Precision farming practices with data-driven analysis and machine learning-based crop and fertiliser recommendation system

G Vijender Reddy[1*], *M* Venkata Krishna Reddy[2], *K*. Spandana[2], *yerragudipadu* subbarayudu[3], *Ali* Albawi[4], *Rakesh* Chandrashekar[5], *Atul* Singla[6] and Praveen[7]

[1]Department of Information Technology, GRIET, Hyderabad, Telangana, India.
[2]Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad, Telangana, India.
[3]Department of CSE-AI&ML, KG Reddy College of Engineering and Technology, Hyderabad, Telangana, India.
[4]Radiology Techniques Department, College of Medical Technology, The Islamic University, Najaf
[5]Department of Mechanical Engineering, New Horizon College of Engineering, Bangalore, Karnataka, India.
[6]Lovely Professional University, Phagwara, Punjab, India
[7]Lloyd Institute of Engineering & Technology, Knowledge Park II, Greater Noida, Uttar Pradesh, India.

Abstract. Agriculture forms a major occupation in countries like India. More than 75% people rely on farming for their daily wages. Food security on a global scale is mostly dependent on agriculture. Hence, achieving good yield in the crops grown by farmers is the major concern. Various environmental factors have a significant impact on the crop yield. One such component that contributes majorly to the crop yield is soil. Due to urbanization and enhanced industrialization, the agricultural soil is getting contaminated, losing fertility, and hindering the crop yield. One exciting new way to maximise crop yields while decreasing input costs is precision farming, which makes use of machine learning (ML) and the IoT. Machine Learning (ML) is employed for agricultural data analysis. The goal of this research is to optimise agricultural practices by presenting an integrated crop and fertiliser recommendation system. The proposed ML based model "Precision Agriculture" aims at predicting the suitable crops that can be grown based on the class which the soil sample belongs to and suggests the fertilizers that can be used to further enhance the fertility of soil. Using proposed model, farmers can make decisions on which crop to grow based on the soil classification and decide upon the nitrogen–phosphorous–potassium (NPK) fertilizers ratio that can be used. Comparison of the SVM algorithm with Naive Bayes, and LSTM has shown that SVM performed with a higher accuracy. Decision support tools that integrate AI and domain

* Corresponding author: gurramvijendarreddy@gmail.com

knowledge are provided by the study, which is a substantial contribution to precision agriculture.

Keywords. Agriculture, Soil Features, Crop recommendations, fertilizer recommendation, Machine Learning models

1 Introduction

The highly populous nation of India is particularly vulnerable to sudden shifts in weather patterns, which can endanger food supplies around the world. Droughts have a disproportionate impact on farmers. Crop output is highly dependent on soil type. Fertiliser advice help farmers make informed selections based on their unique agricultural scenarios. Innovations in agriculture have benefited from the use of science and technology, which are intrinsic to modern living. Without a question, agriculture is the backbone of India's economy[1]. In India, 58% of people work in agriculture in some capacity. Crop production and soil improvement are at the heart of agriculture. It is critical to keep the soil in good condition while cultivating. Soil qualities greatly affect agricultural yield fertility because of their importance in supplying plants with deeper root nutrition. Because of factors such as soil infertility or planting during the incorrect season, crop yields may be low[2]. The importance of using fertilisers correctly and working with the correct soil type to improve crop development and agricultural sustainability is well-known. A lack of intelligent recommendations has long dogged the conventional agricultural system, which relies on broad principles, anecdotal evidence, and little testing [3]. Farmers incur more expenses, resources are not optimally distributed, and environmental consequences are subpar because conventional systems do not take into account the unique requirements of different crops and fields [4]. The agricultural sector is on the cusp of a paradigm shift as a result of these problems. Immediate action is required to transform agricultural methods.

Several studies and Data Mining approaches can be used in conjunction with Information and Communication Technology (ICT) to forecast crop yield. By examining past data, one can improve yield by considering elements like soil type, soil fertility, rainfall, and groundwater availability. Wetlands are more suited for cash crops like wheat and sugarcane, whereas drylands are better suited for other crops. There are fifteen distinct agro-climatic zones in India, and each one is ideal for cultivating a different kind of crop. To accomplish this, farmers must be able to identify and handle possible issues in a timely manner if they want to improve their crop yield. Predicting agricultural yields through research is crucial for food security. Therefore, farmers need guidance on which crops thrive in their specific agroclimate zones. To gain a deeper understanding of crop production, it is essential to analyse massive volumes of data using machine learning algorithms. This will allow for the correct forecast of yield for individual crops. With this data in hand, we can advise farmers on more suitable crop varieties.

A novel machine learning-based recommendation system, "Precision Agriculture," is introduced in this research. The goal of the proposed system is to use a set of established parameters to determine which crop and fertiliser is needed for each farmer according to the soil type. Using machine learning-based recommendation algorithms, a number of criteria were taken into account, including soil type, soil quality, water conditions, crop water requirements, land area, and crop market value. In order to recommend suitable crop and fertilisers, this system analyses the results of soil tests. Using crop and nutrients as inputs, the proposed method generates appropriate crop and fertiliser recommendations.

2 Literature Survey

Many researchers and developers from all around the world have worked on different tactics and tools to improve farming techniques, with a strong emphasis on precision agriculture. A survey of the literature on soil quality prediction, crop suggestion, and fertiliser recommendation is presented here.

Crop damage from climate change and other environmental stresses on farming methods were discussed in [5] by Rajamanickam and Mani. Higher accuracy and lower processing time are provided by the authors' suggested probabilistic neural network for the soil fertility prediction approach. Several artificial intelligence (AI) methods for increasing agricultural harvest yields are included in the work of Katarya et al. [6]. These methods derive from the idea of crop recommender systems, which are important to the precision agriculture (PA) paradigm. Methods including K-nearest neighbour (KNN), neural networks, ensemble learning, and similarity-based classifiers are covered in detail. For the purpose of making the best crop recommendations, the authors present a model that takes into account environmental variables including temperature, precipitation, and soil profile. Greater efficiency in the use of resources and higher yields are possible outcomes. Klerkx et al. [7] offered a thorough synopsis of the new area of digital agriculture, touching on many different subfields such as digital technology adoption on farms, how digitalization affects farmers' identities and abilities, digital agriculture ethics, how digitalization affects agricultural knowledge and innovation systems, and digital agriculture economics. Using these groups as a framework, the research analyses the 17 special issue articles' contributions. It delves into the nexus of digital agriculture with new institutional frameworks, economic diversification on farms, and digital agriculture governance. An AI-enabled low-power embedded system for continuous analysis of plant leaf growth was created by Shadrin et al. [8]. A regular Li-ion battery can power the device autonomously for 180 days thanks to the GPU running an onboard recurrent neural network (LSTM). With the release of the Tomato Growth dataset, the authors of this work have opened up new avenues for intelligent monitoring in agriculture. [9] In 2014, Shivnath Ghosh and colleagues Sample (Different dirt with the same number of attributes but different parameters), Back Propagation Algorithm, and Weight update are the three stages that make up a machine learning system in this study.

Soil samples from various places were used to train neural network and generalised linear model (GLM) models in order to predict soil quality. The samples varied in attributes such as pH, organic matter concentration, and texture. In terms of prediction accuracy, the neural network model surpassed the GLM model in the experiments (2023, 13, 2141 4 out of 23). If direct measurements of aggregate stability are not available, the study's two models can be utilised to make soil assessments more thorough. Through the use of extreme learning machines (ELM), Suchithra and Pai [10] were able to optimise agricultural operations by conducting soil tests and classes. Soil fertility indices for key nutrients are classified village-wise by evaluating the values of soil test reports. In most classifications, the Gaussian radial basis function has an accuracy of over 80%, making it the top performer. Soil health and environmental quality can be improved, profitability can be increased, and fertiliser waste can be reduced through the provided approach, all of which contribute to sustainable agriculture in India. Local farms often have more accurate predictions than farms in various places, and Chambers demonstrated in [11] that the accuracy of soil property predictions can be affected by the type of ML model utilised. A 50-component principal component analysis (PCA) proved useful. In their study, Wu et al. [12] demonstrated that the GRNN model is capable of accurately estimating soil nutrients for Dacrydium pectinatum populations in China. Soil nutrient content and quality grades are evaluated using the GRNN model in conjunction with the KNN and SVM models. Predicting soil fertility and controlling

ecosystems with less human intervention were highlighted in the work of Rose et al. [13] as important uses of ML classifiers and statistical methods. Rajamanickam achieved 99% accuracy with the decision tree technique in his soil fertility prediction study [14], which used KNN and SVM algorithms to analyse data on macro- and micronutrients.

To improve yield forecast accuracy in 2018, Chaudhari and colleagues [15] used a trifecta of clustering k-means, A priori, and Bayes. To create a system that could suggest appropriate crops to grow depending on these particular characteristics, their research centred on analysing a number of variables, such as area, rainfall, and soil type.

E. Manjula et al. [16] The soil supplements are investigated in this study using a combination of Naïve Bayes, Decision Tree, and a hybrid technique that includes Potassium, Nitrogen, Calcium, Magnesium, Sulphur, Iron, Zinc, and so on. We evaluate the categorization algorithms' efficiency and precision by comparing their runtimes and accuracy.

A smart farming strategy that incorporates machine learning and real-time sensor data is suggested by Rehman et al. [17]. Their strategy surpasses the shortcomings of conventional smart farming techniques while also improving precision agriculture through the integration of these two technologies. Based on variables including soil type, pH, soil moisture, temperature, and humidity, Priya et al. [18] proposed employing deep learning algorithms to forecast optimal crop growth conditions. Especially when dealing with unpredictable weather, our crop suggestion system aids farmers in making well-informed selections that boost yield. In their discussion of the Internet of Things (IoT) and data mining, Biradar et al. [19] emphasise the possibility of creating smart systems to improve agricultural water management. An economical method of tracking and regulating water usage, sensor networks can boost agricultural output and ensure a steady supply of food. Thus, it is clear that numerous studies have demonstrated the potential of PA applications and ML-based predictive modelling to enhance agricultural systems. Though there are benefits to each approach, there are also drawbacks. Research on Rwanda's agriculture sector has likewise shown a dearth of action. Numerous studies on ML's many uses in smart agriculture can be found in the literature. On the other hand, such methods are more often discussed in theory than put into practice. While it is helpful to have these theoretical conversations, it is crucial to validate these models in actual agricultural contexts so that we can see how well they work. Prior studies have investigated the feasibility of combining AI and the Internet of Things (IoT) in the agricultural sector. The majority of these research, however, deal with crop recommendations or fertiliser recommendations in isolation. Even when studies don't mention the features employed or the reasoning behind their choice of learning model, it is clear that they don't provide enough information regarding the data source system implementations. In the part that follows, we will talk about a system that would fill all these deficiencies.

The main objective of the proposed approach, 'Precision Agriculture' is to increase agricultural productivity and efficiency through the provision of intelligent technology that can guide farmers in the application of optimal crops and fertilisers. The following is an expanded outline of the article: Our proposed methodology is discussed in Section 3. In contrast, Section 4 details the experimental findings and performance evolution of the suggested system. Section 5 provides a conclusion to the essay by outlining potential future extensions and then lists the sources.

3 Methodology

Based on an analysis of the soil's composition and climatic elements such as temperature, humidity, soil PH, and rainfall, the suggested system will ascertain the crop that is most suited

to a specific plot of land. In Figure 1 the workflow of the suggested model is shown.

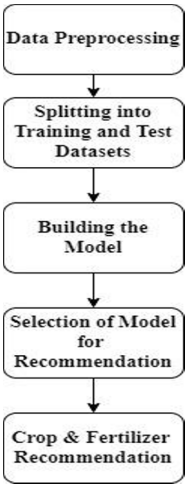

Fig. 1. Proposed System

The model's process flow is structured into five distinct phases: 1) Dataset Collecting 2) Feature extraction and pre-processing 3) Applying number of different ML models 4) Recommending the Best Crop to Use

3.1 Dataset Used

Features in the dataset include nitrogen (N), phosphorus (P), potassium (K), soil pH (PH), relative humidity (RH), precipitation (rainfall), and temperature (T). From the Kaggle website, the data sets were downloaded. The data set contains 1100 occurrences culled directly from the archives.

Out of the eight qualities included in the dataset, one class's value represents the crop grown in relation to the weather conditions. The sample dataset, which shows the relevant values for several variables for a specific crop, is presented in Table 1. Half of the dataset is reserved for testing purposes, and the other half is used for training purposes.

Table 1. Sample Dataset

Label	N	P	K	PH	RF	Temperature	Crop	Fertilizer
1	9	10	11	119	58	32	Rice	Urea
2	8	12	10	118	76	31	Sugarcane	Urea

3.2 Machine Learning Algorithm

The goal of making a prediction is to use a known set of values, or predictors, from the data to anticipate the value of a variable, the predicted variable. In an online classroom, prediction aids in spotting slow learners early on, allowing for more targeted interventions to

boost their performance. Models LSTM, SVM and Naive Bayes, are used for prediction of crop and fertilizer.

3.2.1 LSTM (Long Short-Term Memory Model)

An improvement over Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs) successfully handle the difficulties of sequential data. By representing the whole sequence as a network of linked cells, an LSTM is able to capture the sequential character of the incoming data. The network is able to take into account the context of earlier inputs since new information is added to the model at each timestep and is seamlessly merged with the previously hidden state. The capacity to selectively store or delete data from the cell state through gate-regulated processes is what sets LSTMs apart. The model's ability to handle and take advantage of sequential dependencies in the data is improved by these gates, which act as safeguards and allow for exact control over the flow of information.

3.2.2 SVM (Support vector machines)

The use of kernels in Support Vector Machines (SVMs) increases their capacity to define nonlinear decision limits by transforming data into a higher-dimensional space. In order to accomplish this transformation, the Support Vector Classification (SVC) classifier employs the Gaussian Radial Basis Function (RBF) kernel. In the RBF kernel, we may find the degree of similarity between two points, X1 and X2, by computing their dot product, which is a measure of distance in the original space. Using this method, support vector machines (SVMs) may better grasp complex data patterns and relationships, opening the door to more nuanced decision boundaries than linear classifiers can provide.

3.2.3 Naive Bayes

Naive Bayes is a method for statistical categorization that uses probability theory and has its foundation in Bayes Theorem. The notion of conditional independence is fundamental to the Naive Bayes classifier. This means that the class's performance is evaluated without taking other features into account. When working with a large number of features, Naive Bayes is particularly successful since its "naive" assumption reduces the computing cost of probability computations. Naive Bayes frequently proves its worth in practical situations, despite its simplification assumption, by performing exceptionally well in a variety of applications.

4 Outcome and Results Analysis

4.1 Performance Evolution of the Models

After conducting the data cleaning and visualizations, we implemented our machine learning algorithms on the features of the dataset. Three different algorithms—Navies Bayes, LSTM, and Support Vector Machine (SVM have been applied in our work. The performance of the three algorithms is compared in terms of accuracy and depicted in table 2, and the same is pictorially represented in fig 2.

Table 2. Comparison Results

Method	Precision	Recall	F1-Score	Accuracy
SVM	0.94	0.68	0.78	0.91
NB	0.93	0.63	0.7	0.87
LSTM	0.92	0.95	0.64	0.83

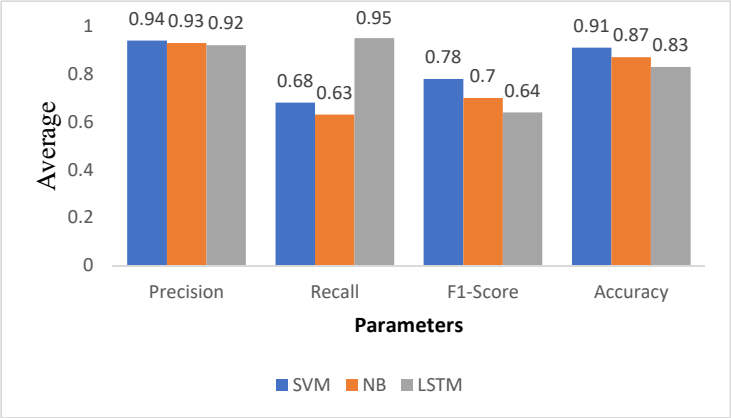

Fig. 2. Comparison of three models and results

The table shows that out of all the classifiers, SVM had the best classification accuracy (91.%), followed by Navies Bayes (87%), and LSTM (83.18%).

5 Conclusion

This research presents a new crop and fertiliser recommendation system, "Precision Farming", tailored to the unique farming conditions. By analysing data and applying machine learning, the technology provides farmers with insights that can guide their fertiliser and crop selection decisions. The models accuracy, stability, and competence were on full display in thorough comparison tests, where it triumphed over other well-known machine learning models. While lowering environmental impact and encouraging cost-effective farming methods, the system has the ability to increase crop yield and quality. Its data modelling and feature extraction process might need some work, and it should take into account more geographical and environmental parameters. Additional geographical and environmental variables, such as humidity, height, temperature, and precipitation, will be the subject of future research. Collaboration between environmental scientists, agronomists, and technologists, as well as the use of more advanced deep learning algorithms, will allow the study to create a system that is more thorough, adaptive, and effective.

References

1. Pawlak, Karolina, and Małgorzata Kołodziejczak. "The role of agriculture in ensuring food security in developing countries: Considerations in the context of the problem of sustainable food production." *Sustainability* 12.13 (2020): 5488.

2. Norton, George W., and Jeffrey Alwang. "Changes in agricultural extension and implications for farmer adoption of new practices." *Applied Economic Perspectives and Policy* 42.1 (2020): 8-20.
3. Khan, Nawab, et al. "Current progress and future prospects of agriculture technology: Gateway to sustainable agriculture." *Sustainability* 13.9 (2021): 4883.
4. Chen, Yuanzhe, et al. "AgriKG: an agricultural knowledge graph and its applications." *Database Systems for Advanced Applications: DASFAA 2019 International Workshops: BDMS, BDQM, and GDMA, Chiang Mai, Thailand, April 22–25, 2019, Proceedings 24*. Springer International Publishing, 2019.
5. Rajamanickam, Jayalakshmi, and Savitha Devi Mani. "Kullback chi square and Gustafson Kessel probabilistic neural network based soil fertility prediction." *Concurrency and Computation: Practice and Experience* 33.24 (2021): e6460.
6. Katarya, Rahul, et al. "Impact of machine learning techniques in precision agriculture." *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*. IEEE, 2020.
7. Klerkx, Laurens, Emma Jakku, and Pierre Labarthe. "A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda." *NJAS-Wageningen journal of life sciences* 90 (2019): 100315.
8. Shadrin, Dmitrii, et al. "Enabling precision agriculture through embedded sensing with artificial intelligence." *IEEE Transactions on Instrumentation and Measurement* 69.7 (2019): 4103-4113.
9. Anushabai, D., et al. "Insights Into Yield Estimation Of Horticultural Crops Using Machine Learning Algorithms."
10. Suchithra, M.S.; Pai, M.L. Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. Inf. Process. Agric. 2020, 7, 72–82. [CrossRef]
11. Chambers, O. Machine Learning Strategy for Soil Nutrients Prediction Using Spectroscopic Method. Sensors 2021, 21, 4208. [CrossRef]
12. Wu, C.; Chen, Y.; Hong, X.; Liu, Z.; Peng, C. Evaluating soil nutrients of Dacrydium pectinatum in China using machine learning techniques. For. Ecosyst. 2020, 7, 30. [CrossRef]
13. Rose, S.; Nickolas, S.; Sangeetha, S. Machine Learning and Statistical Approaches used in Estimating Parameters that Affect the Soil Fertility Status: A Survey. In Proceedings of the 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT), Karnataka, India, 16–18 August 2018; IEEE: New York, NY, USA, 2018; pp. 381–385.
14. Rajamanickam, J. Predictive model construction for prediction of soil fertility using decision tree machine learning algorithm. Infocomp J. Comput. Sci. 2021, 20, 49–55.
15. Bang, Shivam, et al. "Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models." *2019 Twelfth international conference on contemporary computing (IC3)*. IEEE, 2019.
16. Veenadhari, S., Bharat Misra, and C. D. Singh. "Machine learning approach for forecasting crop yield based on climatic parameters." *2014 International Conference on Computer Communication and Informatics*. IEEE, 2014.
17. Rehman, A.; Liu, J.; Li, K.; Mateen, A.; Yasin, Q. Machine Learning Prediction Analysis using IoT for Smart Farming. Int. J. Emerg. Trends Eng. Res. 2020, 8, 1–30.

18. Priya, P.K.; Yuvaraj, N. An IoT Based Gradient Descent Approach for Precision Crop Suggestion using MLP. J. Phys. Conf. Ser. 2019, 1362, 012038. [CrossRef]
19. Biradar, H.B.; Shabadi, L. Review on IOT based multidisciplinary models for smart farming. In Proceedings of the 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 19–20 May 2017; pp. 1923–1926.