https://doi.org/10.48047/AFJBS.6.12.2024.1682-1694

AFJBS

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

A fertilizer recommendation system using light gradient boost regressor

Uditendu Sarkar[1], Gouravmoy Banerjee[2], and Indrajit Ghosh[3*]

[1]Scientist "F" and Senior Director (IT), National Informatics Centre (NIC), Ministry of Electronics & Information Technology, Government of India, West Bengal State Centre, Ground Floor, Vidyut Bhawan, DJ Block, 10 Sector II, Salt Lake, Kolkata, 700091, West Bengal, India

[2]State Aided College Teacher, Department of Computer Science, Ananda Chandra College, Jalpaiguri, 735101, West Bengal, India

[3]Associate Professor, Department of Computer Science, Ananda Chandra College, Jalpaiguri, 735101, West Bengal, India

Corresponding address:

Dr. Indrajit Ghosh,

Associate Professor, Department of Computer Science, Ananda Chandra College, Jalpaiguri, 735101, West Bengal, India

Email: ighosh2002@gmail.com

Article History

Volume 6 Issue 12, 2024

Received: 25 May 2024

Accepted : 30 June 2024

doi: 10.48047/AFJBS.6.12.

2024.1682-1694

Abstract:

Over time, the soil's nutrient level gradually declines. Soil scientists recommend precise doses of fertilizers to compensate for soil nutrient loss. However, soil scientists are scarce, expensive, and inaccessible to marginal farmers in most countries. As a common practice, rural farmers use chemical fertilizers in blind doses without proper scientific knowledge. Such indiscriminate use of chemicals creates a nutrient imbalance and leads to huge crop losses. This work aims to provide a low-cost and near-expert-level recommendation system for three fertilizers, nitrogen (N), phosphorous (P), and potassium (K), for two major crops, paddy and potato, cultivated in the Gangetic alluvial plain of West Bengal, India. We designed the system using a light gradient boost regressor, one of the most preferred machine learning methods for solving various soil-related issues, to suggest the precise doses of N, P, and K. Our designed system recommends fertilizers based on the nutrient contents and other relevant parameters in the arable soil. Experimental results revealed that the system achieved the highest performance (with R^2 = 0.9997 and RMSE = 0.9381). The proposed system provides an elegant alternative to the scarce and expensive soil scientists who recommend the precise dose of the appropriate fertilizer.

Keywords: Fertilizer; Fertilizer recommendation system; Light gradient boost regressor; Soil health card data.

1 Introduction

Over the years, agricultural soil has been pivotal in crop productivity, profitability, and environmental sustainability. The nutrients present in the farm soil directly impact the quantity and quality of crops growing on it (Sindelar, 2015). The lack of nutrients leads to a decrease in crop yield and quality, increasing the production cost. On the other hand, excessive nutrients in the soil have an adverse effect on plant growth. Therefore, the correct proportion of various nutrients in agricultural soil has a critical role in increasing farming yield (Ahmed *et al.*, 2021). In India, most farmers cultivate two crops a year without adopting any fertilizer management strategy with adequate scientific knowledge (Priya and Ramesh, 2018). This traditional practice results in nutrient depletion over time and changes the chemical properties of the soil. Farmers apply chemical fertilizers to compensate for the ongoing nutrient deficiency. Rural farmers, however, due to a lack of scientific knowledge, tend to use these chemical fertilizers indiscriminately as a blanket dose (Sun et al., 2019), without a precise estimation of the quantity of nutrients present in the soil and the nutrients required for the targeted crop that they have to cultivate (Dhaygude and Chakraborty, 2020).

West Bengal is one of the most productive agrarian states in India, and nearly 8% of the population of the country resides in this state. This state covers about 4.67% of the agricultural land in the country. Approximately 7.13 million families in West Bengal are associated with agriculture, of which 96% are marginal farmers (Department of Agriculture, Govt. of West Bengal). Agriculture contributes about 22% of the state's GSVA (Gross state value added) (Economics and Statistics Division, Ministry of Agriculture & Farmers Welfare, Department of Agriculture and Farmers Welfare, Govt. of India). The Gangetic alluvial plain is the most fertile region of the state to produce two major crops, paddy and potato. The major cultivators of this region are the marginal farmers with less than one-hectare land holdings. These layman farmers suffer from inadequate knowledge about the nutrients present in their soil and the precise application of the required fertilizers accordingly. The central government, on the other hand, subsidizes fertilizers for the welfare of marginal farmers throughout the country, and the farmers in West Bengal are one of the largest beneficiaries of the fertilizer subsidy (Sharma and Thaker, 2010). The indiscriminate use of fertilizers also wastes money and burdens the government exchequer unnecessarily.

Fertilizers provide nutrients to the agricultural soil. Nitrogen (N), phosphorous (P), and potassium (K), together as NPK, are macronutrients that are required in large quantities and have a direct impact on plant growth (Hossain *et al.*, 2017). The micronutrients, such as sulphur (S), copper (Cu), zinc (Zn), manganese (Mn), boron (B), and iron (Fe), are required in very small quantities, and the farmers are less concerned about the well-being of a crop (Sillanpää, 1982). The depletion of macronutrients in soil is compensated by applying chemical fertilizers, and a precise application of fertilizers plays a significant role in the better production of crops (Ju *et al*, 2007). However, layman farmers in West Bengal are applying chemical fertilizers of their own choice and local availability without analyzing the existing nutrient level of their soil. In contrast, soil testing laboratories and soil scientists are very scarce, expensive, and inaccessible to marginal farmers for appropriate fertilizer recommendations. Since 2015, the Ministry of Statistics and Programme Implementation, Government of India, has observed a gradual decrease in soil fertility and nutrient level imbalance due to this exasperating scenario (Ministry of Statistics and Programme Implementation, Government of India).

Several scientific approaches have been suggested in the literature for precise estimation of fertilizer requirements for a crop (Samal *et al.*, 2020). Still, each proposed method requires complex mathematical calculations and a comprehensive knowledge of soil science. These prerequisites can be quite challenging for a rural farmer to meet. Therefore, an alternative solution is to be provided to rural farmers to recommend precise doses of fertilizers that are easier to use without expert knowledge.

To mitigate these problems and to increase crop productivity, profitability, and environmental sustainability, site-specific decision support systems using state-of-the-art machine learning techniques provide trustworthy solutions for various management issues in agriculture, such as fertilizer recommendation, irrigation scheduling, pest and disease control, crop yield prediction, etc. (Ahmed *et al.*, 2021; De-Oliveira and De-Silva, 2023). Several machine learning techniques were suggested to design soil analysis and fertilizer recommendation systems. Such methods include fuzzy logic (Indahingwati *et al.*, 2018), artificial neural networks (Moreno *et al.*, 2018), decision trees (Jahan and Shahariar, 2020; Singh *et al.*, 2020), support vector machines (Suchithra and Pai, 2018a), linear regression models (Saïdou *et al.*, 2018), gradient-boosted trees (Qin *et al.*, 2018), random forests (Ransom *et al.*, 2019; Suleymanov *et al.*, 2023), deep learning models (Suchithra and Pai, 2018b), and many more to mention. As a tree-based regression model, the light gradient boost (LGB) regressor is very efficient in designing systems that require less space and time to be trained and can handle large volumes of data with greater accuracy. LGB is one of the most preferred machine learning methods for designing various systems for soil-related issues (Motia and Reddy, 2021). However, no such fertilizer recommendation system has been suggested for precise estimation of N, P, and K fertilizers using the LGB regressor as one of the most preferred machine learning methods. Furthermore, there has been no other reported fertilizer recommendation system for the Gangetic alluvial plain in West Bengal to date. The aim of this work is to provide a low-cost and near-expert-level fertilizer recommendation system using an LGB regressor for three varieties of paddy (IET4094, IET4097, and BORO-4789) and two varieties of potato, Kufri jyoti (high) and Kufri jyoti (low), grown in the Gangetic alluvial plain in West Bengal.

Our proposed system was designed to suggest location-specific fertilizer doses for N, P, and K based on the different nutrient contents and other relevant soil parameters. The relevant dataset of nutrient content and other parameters was collected from the freely available Soil Health Card (SHC) datasets provided by the Indian Council of Agricultural Research, Govt. of India. The system's performance was evaluated in terms of two statistical metrics: the coefficient of determination (R^2) and the root mean square error ($RMSE$). Empirical studies revealed that the system achieved the highest performance (in terms of the highest value of $R^2 = 0.9997$ and the lowest value of $RMSE = 0.9381$).

2 Materials and Methods

2.1 Study Location

The Gangetic alluvial plain is one of the most fertile regions situated in the central part of the state of West Bengal in India. The study area comprises three districts, Burdwan, Hooghly, and Nadia, within this region from 22.47°N to 23.82°N and 86.80°E to 88.69°E. The livelihood of the major population is primarily based on agriculture, resulting in intensive agricultural activities. Three varieties of paddy and two varieties of potato are the major cash crops produced in this area. The gross cultivation land for paddy and potato is 11,78,300 and 2,23,900 hectares, respectively.

The study area was selected for four main reasons: (i) most rural farmers apply fertilizers with a blanket dose as traditional practices without precise estimation of crop nutrient requirement, leading to a nutrient level imbalance (Ministry of Statistics and Programme Implementation, Government of India); (ii) the cropping intensity of this area is very high (Economics and

Statistics Division, Ministry of Agriculture & Farmers Welfare, Department of Agriculture and Farmers Welfare, Govt. of India); (iii) authenticated soil datasets are available for training and testing (Soil Health Card); and finally, (iv) no such system has been proposed for fertilizer recommendation for this study area. Figure 1 presents the geographic map of the study area.

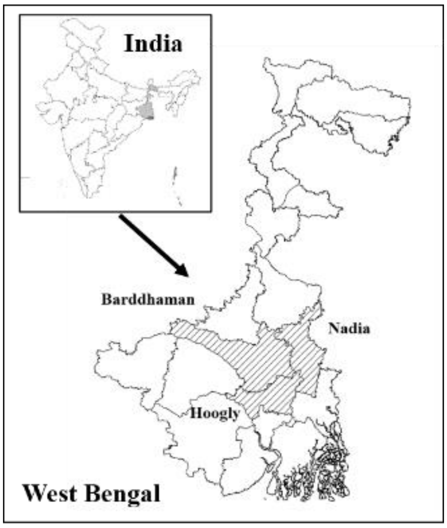


Figure 1. The area selected for the study

2.2 Dataset Used

The datasets used to design the system were obtained from the Soil Health Card (SHC) repository provided by the Department of Agriculture and Farmers Welfare, Govt. of India (Soil Health Card). The Soil Health Card scheme is a flagship program launched in February 2015 and is run by the Government of India for monitoring soil health. In the SHC scheme, uniform norms are followed across different states in India to assist site-specific fertilizer management. The Integrated Nutrient Management Division manages the scheme in the Ministry of Agriculture and Farmers Welfare, Government of India. Several soil testing laboratories across India analyze soil samples collected from various locations according to the norms provided by the authority. The datasets include various soil nutrients and other physical parameters of soil such as pH, soil organic carbon content (OC), electrical conductivity (EC), etc.

For the current study, we have considered six soil and crop-related parameters as inputs to the system, along with the targeted crop yield. These are available nitrogen (N in Kg/ha), phosphorous (P in Kg/ha), and potassium (K in Kg/ha), soil pH (measured on a 14-point scale), electrical conductivity (in dS/m), and soil organic carbon (in %). The dataset was collected from the SHC data repository. A total of 9042 samples were taken from Hooghly, 922 from Burdwan, and 1599 from the Nadia districts, respectively. The dataset was carefully examined to eliminate missing data and was arbitrarily divided into two parts: 70% for training and the rest (30%) for testing the system.

2.3 Reference Values of Fertilizers

The reference values for the exact quantity of fertilizers were calculated using the soil test crop response (STCR) model provided by ICAR (Indian Council for Agricultural Research). Ramamoorthy and others suggested the STCR model, which is based on Liebig's rule of the minimum to estimate the precise quantity of fertilizers needed for a particular variety of crops (Ramamoorthy and Velayutham, 1971). The STCR model was developed through experiments conducted at various places using different fertilizers for a targeted crop to establish the ideal amount of nutrients needed. For the best possible output, the STCR suggests the precise dose of N, P, and K fertilizers for a particular crop. The estimation of the quantity of soil nutrients present in the soil is an important issue because the STCR model suggests the appropriate dose of fertilizer based on the nutrients present in the soil.

The STCR suggests several equations to estimate the amount of various nutrients required for a particular crop. These equations are the outcome of numerous studies carried out at various degrees of soil fertility with respect to various nutrients. The equations were then formulated after evaluating the nutrient uptake of the crops. The primary goal of the STCR model is to assist farmers in applying the best fertilizer dose to achieve the desired crop yield to the maximum possible extent.

In our present study, we have considered two major crops, paddy, and potato, cultivated in Gangetic alluvial plains in West Bengal. We selected three varieties of paddy, namely IET4094, IET4097, and BORO4789, and two varieties of potato, Kufri jyoti (high) and Kufri jyoti (low), as the targeted crops. Table 1 provides the STCR-recommended equations for N, P, and K against these crops.

Table 1. STCR equations for N, P, K against five varieties of paddy and potato

Targeted crops	**Nitrogen required (Kg/Ha)**	**Phosphorous required (Kg/Ha)**	**Potassium required (Kg/Ha)**
Paddy (IET4094)	N = 3.60×T- 0.25×SN	P = 2.29×T- 0.18×SN	K = 2.61×T- 0.19×SK
Paddy (IET4097)	N = 15.34×T-1.62×SN	P = 15.34×T-1.62×SN	K = 2.52×T- 0.28×SK
Paddy (BORO4789)	N = 3.28×T- 0.18×SN	P = 4.80×T- 5.02×SN	K = 2.83×T- 0.54×SK
Potato Kufri jyoti (high)	N = 1.61×T- 0.43×SN	P = 0.95×T- 1.0×SP	K = 0.89×T- 0.34×SK
Potato Kufri jyoti (low)	N = 1.80×T- 033×SN	P = 1.12×T- 1.39×SP	K = 1.51×T- 0.29×SK

SN: Nitrogen present in the soil, SP: Phosphorous present in the soil, SK: Potassium present in the soil, T: Target yield.

2.4 Light Gradient Boost Regressor

The light gradient boost (LGB) regressor was proposed by Ke *et al.*, (2017). This model uses a gradient learning framework based on a decision tree and the concept of boosting. The LGB model follows the leaf-wise (depth) growth of the consecutive learner trees rather than the level-wise (breadth) expansion of the learner trees. It finds the leaves with the highest branching gain from all the leaves and then goes through the branching cycle. Two techniques are employed to boost the model's scalability: gradient-based one-side sampling (GOSS) and exclusive feature bundling (EF-B). The GOSS algorithm only considers samples with more information gain than a predefined threshold, improving the accuracy. The EF-B algorithm is a dimensionality reduction tool for merging or projecting several feature sets. This model has the advantages of taking up less space, taking less time to train, handling large volumes of data, and having incredible accuracy. Ke et al. (2017) describe the mathematical foundation of LGB as follows:

For the given training dataset $X = \{(x_i, y_i)\}_{i=1}^{m}$,

LGB searches for an approximation $\hat{\boldsymbol{f}}(\boldsymbol{x})$ to the function $\boldsymbol{f}^*(\boldsymbol{x})$ for minimizing expected values of specific loss functions $\boldsymbol{L}(\boldsymbol{y}, \boldsymbol{f}(\boldsymbol{x}))$,

$\boldsymbol{L(y, f(x)}: \hat{\boldsymbol{f}}(\boldsymbol{x})\ \mathbf{arg\,min}_{f}\ \boldsymbol{E_{y,X}L\big(y, f(x)\big)}$

It integrates many $\mathbf{T}$ regression trees $\sum_{t=1}^{T} \boldsymbol{f_t(X)}$ for approximating the eventual model, defined as:

$$\boldsymbol{f_T(X)} = \sum_{t=1}^{T} \boldsymbol{f_t(X)} \qquad (2)$$

The regression trees are defined as $\boldsymbol{q} \in \{\mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{N}\}$, where $\mathbf{N}$ represents the number of tree leaves, $\boldsymbol{q}$ is the decision rule of trees, and $\boldsymbol{w}$ is a vector denoting the sample weights of leaf nodes. The model is then trained in the additive form at step $\boldsymbol{t}$ as follows:

$$\boldsymbol{\Gamma_t} \cong \sum_{j=1}^{N} \boldsymbol{L\big(y_i, F_{t-1}(x_i) f_t(x_i)\big)} \qquad (3)$$

The objective function is fast approximated by using Newton's approach. By removing the constant term, Eq. (3) is simplified as:

$$\boldsymbol{\Gamma_t} \cong \sum_{j=1}^{N} \left(\boldsymbol{g_i f_t(x_i)} + \frac{\mathbf{1}}{\mathbf{2}} \boldsymbol{h_i f_t^2(x_i)}\right) \qquad (4)$$

where $\boldsymbol{g_i}$ and $\boldsymbol{h_i}$ are 1st and 2nd order gradient statistical results of loss functions.

Now, if the sample set of leaf *j* is represented by $\boldsymbol{I_j}$, then Eq. (4) can be rewritten as:

$$\boldsymbol{\Gamma_t} = \sum_{j=1}^{J} \left(\left(\sum_{i \in I_j} \boldsymbol{g_i}\right) \boldsymbol{\omega_j} + \frac{\mathbf{1}}{\mathbf{2}} \left(\sum_{i \in I_j} \boldsymbol{h_i} + \boldsymbol{\lambda}\right) \boldsymbol{\omega_j^2}\right) \qquad (5)$$

In terms of the tree structure $\mathbf{q(x)}$, the optimum leaf weights of the leaf nodes $\boldsymbol{\omega_j^*}$ and extreme values of $\boldsymbol{\Gamma_K}$ are obtained by using Eqs. (6) and (7):

$$\boldsymbol{\omega_j^*} = -\frac{\sum_{i \in I_j} \boldsymbol{g_i}}{\sum_{i \in I_j} \boldsymbol{h_i} + \boldsymbol{\lambda}} \qquad (6)$$

$$\boldsymbol{\Gamma_T^*} = -\frac{\mathbf{1}}{\mathbf{2}} \sum_{j=1}^{J} \frac{\left(\sum_{i \in I_j} \boldsymbol{g_i}\right)^{\mathbf{2}}}{\sum_{i \in I_j} \boldsymbol{h_i} + \boldsymbol{\lambda}} \qquad (7)$$

where $\boldsymbol{\Gamma_T^*}$ is the weight function measuring the quality of tree structure $\boldsymbol{q}(x)$.

The objective function is eventually obtained by integrating the split:

$$\boldsymbol{G} = \frac{\mathbf{1}}{\mathbf{2}} \left(\frac{\left(\sum_{i \in I_l} \boldsymbol{g_i}\right)^{\mathbf{2}}}{\sum_{i \in I_l} \boldsymbol{h_i} + \boldsymbol{\lambda}} + \frac{\left(\sum_{i \in I_r} \boldsymbol{g_i}\right)^{\mathbf{2}}}{\sum_{i \in I_r} \boldsymbol{h_i} + \boldsymbol{\lambda}} + \frac{\left(\sum_{i \in I} \boldsymbol{g_i}\right)^{\mathbf{2}}}{\sum_{i \in I} \boldsymbol{h_i} + \boldsymbol{\lambda}}\right) \qquad (8)$$

where $\boldsymbol{I_l}$ and $\boldsymbol{I_r}$ are samples of the left and right branches, respectively.

2.6 Performance Evaluation

Evaluation of the system's performance plays a significant role in any model-building process. To evaluate the prediction performance of our proposed model, we applied the four-fold cross-validation approach. Two widely used statistical metrics, the coefficient of determination (R^2) and the root mean square error (*RMSE*), were used to validate the prediction accuracy. R^2 is a goodness metric that projects the relationship between the actual and predicted values, while *RMSE* is an error metric that projects the errors in the predicted values. The mathematical equations of these two metrics are defined as (Sun *et al.*, 2020; Tao *et al.*, 2018):

$$\boldsymbol{R^2 = \frac{\sum_{i=1}^{n}(R_{Ri} - R_{Pi})^2}{\sum_{i=1}^{n}\left(R_{Ri} - \overline{R_{Pi}}\right)^2}} \tag{9}$$

$$\boldsymbol{RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(R_{Ri} - R_{Pi})^2}} \tag{10}$$

Where R_{Ri} is the reference value of the macronutrient obtained using the STCR equations, R_{Pi} is the model recommended value, $\overline{\boldsymbol{R}}_{\boldsymbol{Pi}}$ is the mean recommended value, and $\boldsymbol{n}$ is the total number of samples. In general, higher values of R^2 and lower values of *RMSE* indicate better model performance.

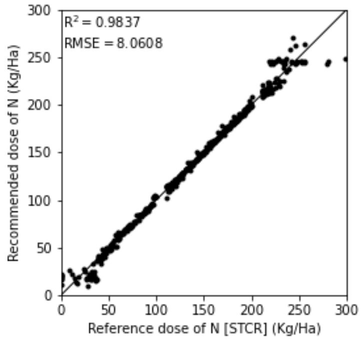
3 Results and Discussions

The deployable version of the LGB regressor system was coded using the LGBM library package, and the results were generated using Python's Scikit-learn library package (Ver. 1.3). As a case study, the system was implemented for the recommendation of three major fertilizers, N, P, and K, in three districts, Hooghly, Burdwan, and Nadia, in the Gangetic alluvial plain of West Bengal, India. Three varieties of paddy (IET4094, IET4097, and BORO4789) and two varieties of potato (Kufri jyoti (high) and Kufri jyoti (low)) were selected as the targeted crops. For each of the crops, the quantity (dose) of NPK to be applied was recommended by the system. The system-recommended doses of NPK were compared with the reference doses suggested by STCR (Ramamoorthy and Velayutham, 1971). To validate the accuracy of the prediction, two performance metrics, the coefficient of determination (R^2) and the root mean square error (*RMSE*), were used as defined in equations 9 and 10. The empirical values of R^2, and *RMSE* obtained for three fertilizers, N, P, and K, against five crop varieties are presented in Table 2.

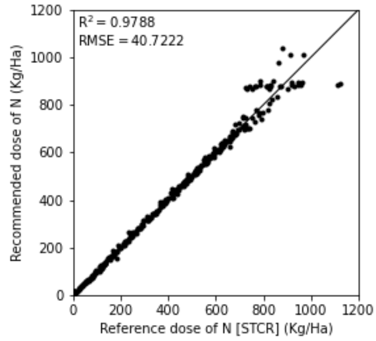
Table 2. Empirical values of the performance metrics for NPK against five crop varieties

Targeted crops	Nitrogen (N)		Phosphorus (P)		Potassium (K)	
	R^2	*RMSE*	R^2	*RMSE*	R^2	*RMSE*
Paddy (IET4094)	0.9837	8.0608	0.9940	13.4147	0.9997	0.9381
Paddy (IET4097)	0.9788	40.7222	0.7262	58.2868	0.9996	1.1273
Paddy (BORO4789)	0.9573	9.4341	0.9609	80.3638	0.9995	2.1135
Potato-Kufri jyoti (high)	0.9750	17.4285	0.9978	18.7500	0.9997	1.6212
Potato-Kufri jyoti (low)	0.9653	15.6291	0.9962	24.6089	0.9972	5.5208

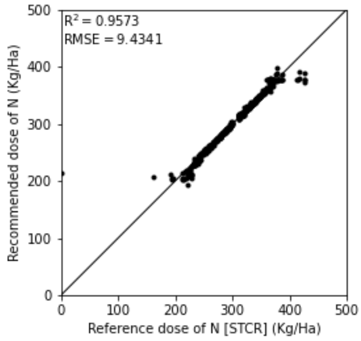
For better representation, the scatter plots of the recommended doses against the reference doses for each of the three fertilizers, N, P, and K, are presented in Figures 2, 3, and 4, respectively. It is to be noted that the ranges of coordinate values of the scatter plots are dissimilar because the range of doses of different fertilizers varies with the targeted crops.



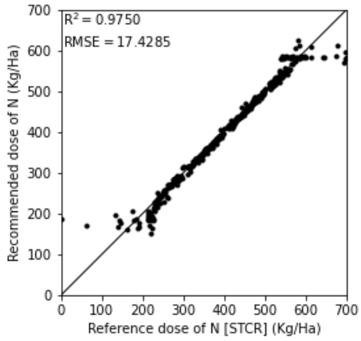
(a) Paddy (IET4094)



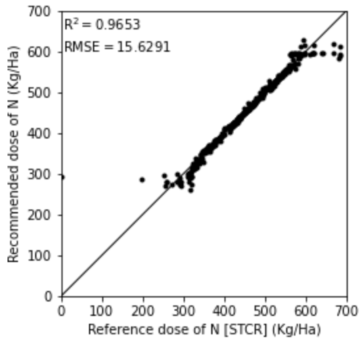
(b) Paddy (IET4097)



(c) Paddy (BORO4789)

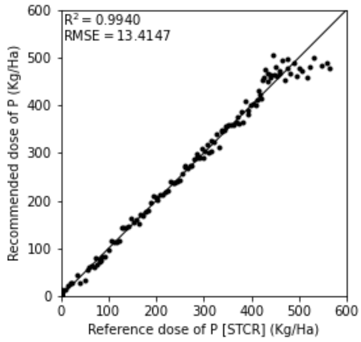


(d) Potato-Kufri jyoti (high)

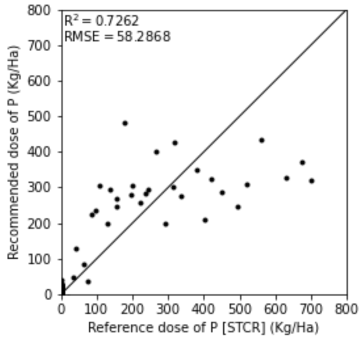


(e) Potato-Kufri jyoti (low)

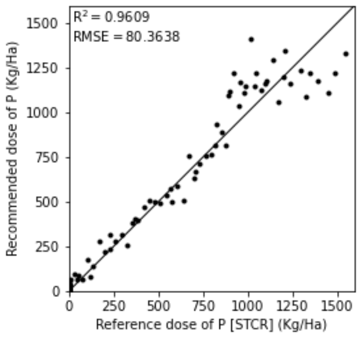
Figure 2. The scatter plots of the recommended values R_{Pi} against the reference values R_{Ri} for N against five crops



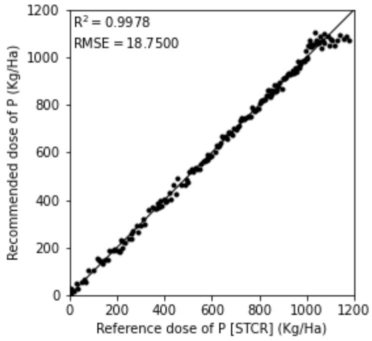
(a) Paddy (IET4094)



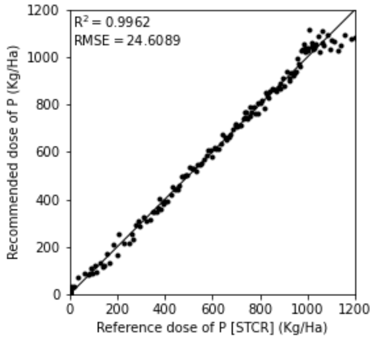
(b) Paddy (IET4097)



(c) Paddy (BORO4789)

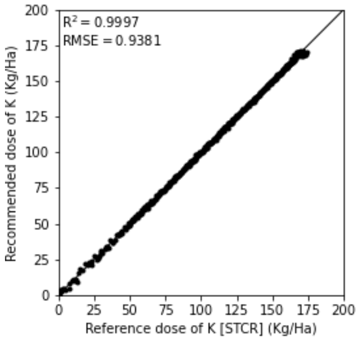


(d) Potato-Kufri jyoti (high)

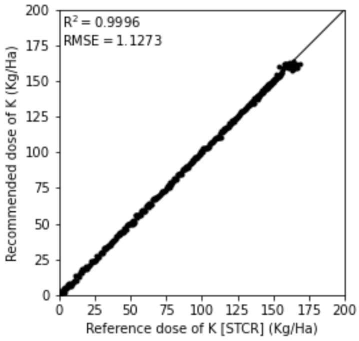


(e) Potato-Kufri jyoti (low)

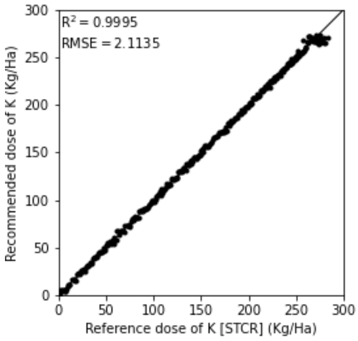
Figure 3. The scatter plots of the recommended values R_{Pi} against the reference values R_{Ri} for P against five crops



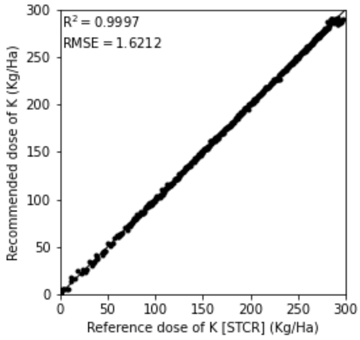
(a) Paddy (IET4094)



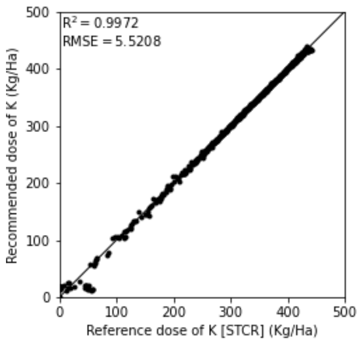
(b) Paddy (IET4097)



(c) Paddy (BORO4789)



(d) Potato-Kufri jyoti (high)



(e) Potato-Kufri jyoti (low)

Figure 4: The scatter plots of the recommended values R_{Pi} against the reference values R_{Ri} for K against five crops

The experimental results reveal that the system's performance in recommending a precise dose of fertilizers (NPK) varies with the targeted crops (paddy, potato). In addition, the performance also varies for the different varieties of the same crop.

Figures 2(a-e) depict that, for N, all five scatter plots for paddy and potato are very identical. The distribution is more scattered at the lower and higher ends of the plots. It signifies that the system's performance degrades when the recommended dose range is very low or high. The overall performance for N is appreciable (with R^2 ranging from 0.9573 to 0.9837 and *RMSE* ranging from 8.0608 to 40.7222).

Table 2 and Figure 3 show that the system showed the highest performance for recommending the dose of P against Potato-Kufri jyoti (high) (with R^2 = 0.9978 and *RMSE* = 18.7500). A nearly equitable performance was obtained for Potato-Kufri jyoti (low) (with R^2 = 0.9962 and *RMSE* = 24.6089). An acceptable performance was observed for the other two varieties of paddy (IET4094 and BORO4789) (with R^2 ranging from 0.9609 to 0.9940 and *RMSE* ranging from 13.4147 to 80.3638). For paddy (IET4097), the system is the worst performer (with R^2 = 0.7262 and *RMSE* = 58.2868).

For the recommendation of K, excellent performance was achieved for three varieties of paddy (IET4094, IET4097, and BORO4789) and a variety of potato (Kufri jyoti (high)) (with R^2 ranging from 0.9995 to 0.9997 and *RMSE* ranging from 0.9381 to 2.1135). However, the distribution is slightly scattered at the higher end of the plot. It signifies that performance degrades when the recommended dose is very high. For potato Kufri jyoti (low), the distribution is slightly uneven at the lower part of the plot (Figure 4(e)). This observation depicts that performance degrades when the recommended dose is very low. Figures 2-4 represent that the recommended doses of three fertilizers, N, P, and K, are consistent with those suggested by STCR (Motia and Reddy, 2021).

5 Conclusion

Recommendation of a precise dose of fertilizer is a prime issue worldwide for preventing the decline of soil fertility, better production of crops, and conservation of soil resources. This research aimed to develop a low-cost pioneering fertilizer recommendation system that might assist rural farmers who apply fertilizer indiscriminately owing to a lack of scientific expertise. It is evident from the experimental results that our target of designing an efficient system to recommend the precise doses of major fertilizers (NPK) for paddy and potato cultivation in the Gangetic alluvial plain of West Bengal in India has been achieved. It is a user-friendly system and can easily be deployed in any rural area using a laptop or desktop.

The system has been designed based on the freely available region-specific SHC datasets and incurs zero cost compared to expensive chemical testing of soil nutrients and expert advice. It can be employed to improve the fertilizer management strategy for sustainable crop production in other regions of the country where SHC data is available.

However, our proposed system provides a reliable solution for fertilizer management in the Gangetic alluvial soil. Still, its efficiency in different agroclimatic soils remains to be explored, which is our future target.

References

Ahmed, U., Lin, J. C.W., Srivastava, G., and Djenouri, Y. (2021). A nutrient recommendation system for soil fertilization based on evolutionary computation. Computers and Electronics in Agriculture, 189, 106407.

De-Oliveira, R. C., and De-Silva, R. D. (2023). Artificial Intelligence in Agriculture: Benefits, Challenges, and Trends. Applied Sciences, 13(13), 7405.

Department of Agriculture, Govt. of West Bengal. Retrieved April 11, 2023, from https://wb.gov.in/departments-details.aspx?id=D170907140022669&page=Agriculture.

Dhaygude, M., and Chakraborty, D. (2020). Rethinking design of digital platforms for emergent users: Findings from a study with rural Indian farmers. In IndiaHCI'20: Proceedings of the 11th Indian Conference on Human-Computer Interaction (pp. 62–69).

Economics and Statistics Division, **Ministry of Agriculture & Farmers Welfare, Department of Agriculture and Farmers Welfare, Govt. of India**. (2022). Agricultural Statistics at a Glance 2022. Retrieved June 18, 2023, from https://agricoop.nic.in/Documents/CWWGDATA/Agricultural_Statistics_at_a_Glance_2022_0.pdf.

Hossain, M. A., Kamiya, T., Burritt, D., Tran, L.-S. P., & Fujiwara, T. (2017). Plant macronutrient use efficiency: Molecular and genomic perspectives in crop plants. Academic Press.

Indahingwati, A., Barid, M., Wajdi, N., Susilo, D., Kurniasih, N., and Rahim, R. (2018). Comparison analysis of TOPSIS and fuzzy logic methods on fertilizer selection. International Journal of Engineering and Technology, 7(2.3), 109–114.

Jahan, N., and Shahariar, R. (2020). Predicting fertilizer treatment of maize using decision tree algorithm. Indonesian Journal of Electrical Engineering and Computer Science, 20(3), 1427–1434.

Ju, X. T., Kou, C. L., Christie, P., Dou, Z., and Zhang, F. (2007). Changes in the soil environment from excessive application of fertilizers and manures to two contrasting intensive cropping systems on the North China Plain. Environmental Pollution, 145(2), 497–506.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30.

Ministry of Statistics and Programme Implementation, Government of India. Soil Nutrient Indices. Retrieved from https://www.mospi.gov.in/sites/default/files/reports_and_publication/statistical_publication/EnviStats/b14_Chapter%202.pdf.

Moreno, R. H., Garcia, O., and Arias R., L. A. (2018). Model of neural networks for fertilizer recommendation and amendments in pasture crops. In 2018 ICAI Workshops (ICAIW) (pp. 1-5). Bogota, Colombia.

Motia, S., and Reddy, S. (2021). Exploration of machine learning methods for prediction and assessment of soil properties for agricultural soil management: A quantitative evaluation. In Journal of Physics: Conference Series (pp. 012037). IOP Publishing.

Priya, R., and Ramesh, D. (2018). Adaboost.RT based soil NPK prediction model for soil and crop specific data: A predictive modelling approach. In Big Data Analytics: 6th International Conference, BDA 2018, Warangal, India, December 18–21, 2018, Proceedings 6 (pp. 322–331). Springer.

Qin, Z., Myers, D.B., Ransom, C.J., Kitchen, N.R., Liang, S. Z., Camberato, J.J., Carter, P.R., Ferguson, R.B., Fernandez, F.G., Franzen, D.W., Laboski, C.A.M., Malone, B.D., Nafziger, E.D., Sawyer, J.E. and Shanahan, J.F. (2018), Application of Machine Learning Methodologies for Predicting Corn Economic Optimal Nitrogen Rate. Agronomy Journal, 110: 2596-2607. https://doi.org/10.2134/agronj2018.03.0222.

Ransom, C. J., Kitchen, N. R., Camberato, J. J., Carter, P. R., Ferguson, R. B., Fernández, F. G., Franzen, D. W., Laboski, C. A. M., Myers, D. B., Nafziger, E. D., Sawyer, J. E., and Shanahan, J. F. (2019). Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations. Computers and Electronics in Agriculture, 164, 104872. https://doi.org/10.1016/j.compag.2019.104872.

Ramamoorthy, B. and Velayutham, M. (1971). Soil test crop response correlation work in India, World soil resources report No. 41: 96-105, FAO, Rome.

Saïdou, A., Balogoun, I., Ahoton, E. L., Igué, A. M., Youl, S., Ezui, G. and Mando, A. (2018). Fertilizer recommendations for maize production in the South Sudan and Sudano-Guinean zones of Benin. Nutrient Cycling in Agroecosystems, 110(3), 361–373. https://doi.org/10.1007/s10705-017-9902-6.

Samal, S., Prasad, L., and Kumar, R. (2020). How to apply fertilizers, based on soil test? A step by step guide. Food and Scientific Reports, 1(6), 51–52.

Sillanpää, M. (1982). Micronutrients and the nutrient status of soils: A global study (Vol. 48). Food & Agriculture Organization.

Sindelar, M. (2015). Soils support agriculture. Journal of Soil Science Society of America. Retrieved from https://www.soils.org/files/sssa/iys/march-soils-overview.pdf.

Singh, P., Garg, C., and Namdeo, A. (2020). Applying machine learning techniques to extract dosages of fertilizers for precision agriculture. In IOP Conference Series: Earth and Environmental Science (p. 012136). IOP Publishing.

Sharma, V. P., and Thaker, H. (2010). Fertiliser subsidy in India: Who are the beneficiaries? Economic and Political Weekly, 45(27), 68–76.

Suchithra, M., and Pai, M. L. (2018a). Improving the performance of sigmoid kernels in multiclass SVM using optimization techniques for agricultural fertilizer recommendation system. In P. Vadakkepat, S. Bakshi, & P. Agarwal (Eds.), Soft Computing Systems: Second International Conference, ICSCS 2018, Kollam, India, April 19–20, 2018, Revised Selected Papers 2 (pp. 857–868). Springer.

Suchithra, M., and Pai, M. (2018b). Impact of Deep Neural Network on Predicting Application Rate of Fertilizers (Focus on Coconut Trees of Kerala Northern Coastal Plain Agro Ecological Unit). International Journal of Pure and Applied Mathematics, 119(10), 451–466.

Suleymanov, A., Gabbasova, I., Komissarov, M., Suleymanov, R., Garipov, T., Tuktarova, I., and Belan, L. (2023). Random Forest Modeling of Soil Properties in Saline Semi-Arid Areas. Agriculture, 13(5), 976. https://doi.org/10.3390/agriculture13050976.

Soil Health Card. Retrieved May 20, 2020, from https://soilhealth.dac.gov.in/

Sun, Y., Hu, R., and Zhang, C. (2019). Does the adoption of complex fertilizers contribute to fertilizer overuse? Evidence from rice production in China. Journal of Cleaner Production, 219, 677–685.

Sun, X., Liu, M., and Sima, Z. (2020). A novel cryptocurrency price trend forecasting model based on LightGBM. Finance Research Letters, 32, 101084.

Tao, H., Diop, L., Bodian, A., Djaman, K., Ndiaye, P. M., and Yaseen, Z. M. (2018). Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in Burkina Faso. Agricultural Water Management, 208, 140–151.