Soil Fertility and Yield Prediction of Coffee Plantation using Machine Learning Technique

Varshitha D. N and Savita Choudhary

Research Journal of Agricultural Sciences
An International Journal

P- ISSN: 0976-1675
E- ISSN: 2249-4538

Volume: 13
Issue: 02

Res. Jr. of Agril. Sci. (2022) 13: 514–518

CARAS

Res. Jr. of Agril. Sci. (Mar-Apr 2022)
13(2): 514–518

ISSN: 0976-1675 (P)
ISSN: 2249-4538 (E)

www.rjas.org

Full Length Research Article

Soil Fertility and Yield Prediction of Coffee Plantation using Machine Learning Technique

Varshitha D. N[1] and Savita Choudhary[2]

Received: 17 Jan 2022 | Revised accepted: 23 Mar 2022 | Published online: 15 Apr 2022

A B S T R A C T

Agriculture is the prime occupation of Indians. Majority in rural India is still dependent on the agriculture. A lot of research works on agriculture and crops happen every day. Agriculture in India is believed to be strengthened through modern agricultural practices. A lot of agricultural commodities are exported from India to all over the world. Indian beverages and spices are most popular agricultural products across the globe. Soil is the vital part of agriculture. Farmers can enhance their yield and income by knowing more about the soil. Farmers must be guided correctly regarding the soil health and crops so that unforeseen conditions can be avoided. If soil is tested and fertility rate of the soil with yield estimation is informed to farmer, it will be beneficial for him to continue the agriculture in right way. Based on soil quality and his other conditions farmer can choose the right crops and fertilizers for his land. In this paper we have discussed about the bagging approach which is a machine learning (ML) technique used for implementing soil fertility and yield prediction. We have considered some soil parameters such as pH, Nitrogen, Phosphorus, Potassium (NPK), humidity, temperature, organic carbon, moisture content and rainfall to predict the soil fertility and crop yield. We have also discussed about the test results and checked with other approaches of soil fertility & yield predictions to compare the performance of models.

***Key words*:** Agriculture, Bagging approach, Prediction, Machine learning, Soil parameters, Soil fertility, Yield

India is the world's largest producer of many varieties of crops. India exports large number of agricultural products across the globe. India is the country which has vast diversity. In India traditions, languages and food of people vary from one place to another depending on the geographical regions. Indian spices and beverages are world famous. Coffee is one such famous commercial crops grown in few eastern coastal regions and hilly areas of south Indians states like Karnataka, Kerala and Tamilnadu. About 80% of coffee grown in India is being exported from India to many parts around the world. Karnataka is the largest grower of the coffee and it is grown in Kodagu, Chikmagalore and Hassan. Robusta and Arabica are the varieties of coffee beans produced in Karnataka. Both species prefer very deep, well drained, non-gravelly, fine-loamy to clayey soils which have slightly - moderately acid reaction, high organic carbon, good water and nutrient retention capacities [1]. Pepper, cardamom, orange, banana and vanilla are few other predominately grown crops alongside coffee.

Since agriculture is the prime occupation of Indians, there is a wide scope of innovations in agricultural domain. For agricultural growth, soil plays vital role. Farmers in India have a lot of difficulties to opt for precision agricultural practices [2-3]. Adopting to new techniques to test soil becomes challenging to farmers due to lack of technical knowledge and support. Any approach which is given for farmer's assistance must be easy and less time consuming for usage. Soil fertility is the ability of soil to provide nutritious environment for the plant growth. pH level, Nitrogen, phosphorus and potassium (NPK), organic carbon, electric conductivity, moisture content, temperature, micro nutrients and macro nutrients are the soil parameters, all these in adequate amount are necessary for any plant growth to make the soil fertile and facilitate the crop growth [4-7]. It becomes very important to know about the soil fertility rate to grow the healthy crops [8]. Yield is the measure of crop produced in any agricultural land. It is directly proportional to the soil fertility rate. If the farmers want to increase the yield, monitoring the soil health becomes the primary concern. There are several schemes and methods employed to test the fertility of the soil which include chemical treatment of soil sample that takes a lot of time for estimation of fertility.

Artificial intelligence and Machine learning techniques can be used to assist the farmers to easily predict soil health and recommend the crops and other agricultural tasks in efficient ways [9-13]. We propose an approach which has Artificial Intelligence (AI), Machine Learning (ML) technique and Internet of Things (IoT) features to predict the soil & crop related information. Since we could collect the real data of

* **Varshitha D. N.**
✉ varshithadn.sirmvit@gmail.com

[1-2] Department of Computer Science & Engineering, Sir M Visvesvaraya Institute of Technology, Bengaluru - 562 157, Karnataka, India

coffee growing soil samples of Kodagu district of Karnataka, we have implemented the bagging which is a ML technique [14] to predict the soil fertility and yield of coffee plantation. Same technique can be employed to other crops too. Further the soil fertility and parameters are used to predict the crops and fertilizers recommendation for the same. Within in less time farmer can fetch the soil related information easily and accurately. In the subsequent sections of this paper we will see the methodology of the proposed system, results of the implementations.

There are many papers published on the topics soil fertility detection, crop yield estimation, crop and fertilizer recommendation, plant disease prediction, pesticide recommendation and so on using artificial intelligence and machine learning. Here we have discussed few concepts of soil fertility detection and crop yield prediction using ML models. Many of the papers have concept of soil fertility addressed as classification problem where the soil will be classified either as fertile or non-fertile. The yield can be estimated using regression models.

Support vector machine (SVM) finds a hyperplane to classify the data points in N dimensional space .data points are categorised based on the side it falls considering the hyperplane. Hence hyperplanes are decision boundaries. SVM is used both for classification and regression. SVM can be used for predicting soil quality and yield [15-21]. A lot of soil quality prediction & classification systems have been proposed based on the concept of SVM

Naïve Bayes algorithm is mostly a classification algorithm. It is based on Bayes' theorem. A Naive Bayes classifier has an assumption that the existence of a particular feature in a class is not related to the presence of other feature. Based on the probability, predictions are done in Naïve Bayes approach. Soil fertility and yield predictions are also done using Naïve Bayes approach [17-18].

Decision trees are supervised learning algorithms which can be also used for predicting soil fertility and estimating yield [16], [21-22]. Decision tree can create a training model that will predict a value by learning the simple rules which are meant for taking decisions understood from training data. Each example should go from the root down to the leaf based on the decision rules, Leaf node provides the classification result. Decision tree construction procedure is recursive

Artificial Neural Network (ANN) and Deep Neural Networks (DNN) are inspired by biological neural network for computation. The complexity of network depends on the number of layers present in the network. Network can have input layer, several hidden layers and output layers. Neurons in hidden layers perform computation. Neural networks can be used for predicting fertility of the soil and yield of crop [15-16], [18-24].

K-Nearest Neighbor (KNN) is the simplest supervised ML model that can be useful in solving problems that depend on identifying objects which are most similar. K-NN algorithm assumes the similarity between the new data and data already present and categorise the new data to most similar class .KNN is used for regression as well as for classification .The soil fertility and yield estimation is possible using KNN model[15-17], [20-22], [24-25].

MATERIALS AND METHODS

Soil fertility detection and yield prediction becomes an important task to assist the farmers to grow healthy crops and more yield. Poor soil quality cannot assure the healthy harvest. Healthy and fertile soil helps the plants grow well by providing all the nutrients required in adequate quantities. Soil parameters determine the fertility rate of any soil sample. We propose a system which we call it as 'Soil fertility and crop friendliness detection and monitoring using AI' where AI and ML techniques are used with some concepts of IOT. This is to provide awareness about the soil health and crops to the farmers. 'Soil Fertility detection and yield prediction' is a significant part of our proposed system with which fertility of the soil and yield could be predicted.

Initially soil parameters are taken from sensors or user input and classified it as known or unknown soil, then if soil type is known, fertility rate of soil along with yield could be predicted for the same, else the best crop suggestion along with soil fertility and yield prediction could be informed. The crops are suggested based on the results of previous module and later fertilizers-nutrients along with their quantities are recommended for the suggested crops. The soil test reports are stored in cloud and are sent to the farmers to their phones in easily readable and understandable format. The overview of 'Soil fertility detection and crop friendliness detection and monitoring using AI' can be seen in the (Fig 1).

Every day, a lot of agricultural researches happen with the application of computer science and technology, many of them fail to reach the farmers due to the inaccurate results and difficulty faced by farmers to get adopted to the technology. If a farmer gets his soil report easily and accurately, it will surely benefit him.

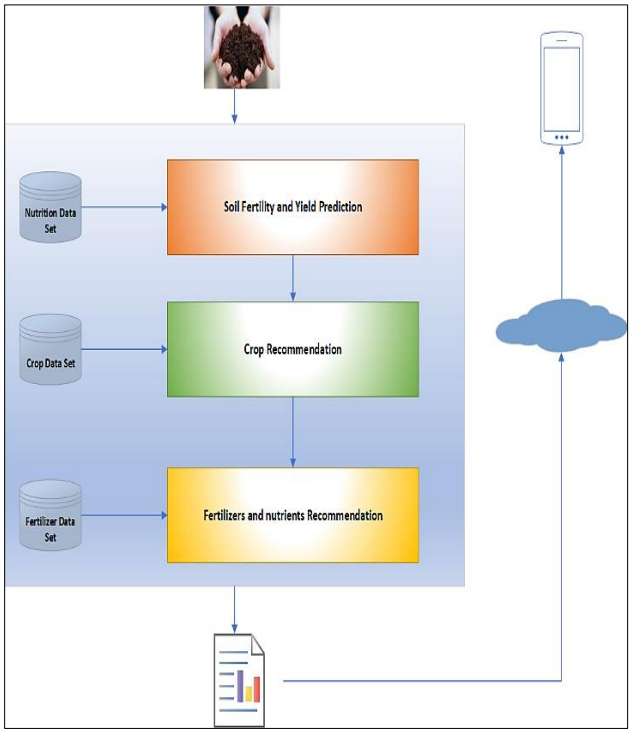

Fig 1 Overview of soil fertility detection and crop friendliness detection and monitoring using AI

Dataset

We could collect the data with the help of soil sensors in some part of Kodagu district of Karnataka using soil sensors and some data from coffee website of Government of India. pH level, organic carbon, Nitrogen-phosphorus- potassium (NPK) values, temperature, moisture content, humidity, rainfall is all considered to estimate and predict the fertility rate of the soil and yield of coffee plantations. 1100 coffee soil samples are collected in total and used for implementation. We have named the dataset as 'Nutrition dataset'. We can see the part of Nutrition dataset in the (Table 1).

Table 1 Part of nutrition dataset stored as CSV file

pH	Organic carbon	Phosphorus	Potassium	Nitrogen	Moisture	Rainfall	Humidity	Temperature	Fertility	Yield
5.7	1.9	4	370	0	21	194.8773	68.4803	25.11411	35	350
5.7	1.9	4	370	0	22.2	134.6804	58.84881	24.02953	35	350
5.7	1.9	4	370	0	23	158.8609	69.71184	26.70898	35	350
5.8	1.9	4	370	0	21.5	126.8074	50.50149	26.71717	35	350
5.4	4.4	3	516	0	32.3	156.681	58.52534	27.57848	40	400
4.3	3.6	2	482	0	31.1	145.1051	57.56696	26.65069	40	400
5.4	4.4	3	516	0	36	142.8611	62.01836	26.97252	40	400
5.4	4.4	3	516	0	23	153.1202	52.97841	23.17125	40	400

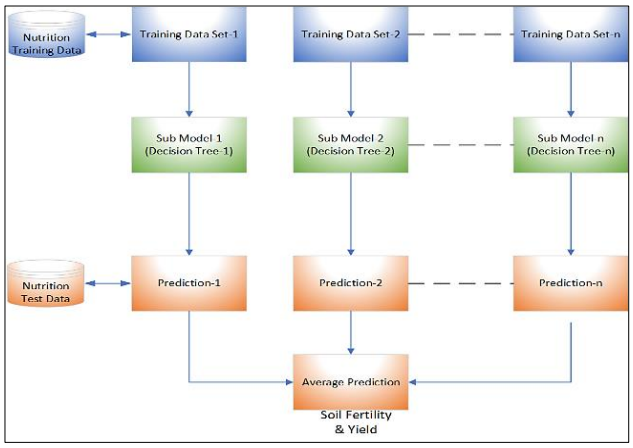

Fig 2 Bagging approach to predict the soil fertility and Yield

Bagging approach for soil fertility and yield prediction

The dataset of coffee nutrition is highly nonlinear and has different parameters with different ranges. Each sample of data is different with that of other sample. Hence normal classification algorithm doesn't work well. Also, we are not classifying the soil as fertile or in fertile like other fertility detection system. Our prediction system aims at providing exact fertility rate to the farmer. Hence, we go for regression instead of using classification technique. Regressions are performed to find the exact fertility rate of the soil and yield of coffee in kilogram per acre of land.

We use bagging technique for predicting the fertility rate and yield of the coffee plantation. In this approach we shall be using ensemble model having random forest technique with 'n' number of base models where each sub model gets trained on a part of shuffled samples of coffee Nutrition dataset and some features of data set. Hence, we can say each sub model gets trained on different samples and different parameters of data set. Base sub models are decision tree classifiers. In general, we can call it as random forest of decision trees. So whenever new prediction has to happen or we have to check the predictions test data, the average of predicted values of all sub models are considered as final predicted output. As the final output is the result of several sub models, we can expect accurate prediction. Random forest is also called as bootstrap method since random sample of the data are given to each model with replacement. Also, random forest bagging technique overcomes the overfitting issues, reduces variance and any new predictions would be highly accurate, so high performance becomes the great advantage of using this approach. Error rate will be minimized because of the models getting trained on random samples. In the (Fig 2), the bagging approach of predicting soil fertility and yield is shown.

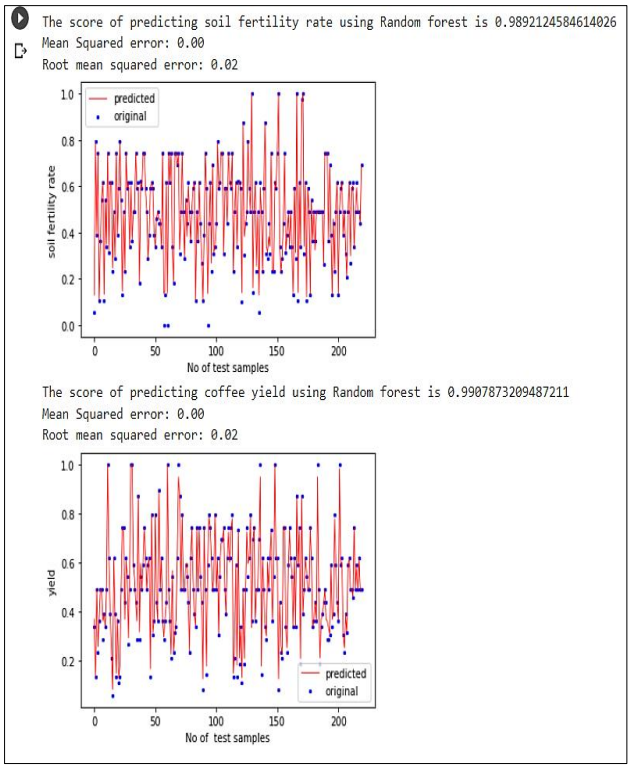

Fig 3 Test data predictions plotted in graph with r2 score for soil fertility and crop yield prediction using bagging approach

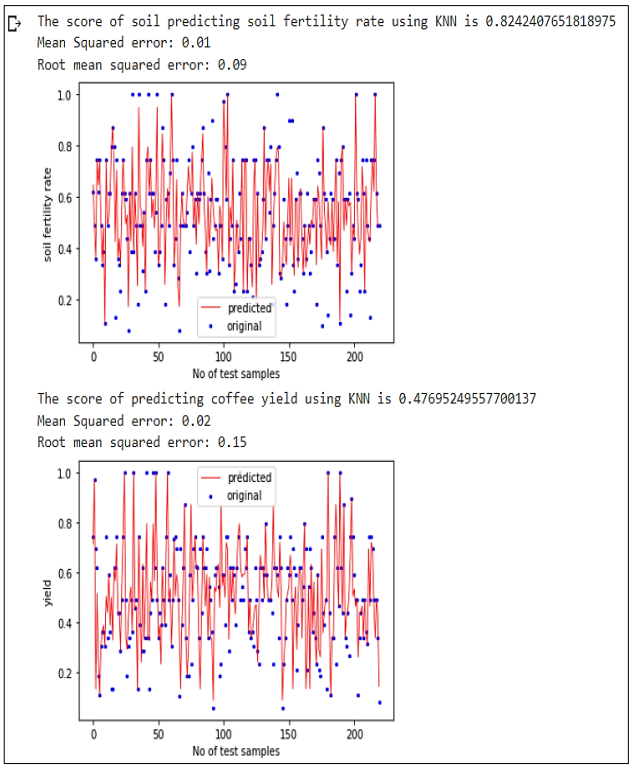

Fig 4 Test data predictions plotted in graph for soil fertility and crop yield prediction using KNN

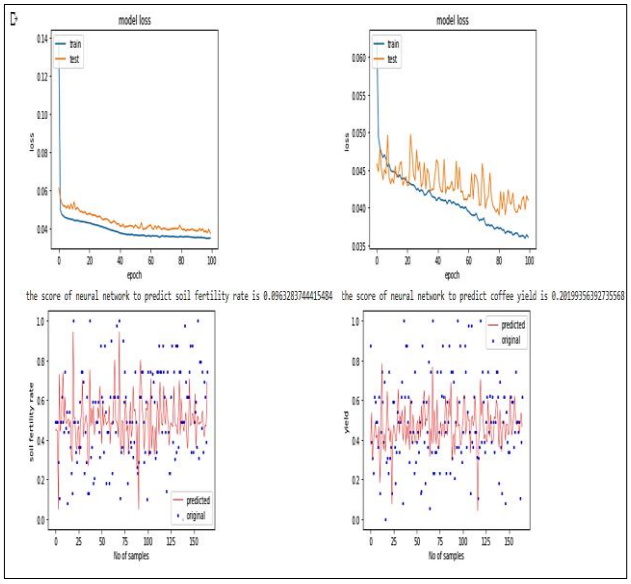

Fig 5 Test data predictions plotted in graph for soil fertility and crop yield prediction using DNN

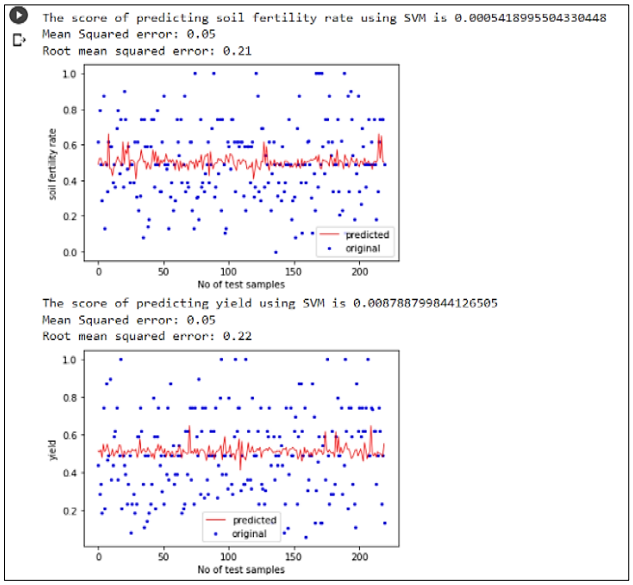

Fig 6 Test data predictions plotted in graph for soil fertility and crop yield prediction using DNN

The Necessary libraries and packages are loaded into out colab file. The soil fertility and yield detection module are implemented in colab platform, The Nutrition dataset is read into our colab file. Data is normalized to put the data in the common scale so that system accuracy will be more. Later parameters are set to input and target. Data is split as 80% training and 20% testing data and regression models are run; data is being fit to the model. Predictions are performed and evaluation metrics are derived and analyzed along with plotting the same on graphs. Score value and mean squared error is analyzed to see the efficiency of the model. Score is nothing but coefficients of determination which is referred as r2 score. For perfect prediction the r2 score value will be approximately equal to 1.

RESULTS AND DISCUSSION

The random forest –bagging approach suggested in our proposed system gives the score of 0.98 for soil fertility rate and 0.99 for yield prediction. The score which is approximately equal to '1' indicates the best accuracy. The predictions and result of test data are shown in (Fig 3), the graphs indicate that almost all the predicted values (red lines) are meeting the original values (blue dots). We have implemented few other models like KNN, SV regression and DNN on same dataset to check the accuracies against bagging approach and we get following results. The KNN shows r2 score as 0.82 for fertility prediction and 0.47 for yield prediction on our dataset and other regression models are working poor. Hence, we can conclude that Bagging Random Forest approach gives the best predictions. The test prediction results using KNN, DNN and SVM are shown in (Fig 4-6) respectively.

We can see the comparison chart shown in (Fig 7). Where all the regression models and their r2 score values for both fertility and yield prediction values are compared.

CONCLUSION

We propose a system for assisting the farmer to know about the soil. Soil fertility and yield prediction becomes very important to guide the farmers to do the agriculture in the right way. If fertility and yield is known then crops & fertilizers can be rightly chosen. We have used Bagging ensemble technique which is a Random Forest of several decision trees for the implementation. We perform regression to get the exact values for results. We have also checked with other models and obtained the results. We obtained very low score for support vector regression and deep neural network and we got score values as 0.47 and 0.82 for fertility of soil and yield respectively for K nearest neighbour models which shows average performance and bagging random forest method shows 0.99 and 0.98 score values which means bagging shows maximum accuracy and best predictions for coffee plantation .Hence we can conclude that bagging approach is the best technique used for predicting soil fertility and yield. Our model gives accurate results for any new predictions. Model is robust to outliers, overfitting and variance problems are also absent.

Regression Models	r2 score (soil fertility)	r2 score(yield)
SV regression	0	0
Deep Neural Network	0	0.2
KNN	0.82	0.47
Bagging	0.98	0.99

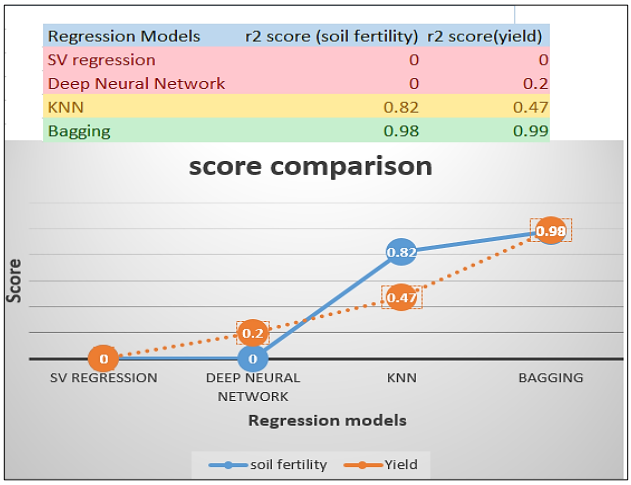

Fig 7 r2 scope comparison chart of the regression models

LITERATURE CITED

1. Shalima Devi GM, Anil Kumar KS. 2009. Characteristics of coffee-growing soils and their organic carbon stocks in Karnataka state. *Agropedology* 19(1): 6-14.
2. Katke K. 2019. Precision Agriculture Adoption: Challenges of Indian Agriculture. *International Journal of Research and Analytical Reviews* 6(1): 863-869.

3. Dev MS. 2014. Small farmers in India: Challenges and opportunities. *Indira Gandhi Institute of Development Research, Mumbai Working Papers 2012-014*, Indira Gandhi Institute of Development Research, Mumbai, India.
4. Warudkar G, Dorle S. 2016. Review on sensing the fertility characteristics of agriculture soils. *International Conference on Information Communication and Embedded Systems*. pp 1-6, doi: 10.1109/ICICES.2016.7518850
5. Masrie M, Rosman MSA, Sam R, Janin Z. 2017. Detection of nitrogen, phosphorus, and potassium (NPK) nutrients of soil using optical transducer. IEEE 4th *International Conference on Smart Instrumentation, Measurement and Application* (ICSIMA). pp 1-4, doi: 10.1109/ICSIMA.2017.8312001.
6. Chitragar AA, Vasi SM, Naduvinamani S, Katigar AJ, Hulasogi TI. 2017. Nutrients detection in the soil: Review Paper. *International Journal on Emerging Technologies* 2017(2): 257-260.
7. Trontelj J, Chambers O. 2021. Machine learning strategy for soil nutrients prediction using spectroscopic method. *Sensors* 21: 4208. https://doi.org/10.3390/ s21124208
8. Amrutha A, Lekha R, Sreedevi A. 2016. Automatic soil nutrient detection and fertilizer dispensary system. *International Conference on Robotics: Current Trends and Future Challenges*. pp 1-5, doi: 10.1109/RCTFC.2016.7893418.
9. Rao TVN, S. Manasa S. 2019. Artificial neural networks for soil quality and crop yield prediction using machine learning. *International Journal on Future Revolution in Computer Science and Communication Engineering* 5(1): 57-60.
10. Zhao J, Guo J. 2018. Big data analysis technology application in agricultural intelligence decision system. *IEEE 3rd International Conference on Cloud Computing and Big Data Analysis*. doi: 10.1109/ICCCBDA.2018.83865
11. Maurya B, Beg MR, Mukherjee S. 2013. Expert system design and architecture for farming sector. *IEEE Conference on Information and Communication Technologies*. pp 10-15. doi: 10.1109/CICT.2013.6558053.
12. Benos L, Tagarakis AC, Dolias G, Berruto R, Kateris D, Bochtis D. 2021. Machine learning in agriculture: A comprehensive updated review. *Sensors (Basel)* 21(11): 3758. doi: 10.3390/s21113758
13. Meshrama V, Patila K, Meshrama V, Hanchateb D, Ramktekec SD. 2021. Machine learning in agriculture domain: A state-of-art survey .*Artificial Intelligence in the Life Science*s Volume 1, December 2021, 100010 https://doi.org/10.1016/j.ailsci.2021.100010
14. Biau G. 2012. Analysis of a random forests model. *Journal of Machine Learning Research* 13(2012): 1063-1095. retrieved from https://www.jmlr.org/papers/volume13/biau12a/biau12a.pdf
15. Gopal M, Bhargavi. 2019. Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms. *Applied Artificial Intelligence: An International Journal* 33(7): doi: https://doi.org/10.1080/08839514.2019.1592343
16. Janez T. 2021. Soil fertility prediction tool for real time precision agriculture. *Mod Concep Dev Agrono.* 8(3). *MCDA.* 000689. 2021. DOI: 10.31031/MCDA.2021.08.000689
17. Wankhede DS. 2021. Analysis and prediction of soil nutrients pH, N, P, K for crop using machine learning classifier: A review. *In*: (Eds) Raj J.S. *International Conference on Mobile Computing and Sustainable Informatics. ICMCSI 2020. EAI/Springer Innovations in Communication and Computing. Springer*, Cham. https://doi.org/10.1007/978-3-030-49795-8_10
18. Jadav J, Chopra S, Vijaylaksmi. 2021. Soil analysis and crop fertility prediction using machine learning. *International Journal of Innovative Research in Advanced Engineering* 3(8): 41-49.
19. Dang C, Liu Y, Yue H, Qian JX, Zhu R. 2021. Autumn crop yield prediction using data-driven approaches: Support Vector Machines, Random Forest, and Deep Neural Network Methods, *Canadian Journal of Remote Sensing* 47(2): 162-181.
20. Sai R, Sathiaseelan JGR. 2018. Comparison of classifiers to predict classification accuracy for soil fertility. *International Journal of Advanced Studies of Scientific Research* 3(9): 75-79.
21. Chaudhari R, Chaudhari S, Shaiak A, Chiloba R, Khabtare. 2020. Soil fertility prediction using data mining techniques. *Mukt Shabd Journal* 9(6): 202-209.
22. Janvier N, Arcade N, Eric N, Jean N. 2021. Machine learning based soil fertility prediction. *International Journal of Innovative Science, Engineering and Technology* 8(7): 2348-7968.
23. Klompenburg TV, Kassahun A, Catal C. 2020. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture* 177: https://doi.org/10.1016/j.compag.2020.105709
24. Reddy DJ, Kumar MR. 2021. Crop yield prediction using machine learning algorithm. *5th International Conference on Intelligent Computing and Control Systems*. pp 1466-1470. doi: 10.1109/ICICCS51141.2021.9432236.
25. Insozhan N, Parthasarathy V. 2017. Evaluation and management of soil fertility international journal of pure and applied mathematics.117(8): 11-15.