2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)

Soil Nutrient Prediction and Crop Recommendation System

[1]Madhumathi R,
Department of Computer Science and Engineering,
Sri Ramakrishna Engineering College, Coimbatore, India
Email: madhumathi.r@srec.ac.in

[2]Ashok Kumar K, [3]Chandra Kumar N , [4]Dhyanesh A
Department of Computer Science and Engineering,
Sri Ramakrishna Engineering College, Coimbatore, India
Email: ashokkumar.2101017@srec.ac.in

[5]Arumuganathan T,
ICAR - Sugarcane Breeding Institute, Coimbatore, India
Email: arumuganathan@gmail.com

2023 International Conference on Circuit Power and Computing Technologies (ICCPCT) | 979-8-3503-3324-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCPCT58313.2023.10245685

***Abstract*— Food is the necessity of humans which is in great demand. Agriculture is the major domain and primary source of food. Soil is the main resource for agriculture where the essential nutrients required for plant growth is available. Agriculture has progressed over time with modern technologies and more equipment's have replaced ancient farming methods. There is a great need to manage the amount of fertilizer to be added for every crop. Without soil testing, it is very difficult to ensure the right application of fertilizers for the crop and get the optimum yield. Analyzing soil samples in a laboratory is a time-consuming process that entails numerous steps, such as gathering soil samples, preparing them for analysis, conducting chemical or physical analysis, interpreting the analysis results, and ultimately providing fertilizer recommendations for crops. In this work, soil nutrient prediction system is developed which predicts the Nitrogen, Phosphorus, Potassium (NPK) values and recommends the crop to be cultivated. In this system, BORUTA regression model gives highest accuracy of 91%. Python is used to develop this prediction and recommendation system.**

Keywords—Smart Agriculture, NPK, Machine Learning, Regression Methods, Nutrient Prediction, Crop Recommandation.

I. INTRODUCTION

Agriculture has been a vital sector of human society for thousands of years, providing food, fiber, and other resources essential for survival. However, the agricultural sector is currently dealing with a number of issues, such as population growth, climate change, and the necessity to produce more food with less resources. To meet these challenges, the agricultural sector is increasingly turning to technology for solutions. From precision agriculture and biotechnology to robotics and big data analytics, technology is being used to improve efficiency, reduce waste and increase productivity.

In recent years, the use of data-driven approaches to agriculture has become increasingly popular. Soil nutrient prediction and crop recommendation systems are examples of such approaches, aimed at providing farmers with recommendations on the optimal crop to plant and the number of fertilizers to apply based on soil nutrient levels. These systems use machine learning algorithms to analyze soil data and historical crop yield data to make predictions about the optimal crop and fertilizer application rate for a given plot of land.

The use of soil nutrient prediction and crop recommendation system has several potential benefits. First, it can help farmers optimize their use of resources such as fertilizers, leading to cost savings and improved sustainability. Second, it can help improve crop yields and quality, leading to higher profits for farmers. Finally, it can help the environmental impact of agriculture by reducing the use of fertilizers and other inputs. The system uses machine learning algorithms to analyze soil data and historical crop yield data. Farmers face several challenges when it comes to soil nutrient prediction, which can impact their ability to optimize fertilizer use and improve crop yields. By providing farmers with reliable information about their soil and crop nutrient requirements, can help the farmers optimize their fertilizer use and improve crop yields. Overall, this work demonstrates the potential of soil nutrient prediction and crop recommendation system to improve the sustainability and profitability of agriculture.

The structure of the paper consists of the following sections. The related research surrounding the analysis of soil nutrients are covered in Section II. The model proposed for the nutrient detection is detailed in Section III. Sections IV and V explains the results and conclusion, respectively.

II. RELATED WORK

Vrunda Kusanur et al. used Convolutional neural networks (CNN) with Transfer Learning (TL) models like Inception V3 to automate the diagnosis of nutrient deficiencies etc. [1]. Sudha Bhatia and colleagues collected soil samples from the farms of Amity University in Dubai to investigate how different pesticide treatments affect two important soil characteristics namely soil nutrient content and soil pH [2]. José Escorcia-Gutierrez et al. recommended deep learning methods for evaluating soil nutrients [3].

Gaganjot Kaur performed study and read several studies and used many frameworks to analyze the negative impacts of a deficit much more quickly than the human eye could. These frameworks employ digital image processing, computer vision, and IoT [4]. Mayuri Sharma et al. employed transfer learning models, a type of CNN-based Deep Learning (DL), in their agricultural study [5].

Hiroyuki Onoyama and co-authors employed a ground-based hyperspectral remote sensing technique in their research to determine the amount of nitrogen present in panicled rice plants, using the GreenNDVI-NDVI equation [6]. Jaafar Abdulridha and colleagues suggested a potential solution for this investigation by proposing to use unsupervised pre-training on unlabeled aerial photos that can be easily obtained through the UAS [7]. Jiating Li et al. introduced a non - invasive remote sensing approach [8]. Nevertheless, Jayme Garcia employed Machine Learning (ML) in the unique instance of proximal pictures [9]. Silvia Sau et al. addressed morpho-colourimetric properties and used it to validate their approach [10].

Anna Chlingaryan et al utilized the AgriNet dataset, which includes 423 categorizations of species of plants and illnesses, 160k agricultural images from more than nineteen different geographic regions and more than 423 plant diseases [11]. An IoT-enabled system using the colorimetry concept was created by Madhumathi et al [12]. The soil nutrients are assessed or evaluated in terms of their quantity, and a fuzzy expert system is used to make recommendations on how much fertilizer should be supplied. Using WSN, in order to remotely track soil health and other soil characteristics, Madhumathi et al. created a technique for a precision farming system that uses cloud and mobile apps [13]. Multiple linear regression (MLR) techniques were used by Madhumathi et al. to develop ML model for figuring out the NPK content of soil [14]. The accuracy rate of the prediction, which uses multiple linear regression is 78%.

In a tomato crop cultivated without soil, A.Elvanidi et al. demonstrated how to use non-contact hyperspectral machine vision way of identifying nitrogen shortage [15]. The findings indicated that the values of (MSAVI) and (OSAVI) changed by more than 0.05 and 0.25, respectively, for nitrogen rises more than 0.20%. Hao-Tian Cai et al. Notable are recent developments in machine learning-based systems for precise agricultural production prediction and nitrogen status assessment. [16]. To create a model for extracting information about soil nutrients, Zahraa Al Sahili intended to integrate the near-infrared spectroscopy technology and the transfer learning algorithm [17]. DL networks and transfer learning techniques were utilized by Borja Espejo-Garcia et al. to identify the signs of nutritional shortage in RGB photographs [18]. The best results were obtained by picture classification using EfficientNetB4's fine-tuning, whose initial weights were from a chaotic student ImageNet training set, with Top-1 accuracies of 98.65% and 98.52% on both datasets.

A method that automatically analyses soil nutrients using regression approaches was proposed by Sirsat et al. [19]. Usman Ahmed et al. proposed nutrient recommendation for soil fertilization using evolutionary methods [20].

III. PROPOSED SYSTEM

Analysis of the soil nutrients is routinely performed to learn more about the soil quality, nutrient content, and changes to key soil characteristics. Nutrient concentration, moisture content, soil texture, pH and many other factors may be measured on soil to learn more about its properties and to decide the appropriate treatment to increase its fertility. Macronutrients like NPK are typically searched for during soil testing processes. A NPK sensor is used in the proposed system to measure the macronutrients, which are then categorized as low, medium, and high. The most effective way to gather and measure data from many sources is through data collection. These characteristics must be present in this dataset. The variables like Soil pH, Temperature, Humidity, Crop data, NPK values are taken into consideration for crop prediction:

After compiling information from several sources, model training must come before dataset preparation. The first phase in the data preprocessing process, which comes after data cleaning, is reading the obtained dataset. Some expendable attributes are removed from the datasets during data cleaning so that crop predictions can be made. Therefore, it is necessary to eliminate pointless attributes and fill in any missing values in datasets with unfavorable nan values in order to increase accuracy.

Machine Learning predictive algorithms have highly enhanced predictions of future results based on taught data. Predictive analytics calculates the likelihood of future outcomes based on prior data using various regression methods and machine learning techniques. Fig.1 illustrates the working flow of the nutrient prediction model.

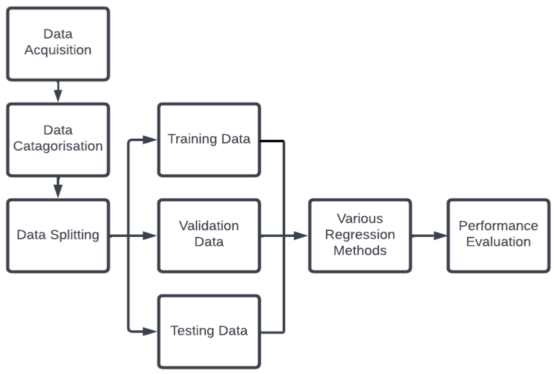

Fig 1. Flow Diagram for Soil Nutrient Prediction System

Algorithm – Soil Nutrient Prediction System:

1. Collect the data set
2. Do preprocesing
3. Split( ) the data to evaluate its performance
4. Validate( )
5. Build models()
   5.1 Multiple Linear Regression
   5.2 Lasso Regression
   5.3 Elasticnet
   5.4 Ridge Regression
   5.5 Polynomial Regression
   5.6 Boruta
   5.7 SVM
6. Evaluate the preformance of the model
7. Predict NPK
8. Plot the graph

A. *Regression Analysis*

Regression analysis is a statistical technique that facilitates quantifying and understanding the relationship between two or more variables. The dependent variable named Y is the variable to predict, while the independent variable named as X is the variable that believe has an influence on the dependent variable.

A mathematical model that can accurately forecast the value of the dependent variable given the independent variable. Once the equation or model is established, predictions about the dependent variable for new values of the independent variables are calculated.

Procedure – Multiple Linear Regression

A statistical method known as MLR is like linear regression predicts the relationship between one dependent and more independent variables and it is shown in eq (1).

$$Y = b_0 + b_1X_1 + b_2X_2 + ... + b_n*X_n + \varepsilon \quad (1)$$

where,

Y - Dependent variable
X_1-X_n -> Independent variables
b_1-b_n -> coefficients for each independent variable
b_0 - Intercept
ε - error term

The polynomial regression is also follows same as equation eq. (1).

Procedure – Lasso Regression

The Lasso regression is more like multiple linear regression, but with an additional penalty term that encourages the coefficients to be as small as possible denoted in eq. (2).

$$\text{Objective function} = SS_{res} + \lambda * \Sigma |bi| \quad (2)$$

where,

λ - hyperparameter that controls the strength of the penalty term
$\Sigma |bi|$ - L1 norm of the coefficients

Procedure – Elasticnet

Elastic net linear regression applies the penalties of both lasso and ridge regression techniques to regulate regression models. The objective function in Elastic Net regression is a combination of the L1 and L2 penalties is shown in eq. (3).

$$\text{Objective function} = SS_{res} + \lambda1 * \Sigma |bi| + \lambda2 * \Sigma b^{2i} \quad (3)$$

Where,

$\lambda1$ and $\lambda2$ - hyperparameters that control the strength of the L1 and L2 penalties, respectively.

Procedure – Ridge Regression

Ridge regression functions by introducing a penalty term to the cost function, which is proportional to the sum of the squares of the coefficients is shown in eq. (4).

$$\text{Objective function} = SS_{res} + \lambda * \Sigma b^{2i} \quad (4)$$

Where,

λ - hyperparameter that controls the strength of the penalty term
Σb^{2i} - the L2 norm of the coefficients.

Procedure – Boruta

Initially, shuffled duplicates of all features are generated to add randomness to the dataset.Next, a random forest classifier is trained on the expanded dataset, and a feature importance measure is computed.

It also follows the eq. (1). At each iteration, the algorithm evaluates whether an actual feature has more significance than the top shadow feature.The algorithm continues until all features are verified, and it concludes when all features are confirmed.

Procedure – Support Vector Machine (SVM)

SVM is known for its ability to handle large datasets with high accuracy and good generalization performance. It is also known for its robustness to noise and outliers in the data. The eq. (5) for the hyperplane is defined as,

$$w^T x + b = 0 \quad (5)$$

where,

w – normal vector
x – feature vector
b - offset

The accuracy of the model is found using the formula eq. (6),

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (6)$$

where,

SS_{res} - Sum of the squared residuals.
SS_{tot} - Total sum of squares.

Fig.2 illustrates the working flow of the crop recommendation model.

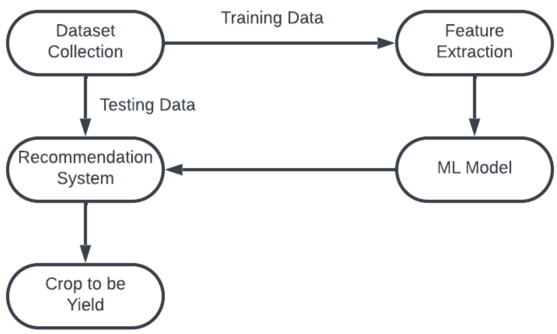

Fig 2. Flow Diagram for Crop Recommendation System

Algorithm: Crop Recommendation System

1. Collect data set
2. Do preprocessing
3. Split( ) the data to evaluate its performance
4. Validate( )
5. Build()
   5.1 KNN (K-Nearest Neighbour) model
6. Evaluate() the preformance

7. Predict()

Procedure - K-Nearest Neighbour

- Assign the neighbors' K-numbers.
- Calculate Euclidean distance K()
- Select the K closest neighbors
- Count data points based on k-neighbors.
  IF K count is high ->assign the fresh data points
- Build recommendation model

IV. RESULT

In this work, various regression methods are used for predicting the NPK values. With respect to the pH, EC, the NPK value is predicted. The graph is plotted for each method and the comparison graph based on the accuracies is also shown. Nutrient prediction model is developed using various regression models.

Fig. 3 shows actual and predicted values in multiple linear regression. The accuracy of this model is 85%.

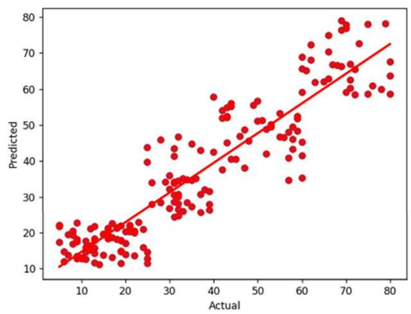

Fig 3. Multiple Linear Regression

Fig. 4 illustrates actual and predicted value of Lasso Regression. The accuracy of this model is 84%.

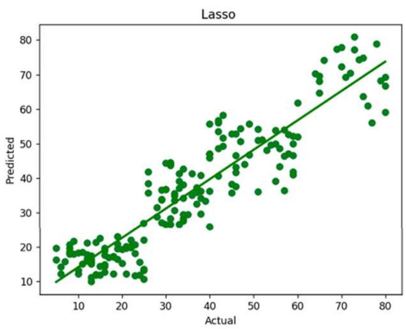

Fig 4. Lasso Regression

Fig. 5 shows the graph plotted between actual and predicted value in Elasticnet. The accuracy of this model is 81%.

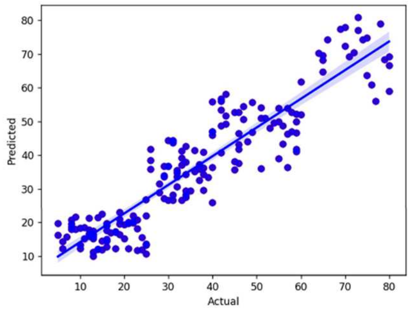

Fig 5. Elasticnet

Fig. 6 shows the actual and predicted values of Ridge Regression. The accuracy of this model is 83%.

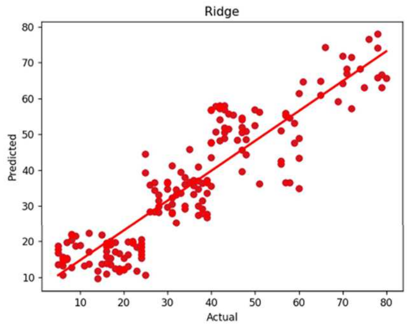

Fig 6. Ridge Regression

Fig. 7 shows the actual and predicted values of polynomial regression. The accuracy of this model is 84%.

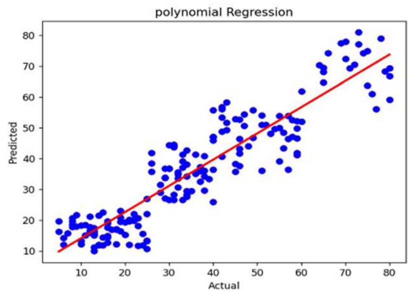

Fig 7. Polynomial Regression

Fig. 8 demonstrates the actual and predicted values of Boruta. The accuracy of this model is 91%.

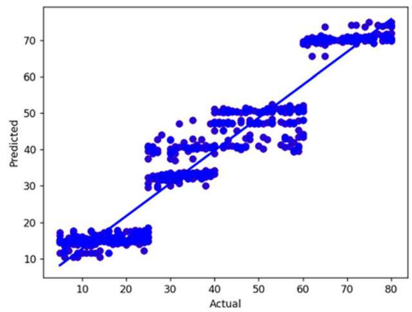

Fig 8. Boruta

Fig. 9 shows the graph plotted between actual and predicted value in SVM. The accuracy of this model is 77%.

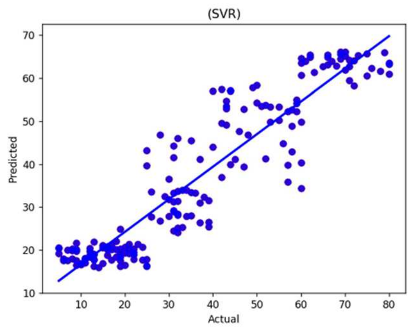

Fig 9. SVM

Fig. 10 shows the comparison between various regression methods with their accuracies.

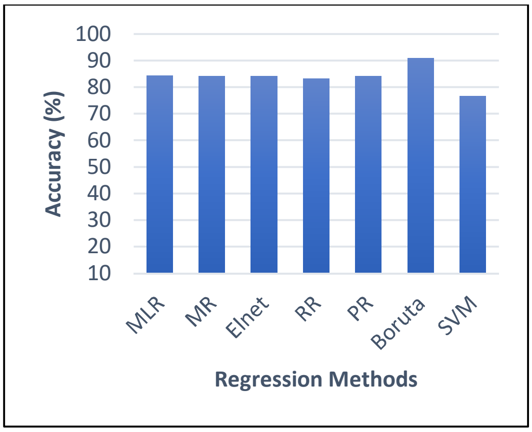

Fig 10. Various Regression Methods with accuracy

NPK values are predicted using various regression methods and got highest accuracy of 91% in Boruta.

The Crop recommendation system is developed using the K-Nearest Neighbor method and got accuracy of 91%. Based on N, P, K, pH, temperature, humidity the crops like rice, maize, chickpea, black gram and lentil are recommended. The implementation of the crop recommendation system for rice crop is shown in Fig. 11.

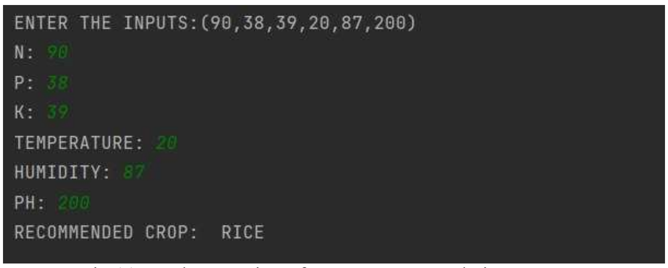

Fig 11. Implementation of Crop Recommendation System

Fig. 12 shows the confusion matrix of the crop recommendation system.

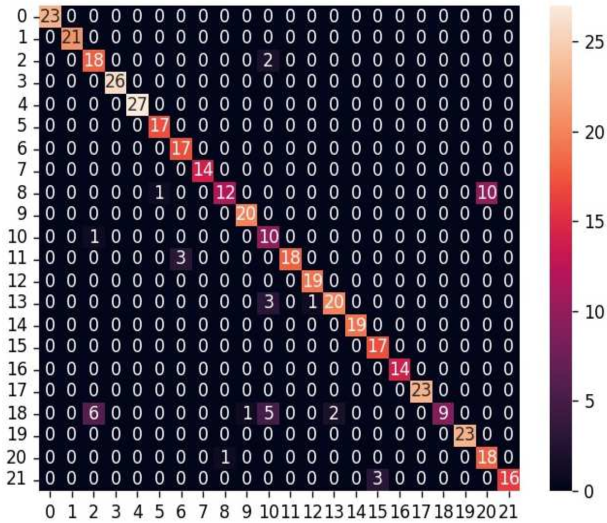

Fig 12. Confusion Matrix of the Crop Recommendation System

V. CONCLUSION

Different regression methods are employed to conduct a soil nutrient prediction. The findings revealed that the BORUTA model demonstrated higher efficiency with an accuracy rate of 91%. The development of a Python program enabled the creation of a crop recommendation system, which suggests appropriate crops based on the NPK value. Furthermore, a crop nutrient recommendation system is created to suggest the appropriate fertilizer amount required for the crop, based on the measured nutrient value. However, the fertilizer recommendation system in this model is only proposed for macronutrients in various crops. Future upgraded models will be developed to improve crop recommendation systems with the integration of real-time data. This could include weather data, soil moisture data, and other data sources that can help provide more accurate and up-to-date recommendations to farmers.

ACKNOWLEDGMENT

The authors acknowledged the DST –SERB (TARE) scheme, Sri Ramakrishna Engineering College and ICAR – Sugarcane Breeding Institute (SBI), Coimbatore for providing the facilities required for conducting this study.

REFERENCES

[1] Vrunda Kusanur and Veena S and Chakravarthi, “Using Transfer Learning for Nutrient Deficiency Prediction and Classification in Tomato Plant”, *International Journal of Advanced Computer Science and Applications, Vol. 12, No. 10, pp. 784-790, 2021*

[2] Sudha Bhatia, Reshmi S Nair and Ved Prakash Mishra, “Nutrient Analysis of Soil Samples Treated with Agrochemicals”, *International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), pp.64-67, 2021.*

[3] José Escorcia-Gutierrez et al., “Intelligent Agricultural Modelling of Soil Nutrients and pH Classification Using Ensemble Deep Learning Techniques”, *Journal of Agriculture, MDPI, pp.1-16,* 2022.

[4] Gaganjot Kaur, “Automated Nutrient Deficiency Detection in Plants: A Review”, *Journal of Archaeology of Egypt/Egyptology, vol. 17, no. 6, pp. 5894-5901, 2020.*

[5] Mayuri Sharma et al., "Ensemble Averaging of Transfer Learning Models for Identification of Nutritional Deficiency in Rice Plant", *Electronics, MDPI, vol. 11, pp.1-16, 2022.*

[6] Hiroyuki Onoyama et al. "Potential of Hyperspectral Imaging for Constructing a Year-invariant Model to Estimate the Nitrogen Content of Rice Plants at the Panicle Initiation Stage*", 4th IFAC Conference on Modelling and Control in Agriculture, Horticulture and Post Harvest Industry, Espoo, Finland, pp.219-224, 2013.*

[7] Jaafar Abdulridha et al., "Evaluating the Performance of Spectral Features and Multivariate Analysis Tools to Detect Laurel Wilt Disease and Nutritional Deficiency in Avocado", *Elsevier B.V, pp.203-211, 2018.*

[8] Jiating Li et al., "Improving Model Robustness for Soybean Iron Deficiency Chlorosis Rating by Unsupervised Pre-Training on Unmanned Aircraft System Derived Images", *Elsevier B.V, Computers and Electronics in Agriculture, vol. 175, pp.1-11, 2020.*

[9] Jayme Garcia Arnal Barbedo, "Detection of nutrition deficiencies in plants using proximal images and machine learning: A review", *Computers and Electronics in Agriculture, Elsevier, vol. 162, pp.482-492, 2019.*

[10] Silvia Sau et al., "Potential use of seed morpho-colourimetric analysis for Sardinian apple cultivar characterisation", *Computers and Electronics in Agriculture, Elsevier, vol. 162, pp. 373–379, 2019.*

[11] Anna Chlingaryan et al., "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review", *Computers and Electronics in Agriculture 151, pp. 61–69, 2018.*

[12] R. Madhumathi, T. Arumuganathan, and R. Shruthi, "Soil Nutrient Detection and Recommendation Using IoT and Fuzzy Logic", *Computer Systems Science and Engineering (CSSE), vol. 43, Issue no. 2, pp. 455-469, 2022.*

[13] R. Madhumathi, T. Arumuganathan, and R. Shruthi, "Soil NPK and Moisture analysis using Wireless Sensor Networks," *11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-6, 2020.*

[14] R. Madhumathi, T. Arumuganathan, and Sneha Iyer, "Soil Nutrient Analysis Using Machine Learning Techniques", *National E-Conference on "Comunication, Computation, Control and Automation (CCCA-2020)", pp. 1-16, 2020.*

[15] A. Elvanidi et al., "Crop reflectance measurements for nitrogen deficiency detection in a soilless tomato crop", Bio Systems Engineering, *Elsevier, vol. 176, pp. 1-11, 2018.*

[16] Hao-Tian Cai et al., "Soil nutrient information extraction model based on transfer learning and near infrared spectroscopy", *Alexandria Engineering Journal, Elsevier, vol. 60, pp. 2741–2746, 2021.*

[17] Zahraa A Sahili1 and Mariette Awad, "The Power of Transfer Learning in Agricultural Applications: AgriNet", Technical Advances in Plant Science, vol. 13, *pp. 1-19, 2022.*

[18] Borja Espejo-Garcia et al., "Using EfficientNet and transfer learning for image-based diagnosis of nutrient deficiencies", *Computers and Electronics in Agriculture, Elsevier, vol. 196, pp. 1-11, 2022.*

[19] M.S. Sirsat, E. Cernadas, M. Fernández-Delgado and S. Barro, "Automatic prediction of village-wise soil fertility for several nutrients in India using a wide range of regression methods", *Computers and Electronics in Agriculture, vol. 154, pp. 120-133, 2018.*

[20] Usman Ahmed et al., "A nutrient recommendation system for soil fertilization based on evolutionary computation", *Computers and Electronics in Agriculture, vol. 189, pp. 1-7 Elsevier, 2021.*