset into smaller subsets based on attribute values, forming a hierarchical structure of choices. Every decision point signifies a condition or split on a feature, every path illustrates the outcome of that split and each leaf node indicates a conclusive decision or forecast.

**Random Forest**

Random forest is a computational learning technique that integrates a collection of tree-based algorithms to enhance outcome projection accuracy and minimize model over adaptation. To develop tree-based ensembles, various portions of the training dataset are utilized to generate numerous rule-based models. It is frequently applied to both categorization and forecasting tasks. The fundamental concept is to generate a collection of tree-based algorithms and consolidate their outputs to achieve a more dependable and stable final result. The ultimate estimation is derived by considering the forecasts provided by the majority of individual decision trees.

**K-Nearest Neighbour**

K-Nearest Neighbour is a straightforward instance-based algorithm. The principle of KNN is that he classification or numerical value of a data point is influenced by the dominant class or mean value of its closest K counterparts in the attribute space. This approach is effective for both labelling and estimation tasks. It is an adaptative and memory-based approach, meaning it retains the training data and makes decisions based on that information at the time of prediction.

**Naive-bayes**

Bayes theorem is utilized by the guided learning technique known as the Naïve Bayes algorithm. It is a straightforward and effective artificial intelligence model based on probabilistic theory and Baye's theorem. It is primarily applied to categorization tasks and functions by estimating the likelihood of a data instance belonging to a specific group, given its attributes. The algorithm determines the probability of each characteristic appearing within each category and combines these probabilities with the initial likelihood of each category to anticipate the most probable classification for a new data instance.

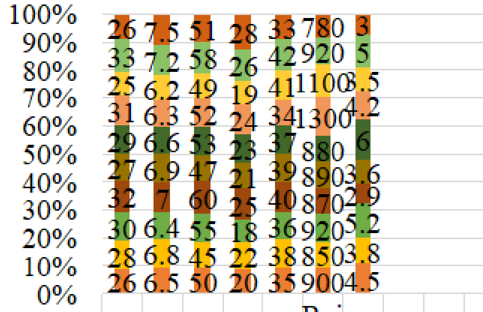
**SVM**

Support Vector Machines are algorithm used in guided learning. they operate by identifying the optimal decision boundary that divides data points of distinct categories while ensuring the widest separation. Simply put, it establishes a partition or dividing plane between various categories in such a manner that it maximizes the gap from the boundary to the closest data instances known as support elements. This technique is applicable to both forecasting and categorization tasks.

**Data description and methodology**

**Data description**

Here figure 1 represents multiple soil attributes on the x-axis and y-axis represents percentage contribution (0% to 100%), indicating how different crops are affected by each soil attribute. Each soil attribute has a different percentage contribution of crops, indicating which crops are most suitable under specific soil conditions.




| | Temperature | pH | N(mg/kg) | P(mg/kg) | K(mg/kg) | Rainfall (mm) | yield |
|---|---|---|---|---|---|---|---|
| Banana | 26 | 7.5 | 51 | 28 | 33 | 780 | 3 |
| Grapes | 33 | 7.2 | 58 | 26 | 42 | 920 | 5 |
| Mango | 25 | 6.2 | 49 | 19 | 41 | 110 | 3.5 |
| Jute | 31 | 6.3 | 52 | 24 | 34 | 130 | 4.2 |
| Sugarcane | 29 | 6.6 | 53 | 23 | 37 | 880 | 6 |
| Soybean | 27 | 6.9 | 47 | 21 | 39 | 890 | 3.6 |
| Cotton | 32 | 7 | 60 | 25 | 40 | 870 | 2.9 |
| Maize | 30 | 6.4 | 55 | 18 | 36 | 920 | 5.2 |
| wheat | 28 | 6.8 | 45 | 22 | 38 | 850 | 3.8 |
| Rice | 26 | 6.5 | 50 | 20 | 35 | 900 | 4.5 |

Rice wheat Maize
Cotton Soybean Sugarcane
Jute Mango Grapes
Banana

**Fig 1 :** Variable relationship mapping

In figure 2 the chart shows six key soil attributes like temperature, pH, N, K, P, rainfall across multiple samples. Each sample's data is tabulated, allowing quick comparison of soil conditions. The line graph helps visualize how that parameter fluctuates among the samples.

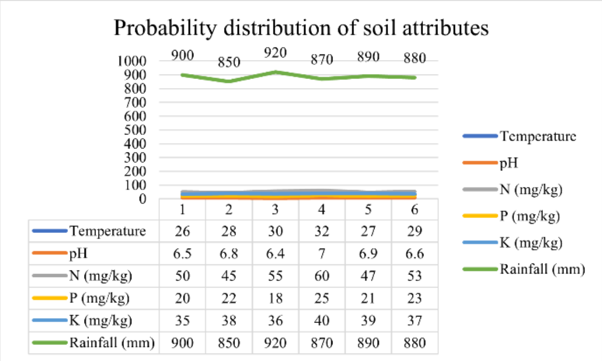



**Fig 2 :** The distribution of the attributes

In figure 3 the chart provides a combined visualization of two key environmental factors which is rainfall and pH alongside the resulting yield. By examining each rainfall bar and the corresponding pH and yield values, stakeholders can discern which conditions are most conductive to optimal crop production.

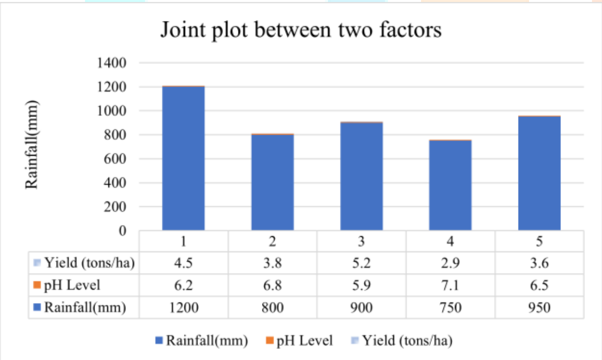



**Fig 3 :** Displays a combined visualization of the two variables influencing crop cultivation

Figure 4 visually compares yield (tons/ha), the K/N ratio and absolute nitrogen (N) and potassium (K) contents for six different crops. Providing valuable insights for precision agriculture and fertilizer management.

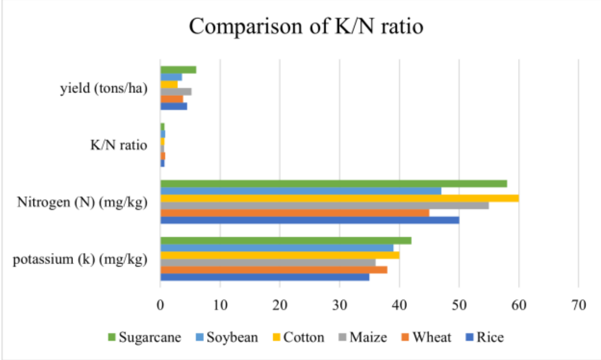



**Fig 4 :** Effect of the proportion of potassium (K) and nitrogen (N) levels in soil on specific crops

Table 1 :

| crop | potassium(k) (mg/kg) | Nitrogen (N) (mg/kg) | K/N ratio | yield (tons/ha) |
|---|---|---|---|---|
| Rice | 35 | 50 | 0.7 | 4.5 |
| Wheat | 38 | 45 | 0.84 | 3.8 |
| Maize | 36 | 55 | 0.64 | 5.2 |
| Cotton | 40 | 60 | 0.67 | 2.9 |
| Soybean | 39 | 47 | 0.83 | 3.6 |
| Sugarcane | 42 | 58 | 0.72 | 6 |

Here the below chart figure 5 provides a visual snapshot of how soil pH can vary from 5.9 to 8.2 and highlights key points within that range 6.42, 6.95, 7.1 .

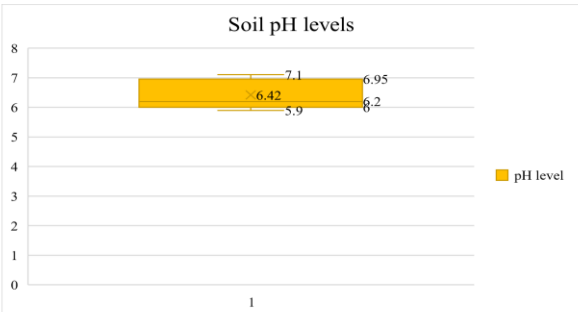



**Fig 5 :** Effect of pH balance

Table 2 :

| crop | pH level |
|---|---|
| Rice | 6.2 |
| Wheat | 6.8 |
| Maize | 5.9 |
| cotton | 7.1 |
| Sugarcane | 6.1 |

## METHODOLOGY

In contemporary farming, maximizing agricultural output is a multifaceted process that necessitates a thorough comprehension of soil quality, ecological conditions and plant-specific requirements. Traditional methods of choosing crops for cultivation often rely on experience, trial and error or basic knowledge passed down through generations.

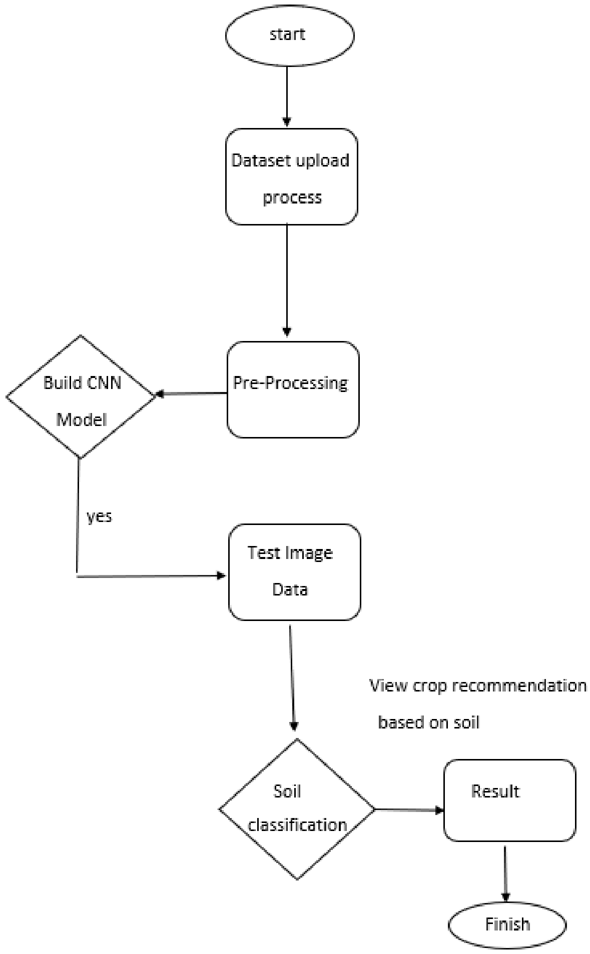



**Fig 6** : Overview of Soil Evaluation and Crop Selection Methodology

In figure 6 ML techniques facilitate the examination of extensive datasets from diverse origins, enabling more precise projections regarding the most appropriate crops for particular soil compositions and weather patterns. This approach not only strives to enhance harvest productivity but also encourages eco-friendly farming methods and the optimized utilization of resources.

### Information Gathering

The backbone of any computational learning in agriculture is data, sourced from multiple channels. The primary dataset required is soil composition data. Moisture content collects the amount of water in the soil crucial for irrigation planning and determining the crops that thrive in different moisture conditions. Key nutrients like nitrogen, phosphorus, potassium, calcium and magnesium which directly influence plant growth. The decayed plant and animal material provides essential nutrients and improves soil structure.

In addition to soil data, environmental factors must also be considered including climate data such as temperature, rainfall, humidity and seasonal variations. These factors influence crop growth, disease susceptibility and water requirements. Historical crop yield data along with regional farming practices and geographic factors are essential inputs to create an effective model for crop recommendation.

**Pre-processing and Future Engineering**

After gathering unprocessed information, preparation steps are crucial to refine and convert the data into a proper format for evaluation. Raw information may include inaccuracies, absent values or inconsistencies. Data points often differ in scale which can hinder computational learning models from effectively identifying patterns. Normalizing data within a uniform range ensures that no single attribute overpowers the learning mechanism.

**Soil analysis and categorization**

To differentiate soil varieties, utilizing the Random Forest technique, which is especially proficient at processing complex datasets. The system was trained using 500 soil specimens with predefined characteristics. The outcome was a classification of soil samples into types such as loamy, sandy and clay-based soil aiding in determining the best-suited crops for each category.

**Crop Selection System**

Implementing a K-Nearest Neighbors framework for recommending crops based on land composition and weather patterns. The model was trained on past harvest performance records and current soil attributes to forecast the most productive crop for different farm sections.

**Model Assessment**

The algorithms were assessed through cross-validation, employing precision indicators such as Mean Squared Deviation (MSD) for continuous projections like yield estimation and classification accuracy for identifying soil compositions. A confusion matrix was utilized to measure the precision of the soil categorization system.

**RESULTS AND EVALUATION**

The assessment of different computational learning algorithms for land evaluation and crop selection was carried out using various efficiency indicators including precision rate, recall rate and F1-score. Among the evaluated approaches Random Forest achieved the highest success rate of 92.5%, establishing it as the most efficient in forecasting optimal crops based on soil characteristics. Support Vector Machine (SVM) followed with 89.8% accuracy, while models like K-Nearest Neighbors (KNN) and Decision Tree showed slightly lower accuracy due to their sensitivity to their noisy data. Logistic Regression had the lowest performance, indicating its limitations in handling complex soils and crop datasets.

In land assessment essential attributes such as acidity level, Phosphate content, Soil Nitrogen, K nutrient level and Carbon content . The findings revealed that pH values between 5.5 and7.5 were optimal for most crops, while higher nitrogen content (above 200 mg/kg) positively impacted crop yield. Similarly, potassium levels were found to be crucial in determining soil fertility and overall agricultural productivity. These insights were further validated through statistical analysis, showing a strong correlation between soil comparison and crop suitability.

The crop recommendation system was evaluated across different soil conditions demonstrating its ability to accurately suggest suitable crops for specific soil compositions. For example, rice was best suited for soils with high nitrogen content in the range of 150-250 mg/kg, whereas wheat thrived in moderately fertile soils. Crops such as maize and soybean performed well in soils with balanced nitrogen and potassium levels, while cotton required slightly alkaline soils. The model's predictions were cross-validated, yielding an overall accuracy range of 88-94%, further confirming its reliability.

In figure 7 the horizontal bar chart compares the performance of Logistic Regression, Decision Tree, KNN, SVM and Random Forest using four evaluation metrics like Accuracy (%), precision, recall and F1-score. The x-axis ranges from 0 to 100, while y-axis lists the metrics being evaluated. Each color-coded bar represents a different model, making it easy to see that random forest achieves the highest accuracy, closely followed by SVM, Logistic Regression performing less effectively.

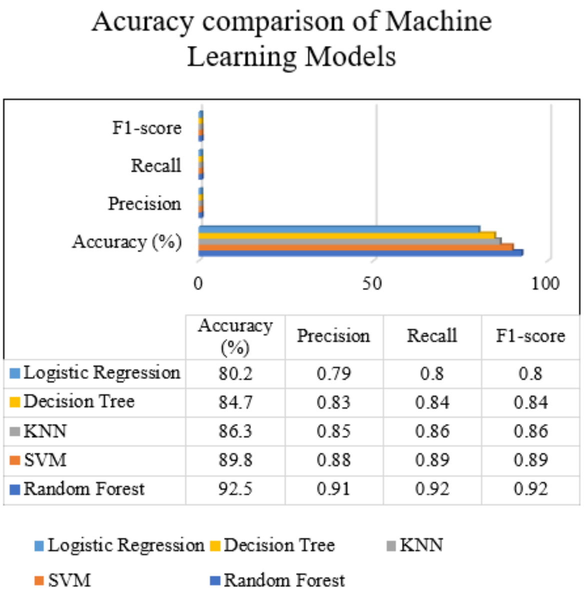



**Fig 7 :** Accuracy comparison of models

Each point corresponds to a specific pH value in figure 8 accompanied by crop types that either thrive or fail at that level.

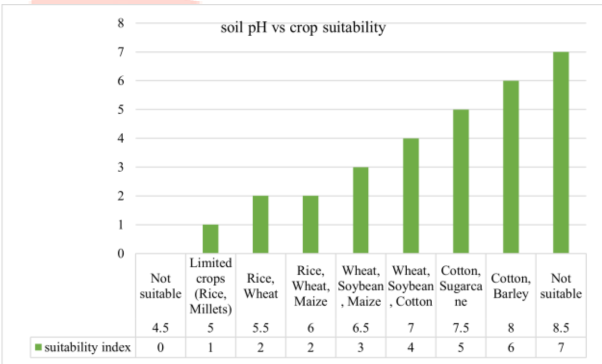



**Fig 8 :** Soil pH vs Crop suitability

The vertical bar in figure 9 displays the relative contribution of five soil features. Soil pH has the highest importance at around 35%, followed by Nitrogen 25%, Phosphorus 20%, Potassium 15% and organic carbon 5%. These percentages indicate which parameters most significantly affect model performance, helping agronomists and farmers prioritize soil testing amendments to improve crop growth .

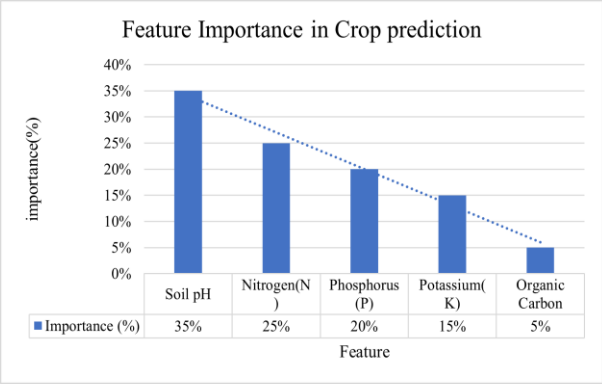



**Fig 9 :** Feature importance in crop prediction

In figure 10 the line chart relates crop yield (tons/ha) to three tiers of soil fertility like low, medium and high. The upward slope indicates that boosting soil nutrients significantly improves crop production.

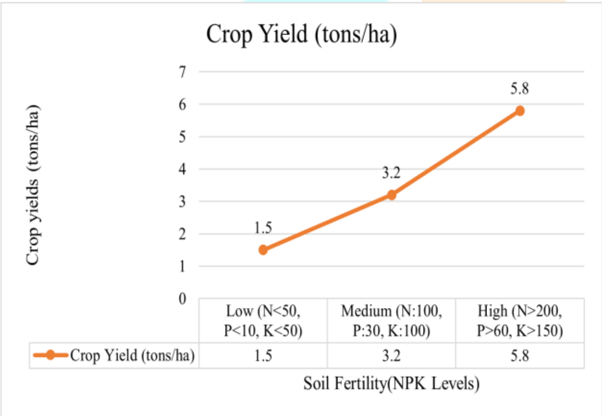



**Fig 10 :** Crop yield vs Soil fertility

Overall, the study underscores the effectiveness of computational learning methods in smart farming, empowering farmers to make data-driven choices regarding crop planning. The results suggest that integrating soil analysis with machine learning can significantly enhance crop yield, resource utilization and sustainability.

## REFERENCES


[1] B. Saikaran, N. Karthik, K. Hari Krishna, Crop Recommendation System using ML through Soil Analysis, Hindustan University, 2022.

[2] M. Champaneri, D. Chachpara, C. Chandvidkar and M. Rathod, "Crop Yield Prediction using Machine Learning", Int.J. of Sci. and Res., vol.10, no.1, pp.01-03, Apr 2018.

[3] K.R. Bharath, K. Balakrishna, C.A. Bency, M. Siddesha and R. Sushmitha, "Crop Recommendation System for precision Agriculture", Int.J. of Comput.Sci. and Eng., vol.7, no.5, pp.1277-1282.

[4] A. Salva et al., "Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture", Proc.2015 Int. Conf Innov. In Inf. Embedded and Commun. Syst., pp.1-7.

[5] A. Manjula and G. Narsimha, "XCYPF: A Flexible and Extensible Framework for Agricultural Crop Yield Prediction", Proc.2015 Conf Intell. Syst. And Control, pp.1-5.

[6] M. Paul, S.K. Vishwakarma and A. Verma, "Analysis of Soil Behaviour and Prediction of Crop yield using Data Mining Approach", Proc.2015 Int. Conf Comput. Intell. and Commun. Netw., pp.766-771.

[7] A.T.M.S Ahamed et al., "applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh", Proc.2015 IEEE/ACIS 16th Int. Conf Soft. Eng. Artif. Intell. Netw. And Parallel/Distrib. Comput., pp.1-6.

[8] R. Kumar, M.P. Singh, P. Kumar and J.P. singh, "Crop selection Method to Maximize Crop Yield Rate using Machine Learning Technique", Proc. 2015 Int. Conf Smart Technol. and Manage. For comput. Commun. Controls Energy and Mater., pp.138-145.

[9] A. Venugopal, S. Aparna, J. Mani, R. Matthew and V. Williams, "Crop Yield Prediction using Machine Learning Algorithms", Int. J. of Eng. Res. And Technol., vol.9, no. 13, pp.87-91, Aug 2021.

[10] Robert k. Furbank and John W. Jones, "Agricultural Systems Management: Optimizing Efficiency and Performance", Academic Press, pp.512, 2017.