agriculture

MDPI

Article

A Comprehensive Analysis of Machine Learning-Based Assessment and Prediction of Soil Enzyme Activity

Yogesh Shahare [1], Mukund Partap Singh [2], Prabhishek Singh [2], Manoj Diwakar [3], Vijendra Singh [4], Seifedine Kadry [5,6,7,8] and Lukas Sevcik [9,*]

[1] Department of Information Technology, Mahatma Gandhi Mission's College of Engineering and Technology (MGMCET), Navi Mumbai 410 209, India
[2] School of Computer Science & Engineering Technology, Bennett University, Greater Noida 201310, India
[3] Computer Science and Engineering Department, Graphic Era (Deemed to be University), Dehradun 248002, India
[4] School of Computer Science, University of Petroleum and Energy Studies, Dehradun 248007, India
[5] Department of Applied Data Science, Noroff University College, 4612 Kristiansand, Norway
[6] Artificial Intelligence Research Center (AIRC), Ajman University, Ajman 346, United Arab Emirates
[7] Department of Electrical and Computer Engineering, Lebanese American University, Byblos 13-5053, Lebanon
[8] MEU Research Unit, Middle East University, Amman 11831, Jordan
[9] University of Zilina, 010 26 Zilina, Slovakia
* Correspondence: lukas.sevcik@uniza.sk

Abstract: Different soil characteristics in different parts of India affect agriculture growth. Crop growth and crop production are significantly impacted by healthy soil. Soil enzymes mediate almost all biochemical reactions in the soil. Understanding the biological processes of soil carbon and nitrogen cycling requires defining the significance of prospective elements at the play of soil enzymes and evaluating their activities. A combination of Multiple Linear Regression (MLR), Random Forest (RF) models, and Artificial Neural Networks (ANN) was employed in this study to assess soil enzyme activity, including amylase and urease activity, soil physical properties, such as sand, silt, clay, and soil chemical properties, including organic matter (SOM), nitrogen (N), phosphorus (P), soil organic carbon (SOC), pH, and fertility level. Compared to other methods for estimating soil phosphatase, cellulose, and urease activity, the RF model significantly outperforms the MLR model. In addition, due to its ability to manage dynamic and hierarchical relationships between enzyme activities, the RF model outperforms other models in evaluating soil enzyme activity. This study collected 3972 soil samples from 25 villages in the Bhandara district of Maharashtra, India, with chemical, physical, and biological parameters. Overall, 99% accuracy was achieved for cellulase enzyme activity and 94% for N-acetyl-glucosaminidase enzyme activity using the Random Forest model. Crops have been suggested based on the best performance accuracy algorithms and evaluation performance metrics.

Keywords: soil organic matter (SOM); soil enzyme activity (SEA); soil organic carbon (SOC); physical soil features; chemical soil features; machine learning (ML); Artificial Neural Network (ANN)

check for updates

Citation: Shahare, Y.; Singh, M.P.; Singh, P.; Diwakar, M.; Singh, V.; Kadry, S.; Sevcik, L. A Comprehensive Analysis of Machine Learning-Based Assessment and Prediction of Soil Enzyme Activity. *Agriculture* **2023**, *13*, 1323. https://doi.org/10.3390/agriculture13071323

Academic Editor: Francesco Marinello

Received: 4 March 2023
Revised: 20 April 2023
Accepted: 21 April 2023
Published: 28 June 2023

1. Introduction

Various factors, including agricultural soil, soil management, soil productivity, irrigation, fertilizer, and climate, impact the agriculture sector to produce a good quantity of crops. The primary determinant of an agricultural field is the soil. The capacity of agricultural soil to develop crops depends on the nutrients it contains. Each soil has a variety of physical, chemical, and biological components. Several researchers have been researching agricultural soil to improve soil quality and other factors, but they have not yet achieved suitable outcomes. Artificial intelligence techniques are more useful and innovative. This technique has the best results for improving and growing crops in the agricultural field, which is helpful to the farmers. Soil quality and the amount of farmed land in Maharashtra agriculture have continued to decline due to a lack of expertise and a harsh environment, which significantly impacts economics and crop production [1,2]. Climate change is affecting the agricultural sector, decreasing crop yield, diminishing soil organic carbon (SOC), changing acceptable cropping areas latitudinal, changing growth time, and causing soil degradation [3]. Due to constant degradation and changes in its composition, the state of soil changes over time [4]. Sustaining the productivity of soil for agriculture requires appropriate stability of physical, chemical, and biological elements [5].

Due to a lack of knowledge about agriculture, farmers are unable to identify the deficiency of important nutrients in the soil that are conducive to growing crops. In terms of the farmers' situation, this work has been proposed. This work uses state-of-the-art artificial technology to develop the prediction of a soil fertility and soil enzyme activity model using a soil dataset from Bhandara district, Maharashtra state, which is helpful to farmers in identifying the nutrient deficiency that is present in the soil. This model was developed by using Python programming with Jupiter Notebook.

The assessment of soil minerals is required for compaction characteristics monitoring. Microorganisms make up a large portion of the biological ingredients of soil and contribute more to its strength than physical or chemical constituents. Microorganisms respond quickly to changes in soil structure and become utilized in their surroundings [6]. Farmers are not gaining the appropriate level of crop productivity as an outcome of weather change and biological activity. In this case, soil biochemical analysis, together with soil chemical and physical features, is critical for minimizing and decomposing the nutrient cycle and providing for the crop [7].

Soil enzymes play a vital role in the biogeochemical cycle of "carbon (C), nitrogen (N), and phosphorus (P)" in the soil and can be employed as early indicators of nutrient imbalances caused by climate change [8,9]. Soil enzyme activities include carbon cycle transformations such as C-glucosidase and invertase, as well as general enzyme activity such as dehydrogenase and catalase, and nitrogen cycle transformations such as urease has a direct impact on the nitrogen supply rate in soil, which is commonly employed as a measure of nitrogen deficiency, N-acetyl-glucosaminidase, and protease [10,11]. Soil enzyme activity can help researchers better understand the biological mechanisms of "carbon and nitrogen" transformation and provide guidelines for assessing soil quality in specific areas [12,13].

Several studies have been created and implemented that are connected to estimating soil enzyme activity utilizing various approaches to acquire the results [14]. Based on this research, it is possible to determine the research gaps of soil enzyme activity and how we can increase the biological mechanism associated with specific enzymes in the carbon and nitrogen cycle in the interest of increasing agricultural yield. Multivariate linear regression (MLR) is the most widely used method for estimating soil parameters due to its simple design, quick calculation, and interpretation. On the other hand, MLR is unable to detect nonlinear relationships between responses and environmental variables. As a result, machine learning methods like "Artificial Neural Network (ANN), Support Vector Machine (SVM), Classification and Regression Tree (CART), and Classification Regression Tree (CART)" are increasingly being used in soil property assessment. Machine learning approaches can represent linear and nonlinear correlations between responses and environmental factors and have a simple structure, good fitting ability, and high prediction accuracy [15].

This method offers a novel and more convenient way to assess and estimate soil parameters such as "soil texture, salinity, soil organic carbon, and nitrogen". Random Forest (RF) is a data mining approach developed as an extension of CART [16–18]. The RF model has several advantages over additional statistical modeling techniques, including the capability to represent extremely nonlinear dimensional associations, resistance to "overfitting", relative dependability in the presence of noise data, the creation of an Unbiased Error Rate measure, and the ability to discern the significance of the variables used. As a result, the

RF model has been frequently used to estimate soil properties in multivariate nonlinear data processing. The main contribution of this research paper is as follows:

1. Predict the activity of soil enzymes based on chemical and physical soil parameters.
2. Evaluate the optimal performance model by correlating all factors.
3. Select and determine the optimal artificial approach algorithm for estimating enzyme activity.
4. Compare the performance model-finding algorithm (PSEA-ML and PSEA–ANN).

2. Materials and Methods

The Bhandara district is in the eastern plateau and hills region of the Maharashtra state of India. The study area is located at 20°44′59.99″ N latitude and 79°52′59.99″ E longitude. The elevation of the study area is probably 2000 m, and the average annual temperature and precipitation at the site are 59.6 degrees Celsius and 250 mm [19], respectively. Different types of soils are available in this region, ranging from deep loamy to clay soil mixed with red and black soils. According to the study of soil analogies, different soils are classified as sand, silt, and clay. This study collects various forms of "sand, silt, soil, pH, nitrogen, phosphorus, soil organic matter (SOM), soil organic carbon (SOC), and soil enzyme activity". Soil is used to assess the quality and quantity of each area of land, to determine whether it is balanced or unbalanced, based on the presence of each feature [20].

2.1. Soil Dataset

The soil data used for this research are from different formers with different blocks. There are three types of soil components available in agricultural soil, i.e., physical soil, chemical soil, and organic soil, which are more important for growing crops. Physical soils are identified by properties such as soil texture, soil structure, soil density, and soil temperature. Soil texture is the main property of physical soil. Therefore, soil texture has been considered in this research. The texture of the soil consists of sand, clay, silt, and depth. Many chemical components are available in chemical soil properties; only nitrogen and phosphorus are included here. Similarly, biological soil properties contain many factors, but here only enzymes are included. Some other components like pH value, soil organic carbon (SOC) and soil organic matter (SOM) are included, which is significant to identify the soil enzyme activities. Two types of soil data have been taken in this research; the first type of soil data is for soil fertility, and the second type of soil data are for soil enzyme activity. Soil data are taken from 25 different villages in the Bhandara district of Maharashtra state, India.

Each soil sample presents different chemical, physical, and biological components. The first parameter is Ph value, which is broadly categorized into three categories—neural, acidic, and alkaline—and represented on a scale from 0 to 10. Ph neutral value is around 7, below 7 is acidic, and above 7 Ph value is alkaline. Sand, silt, clay, nitrogen, phosphorus, soil organic matter, and soil organic carbon parameters are represented by their percentage of content available in each soil sample. Based on content analysis of all the parameters identified, the soil fertility level is considered either low, medium, or high. Soil depth represents how much depth is required to remove the soil for soil testing. Soil depth is used for collecting the soil sample for soil testing. The maximum depth of the soil sample is more effective for finding soil enzymes because more soil organisms are available at a greater depth in the soil. A total of 3972 soil sample data were taken with 11 parameters, as shown in Table 1. Soil was collected from each farmer for soil testing of each soil sample, using the farmer's identity to identify which farmer's soil is deficient of nutrients. For verifying and repeatability of soil samples, preprocessing techniques like finding the missing values, repeated values, and converting from categorical values to numerical values were implemented. Based on this assessment and analysis of soil, a soil dataset was generated for developing the proposed research work.

Table 1. Soil sample dataset of Bhandara district, Maharashtra.

S. No.	Ph	Sand (%)	Silt (%)	Clay (%)	N (ppm)	P (ppm)	SOM (ppm)	SOC (ppm)	Fertility Level	Depth (cm/mm)	Soil Enzyme
1	7.70	20.00	43.00	36.00	40.60	12.60	12.60	9.70	Medium	30	Urease
2	6.58	20.00	43.00	36.00	40.60	12.60	12.60	9.70	Medium	30	Urease
3	6.12	20.00	43.00	36.00	40.60	12.60	12.60	9.70	Medium	30	Invertase
4	6.50	20.00	43.00	36.00	40.60	12.60	12.60	9.70	Medium	30	Invertase
5	6.12	20.00	43.00	36.00	40.60	12.60	12.60	9.70	Medium	30	Acid phosphatase
6	6.42	20.00	43.00	36.00	40.60	12.60	12.60	9.70	Medium	30	Acid phosphatase
7	6.24	20.00	43.00	36.00	36.15	8.23	8.23	9.70	Low	30	Urease
8	6.84	20.00	43.00	36.00	0.87	0.87	1.23	17.28	Low	20	Protease
9	6.84	20.00	43.00	36.00	1.20	1.20	1.23	17.28	Medium	20	Protease
10	6.84	33.00	21.00	46.00	1.37	1.37	1.23	17.28	Medium	20	Protease

2.2. Pre-Processing of Soil Dataset

Before applying machine learning algorithms, pre-processing techniques are required to clean the data and convert the non-numeric data into numeric ones; for example, converting the enzyme classification to a numerical form such that 1 indicates the presence of urease and 0 indicates the absence. Similarly, all enzymes need to convert 1 s and 0 s into numerical form. The soil fertility characteristic has three levels—namely low, medium, and high—which also need to be converted into a numerical form, such as low level indicates 0 values, medium level indicates 1 value, and high level indicates 2 values. After completing the pre-processing techniques, 80% of the data is used for the training model and 20% of the data for the testing model.

2.3. Proposed Methodology

This paper proposes a methodology to predict soil enzyme activities using machine learning algorithms (Multiple Linear Regressions (MLR), Random Forest (RF), Extremely Randomized Tree Classifier (ERTC), and Artificial Neural Network (ANN)) by analyzing physical soil characteristics and chemical characteristics. The block diagram of the proposed methodology (PSEA-PC) predicts the soil enzyme activity and crops, as shown in Figure 1.

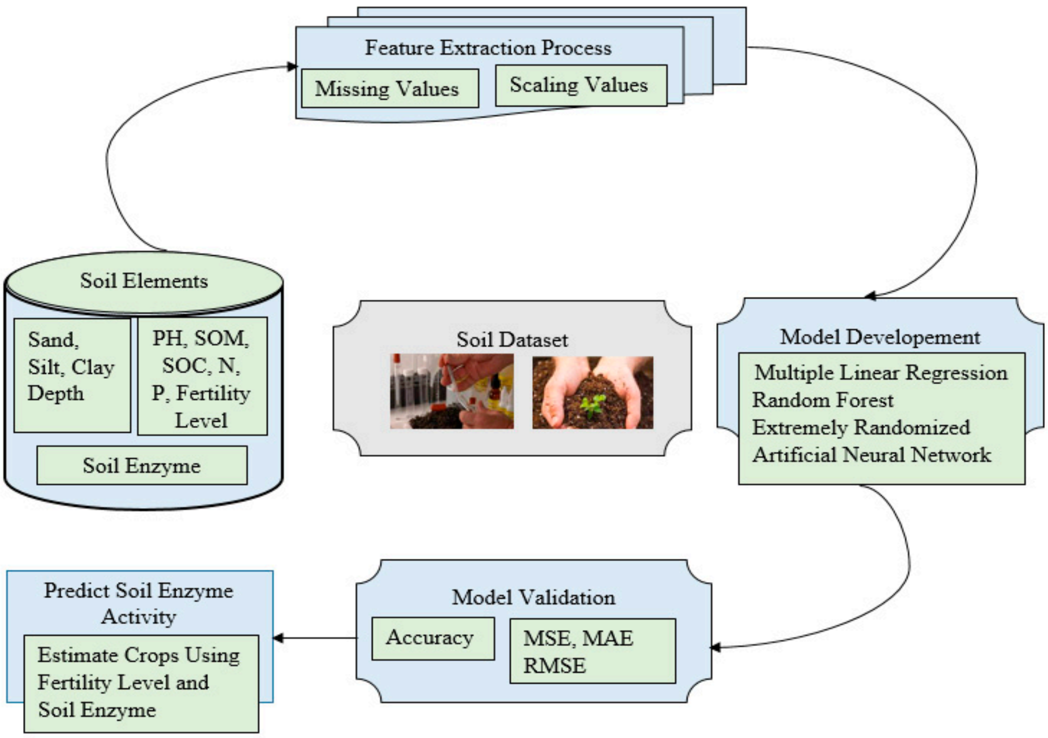

Figure 1. Block diagram of proposed methodology (PSEA-PC).

2.3.1. Development of the ML Model

This section interpreted the different machine learning algorithms with soil enzyme activity.

A. Multiple linear regressions for soil enzyme activity

MLR is a type of supervised machine learning regression technique. Multiple linear regression models are the most suitable technique for predicting soil characteristics [21,22]. Consider a linear regression relationship between numerous independent variables, such as $x_1,\ x_2, x_3 \ldots\ldots x_n$, and a dependent variable used (y_{pred}); $\in$ is denoted as the model error term provided in Equation (1):

$$y_{pred} = \sum_{i=1}^{n} b_i x_i + \in = b_0x_0 + b_1x_1 + \ldots b_nx_n + \in \quad (1)$$

where Y is the dependent variable or outcome, $x_i(i$ = 0, 1, 2, 3, . . . , $n)$ are independent variables, c is an intercept, b_i (i = 0, 1, 2, 3, . . . , n) is the regression coefficient, and ϵ is the residual of regression or error. The cost function (K) is used to find and minimize the error from dependent and independent variables; the best-fit line is provided in Equation (2). The optimized best-fit line is determined using gradient descent utilizing Equation (3), which uses a convergence algorithm for calculating the gradient descent (b_k); detailed discussion is given in Algorithm 1.

$$\mathrm{K}(b_0, b_1) = \frac{1}{2n} \sum_{i=1}^{n} (y\_pred - y)^2 \quad (2)$$

$$\text{Gradient\_descent } (b_k) = b_k - \alpha \frac{\partial}{\partial b_k} k\ (b_0, b_1) \quad (3)$$

α implies the learning rate and it could be considered a small range like 0.001, and k implies the feature index number $k = (0, 1, 2, 3, \ldots\ldots\ldots n)$. From Equations (2) and (3), we obtain Equation (4):

$$b_k = b_k - \frac{\alpha}{n} \sum_{i=1}^{n} (\ y\_pred - y) \quad (4)$$

Algorithm 1 PSEA–MLR-I (Predict the soil enzyme activity using Multiple Linear Regression)

Target: Optimal combination of response variables and enzyme activity in the soil
Input: N = (PH, Sand, Silt, Clay, N, P, SOM, SOC, Fertility level, and Depth)
Output: K = (Predict soil enzyme activity)
1: Initialization of all N and K soil data parameters
2: Pre-processing of the soil dataset with N and K parameters
3: Randomly select 80% soil dataset for training and 20% soil dataset for testing purposes
4: Apply MLR-supervised ML algorithms on a given data set
5: Compute the Accuracy, MSE, RMSE, and MAE of the model
6: Predict soil enzyme activity
7: End

B. Random Forest for soil enzyme activity

The Random Forest model is a multivariate technique that was created to improve the efficiency and accuracy of Classification and Regression Trees (CART). This model combines numerous Classification and Regression Tree algorithms and random variable selection and bagging to make each Classification and Regression Tree more fulfilled. Simultaneously, random feature extraction and bagging techniques cause every factor in the Random Forest to have a smaller correlation [23,24]. Calculate information gain (IG) using the entropy method of all splitting feature data given in Equation (5) and Algorithm 2.

$$\text{IG (IDV, DV)} = \text{Entropy (IDV)} - \text{Entropy (IDV, DV)} \quad (5)$$

For Binary Tree Gini, the importance of two child nodes is provided in Equation (6).

$$mi_{j=}We_jC_j - We_{left(j)}\ C_{left(j)} - We_{right(j)}\ C_{right(j)} \tag{6}$$

The importance of each feature of the decision tree can be calculated in Equation (7)

$$Fe_{i=}\frac{\sum_{i=1}^{n} node\ j\ splits\ on\ feature\ i\ mi_j}{\sum_{k\ \in all\ nodes}\ mi_j} \tag{7}$$

IDV implies an independent variable (x_i features), DV is a dependent variable (y_i features), mi_j implies the importance of node j, $We_{(j)}$ considers the weight number of samples reaching node j, $We_{left(j)}$ represents a left child node that is split on node j, $C_{(j)}$ implies an impurity value of node j, and Fe_i is the importance of features i; see i Equations (8) and (9).

$$Narm_{i=}\frac{Fe_i}{\sum_{k\ \in all\ nodes}\ Fe_i} \tag{8}$$

$$RFfe_i = \frac{\sum_{k\ \in all\ nodes}\ Narm_i}{\text{IDV}} \tag{9}$$

Algorithm 2 PSEA–RF-II (Predict soil enzyme activity using Random Forest algorithms)

Target: Optimal combination of response variables and enzyme activity in the soil
Input: N = (PH, Sand, Silt, Clay, N, P, SOM, SOC, Fertility level, and Depth)
Output: K = (Predict soil enzyme activity)
1: Initialization of all N and K soil data parameters
2: Pre-processing of the soil dataset with N and K parameters
3: Randomly select 80% soil dataset for training and 20% soil dataset for testing purposes
4: Apply RF-supervised ML algorithms on a given data set
5: Compute the Accuracy, MSE, RMSE, and MAE of the Model
6: Predict soil enzyme activity
7: End

This algorithm has been used for predicting soil enzyme activity using Random Forests. First, initialize the soil parameters including the chemical, physical, and biological components. Then, use the pre-processing techniques for cleaning the data, converting the categorical to numerical, and determining the missing and null data. Next, select 80% of the data for training and 20% for testing and use the Random Forest model to identify the best accuracy using evaluation metrics like MSE, RMSE, and MAE.

C. Extremely Randomized Trees Classifiers for soil enzyme activity

Extremely Randomized Trees Classifiers are a form of ensemble classification algorithm that outputs a classification result by combining the outcomes of several de-correlated decision trees aggregated in a "forest." (Algorithm 3). It is conceptually identical to a Random Forest Classifier, apart from how the decision trees in the forest are constructed. Create an additional tree classifier based on each decision tree's original dataset [25]. Using mathematical notation, randomly choose n features from a collection of all features offered by each tree for splitting the data and obtaining the best feature of all trees (Gini Index). This decision tree creates a multi-correlated feature from various random samples. First, we need to calculate the entropy of each feature based on mathematical Equation (10):

$$\text{Entropy(s)} = \sum_{i=1}^{m} -p_i log_2\ (p_i) \tag{10}$$

where Entropy(s) is a random sample of each feature of the tree, m is the number of unique classification labels, p_i is the proportion of each row with the target label i.

Algorithm 3 PSEA–ERT-III (Predict soil enzyme activity using extremely randomized tree classifiers)

Target: Optimal combination of response variables and enzyme activity in the soil
Input: N = (PH, Sand, Silt, Clay, N, P, SOM, SOC, Fertility level, and Depth)
Output: K = (Predict soil enzyme activity)
1: Initialization of all N and K soil parameters
2: Pre-processing of the soil dataset with N and K parameters
3: Randomly select 80% soil dataset for training and 20% soil dataset for testing purposes
4: Apply ERT-supervised ML algorithms on a given data set
5: Compute the Accuracy, MSE, RMSE, and MAE of the Model
6: Predict soil enzyme activity
7: End

2.3.2. Artificial Neural Network for Soil Enzyme Activity

An Artificial Neural Network (ANN) is a system that divides artificial neurons into three layers (input, hidden, and output). During the training of the ANN approach when used for regression analysis, the basic parameters of artificial neurons, such as weight, threshold, and activation functions, were tuned [26–28]. Given input soil properties, this ANN technique predicts soil enzyme activity (Algorithm 4). The ANN model with "relu" and "sigmoid" activation functions was utilized to add a hidden layer. The soil enzyme activities convert the cycle of carbon (C), nitrogen (N), and phosphorus (P) with C-glucosidase, invertase, dehydrogenase, catalase, urease, N-acetyl-glucosaminidase, protease, which are predicted using 80% of the training dataset and 20% of the testing dataset. Figure 2 shows the structure of an Artificial Neural Network Model. An Artificial Neural Network is divided into two techniques for passing data known as forwarding propagation; the forward propagation considered a perceptron is provided in Equations (11)–(15).

$$\sum = (x_1 * we_1) + (x_2 * we_2) + \ldots\ldots.. + (x_n * we_n) \tag{11}$$

$$x\ we_i = (x_1 * we_1) + (x_2 * we_2) + \ldots\ldots.. + (x_n * we_n) \tag{12}$$

$$\sum = x\ we_i \tag{13}$$

$$Z = x\ we_i + B \tag{14}$$

$$y_{pred} = \sigma(z) = \frac{1}{1 + e^{-z}} \tag{15}$$

where σ indicates an activation function of the neural model, we_i is the identified weight feature value, y_{pred} represents a predicted value, and B implies a bias of the neural model. For implementing the Artificial Neural Network, various parameters are required like the input layer, hidden layer, and output layer. Equation (11) shows the addition of all the input features with weight to calculate the average of features. Equation (12) is determined to simplify all the features with feature weight values. Equation (13) is determined to optimize the feature along with weight values in a single process. Equation (14) produces the outcomes by adding the bias (B) with all features and weight values. Equation (15) shows the prediction of the result based on all features using the sigmoid function.

Algorithm 4 PSEA–ANN-IV (Predict soil enzyme activity using ANN)

Target: Optimal combination of response variables and enzyme activity in the soil
Input: N = (Sand, Silt, Clay, PH, SOM, SOC, Available Nitrogen, Available Phosphorus, Depth)
Output: K = (Predict soil enzyme activity)
1: Initialization of all N and K parameters
2: If (Sn ! = Sc)
3: Then pre-process and scale the data
4: Otherwise, go to step 5
5: Choose and select x and y variable
6: Split 80% data for training and 20% data for testing
7: Add first and second hidden layer (activation function = relu)
8: Add output hidden layer (activation function = sigmoid)
9: Compile and validate data
10: Select the epoch and calculate the accuracy
11: Predict soil enzyme activity
12: End
(Where Sn = scale features from soil dataset, Sc = scale soil enzyme features of the dataset)

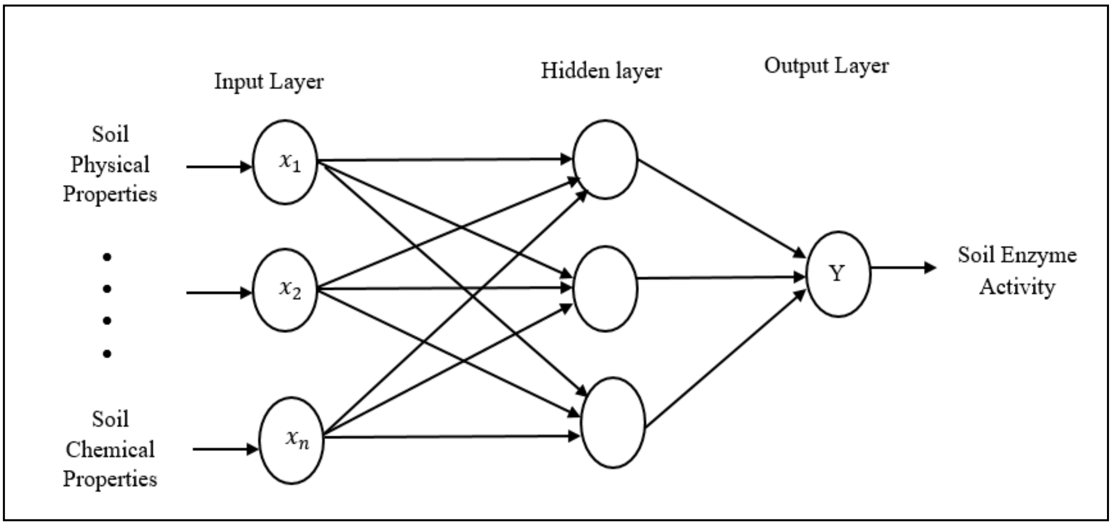

Figure 2. Structure of ANN Model.

3. Result and Discussion

The descriptive statistics of soil enzyme activity with soil properties datasets are shown in Table 2. The descriptive statistics summary includes the following parameters: Minimum (Min), Maximum (Max), Standard Mean Value (Mean), and Standard Deviation (SD). Python with jupyter notebook was used to implement the descriptive statistical analysis, machine learning model, and Artificial Neural Network on Windows 10. In this summary, calculate the descriptive statistics of PH, Sand, Silt, Clay, Available Nitrogen, Available Phosphorus, SOM, SOC, Depth, Soil Fertility Level, and Soil Enzyme Activity. The PH standard deviation value was smaller than the mean (SD Mean), and the soil enzyme activity was SD > Mean. The mean value of available nitrogen and depth soil parameters was extremely covariate (Mean > 100%) compared to other parameters, while the standard deviation of available nitrogen and available phosphorus was highly covariate (SD > 100%) compared to certain other factors [29]. Figure 3 illustrates the very positive and strongly negative correlations of a soil dataset in matrix format and according to this correlation, the correlation between sand depth and clay soil qualities is substantially negatively correlated.

Table 2. Statistics summary of soil physical, chemical features, and soil enzyme activity.

Soil Properties	Mean	Min	Max	SD
PH	6.3314	3.8000	8.7900	0.7280
Sand	38.1617	8.0000	90.0000	19.3332
Silt	31.7412	4.0000	67.0000	14.5515
Clay	32.1268	11.0000	47.0000	10.5310
Available Nitrogen	19.1986	0.6900	317.0000	51.6195
Available Phosphorus	9.9923	0.6900	146.5000	23.0569
SOM	5.1652	0.6900	127.2000	11.1173
SOC	7.6628	0.6900	127.2000	14.1052
Depth	19.2412	10.0000	30.0000	5.4071
Soil Fertility Level (Low, Medium, and High)	1.2921	0.0000	2.0000	0.8235
Soil Enzyme	9.3529	0.0000	19.0000	18.0000

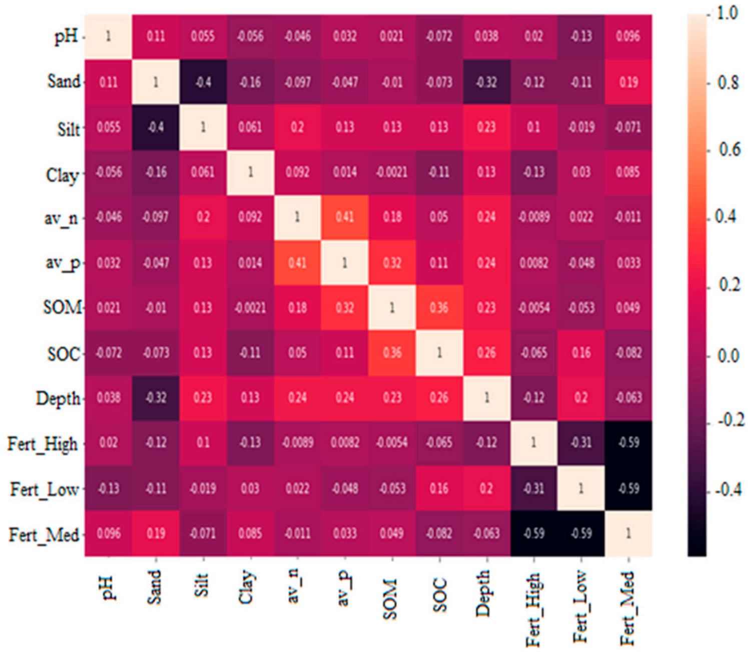

Figure 3. Heatmap of the positive and negative correlation of all soil features.

Model Validation

In this study, the dataset collected approximately 3972 soil sample datasets, of which we used the training dataset of randomly selected records, accounting for approximately 80% of the total records, to develop Multiple Linear Regression, Random Forest, extra tree classifier, and Artificial Neural Network models, and the testing dataset included the remaining 20% of the records to verify the model's estimation accuracy for soil enzyme activities [30]. The performance of the ML and ANN models was assessed using the coefficient of determination (R2), mean absolute error (MAE), root mean square error (RMSE), classification report, and confusion matrix in Table 3. The following are the evaluation performance indices calculated to validate the models provided in Equations (16)–(18).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (pre_i - \overline{obs_i})^2}{\sum_{i=1}^{n} (obs_i - \overline{obs_i})^2} \tag{16}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | pre_i - obs_i | \tag{17}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{1=1}^{n}(obs_i - pre_i)^2} \quad (18)$$

where pre_i and obs_i are the predicted and observed values, respectively, from a random sample of i, $\overline{obs_i}$ is the mean value observation, and n is the total number of sample records.

Table 3. Confusion matrix with classification metrics report.

		Predicted Soil Enzyme Activity Data		
		Positive	**Negative**	
Actual Soil Enzyme Activity Data	Positive	True Positive (TP)	False Negative (FN)	Sensitivity $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP)	True Negative (TN)	Specificity $\frac{TN}{(TN+FP)}$
Classification Metrics		Precision $\frac{TP}{(TP+FP)}$	Negative Predictive Value $\frac{TN}{(TN+FN)}$	Accuracy $\frac{TP+TN}{(TP+TN+FP+FN)}$

These tables have identified the confusion matrix with classification reports. There are four parameters: true positive, true negative, false positive, and false negative. Based on these parameters, the actual observation and predicted observation has been measured. This study identifies the actual soil enzyme activity and predicts the soil enzyme activity based on the confusion matrix.

This correlation aims to identify important features that may be used to implement and estimate soil enzyme activity using numerical data from the training and testing datasets. Figure 4 depicts the dataset's summary of soil enzyme activity. In this study, 3972 soil samples were collected from the Bhandara district of Maharashtra to predict the activity of each enzyme factor related to the carbon, nitrogen, and phosphorus nutrient cycle, including urease, acid phosphatase, invertase, alkaline phosphatase, phosphatase, protease, cellulose, N-acetyl-glycosaminidases, and C-glucosidase [31,32]. This study used a large amount of urease and cellulose soil enzyme activity for prediction, with 80% of the dataset being used for training and 20% being used for testing to determine the best soil enzyme activity solution.

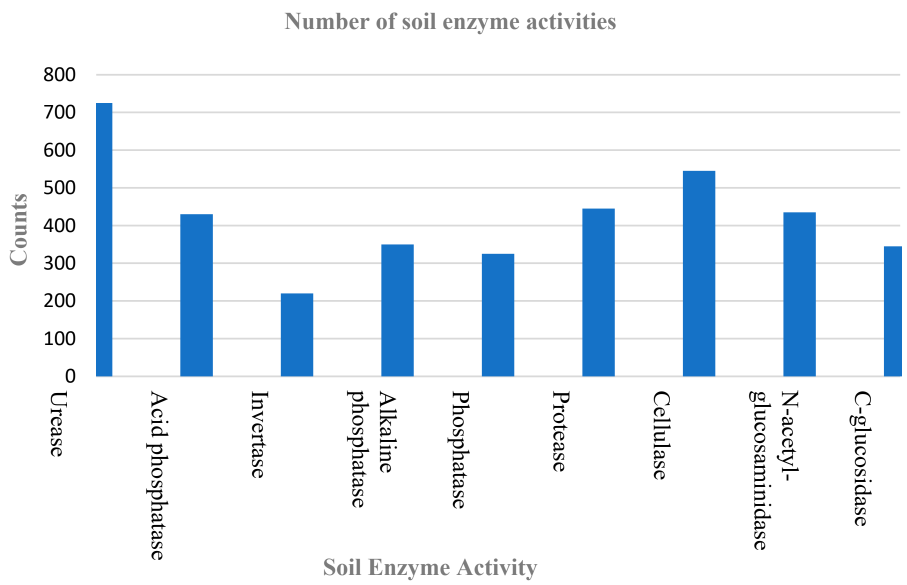

Figure 4. Summary of soil enzyme activities in the dataset.

Figure 5 depicts the graphical representations of soil samples, including the soil's physical, chemical, and enzyme activity. This is used to forecast soil enzyme activity based on each soil component's amount, determining what proportion of soil components are available in the soil, as well as soil enzyme activity and fertility level. Based on the summary of all values of soil factors, levels of urease, acid phosphatase, invertase, alkaline phosphatase, phosphatase, protease, cellulose, N-acetyl-glucosaminidase, and C-glucosidase are predicted.

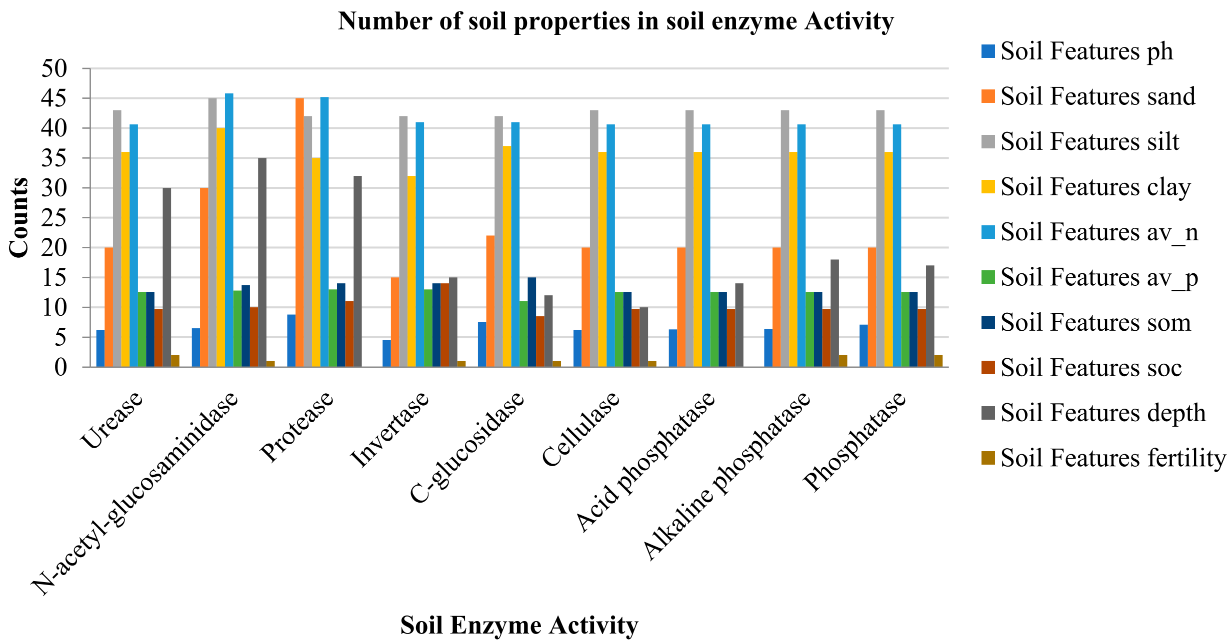

Figure 5. Summary of soil enzyme activities with soil properties.

Figure 6 shows a graphical depiction of the machine learning model classification, which includes Random Forest, additional tree classifier, and regression—including Multiple Linear Regression and Artificial Neural Network—to predict soil enzyme activity. In comparison to other classifications, urease soil enzyme activity found the RF and additional tree models had good accuracy. The RF model was found to be superior in terms of acid phosphatase activity, invertase, alkaline phosphatase, phosphatase, protease, and N-acetyl-glucosaminidase. The MLR model was best suited for cellulose, and the ANN model seemed good for C-glucosidase. According to the implementation results, the Random Forest model outperformed other models in terms of identifying soil quality and enhancing agricultural productivity for a specific location. Using Multiple Linear Regression with training and testing datasets, we were able to estimate the cellulose soil enzyme activity with high accuracy.

Classification and regression approaches were utilized in this work to determine the most effective approach for estimating soil enzyme activity. For predicting the activity of urease, acid phosphatase, invertase, alkaline phosphatase, phosphatase, protease, cellulose, N-acetyl-glucosaminidase, and C-glucosidase soil enzymes, RF, MLR, Extra Tree, and ANN models were used in classification, while RF and MLR were used in regression.

Multiple Linear Regression and Random Forest models were employed in this investigation to determine which soil enzyme activity had the best performance, including MSE, MAE, and RMSE characteristics. Figure 7 shows how MLR's evaluation performance metrics are represented. MSE's urease and N-acetyl-glucosaminidase soil enzyme activity was found to be good, meaning there was minimal error compared to others (0.0549 and 0.0449). For MAE, a slight error of phosphatase and C-glucosidase (0.111567) was discovered rather than the activity of other enzymes. C-glucosidase (0.1342) had a small error in RMSE compared to other enzymes. Random Forest regression techniques are the best appropriate model for predicting soil enzyme activity. In Table 4, we estimate the soil enzyme activity by analyzing the MSE, RMSE, and MAE of the Multiple Linear Regression approach.

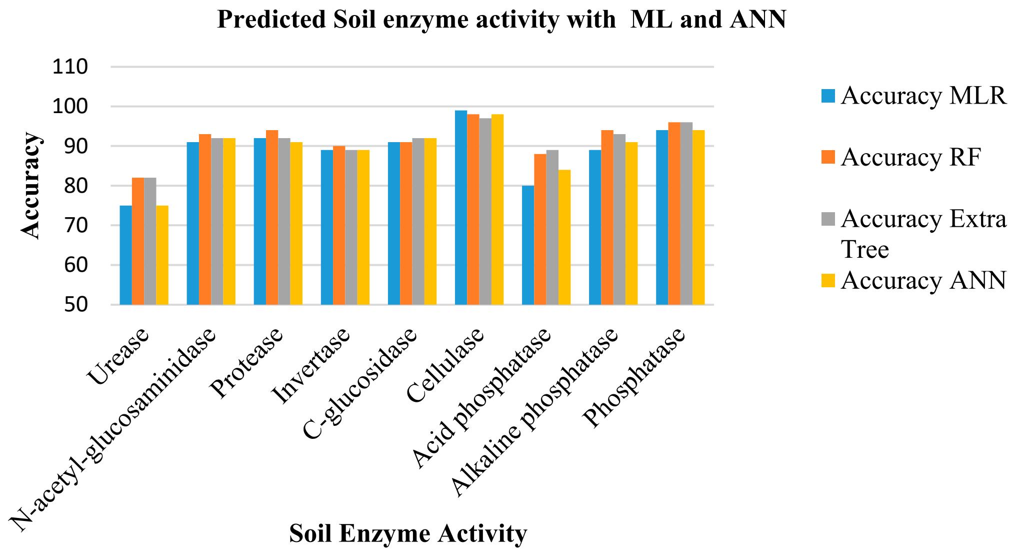

Figure 6. Predicted soil enzyme activity with ML and ANN.

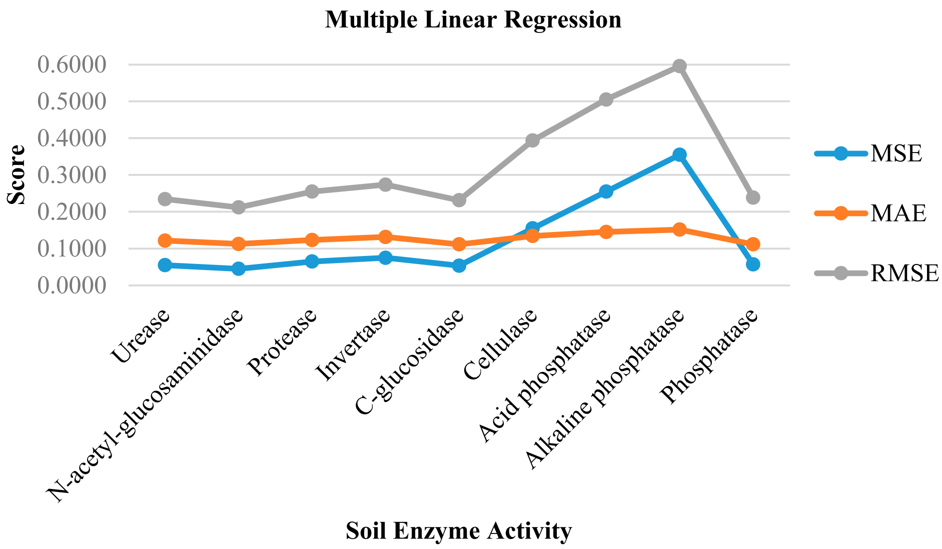

Figure 7. Evaluation performance metrics of MLR.

Figure 8 shows that the MAE, RMSE, and MSE parameters were used to construct the evaluation performance metrics. Alkaline phosphatase soil enzyme activity was found to be an excellent dependent variable for MSE prediction outcomes, with a lower error (0.0146) than other variables. C-glucosidase (0.0162) soil enzyme activity was discovered to have a lower error than the others in MAE performance criteria. Alkaline phosphatase soil enzyme activity had a lower error (0.1209) than the others in terms of RMSE performance metrics.

Table 4. Compare the MSE, MAE, and RMSE of MLR for PSEA-ML.

Soil Enzyme	MSE	MAE Multiple Linear Regression	RMSE
Urease	0.0549	0.1216	0.2343
N-acetyl-glucosaminidase	0.0449	0.1126	0.2119
Protease	0.0649	0.1232	0.2547
Invertase	0.0749	0.1316	0.2737
C-glucosidase	0.0535	0.1116	0.2313
Cellulase	0.1549	0.1342	0.3936
Acid phosphatase	0.2549	0.1452	0.5049
Alkaline phosphatase	0.3549	0.1516	0.5957
Phosphatase	0.0569	0.1116	0.2385

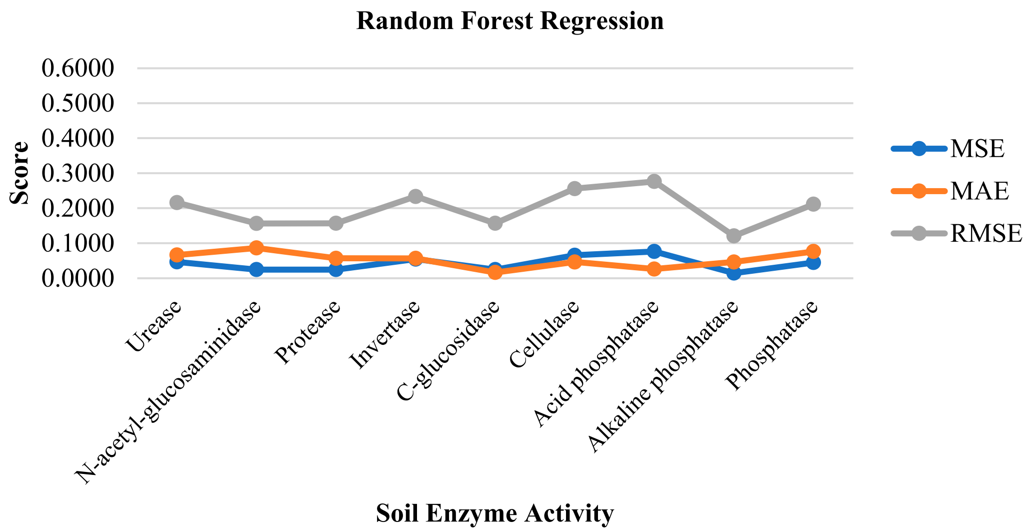

Figure 8. Evaluation performance metric of Random Forest.

In Table 5, the soil enzyme activity is estimated by analyzing MSE, RMSE, and MAE of the Random Forest regression approach. The result demonstrated the identification of the best optimal outcome of soil enzyme activity.

Table 5. Compare the MSE, MAE, and RMSE of the Random Forest approach for PSEA-ML.

Soil Enzyme	MSE	MAE RF	RMSE
Urease	0.0466	0.0662	0.2159
N-acetyl-glucosaminidase	0.0245	0.0862	0.1564
Protease	0.0246	0.0569	0.1569
Invertase	0.0545	0.0566	0.2334
C-glucosidase	0.0246	0.0162	0.1569
Cellulase	0.0655	0.0462	0.2559
Acid phosphatase	0.0765	0.0262	0.2765
Alkaline phosphatase	0.0146	0.0462	0.1209
Phosphatase	0.0446	0.0762	0.2112

In Table 6, the measured MAE, RMSE, and MSE parameters were used to construct the evaluation performance metrics. Cellulase soil enzyme activity was found to be an excellent dependent variable for MSE prediction outcomes, with a lower error (0.0259) than other variables. Cellulase soil enzyme activity was discovered to have a lesser error (0.0752) than the others in the MAE performance criteria. Cellulase and N-acetyl-glucosaminidase soil enzyme activity had less error (0.1609 and 0.2493, respectively) than the others in terms of RMSE performance metrics.

Table 6. Compare the MSE, MAE, and RMSE of the extra tree regressor approach for PSEA-ML.

Soil Enzyme	MSE	MAE Extra Tree Regressor	RMSE
Urease	0.1761	0.7610	0.4196
N-acetyl-glucosaminidase	0.0622	0.2746	0.2493
Protease	0.0777	0.0777	0.2787
Invertase	0.1036	0.1036	0.3219
C-glucosidase	0.0829	0.0829	0.2879
Cellulase	0.0259	0.0752	0.1609
Acid phosphatase	0.0907	0.1473	0.3012
Alkaline phosphatase	0.0618	0.0999	0.2486
Phosphatase	0.0500	0.0814	0.2236

The Artificial Neural Network model implemented using Python programming with Keras and TensorFlow library in Python was built using three layers: input layer, hidden layer, and output layer. Programming used dense layers for the fully connected neural network. The dense layer considered an input layer with 10 units (input features) with the 'relu' activation function, dense_1 represented the hidden layer with the 'relu' activation function, and the output layer used 1 unit (output features) with the 'sigmoid' function. For compiling, the ANN model used an 'adam' optimizer for reducing the error/loss with 'binary_crossentropy' loss. This model executed 32 batch sizes and 100 epochs for calculating the loss score, training, and validating accuracy. This model achieved 99% accuracy in cellulase enzyme activity. Figure 9 shows the ANN model summary and Figure 10 shows the epoch generation with ANN model loss and validation accuracy.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 10)                130
_________________________________________________________________
dense_1 (Dense)              (None, 10)                110
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 11
=================================================================
Total params: 251
Trainable params: 251
```

Figure 9. Artificial Neural Network Model Summary.

Table 7 measures the MAE, RMSE, and MSE parameters used to construct the evaluation performance metrics. Protease soil enzyme activity was found to be an excellent dependent variable for MSE prediction outcomes, with a lower error (0.1384) than other variables. C-glucosidase soil enzyme activity was discovered to have a lower error (0.1443) than the others in the MAE performance criteria. C-glucosidase and protease soil enzyme activity had a lower error (0.3799) than the others in terms of RMSE performance metrics.

```
Epoch 1/100
25/25 [==============================] - 3s 4ms/step - loss: 0.6193 - accuracy: 0.7355
Epoch 2/100
25/25 [==============================] - 0s 7ms/step - loss: 0.4433 - accuracy: 0.9342
Epoch 3/100
25/25 [==============================] - 0s 6ms/step - loss: 0.3224 - accuracy: 0.9884
Epoch 4/100
25/25 [==============================] - 0s 4ms/step - loss: 0.2390 - accuracy: 0.9923
Epoch 5/100
25/25 [==============================] - 0s 8ms/step - loss: 0.1805 - accuracy: 0.9923
Epoch 6/100
25/25 [==============================] - 0s 4ms/step - loss: 0.1382 - accuracy: 0.9923
Epoch 7/100
25/25 [==============================] - 0s 4ms/step - loss: 0.1083 - accuracy: 0.9923
Epoch 8/100
25/25 [==============================] - 0s 6ms/step - loss: 0.0865 - accuracy: 0.9923
Epoch 9/100
25/25 [==============================] - 0s 4ms/step - loss: 0.0711 - accuracy: 0.9923
Epoch 10/100
```

Figure 10. ANN model of epoch generation with loss and validation accuracy.

Table 7. Compare the MSE, MAE, and RMSE of the Artificial Neural Network approach for PSEA-ANN.

Soil Enzyme	**MSE**	**MAE ANN**	**RMSE**
Urease	0.1818	0.3464	0.4263
N-acetyl-glucosaminidase	0.2353	0.4732	0.4850
Protease	0.1383	0.3421	0.3718
Invertase	0.3328	0.5687	0.5768
C-glucosidase	0.1443	0.1443	0.3798
Cellulase	0.3826	0.3826	0.6185
Acid phosphatase	0.1913	0.4138	0.4373
Alkaline phosphatase	0.3402	0.5735	0.5832
Phosphatase	0.4263	0.6417	0.6529

Table 8 lists the specific crops depending on all soil enzyme activity and soil fertility levels (low, medium, and high). Because some soil components are particularly low in low fertility levels, a balanced number of factors, as well as enzyme activity, is required. The soil enzyme activity—such as urease, invertase, C-glucosidase, and acid phosphatase, which are associated with low fertility levels—is estimated. Based on this prediction, cucumber, maize, peanut pepper, soybean, and sugarcane were identified as crops that are useful for harvesting and increasing crop productivity [33–35]. Based on the analysis of soil fertility level and soil enzyme activity using machine learning algorithms, specific crops are suggested for increasing crop productivity. Each crop is suggested for a different soil fertility level—low, medium, and high—for each soil enzyme activity. For example, potato crop requires a high fertility level with protease and phosphatase soil enzyme activity.

According to this prediction, the activity of soil enzymes such as N-acetyl-glucosaminidase, cellulase, and alkaline phosphatase—which correspond to a medium fertility level—was estimated. The activity of the N-acetyl-glucosaminidase soil enzyme indicated crops such as chickpea cotton, rice wheat, peanut, and soybean. Cucumbers, maize, peanut, pepper, soybean, and sugarcane are likely have cellulase and alkaline phosphatase soil enzymes. The activity of soil enzymes such as protease and phosphatase—which are associated with high fertility levels—is estimated, and specific crops are chosen based on this prediction. Potato, cotton, sugarcane, maize, soybean, and pear were indicated by protease, and phosphatase soil enzyme activity indicated the crops potato, cotton, sugarcane, maize, soybean, and pear.

Table 8. List of specific crops based on soil fertility level and soil enzyme activity.

Fertility Level	Soil Enzyme Activity	Crops					
Low	Urease	Cucumber	Maize	Peanut	Pepper	Soybean	Sugarcane
Medium	N-acetyl-glucosaminidase	Chickpea	Cotton	Rice	Wheat	Peanut	Soybean
High	Protease	Potato	Cotton	Sugarcane	Maize	Soybean	Pear
Low	Invertase	Cucumber	Maize	Peanut	Pepper	Soybean	Sugarcane
Low	C-glucosidase	Cucumber	Maize	Peanut	Pepper	Soybean	Sugarcane
Medium	Cellulase	Cucumber	Maize	Peanut	Pepper	Soybean	Sugarcane
Low	Acid phosphatase	Cucumber	Maize	Peanut	Pepper	Soybean	Sugarcane
Medium	Alkaline phosphatase	Cucumber	Maize	Peanut	Pepper	Soybean	Sugarcane
High	Phosphatase	Potato	Cotton	Sugarcane	Maize	Soybean	Pear

4. Conclusions

This research was performed to evaluate soil enzyme activities, which included nine target features of soil enzymes such as urease, acid phosphatase, invertase, alkaline phosphatase, phosphatase, protease, cellulose, N-acetyl-glucosaminidase, and C-glucosidase, as well as chemical factors such as PH, SOC, SOM, available nitrogen, and available phosphorus; the physical factors were sand, silt, clay, and depth of soil for soil testing. Machine learning models such as MLR, RF, and extra tree classification techniques were compared with the ANN model for estimating soil enzyme activity. The best model was determined using a classification report, confusion matrix, and evaluation performance regression techniques such as MSE, MAE, and RMSE. According to the experimental results, the Random Forest model seems to be the most suitable model for determining the optimal soil enzyme activities as compared to other classification models. MAE, RMSE, and MSE were used to obtain good results in the MLR and RF regression techniques. Specific crops were recommended based on soil fertility levels, which are divided into three categories: low, medium, and high. Each soil level revealed a varied soil enzyme activity with a given crop, which is extremely beneficial to farmers in terms of enhancing crop output and determining soil quality.

Future work will include collecting additional soil enzyme activity classification of soil samples from various regions, estimating activities using various artificial methodologies, and recommending certain crops with fertilizer doses.

Author Contributions: Conceptualization, Y.S. and M.P.S.; methodology, P.S. and M.D.; software, V.S.; validation, S.K.; formal analysis, L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Luis Moreno, J.; Bastida, F.; Díaz-López, M.; Li, Y.; Zhou, Y.; López-Mondéjar, R.; Benavente-Ferraces, I.; Rojas, R.; Rey, A.; García-Gil, J.C.; et al. Response of soil chemical properties, enzyme activities and microbial communities to biochar application and climate change in a Mediterranean agroecosystem. *Geoderma* **2022**, *407*, 115536. [CrossRef]
2. Tajik, S.; Ayoubi, S.; Lorenz, N. Soil microbial communities affected by vegetation, topography and soil properties in a forest ecosystem. *Appl. Soil Ecol.* **2020**, *149*, 103514. [CrossRef]
3. Xie, X.; Wu, T.; Zhu, M.; Jiang, G.; Xu, Y.; Wang, X.; Pu, L. Comparison of random forest and multiple linear regression models for estimation of soil extracellular enzyme activities in agricultural reclaimed coastal saline land. *Ecol. Indic.* **2021**, *120*, 106925. [CrossRef]
4. Pukalchik, M.A.; Katrutsa, A.M.; Shadrin, D.; Terekhova, V.A.; Oseledets, I.V. Machine learning methods for estimation the indicators of phosphogypsum influence in soil. *J. Soils Sediments* **2019**, *19*, 2265–2276. [CrossRef]
5. Zhang, L.; Chen, X.; Xu, Y.; Jin, M.; Ye, X.; Gao, H.; Chu, W.; Mao, J.; Thompson, M.L. Soil labile organic carbon fractions and soil enzyme activities after 10 years of continuous fertilization and wheat residue incorporation. *Sci. Rep.* **2020**, *10*, 11318. [CrossRef]

6. Jha, S.K.; Ahmad, Z. Soil microbial dynamics prediction using machine learning regression methods. *Comput. Electr. Agric.* **2018**, *147*, 158–165. [CrossRef]
7. Archana, K.; Saranya, K.G. Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier. *SSRG Int. J. Comput. Sci. Eng.* **2020**, *7*, 5–8.
8. Aponte, H.; Meli, P.; Butler, B.; Paolini, J.; Matus, F.; Merino, C.; Cornejo, P.; Kuzyakov, Y. Meta-analysis of heavy metal effects on soil enzyme activities. *Sci. Total Environ.* **2020**, *737*, 139744. [CrossRef]
9. dos Santos Teixeira, A.F.; Silva, S.H.G.; de Carvalho, T.S.; Silva, A.O.; Guimarães, A.A.; de Souza, F.M. Soil physicochemical properties and terrain information predict soil enzymes activity in phytophysiognomies of the Quadrilátero Ferrífero region in Brazil. *Catena* **2021**, *199*, 105083. [CrossRef]
10. Curtright, A.J.; Tiemann, L.K. Meta-analysis dataset of soil extracellular enzyme activities in intercropping systems. *Data Brief* **2021**, *38*, 107284. [CrossRef]
11. Jian, S.; Li, J.; Chen, J.; Wang, G.; Mayes, M.A.; Dzantor, K.E.; Hui, D.; Luo, Y. Soil extracellular enzyme activities, soil carbon and nitrogen storage under nitrogen fertilization: A meta-analysis. *Soil Biol. Biochem.* **2016**, *101*, 32–43. [CrossRef]
12. Piotrowska-Długosz, A.; Kobierski, M.; Długosz, J. Enzymatic activity and physicochemical properties of soil profiles of luvisols. *Materials* **2021**, *14*, 6364. [CrossRef] [PubMed]
13. Tajik, S.; Ayoubi, S.; Nourbakhsh, F. Prediction of soil enzymes activity by digital terrain analysis: Comparing artificial neural network and multiple linear regression models. *Environ. Eng. Sci.* **2012**, *29*, 798–806. [CrossRef]
14. Zhang, H.; Wu, P.; Yin, A.; Yang, X.; Zhang, M.; Gao, C. Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison of multiple linear regressions and the random forest model. *Sci. Total Environ.* **2017**, *592*, 704–713. [CrossRef]
15. Karthigadevi, K. Random Forest Classification Algorithm for Agricultural Data Analysis in Tirunelveli District. *J. Xi'an Univ. Archit. Technol.* **2020**, *12*, 418–432.
16. Hemageetha, N.; Nagalakshmi, N. Classification Techniques in Analysis of Salem District Soil condition for Cultivation of Sunflower. *Int. J. Comput. Sci. Eng.* **2018**, *6*, 642–646. [CrossRef]
17. Rajamanickam, J.; Savitha, D.M. Predictive model construction for prediction of soil fertility using decision tree machine learning algorithm. *Kongunadu Res. J.* **2021**, *8*, 30–35.
18. Zhao, D.; Arshad, M.; Li, N.; Triantafilis, J. Predicting soil physical and chemical properties using Vis-NIR in Australian cotton areas. *Catena* **2021**, *196*, 104938. [CrossRef]
19. Kashiwar, S.R.; Kundu, D.M.C.; Dongarwar, U.R. Soil fertility appraisal of Bhandara block of Maharashtra using geospatial techniques. *Int. J. Chem. Stud.* **2020**, *8*, 2570–2576. [CrossRef]
20. Geng, X.; Zhu, C.; Zhang, J.; Xiong, Z. Prediction of Soil Fertility Change Trend Using a Stochastic Petri Net. *J. Sign. Process. Syst.* **2021**, *93*, 285–297. [CrossRef]
21. Pinheiro, H.S.K.; de Carvalho, W., Jr.; da Silva Chages, C.; dos Anjos, L.H.C.; Owens, P.R. Prediction of topsoil texture through regression trees and multiple linear regressions. *Rev. Bras. Cienc. Solo* **2018**, *42*, e0170167. [CrossRef]
22. Zhang, H.; Yin, S.; Chen, Y.; Shao, S.; Wu, J.; Fan, M.; Chen, F.; Gao, C. Machine learning-based source identification and spatial prediction of heavy metals in soil in a rapid urbanization area, eastern China. *J. Clean. Prod.* **2020**, *273*, 122858. [CrossRef]
23. Deng, X.; Ma, W.; Ren, Z.; Zhang, M.; Grieneisen, M.L.; Chen, X.; Fei, X.; Qin, F.; Zhan, Y.; Lv, X. Spatial and temporal trends of soil total nitrogen and C/N ratio for croplands of East China. *Geoderma* **2020**, *361*, 114035. [CrossRef]
24. Xu, Y.; Wang, X.; Bai, J.; Wang, D.; Wang, W.; Guan, Y. Estimating the spatial distribution of soil total nitrogen and available potassium in coastal wetland soils in the Yellow River Delta by incorporating multi-source data. *Ecol. Indic.* **2020**, *111*, 106002. [CrossRef]
25. Sirsat, M.S.; Cernadas, E.; Fernández-Delgado, M.; Khan, R. Classification of agricultural soil parameters in India. *Comput. Electr. Agric.* **2017**, *135*, 269–279. [CrossRef]
26. Kalkhajeh, Y.K.; Arshad, R.R.; Amerikhah, H.; Sami, M. Comparison of multiple linear regressions and artificial intelli-gence-based modeling techniques for prediction the soil cation exchange capacity of Aridisols and Entisols in a semi-arid region. *Aust. J. Agric. Eng.* **2012**, *3*, 39–46.
27. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electr. Agric.* **2020**, *177*, 105709. [CrossRef]
28. Were, K.; Bui, D.T.; Dick, Ø.B.; Singh, B.R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecol. Indic.* **2015**, *52*, 394–403. [CrossRef]
29. Swapna, B.; Manivannan, S.; Nandhinidevi, R. Prediction of Soil Reaction (Ph) and Soil Nutrients Using Multivariate Statistics Techniques for Agricultural Crop and Soil Management. *Int. J. Adv. Sci. Technol.* **2020**, *29*, 1900–1912.
30. Munawar, A.A.; Yunus, Y.; Devianti; Satriyo, P. Calibration models database of near-infrared spectroscopy to predict agricul-tural soil fertility properties. *Data Brief* **2020**, *30*, 105469. [CrossRef]
31. Kivlin, S.N.; Treseder, K.K. Soil extracellular enzyme activities correspond with abiotic factors more than fungal community composition. *Biogeochemistry* **2014**, *117*, 23–37. [CrossRef]
32. Weintraub, S.R.; Wieder, W.R.; Cleveland, C.C.; Townsend, A.R. Organic matter inputs shift soil enzyme activity and allocation patterns in a wet tropical forest. *Biogeochemistry* **2013**, *114*, 313–326. [CrossRef]

33. Panchamurthi, S. Soil Analysis and Prediction of Suitable Crop for Agriculture using Machine Learning. *Int. J. Res. Appl. Sci. Eng. Technol.* **2019**, *7*, 2328–2335. [CrossRef]
34. Priya, P.; Muthaiah, U.; Balamurugan, M. Predicting yield of the crop using machine learning algorithms. *Int. J. Eng. Sci. Res. Technol.* **2018**, *7*, 1–7.
35. Sukhadia, K.; Chaudhari, M.B. A Survey on Rice Crop Yield Prediction in India Using Improved Classification Technique. Int. *J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **2019**, *5*, 501–507.