IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 July 2023, accepted 17 August 2023, date of publication 4 September 2023, date of current version 13 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3311827

APPLIED RESEARCH

Explainable AI for Soil Fertility Prediction

HARSHIV CHANDRA[1], (Student Member, IEEE), PRANAV M. PAWAR[1], (Member, IEEE), ELAKKIYA R.[1], TAMIZHARASAN P S[1], (Member, IEEE), RAJA MUTHALAGU[1], AND ALAVIKUNHU PANTHAKKAN[2], (Senior Member, IEEE)

[1]xPERT Research Group, Department of Computer Science, BITS Pilani, Dubai Campus, Dubai, United Arab Emirates
[2]College of Engineering and IT, University of Dubai, Dubai, United Arab Emirates

Corresponding authors: Alavikunhu Panthakkan (apanthakkan@ud.ac.ae) and Pranav M. Pawar (pranav@dubai.bits-pilani.ac.in)

ABSTRACT Soil fertility refers to the ability of soil in a particular area to provide favorable chemical, physical and biological characteristics that help the plant in its growth. It is affected by multiple parameters, from the available concentration of Nitrogen in the soil to the concentration of Organic Carbon in the soil. This paper discusses the implementation of an explainable AI (XAI) model based on a Random Forest classifier. The developed model reliably predicts the relative soil fertility of a given soil using its various physiochemical properties, and explain the reasons behind the model's soil fertility indicator prediction using user friendly graphs. The model shows 97.02% accuracy in comparison with state-of-the-art machine learning models. The paper also discusses applications of developed model in providing possible solutions to further improve upon soil fertility in the short term and long term.

INDEX TERMS Explainable AI, machine learning, random forest classifiers, soil fertility.

I. INTRODUCTION

Agriculture has always played a vital role in human society. It has had a significant impact on the development of human civilization over the centuries, having influenced and provided human civilizations with the basis of real development in economic terms [1]. It has traditionally needed the presence of favorable environmental conditions. Today, due to a rapid rise in requirements over the past decades, agricultural produce demand has skyrocketed, whilst increasing urbanization has simultaneously led to a decrease in usable arable land for agricultural purposes [2]. This is putting pressure on nations to find a solution to the dual challenge of rising demand with a reduction in available land for agricultural use. In such a scenario, there is a need to adopt sustainable and optimized methods to improve agricultural yields from available arable land, without harming the environment. An increase in soil nutrient removal, due to increased cultivation of land, has led to an overall depletion of soil fertility [3]. This may lead to an increased risk of future food crises for the world's inhabitants.

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wang.

Modern, computer-based optimization and forecasting techniques can be used to help prevent such a scenario. Various computational techniques (traditional as well as AI-based) are being used to compute the fertility parameters of soil samples. The main issues with these methods are the lack of transparency and high capital expenditures. Through the explainable AI (XAI)-based model proposed in this paper, we aim to improve and address these shortcomings, by providing the farmers with the ability to predict and interpret soil fertility using a lucid waterfall plot, without the need for complex traditional analysis techniques, while at the same time, providing the farmer with a clear explanation on the reasons behind the same.

A. SOIL FERTILITY

Soil fertility refers to the ability of soil in a particular area to provide favorable chemical, physical and biological characteristics [4], that help plants in their growth. The presence of fertile soil could be beneficial to the environment, as it improves vegetation restoration, providing an opportunity to develop a carbon-neutral ecosystem [5], if further organic soil amendments are incorporated. It is an important metric as soil fertility is directly related to a plant's nutrition [6],

and a continuous nutrient supply during the crop growth phase can maximize crop productivity [7].

B. SOIL FERTILITY AND ITS ALIGNMENT WITH UNSDG

Improving Soil fertility is a part of the United Nations Sustainable Development Goals (UNSDG). The 2030 Sustainable Development Agenda identifies several goals that are directly or indirectly linked to soil fertility, including SDG 2 (Zero Hunger), SDG 13 (Climate Action), and SDG 15 (Life on Land). Thus, it can aid efforts to end hunger, mitigate climate change, protect the environment, and contribute to an overall improvement in the health and wellbeing of the populace.

C. UNDERSTANDING SOIL PARAMETERS

Soil types differ according to different geographical conditions [9] often characterized by elevation and slope, and some specific soil types typically require specific climatic conditions to exist. Therefore, there must be a standardized set of parameters to consider when looking into soil fertility for different soil samples. Reference [10] describes different parameters that are required for soil fertility prediction in a traditional setup, and how to gather them using lab-based techniques. [10, Table 1] summarizes these parameters and the techniques used to extract them.

Each metric in Table 1 has its purpose in the soil samples, described below,

- SOM provides nutrients to the soil through recycling methods like littering inputs from ash deposits, and mineralization of plant-based remains [11], as well as through the action of living organisms. It is critical for the stabilization of the soil structure and provides mechanisms for the retention and release of plant nutrients and maintenance of water-holding capacity. [12].
- OC is a valuable indicator of soil quality [13], and is a part of SOM. A higher concentration of OC promotes soil structure, leading to greater physical stability, leading to lower chances of erosion and nutrient leaching from the soil.
- pH indicates the acidity or basicity of a soil type, and has an enormous influence on soil biogeochemical processes [14].
- EC indicates the salinity status of the soil and is influenced by both natural and anthropogenic factors [15].
- N is required for plant growth, and plant food processing. It is affected by the changes in the organic matter content of the soil [16].
- K is an essential cation and plays a vital role in the physiological processes in plants, and Na acts as a promoter of plant growth [17].
- Soil micronutrients like Cu, Zn, Fe, and Mn participate in the enzyme activation processes in plants [18].
- Both Ca and Mg help neutralize organic acids, which form during plant cell metabolism. Ca is also required for cell wall formation and normal cell division and participates in the enzyme activation processes in plants. Mg is also an essential component of chlorophyl and acts as a phosphorous carrier in plants [19].
- B plays an important role in structural integration and cell wall synthesis. It also participates in the nitrogen and carbohydrate mechanism and is responsible for sugar transport [20].
- S acts as a signaling molecule in stress management and normal metabolic processes [21].

TABLE 1. Traditional soil chemical parameters.

Soil Parameters	Extracting Methodology
SOM (%)	Gravimetrically measured using LOI at 550 ◦C for 5 h in a muffle furnace.
OC	Chromic acid digestion.
pH1:2	1:2 soil: water ratio via a standard digital pH meter.
EC1:2	1:2 soil: water ratio using a digital EC meter.
Available N	Analysed by the alkaline potassium permanganate method.
Available K and Na	Extracted by ammonium acetate and measured using a flame photometer.
Plant available micronutrients (Zn, Cu, Fe, Mn)	Extracted by DTPA (1:2) subsequently quantified via AAS.
Available Ca and Mg	Extracted using ammonium acetate solution and subsequently measured via AAS.
Available B	Extracted using hot water extraction method with a dilute CaCl2 solution and subsequently colorimetrically measured using azomethine H.
Available S	Extracted via CaCl2 and the resulting turbidity was measured at 440 nm using an UV-vis spectrophotometer.

Key:
SOM : Soil Organic Matter
LOI : Loss on Ignition
OC: Oxidizable Soil Organic Carbon
EC : Electrical Conductivity
N: Nitrogen
K: Potassium
Na: Sodium
Zn: Zinc
Cu: Copper
Fe: Iron
Mn: Manganese
DTPA: Diethylenetriaminepentaacetic acid
AAS: Atomic Absorption Spectrophotometer
Ca: Calcium
Mg: Magnesium
B: Boron
S: Sulphur
UV-vis: ultraviolet-visible

When considered in a composite fashion, these parameters can determine the relative soil fertility of a given soil sample. [10, Table 2] describes experimentally recorded ranges and concentration indicators (either present in low, medium, or high quantities).

D. SOIL FERTILITY PREDICTION

Various techniques such as VNIR (Visible and Near Infrared) spectroscopy, XRF (X-ray fluorescence) spectroscopy, and laser-induced breakdown spectroscopy, [22] have been used to generate datasets for conventional prediction of

TABLE 2. Classification of soil nutrient concentrations.

Element	Low	Medium	High
N (in kg ha^{-1})	<281	281-560	>560
K (in kg ha^{-1})	<151	151-250	>250
Ca (in mg kg^{-1})	<2000	2000-4000	>4000
Mg (in mg kg^{-1})	<396	396-996	>996
Zn (in mg kg^{-1})	<1.21	1.21-2.4	>2.4
Cu (in mg kg^{-1})	<0.41	0.41-1.2	>1.2
Fe (in mg kg^{-1})	<9.1	9.1-27.0	>27.0
Mn (in mg kg^{-1})	<4.1	4.1-16.0	>16.0
B (in mg kg^{-1})	<1.0	1.0-2.0	>2.0

Key:
kg ha^{-1}: Kilograms per hectare
mg kg^{-1}: Milligrams per kilogram

soil fertility. These datasets are important in deducing functions required to predict soil fertility. A mathematical model has also been proposed, which involves the use of differential equations to predict soil fertility [23]. It considers 8 input metrics (or features), and can be represented by [23, eq1],

$$\frac{dPi}{dt} = f_i(humus, N, F, K, pH, W, \Phi_i, CO_2) \quad (1)$$

where P_i measures fertility, N denotes Nitrogen concentration in the soil, F denotes Fluorine concentration in the soil, K denotes Potassium concentration in the soil, pH indicates the potential of Hydrogen that is used to compute how acidic or basic the given soil sample is, W denotes soil moisture, Φ_I denotes the rate of photosynthesis, determined by the intensity of the diffusion flux of the CO2 to chloroplasts from the atmos, and CO_2 denotes the concentration of carbon dioxide in the soil [23].

With differential equations, it can be deduced that the theoretical soil fertility of a particular soil sample can be given by [23, eq2] and [23, eq4], a_1, $a_2$and a_3 being mathematical constants dependent on the soil parameters, c being a random constant, and t denoting a time T for which soil fertility is computed.

$$P = \frac{1}{2a_3}\left[A\left(1 - \frac{2}{e^{A(t+c)}}\right) - a_2\right] \quad (2)$$

$$A = \sqrt{a_2^2 + 4a_1a_3} \quad (3)$$

It is a very large model, requiring various calculations at every stage, making it a complex operation, due to the time requirements. Also, compared to lab-based calculations of soil fertility, the model does not seem to be accurate in predicting soil fertility. However, one advantage of this approach is transparency, as it can justify the increase and decrease of soil fertility via a well-structured relationship – in this case, the fertility decreases with the decrease of humus, photosynthesis rate, the intensity of the diffusion flow of the CO2, calcium, phosphorus, and soil moisture [23].

E. FEASIBILITY OF USING MACHINE LEARNING TO PREDICT SOIL FERTILITY

To simplify the complexity of predicting soil fertility, appropriate machine learning (ML) models can be used to implement a fertility prediction model with given datasets. This would reduce time to derive approximate values of soil fertility, given certain input variables. It has already been achieved whilst addressing other challenges in agriculture (such as crop management, water management, and livestock management). Reference [24] describes a method to predict crop yields using the Naïve Bayes algorithm, using variables such as soil moisture levels, humidity, and temperature. It can predict a particular crop yield with an accuracy of 97%, using a dataset with a training testing split of 70%:30%. Reference [25] has been able to implement a sugarcane yield grade forecasting model, using the Random Forest (RF) machine learning classifier, with variables such as fertilizer type, and soil type. The proposed implementation achieves an accuracy of 71%, an improvement over previous baselines of 50% accuracy. Thus, Machine Learning has already been successfully applied in other domains of agriculture, and there also exist implementations that predict soil fertility using machine learning models, that have been discussed later in this paper.

TABLE 3. Differences between white-box and black-box models.

White Box Models	Black Box Models
The model's internal workings are known and transparent.	The model's internal workings are unknown and opaque.
Easier to debug and explain any behavior of the model	Tough to debug and explain any behavior of the model
Lower accuracy compared to black box models, as they lack the ability to learn complex relationships between input and output variables	Higher accuracy compared to white box models, as they can learn complex relationships between input and output variables

F. DISADVANTAGES OF USING ML FROM THE END USER PERSPECTIVE

One of the main disadvantages when exploring solely ML solutions to agricultural problems is the inherent lack of transparency regarding the decisions taken by the model during prediction of the output [26]. Another aspect of the lack of transparency in traditional ML and DL models is the loss of trust in the model, as it cannot provide a logical reasoning of its predictions, due to the traditional black box approach to solving problems. This is where we can consider the application of Explainable AI onto the model, as it provides the developer with a transparency design that aptly describes how the model functions, from its structure, singular components, and its training algorithms [26], and provides the end user with a post hoc explanation, in terms of analytics, visualizations, and examples [26]. This leads to an increase in trustworthiness in the model as well, due to the white-box approach [27] taken by such models. [28, Table 3] provides a brief comparison between white-box and black-box models.

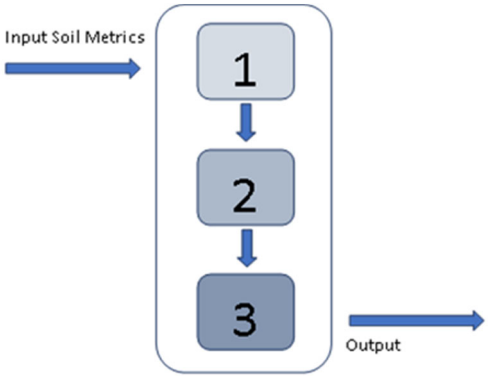

FIGURE 1. High Level Model Abstraction.

G. A HIGH-LEVEL DESCRIPTION OF THE PROPOSED MODEL

The higher-level description of the model is shown in Figure 1. It consists of three layers. This paper discusses the implementation of all three layers of the proposed model, and compares the proposed methods with previous approaches to predicting soil fertility. The functions of each layer are listed below,

- Layer 1: This layer describes the K-means clustering layers, that takes in the input dataset and clusters it accordingly into 5 different categories in terms of input variable concentrations and accordingly classifies each sample in the input dataset into 3 categories depicting relative soil fertility.
- Layer 2: This layer describes the RF classifier model, which takes in input soil variables and generates an approximate label describing the relative soil fertility of the input sample. The implementation of this model would be discussed later in this paper.
- Layer 3: This layer would consist of an Explainable AI layer, which attempts to give a human interpretable understanding of the reasoning behind the value predicted by layer 2, providing it with a transparent design [27].

The proposed approach differs from state-of-the-art approaches in the use of a XAI layer to provide reasoning for the output prediction, whilst also converting the model from a black-box to a white-box based approach [26]. This makes the proposed model's output much more user friendly, as it provides a user interpretable waterfall plot explaining each metrics contribution to the rating of the soil in terms of soil fertility.

The remaining part of the paper is organized into four sections. Section II discusses the related work in area of intelligent soil prediction. Section III provided gives details about proposed XAI model for soil fertility prediction. Section IV provided results and discussion, along with further application of model in real world scenario. Finally, the paper concludes with future directions and references.

II. RELATED WORKS

The selected papers subjects are related to soil fertility and machine learning. Table 4 compares the articles found regarding the machine learning algorithm implemented, the advantages/disadvantages to the approach, and the accuracy to the approach, and the level of transparency and post facie analysis compared to the proposed model.

Reference [29] discusses the implementation of Partial Least Squares (PLS) regression to make predictions of soil fertility and crop yield from a procedurally generated dataset, involving various entities as described in the AgroXML standard. For predicting soil fertility, 2 different PLS models were constructed, with one involving organic matter calibration and the other involving clay calibration data, and the final models had Pearson correlation coefficient (R^2) scores of 0.94 and 0.92, mean square error of calibration (RMSEC) scores of 0.36 and 3.36, and mean square error of cross-validation (RMSECV) scores of 0.54 and 5.28 respectively, an improvement over previous baseline models with smaller datasets involved.

Meanwhile, [30] describes a model that can appropriately determine the suitable algorithm for predicting soil fertility. It shows that linear regression is an efficient algorithm when it comes to grading soils based on their properties, due to its small (RMSE) score of 0.0617, and that the most suitable algorithm for classifying these soils based on measured fertility variables is the RF algorithm, which achieves an accuracy of 72% compared to other algorithms applied on the same dataset (Support Vector Machines (SVM) (linear kernel) and Gaussian Naïve Bayes (GNB) each having an accuracy of 63 % and 50.78% respectively).

Reference [31] compares the performance of the Generalized Linear Model (GLM) and RF algorithm in predicting soil fertility using (PXRF) soil data. It is shown that RF performs better than GLM, providing higher values of R^2, residual prediction deviation (RPD) and ratio of performance to inter-quartile distance (RPIQ) whilst simultaneously having lower values of mean absolute error (MAE), RMSE and normalized root mean square error (nRMSE).

Reference [32] proposes an implementation of various regression algorithms such as RF, Gradient Boosted Machine (GBM), and Bayesian Additive Regression (BAR) tree towards the prediction of soil fertility metrics. It is observed that a RF regression implementation with feature selection achieves the highest R^2 value of 0.70 indicating very good to excellent correlation, as it achieves values closest to the true fertility index of the soil samples tested.

Reference [33] describes an implementation of the SVM algorithm that uses both soil and crop datasets to predict soil fertility and then use that prediction to further predict suitable crops for a given soil type. The paper describes that its proposed model performs better than previous baseline models implemented on similar datasets. (94.95 % accuracy compared to previous baseline implementations of 91.90% accuracy and 92.30 % accuracy respectively).

TABLE 4. Comparison of different soil fertility models.

Criterion	Helfer et al, 2020 [29]	Kumar et al, 2019 [30]	Benedet et al, 2021 [31]	Sirsat et al, 2018 [32]	Rahman et al, 2018 [33]	Our Model
ML Algorithm Implemented	PLS Regression	Various Regression & Classification Algorithms, like SVM, GNB and RF	GLM & RF Algorithm	Various Regression techniques like GBM, BAR and RF	SVM	RF Algorithm
Methodology	Three phases – Data collection and processing, training, and testing. Uses a procedurally generated soil dataset.	Three phases – Data pre-processing, training, and testing. Uses 1 soil dataset, and 2 modules in the model.	Three phases – Soil Sampling and lab analyses, data processing and modelling, spatial application. Uses 1 soil dataset.	Three phases – hyperparametric tuning, training, and testing. Uses one soil dataset	Two phases – training and testing. Uses 2 datasets – soil dataset and crop dataset	Three phases – training, testing, and explaining the results. Uses 1 Soil dataset.
Evaluation of Metrics	R^2, RMSEC, RMSECV	Accuracy %, RMSE	R^2, RMSE, MAE, nRMSE, RPD, RPIQ	R^2, RMSE	Accuracy %	Accuracy %, F1 Scoring, ROC-AUC Score
Level of transparency	Low – no methods mentioned to describe reasons behind model output	Low – no methods mentioned to describe reasons behind model output	Low – no methods mentioned to describe reasons behind model output	Low – no methods mentioned to describe reasons behind model output	Low – no methods mentioned to describe reasons behind model output	High, due to presence of an Explainable AI layer describing implementation of the model in detail

Key:
BAR: Bayesian Additive Regression
GBM:Gradient Boosted Machine
GLM: Generalised Linear Machine
GNB: Gaussian Naïve Bayes
MAE: Mean Absolute Error
nRMSE: normalized Root Mean Square Error
PLS: Partial Least Squares
R^2: Pearson correlation coefficient
RF: Random Forest
RMSE: Root Mean Square Error
RMSEC: Root Mean Standard Error of Calibration
RMSECV: Root Mean Standard Error for Cross Validation
ROC-AUC: Receiver Operating Characteristic - Area Under the Curve
RPD: Residual Prediction Deviation
RPIQ: Ratio of Performance to Inter-Quartile distance
SVM: Support Vector Machine

The proposed model in this paper shall implement a RF algorithm to classify various physiochemical properties of given soil samples to predict a composite soil fertility metric, and provide reasons as to why the soil is fertile.

III. PROPOSED METHODOLOGY

This section describes the implementation of the proposed model in this paper, from the dataset used to the model used. The model has been implemented using the Python programming language and the scikit-learn module. Figure 2 gives a visual understanding of the working of the model.

A. SOIL FERTILITY PREDICTION USING XAI

- Data Preprocessing: The input raw data from the LUCAS 2018 topsoil dataset (which consists of pH_$CaCl_2$, pH_H_2O, EC, OC, $CaCO_3$, N, P and K attributes) is cleaned, and all null values are replaced using the IterativeImputer function present in the scikit-learn module of python.
- Model Preparation and Implementation: As discussed earlier, the proposed model would consist of 3 different models acting as nodes or layers, in which the first node (consisting of a K-Means model) would provide labels to the data, the second node consisting of a RF Classifier would train on the labelled data and generate predictions on given test data, and the third node would consist of an TreeExplainer layer, that would provide explanations of the output from the second node.
- Model Evaluation: The performance of the RF Classifier model will be evaluated using the accuracy and F1 scoring metrics calculated on the predicted output.

B. DATASET USED

The dataset used for this model is the LUCAS 2018 Topsoil dataset [40], that has 18984 sample points of soils taken all over Europe. It is the largest open source database of its kind, spanning across the entire geographic region of the European union and the United Kingdom. It has been stored in the .csv format, and for each datapoint, the proposed model derived upon the above dataset would use the input parameters given in Table 5. Table 6 summarizes the datasets used by the selected papers and the proposed model, alongside the regions and the metrics used for each dataset.

C. DATA PREPROCESSING

As shown in Table 6, some parameters of the raw dataset have < LOD (Limit of Detection) as values, and there are also some instances of NaN values in the dataset. To eliminate problems that stem from these issues, NaN values of the given dataset were filled using the IterativeImputer function in the scikit-learn module, which works by first fitting itself on the given dataset, and then predicting any missing values in the dataset wherever necessary. Table 7 gives a brief

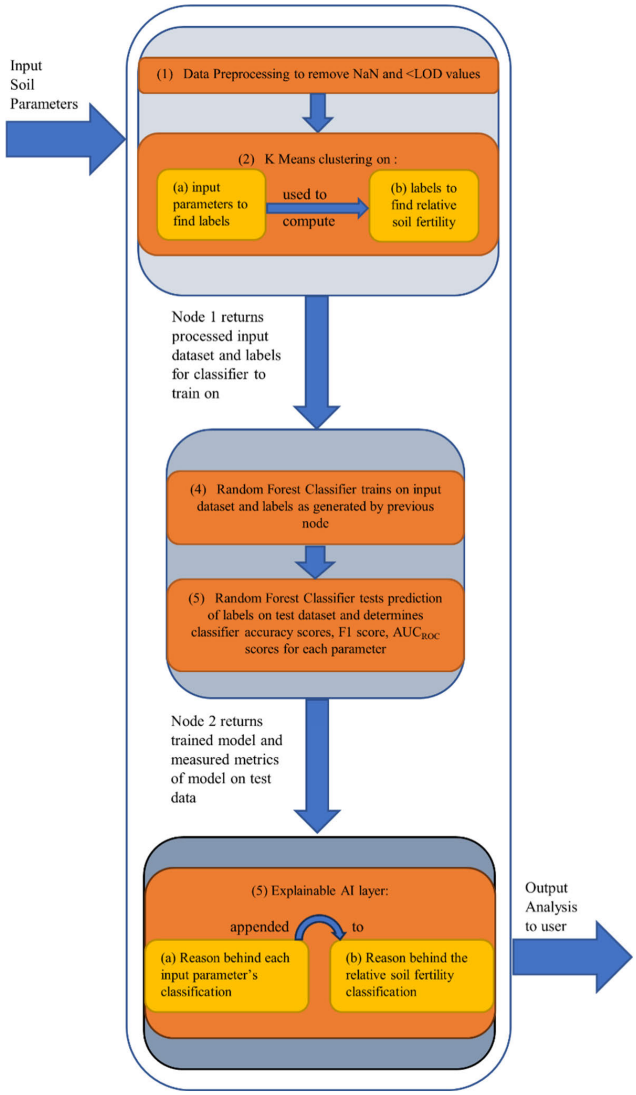

FIGURE 2. A flowchart describing the working of the model.

TABLE 5. Input parameters within given model.

Parameters	DESCRIPTION
pH_$CaCl_2$	pH of soil sample measured in calcium chloride
pH_H_2O	pH of soil sample measured in water
EC	Electrical Conductivity of soil sample (mS m^{-1})
OC	Organic Carbon content of soil sample (g kg^{-1})
$CaCO_3$	Calcium Carbonate content of soil sample (g kg^{-1})
N	Total Nitrogen content in soil sample (g kg^{-1})
P	Total Phosphorous content in soil sample (g kg^{-1})
K	Extractable Potassium content in soil sample

description of the data after processing. Also, for each category, all < LOD values were replaced with the lower bound of the limits of detection asgiven in [34]. Since there were no labels generated previously for this dataset, a clustering model is used to generate labels for the classification model to train on. The clustering model uses the K-Means algorithm, where the K-value is first computed using a within-cluster sum of squares (WSS) algorithm, that evaluates the input data and computes the necessary number of clusters for the input data using the elbow technique, The Elbow Point is the value of k at which the WSS starts to level off and the rate of decrease slows down. It typically forms a sharp bend, resembling an elbow. The Elbow Point indicates the optimal k value, as it strikes a balance between capturing meaningful clusters (low WSS) without overfitting or introducing noise. Figure 3 demonstrates the results of the WSS algorithm that are used to choose the K-values for the clusters in the dataset.

After selecting the appropriate cluster count for the input parameters, the WSS algorithm is also used to compute the appropriate k value for the relative soil fertility labelling on the dataset. Figure 4 shows the results for the same.

Based on the results from the above computations, the K-Means model individually categorizes each parameter of a given datapoint into 5 distinct categories, depending on their relative values, and then collectively clusters them into 3 different relative soil fertility categories. The results of this are shown in Figure 5, in which, the x-axis values are Point IDs, which are unique references to each soil sample in the dataset, and y-axis denotes the respective values for the given metric.

To further understand the monotonic relationships between the parameters and the relative soil fertility, the Pearson correlation coefficient was computed in python using the pandas library. This coefficient is a dimensionless value, and is used to depict whether a particular variable is related to another variable, and usually has a value between −1 to 1 [35], where −1 depicts a negative correlation, 0 depicts no correlation, and 1 depicts a positive correlation between any two variables. Table 8 shows the results of this computation, depicting the relationship between the concentrations of a particular parameter and the relative soil fertility, and its relationships with the categories generated using the K-Means algorithm during the clustering step. It is also imperative to understand the importance of feature selection, and what features are selected by the RF classifier to predict the soil fertility. Thus, the feature importance metric was examined on the input dataset, and the results are shown in Figure 6. Algorithm 1 shows the complete process involved during data preprocessing and clustering.

D. MODEL PREPARATION AND IMPLEMENTATION

In the proposed model, the processed data goes through 2 primary stages, namely, a multi-class multi-output classification model, that formulates relationships between the classification labels and the dataset, and attempts to predict soil fertility categories and soil parameter categories simultaneously for a given dataset, given input real values. The second layer, an explainable AI (XAI) analysis layer, attempts to explain the relative soil fertility based on the results from the above 2 models. A K-Means cluster acts as a data preprocessing

TABLE 6. Comparison of different soil fertility datasets.

Criterion	Helfer et al, 2020 [29]	Kumar et al, 2019 [30]	Benedet et al, 2021 [31]	Sirsat et al, 2018 [32]	Rahman et al, 2018 [33]	Proposed Model
Dataset	Soil Dataset [29]	ICRISAT V3 Soil Dataset [42]	PXRF based Soil Dataset [31]	Soil Dataset [44]	SRDI Soil Dataset [43]	LUCAS 2018 Topsoil Dataset [40]
Source	Soil data constantly generated from sensors deployed on 450 soil samples in Vale do Rio Pardo/RS, Brazil	Soil data compiled from soil samples of farm fields across 13 districts of Andhra Pradesh, India	Soil data compiled from soil samples of 7 Brazilian states	Soil data compiled from soil samples from ten villages across 3 districts of the region of Marathwada, Maharashtra, India	Soil data compiled from 6 upazillas of Khulna district, Bangladesh.	Soil data compiled from samples collected across the European Union and UK
Features of Dataset	pH, EC, OM, luminosity, rainfall, clay, pressure, temperature humidity	pH, EC, OC, S, K, Zn, Mn, B, Soil Type	Al, Ca, Ce, Cl, Cr, Cu, Fe, K, Mn, Nb, Ni, P, Pb, Rb, Si, Sr, Ta, Ti, V, Y, Zn, Zr	EC, OC, N_2O, P_2O_5, K_2O, SO_4 , Cu, Fe, Mn, B, Zn	pH, Salinity, Organic Matter %, K, S, Zn, B, Ca, Mg, Cu, Fe, Mn	pH_$CaCl_2$, pH_H_2O, EC, OC, N, P, K, $CaCO_3$

Key:
EC: Electrical Conductivity
OC: Organic Carbon
OM: Organic Matter
pH: Potential of Hydrogen
$CaCO_3$: Calcium Carbonate
pH_$CaCl_2$: pH measured in Calcium Chloride
K_2O: Potassium Oxide
SO_4: Sulphate
pH_H_2O: pH measured in water
P_2O_5: Phosphorous Pentoxide

TABLE 7. Some statistics of raw input dataset.

Parameter	Mean	Std	Min	Max
pH_CaCl2	5.71	1.40	0.00	9.80
pH_H2O	6.26	1.32	0.00	10.43
EC	18.38	25.56	0.00	1295.60
OC	47.52	81.60	< LOD	723.90
CaCO3	56.82	135.08	< LOD	926.00
N	3.15	3.72	< LOD	46.50
P	25.57	28.16	< LOD	515.00
K	204.03	207.06	< LOD	7578.80

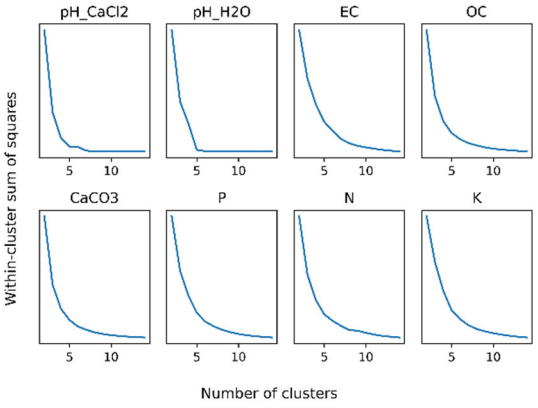

FIGURE 3. A plot showing the optimal number of clusters for the input parameters to the model.

layer for the model, wherein, unlabeled data is provided labels for the rest of the model to compute its predictions from. Table 9 shows the numeric label clusters formed in this step, alongside their equivalent human interpretable discrete label assignment and color (used to depict datapoints in Figure 5). This clustering technique works on the concept of dividing multiple datapoints in N dimensions into K clusters so that the sum of squares within points in a cluster is minimized [36]. Thus, each datapoint is assigned a set of integer labels based on its input attributes. This K-Means cluster uses the implementation present in the scikit-learn module [38] from python. When the labels are generated, the data is stored in .csv files, and then the data is loaded into the program for the classification model.

Each of these algorithms are based on practical implementations of mathematical formulae. The RF Classifier works by computing impurity coefficients respectively to determine how nodes branch on a given tree. By default, the gini impurity coefficient is used, which is given by (4),

$$GiniImpurity = 1 - \sum_{i=1}^{C} (p_i)^2 \tag{4}$$

where P_i denotes the relative frequency of a particular class in a dataset, and C represents the number of classes. The XAI model used uses a SHAP Tree Explainer layer [39],

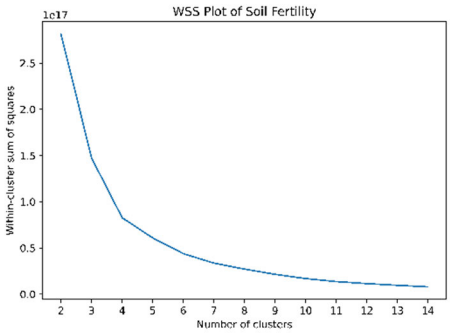

FIGURE 4. A plot showing the optimal number of clusters for generating relative soil fertility labels.

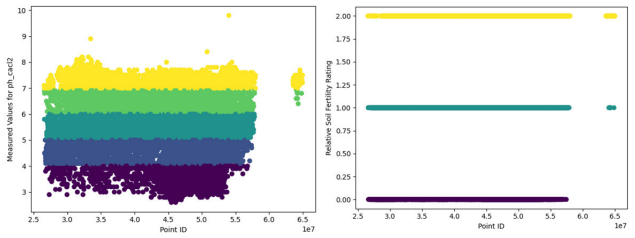

FIGURE 5. A Scatter plot of clusters formed for each soil point on (a) pH_CaCl2 values (b) combined relative soil fertility.

TABLE 8. Some statistics of processed input dataset.

Parameter	Mean	Std	Min	Max
pH_$CaCl_2$	5.71	1.40	2.60	9.80
pH_H_2O	6.26	1.32	3.34	10.43
EC	18.39	25.55	0.24	1295.60
OC	47.52	81.60	0.00	723.90
$CaCO_3$	96.14	126.54	1.00	926.00
N	3.15	3.72	0.20	46.50
P	28.23	26.02	0.00	515.00
K	204.06	207.04	6.20	7578.80

that attempts to compute SHAP values for Tree based models (exclusively) in polynomial time. SHAP values are traditionally computed using (5),

$$\phi_i(p) = \sum_{S \in N[i]} \frac{|S|!(M - |S| - 1)!}{M!} [p(S \cup \{i\} - p(S))] \tag{5}$$

where, φ_i denotes the Shapley value for any feature i (out of a total of N features), M denotes the overall number of features, p denotes the prediction given by the model, and S denotes a set containing non-zero indexes for the features [41].

The input data has a training-testing split of 80%-20%.

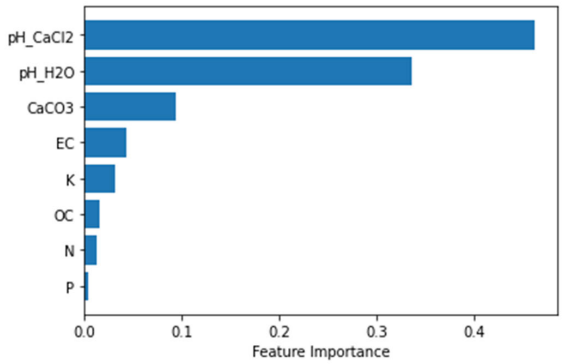

FIGURE 6. Feature Importance for the RF Classifier.

Algorithm 1 Data Preprocessing & Clustering

Input: Raw .csv files containing LUCAS 2018 Topsoil dataset
Step 1: Load & clean dataset D_i using defined classes
Step 2: Define Clusters C_1 & C_2, having 5 & 3 classes respectively.
Step 3: Fit dataset D_i to clusters C_1 and C_2, & combine generated label sets L_1 and L_2 with D_i
Step 4: compute_correlation(Data, RelSoilFertility)
Step 5: compute_feature_importance(Data)
Step 6: write_back(Data, filename ← 'data.csv')

```
import the Numpy library as np and the Pandas library as pd
function clean_data(dataset)
  data_new ← ""
  for i in dataset do
    column ← np.asarray(i).reshape(i.shape[0],1)
    imp ← IterativeImputer(max_iterations = 10)
    col_new = imp.fit_transform(m)
    if type(data_new) = string then
      data_new = col_new
      continue
    end if
    data_new = np.c_[data_new, col_new]
  end for
  data_new = pd.DataFrame(data_new)
  return data_new
end function
function load_data(filename)
  load ← pd.readcsv(filename)
  return load
end function
function combine(dataset1,dataset2)
  return np.c_[dataset1,dataset2]
end function
procedure compute_correlation(dataset)
  r ← dataset[:8].columns
  p = dataset[9:]
  for i in r do
    correlation ← p[i].corr(p['Relative Soil Fertility'])
    corr_itself ← r[i].corr(p[i])
    display correlation
    display corr_itself
  end for
end procedure
procedurecompute_feature_importance(dataset)
    Define the classifier model
    Fit the classifier to the data
    sort ← clf.feature_importances_.argsort()
    Plot a bar graph of sort, labelled Feature Importance
end procedure
```

Output: Processed data, stored in data.csv, with generated labels.

E. CLASSIFICATION MODEL

The classification model is built on the RF classification algorithm [37], a bagging ensemble learning technique that

TABLE 9. Cluster categories.

Criterion	Label map, Integer map and equivalent colour map				
Metrics	Very Low: 0	Low: 1	Medium: 2	High: 3	Very High :4
Relative Soil Fertility	Low: 0		Normal: 1		Excess: 2

involves an additional layer of randomness. It is a supervised learning algorithm, which means that it is trained on a labeled dataset. Each decision tree in the RF makes a prediction based on the features of an example, and the predictions of all the trees are combined to make the final prediction for the RF. This model uses the implementation of the RFClassifier class, present in the scikit-learn module [38] from Python. This model uses the labeled data generated in the previous step to understand and thus predict the relative soil fertility category that each datapoint belongs to. Algorithm 2 shows the step by step working of RF model preparation in abstract form.

Algorithm 2 RF Model

Input: Processed data, stored in data.csv, with generated labels.
Step 1: Fetch data, & split it into training & testing set (80%-20%).
Step 2: Define a RFClassifier from the scikit-learn library
Step 3: Fit the training data variables to the RF Classifier
Step 4: Get Predictions from the RF Classifier based on testing data
Step 5: Compute and display Models F1 score
Step 6: Compute and display Models Accuracy
Output: Trained Multiclass, Multioutput RF Model.

F. EXPLAINABLE AI (XAI) ANALYSIS LAYER

To understand the correlation between the relative soil fertility classification and the soil parameters (like pH_CaCl2, pH_H2O etc.), the relationships formed by the ensemble classifier must be explained using the SHAP (SHapley Additive exPlanations) module from Python. It is a game theoretic approach [39] that allows optimal explanations of the global models output based on understanding the cumulative local explanations of the model's predictions. Figure 7 shows three different waterfall plots from the SHAP module, which describes the impact of the metrics on relative soil fertility categorization. It can be observed in Figure 7 that the pH_H_2O and pH_$CaCl_2$ have a huge impact on the fertility of the soil. A low or high pH_H_2O and pH_$CaCl_2$ value can impact the relative soil fertility rating of a particular sample of soil. Other metrics like P, N, OC, K, $CaCO_3$, and EC also have a significant impact on the relative soil fertility rating of any given soil sample. Table 10 gives the selected datapoints for each waterfall plot, and their categorization. Algorithm 3 shows the stepwise brief overview of XAI model preparation and generating lucid waterfall plots.

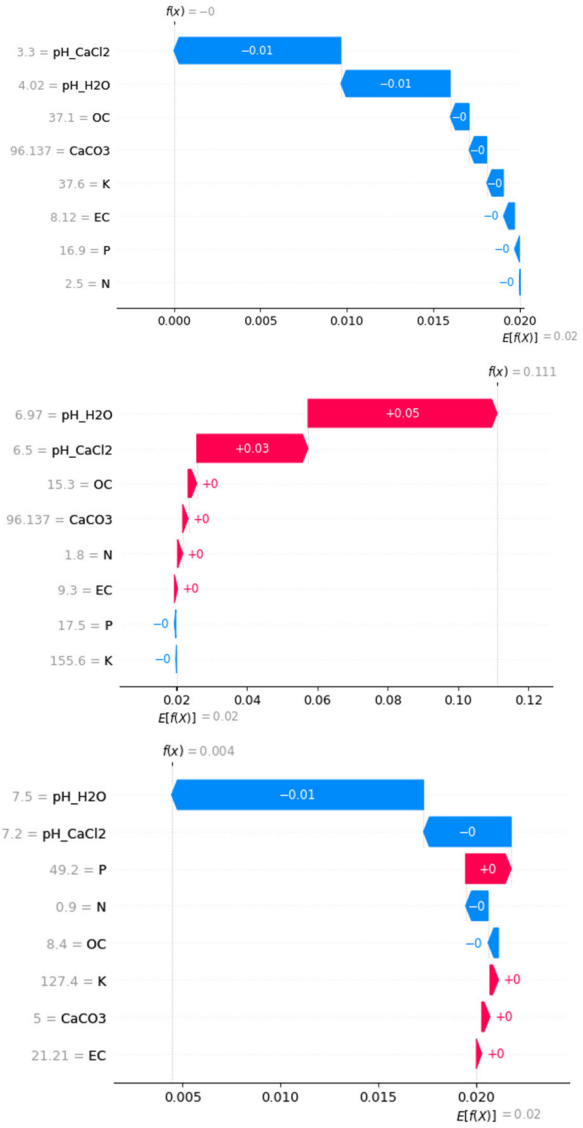

FIGURE 7. A datapoint with (a) low relative soil fertility, (b) medium relative soil fertility and (c) high relative soil fertility.

IV. RESULT, DISCUSSION, & APPLICATION SCOPE OF MODEL

A. EVALUATION METRICS

The performance of the RF model is measured using the Accuracy and F1 scoring for the classification of each metric. The Accuracy metric is computed using (6), (whose variables are explained in Table 11),

$$Accuracy\% = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Alongside this, an F1 score is computed on the classifier's output, for each metric, using (7),

$$F1 = \frac{2 * precision * recall}{precision + recall} \tag{7}$$

TABLE 10. Selected datapoints for waterfall plot.

Metrics	Low Relative Soil Fertility		Normal Relative Soil Fertility		Excess Relative Soil Fertility	
pH_CaCl2	3.30	0	6.97	3	7.50	4
pH_H2O	4.02	0	6.50	3	7.20	4
EC	8.12	0	9.30	0	21.21	1
OC	37.10	1	15.30	0	8.40	0
CaCO3	96.14	1	96.14	1	5	0
N	16.90	0	17.50	0	49.20	2
P	2.50	0	1.80	0	0.90	0
K	37.60	0	155.60	0	127.40	0

Algorithm 3 XAI Model

Input: Trained Multiclass, Multioutput RF Model; Test dataset
Step 1: Load Model into a TreeExplainer Function
Step 2: Create a Tree Explainer class based on Model
Step 3: Load Test Dataset D_{test}
Step 4: Generate SHAP values using TreeExplainer on D_{test}
Step 5: Project SHAP values onto a waterfall plot
Output: Lucid waterfall plot depicting the effect of individual soil fertility metrics on the relative soil fertility for a given soil sample.

TABLE 11. Meaning of the variables.

Name	Function (Predictions)
TP	Number of True Positives
TN	Number of True Negatives
FP	Number of False Positives
FN	Number of False Negatives

where, both precision and recall are quantities that are computed using (8) and (9), assuming the variables as defined in Table 11,

$$precision = \frac{TP}{TP + FP} \tag{8}$$

$$recall = \frac{TP}{TP + FN} \tag{9}$$

The AUC_{ROC} score is also calculated, using the scikit-learn roc_auc_score function, with the inputs being y_test and the prediction probabilities of the model, in the One-vs-Rest format.

Based on calculations using these formulae, it is found that the proposed model achieves a high score of 96.97% in terms of accuracy % and 0.90 in terms of the F1 score. The results

TABLE 12. Model performance.

Metrics	Accuracy	Precision	Recall	F1 Score	AUC_{ROC} Score
pH_CaCl_2	99.97	0.99	0.99	0.99	0.99
pH_H_2O	99.66	0.99	0.99	0.99	0.99
EC	98.68	0.77	0.68	0.71	0.62
OC	99.34	0.98	0.97	0.97	0.99
$CaCO_3$	79.83	0.84	0.82	0.83	0.96
P	97.60	0.76	0.69	0.72	0.96
N	99.55	0.98	0.97	0.97	0.99
K	98.60	0.78	0.63	0.65	0.89
Soil Fertility	100	1.0	1.0	1.0	1.0
Avg.	97.02	0.90	0.86	0.87	0.93

TABLE 13. Performance comparison with other implementations.

Criterion	Algorithm Used	Accuracy
Proposed Model	RF Classifier	97.02%
Kumar et al, 2019 [30]	RF Classifier	72.74%
	SVM	63.33%
	GNB	50.78%
Rahman et al, 2018 [33]	Gaussian SVM	94.95%

are further demonstrated in table 12, and compared with similar implementations in table 13.

Similar implementations of RF Classifiers in other papers (like [30]) achieve an accuracy of 72% whilst predicting soil fertility, implying that the proposed model can outperform existing implementations whilst introducing a layer of transparency often missing from other models. Also, due to the vast preprocessing involved and the self-generation of labels, the proposed model may also have less bias in the data compared to other implementations, that use human defined boundaries as constraints for labels. Figure 8 shows confusion matrices for each metric, generated on a sample size of 3797 datapoints.

The model can classify a sample of soil into a relative soil fertility category, and explain the reasoning behind the classification of the soil in that category. This has tremendous uses in understanding the relationships between the concentrations of a soil's physiochemical properties and the relative soil fertility. To apply our understanding of these relationships, further work must be done to predict the rating from a short term and long-term perspective. This would be useful for farmers to understand both the short-term implications and long-term implications (in years, for example) of using a particular fertilizer in terms of soil fertility rating. A future application could potentially be used in the real world as a soil

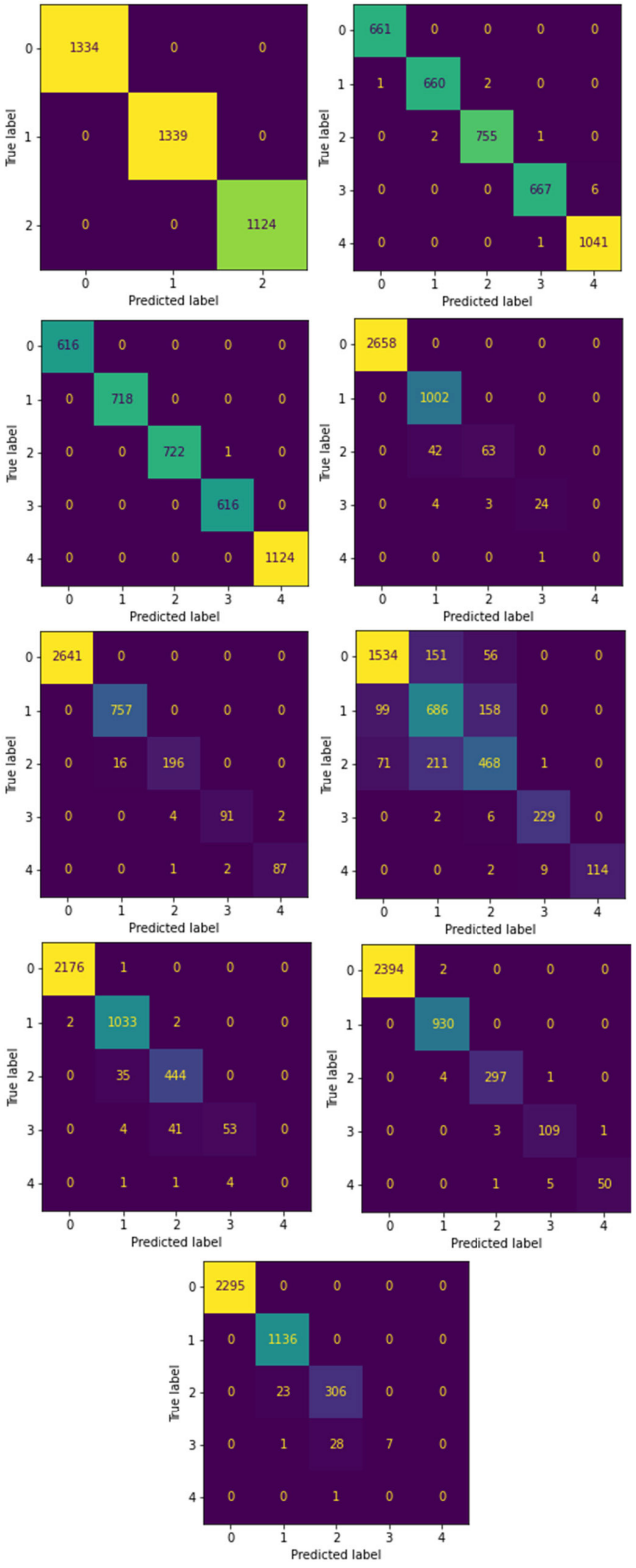

FIGURE 8. Confusion Matrix for (a) Relative Soil Fertility, (b) pH_H2O, (c) pH_CaCl2, (d) EC, (e) OC, (f) CaCO3, (g) P, (h) N, (i) K.

fertility improvement guide, which could improve yields of certain crops over time. Figure 9 demonstrates such a use case in a real-world scenario.

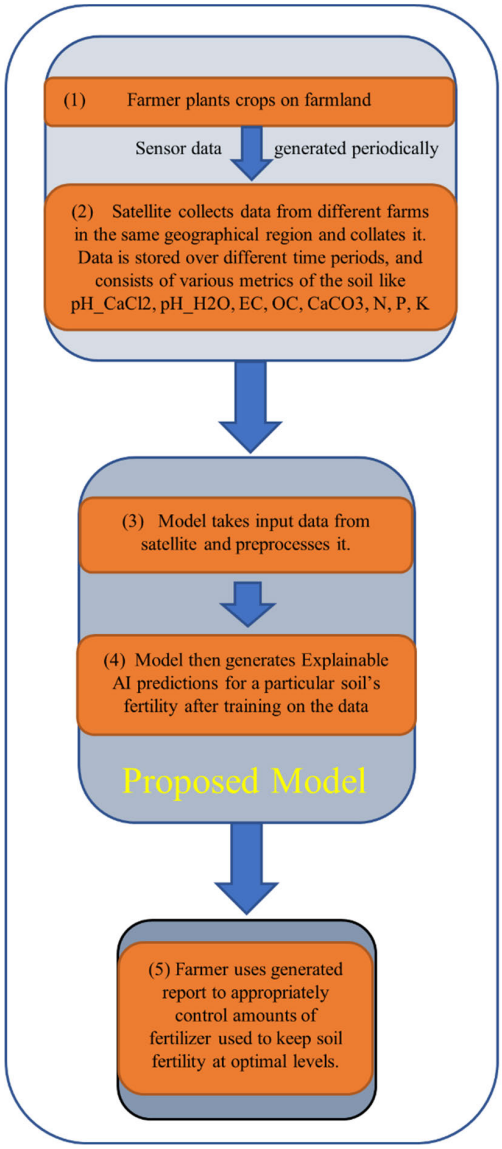

FIGURE 9. A real world scenario for the model.

V. CONCLUSION

Soil fertility is a crucial factor in determining the quality and quantity of crops produced. As agriculture continues to play a vital role in feeding the world's population, it is essential to understand and address issues related to soil fertility. With the advancements in technology, the use of Explainable AI (XAI) based models, like the proposed model, can aid in assessing and identifying reasons for variations in soil fertility over time. In this study, we investigated the application of this model on real-world data collected in the European Union for predicting relative soil fertility. Additionally, we analyzed the factors contributing to specific levels of soil fertility. To enhance interpretability, we presented the results using user-friendly graphs, which demystify the functioning of the model. Such models can assist farmers in understanding soil deficiencies and implementing sustainable solutions to improve fertility and ultimately optimize crop yields. It is

imperative for further research in this field to be conducted to fully harness the potential of these models in improving global food security.

ACKNOWLEDGMENT

The LUCAS 2018 topsoil dataset used in this work was made available by the European Commission through the European Soil Data Centre managed by the Joint Research Centre (JRC).

REFERENCES

[1] J. Gowdy, "Our hunter-gatherer future: Climate change, agriculture and uncivilization," *Futures*, vol. 115, Jan. 2020, Art. no. 102488.
[2] J. Timsina, J. Wolf, N. Guilpart, L. G. J. van Bussel, P. Grassini, J. van Wart, A. Hossain, H. Rashid, S. Islam, and M. K. van Ittersum, "Can Bangladesh produce enough cereals to meet future demand?" *Agricult. Syst.*, vol. 163, pp. 36–44, Jun. 2018.
[3] A. E. Hartemink, "Assessing soil fertility decline in the tropics using soil chemical data," *Adv. Agronomy*, vol. 89, pp. 179–225, Jan. 2006.
[4] Food and Agriculture Organization of the United Nations. (2020). *Soil Fertility*. Global Soil Partnership. [Online]. Available: http://www.fao.org/global-soil-partnership/areas-of-work/soil-fertility/en
[5] F. Wang et al., "Technologies and perspectives for achieving carbon neutrality," *Innovation*, vol. 2, no. 4, Nov. 2021, Art. no. 100180.
[6] H. D. Foth and B. G. Ellis, *Soil Fertility*. Boca Raton, FL, USA: CRC Press, 2018.
[7] M. Koch, M. Naumann, E. Pawelzik, A. Gransee, and H. Thiel, "The importance of nutrient management for potato production part I: Plant nutrition and yield," *Potato Res.*, vol. 63, no. 1, pp. 97–119, Mar. 2020.
[8] *Transforming Our World: The 2030 Agenda for Sustainable Development*, Department of Economic and Social Affairs, New York, NY, USA, 2015.
[9] J. J. Stoorvogel, M. Bakkenes, A. J. A. M. Temme, N. H. Batjes, and B. J. E. T. Brink, "S-World: A global soil map for environmental modelling," *Land Degrad. Develop.*, vol. 28, no. 1, pp. 22–33, Jan. 2017.
[10] S. Dasgupta, S. Chakraborty, D. C. Weindorf, B. Li, S. H. G. Silva, and K. Bhattacharyya, "Influence of auxiliary soil variables to improve PXRF-based soil fertility evaluation in India," *Geoderma Regional*, vol. 30, Sep. 2022, Art. no. e00557.
[11] H. Tiessen, E. Cuevas, and P. Chacon, "The role of soil organic matter in sustaining soil fertility," *Nature*, vol. 371, no. 6500, pp. 783–785, Oct. 1994.
[12] C. Lefèvre, F. Rekik, V. Alcantara, and L. Wiese, *Soil Organic Carbon: The Hidden Potential*. Rome, Italy: Food and Agriculture Organization of the United Nations (FAO), 2017.
[13] H. Liu, J. Zhang, Z. Ai, Y. Wu, H. Xu, Q. Li, S. Xue, and G. Liu, "16-year fertilization changes the dynamics of soil oxidizable organic carbon fractions and the stability of soil organic carbon in soybean-corn agroecosystem," *Agric. Ecosyst. Environ.*, vol. 265, pp. 320–330, Oct. 2018.
[14] D. Neina, "The role of soil pH in plant nutrition and soil remediation," *Appl. Environ. Soil Sci.*, vol. 2019, Nov. 2019, Art. no. 5794869.
[15] O. Dikinya and N. Mufwanzala, "Chicken manure-enhanced soil fertility and productivity: Effects of application rates," *J. Soil Sci. Environ. Manag.*, vol. 1, no. 3, pp. 46–54, 2010.
[16] M. Kumar Bhatt, R. Labanya, and H. C. Joshi, "Influence of long-term chemical fertilizers and organic manures on soil fertility—A review," *Universal J. Agricult. Res.*, vol. 7, no. 5, pp. 177–188, Sep. 2019.
[17] E. Adams and R. Shin, "Transport, signaling, and homeostasis of potassium and sodium in plants," *J. Integrative Plant Biol.*, vol. 56, no. 3, pp. 231–249, Mar. 2014.
[18] N. K. Fageria, *The Use of Nutrients in Crop Plants*. Boca Raton, FL, USA: CRC Press, 2016.
[19] C. Syndor and B. Thompson. *Secondary Nutrients—Nutrient management*. Mosaic. [Online]. Available: http://www.cropnutrition.com/nutrient-management/secondary-nutrients
[20] F. Shireen, M. Nawaz, C. Chen, Q. Zhang, Z. Zheng, H. Sohail, J. Sun, H. Cao, Y. Huang, and Z. Bie, "Boron: Functions and approaches to enhance its availability in plants for sustainable agriculture," *Int. J. Mol. Sci.*, vol. 19, no. 7, p. 1856, Jun. 2018.
[21] O. P. Narayan, P. Kumar, B. Yadav, M. Dua, and A. K. Johri, "Sulfur nutrition and its role in plant growth and development," *Plant Signaling Behav.*, vol. 17, Feb. 2022, Art. no. 2030082.
[22] T. R. Tavares, J. P. Molin, L. C. Nunes, M. C. F. Wei, F. J. Krug, H. W. P. de Carvalho, and A. M. Mouazen, "Multi-sensor approach for tropical soil fertility analysis: Comparison of individual and combined performance of VNIR, XRF, and LIBS spectroscopies," *Agronomy*, vol. 11, no. 6, p. 1028, May 2021.
[23] Y. Rustamov, T. Gadjiev, and S. Askerova, "A mathematical model of soil fertility," in *Proc. 14th Int. Conf. Manage. Sci. Eng. Manage.*, vol. 1. Cham, Switzerland: Springer, 2020, pp. 503–510.
[24] M. Kalimuthu, P. Vaishnavi, and M. Kishore, "Crop prediction using machine learning," in *Proc. 3rd Int. Conf. Smart Syst. Inventive Technol. (ICSSIT)*, Aug. 2020, pp. 926–932.
[25] P. Charoen-Ung and P. Mittrapiyanuruk, "Sugarcane yield grade prediction using random forest with forward feature selection and hyper-parameter tuning," in *Proc. 14th Int. Conf. Comput. Inf. Technol. (IC2IT)*. Cham, Switzerland: Springer, 2019, pp. 33–42.
[26] F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, and J. Zhu, "Explainable AI: A brief survey on history, research areas, approaches and challenges," in *Proc. 8th CCF Int. Conf. Natural Lang. Process. Chin. Comput. (NLPCC)*. Dunhuang, China: Springer, Oct. 2019, pp. 563–574.
[27] A. Das and P. Rad, "Opportunities and challenges in Explainable Artificial Intelligence (XAI): A survey," 2020, *arXiv:2006.11371*.
[28] O. Loyola-González, "Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view," *IEEE Access*, vol. 7, pp. 154096–154113, 2019.
[29] G. A. Helfer, J. L. V. Barbosa, R. dos Santos, and A. B. da Costa, "A computational model for soil fertility prediction in ubiquitous agriculture," *Comput. Electron. Agric.*, vol. 175, Aug. 2020, Art. no. 105602.
[30] T. G. K. Kumar, C. Shubha, and S. A. Sushma, "Random forest algorithm for soil fertility prediction and grading using machine learning," *Int. J. Innov. Technol. Exploring Eng.*, vol. 9, no. 1, pp. 1301–1304, Nov. 2019.
[31] L. Benedet, S. F. Acuña-Guzman, W. M. Faria, S. H. G. Silva, M. Mancini, A. F. dos Santos Teixeira, L. M. P. Pierangeli, F. W. A. Júnior, L. R. Gomide, A. L. P. Júnior, I. A. de Souza, M. D. de Menezes, J. J. Marques, L. R. G. Guilherme, and N. Curi, "Rapid soil fertility prediction using X-ray fluorescence data and machine learning algorithms," *Catena*, vol. 197, Feb. 2021, Art. no. 105003.
[32] 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," *Comput. Electron. Agric.*, vol. 154, pp. 120–133, Nov. 2018.
[33] S. A. Z. Rahman, K. C. Mitra, and S. M. M. Islam, "Soil classification using machine learning methods and crop suggestion based on soil series," in *Proc. 21st Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2018, pp. 1–4.
[34] O. Fernandez-Ugalde, S. Scarpa, A. Orgiazzi, P. Panagos, M. Van Liedekerke, A. Marechal, and A. Jones, *LUCAS 2018 Soil Module*, document EUR 31144, Presentation of Dataset and Results, European Commission, Brussels, Belgium, 2022.
[35] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesthesia Analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018.
[36] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A *K*-means clustering algorithm," *J. Roy. Stat. Soc. C*, vol. 28, no. 1, pp. 100–108, Jan. 1979.
[37] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "*Scikit-learn*: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 10, pp. 2825–2830, Jul. 2017.
[39] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020.
[40] European Commission. (2020). *Lucas 2018 Topsoil Data ESDAC*. [Online]. Available: https://esdac.jrc.ec.europa.eu/content/lucas-2018-topsoil-data
[41] P. Gupta, S. Maji, and R. Mehra, "Predictive modeling of stress in the healthcare industry during COVID-19: A novel approach using XGBoost, SHAP values, and tree explainer," *Int. J. Decis. Support Syst. Technol.*, vol. 15, no. 1, pp. 1–20, Dec. 2022.

[42] *Soil Health Data—Andhra Pradesh Primary Sector Mission*, vol. 3, ICRISAT Dataverse, ICRISAT Develop. Center Government Andhra Pradesh, Andhra Pradesh, India, 2016, doi: 10.21421/D2/K3BPKW.
[43] Government of Bangladesh and Soil Resource Development Institute (SRDI). Accessed: Jul. 18, 2023. [Online]. Available: http://www.srdi.gov.bd/
[44] Government of India, Open Government Data Platform INDIA. Accessed: Jul. 19, 2023. [Online]. Available: https://data.gov.in/

HARSHIV CHANDRA (Student Member, IEEE) is currently pursuing the B.E. degree in computer science engineering with the Birla Institute of Technology and Science, Pilani, Dubai Campus, Dubai, United Arab Emirates. He was a Research Intern with the Indian Institute of Technology, Roorkee, India. His research interests include machine learning, deep learning, VLSI design, and quantum computing. His awards include the first place from the Annual STEM Programming Competition (Abu Dhabi University) and the Wildcard Finalist (Tata Crucible).

PRANAV M. PAWAR (Member, IEEE) received the degree in computer engineering from Dr. Babasaheb Ambedkar Technological University, Maharashtra, India, in 2005, the master's degree in computer engineering from Pune University, in 2007, and the Ph.D. degree in wireless communication from Aalborg University, Denmark, in 2016. He is an IBM DB2 and an IBM RAD certified professional and completed NPTEL certification in different subjects. From 2006 to 2007, he was a System Executive with POS-IPC, Pune, India. He was an Associate Professor with the Department of Information Technology, STES's Smt. Kashibai Navale College of Engineering, Pune, from 2008 to 2018; and MIT ADT University, Pune, from 2018 to 2019. He is currently an Assistant Professor with the Department of Computer Science, Birla Institute of Technology and Science (BITS), Dubai. Before joining BITS, he was a Postdoctoral Fellow with Bar-Ilan University, Israel, from March 2019 to October 2020, in the areas of wireless communication and deep leaning. He received the Recognition from Infosys Technologies Ltd., for contribution in Campus Connect Program and different funding for research and attending conferences at international level. He has published more than 40 papers at national and international levels. His research interests include energy-efficient MAC for WSN, QoS in WSN, wireless security, green technology, computer architecture, database management systems, and bioinformatics. His Ph.D. thesis received nomination for the Best Thesis Award from Aalborg University. He was a recipient of the Outstanding Postdoctoral Fellowship from the Israel Planning and Budgeting Committee.

ELAKKIYA R. received the Ph.D. degree from Anna University, Chennai, in 2018. Her Ph.D. research was focused on sign language recognition. When she is not coding or in her research work, she is most probably cubing. She is currently an Assistant Professor with the Department of Computer Science, Birla Institute of Technology and Science, Pilani, Dubai Campus. During 2020 pandemic, she developed an artificial intelligence-based COVID screening tool for preliminary screening and deployed the tool as open source in three government hospitals at Tamil Nadu, India. She owns three patents and has published two books and more than 50 research articles in reputable journal venues, including IEEE, Elsevier, and Springer. She received many extra-mural funded projects from various government and non-government agencies, served as a machine learning and data analytics consultant, and delivered many products to different industry verticals. Her research interests include addressing the trending issues and huge need for ML and DL, by filling a gap within a multidisciplinary field to include computer science, mathematics, engineering, and alike application domains. She is a member of the Association of Computing Machinery and a life-time member of the International Association of Engineers. She secured the University First Rank and awarded the Gold Medal for her postgraduation in software engineering from CEG Campus, Anna University. She won the iDEX - DISC 4 challenge and received the Grant Award, in 2021, from DIO and DRDO; and the Young Achiever Award, in 2019, from INSc. She is an Editor of *Information Engineering and Applied Computing* journal.

TAMIZHARASAN P S (Member, IEEE) received the master's degree from Anna University, Chennai, and the Ph.D. degree in computer science and engineering from the National Institute of Technology, Tiruchirappalli, India. He is currently an Assistant Professor with the Department of Computer Science, BITS Pilani, Dubai Campus. He has published papers in reputed journals and conferences, such as IEEE Access, *Journal of Artificial Intelligence*, *The Journal of Supercomputing*, and ACM Sensys conference. His research interests include high-performance computing, deep learning, and edge inference.

RAJA MUTHALAGU received the B.E. and M.E. degrees in electronics and communication engineering from Anna University, Chennai, in 2005 and 2007, respectively, and the Ph.D. degree in wireless communication from the National Institute of Technology (NIT), Tiruchirappalli, India, in 2014. He was a Postdoctoral Research Fellow with the Air Traffic Management Research Institute, Nanyang Technological University, Singapore, from 2014 to 2015. He is currently an Associate Professor with the Birla Institute of Technology and Science, Pilani, Dubai Campus, Dubai, United Arab Emirates. He has published more than 55 research articles in reputed journals and conferences. His current research interests include wireless communications, signal processing, aeronautical communications, cyber security, applying intelligent techniques for detecting and mitigating a security attack in the IoT, SDN, and other computer networks. He was a recipient of the Canadian Commonwealth Scholarship Award 2010 for Graduate Student Exchange Program from the Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK, Canada.

ALAVIKUNHU PANTHAKKAN (Senior Member, IEEE) received the Ph.D. degree in electronics engineering. Overall, he is a highly respected and accomplished Research Scientist of AI-based image signal processing, with a passion for advancing the state-of-the-art in his field through innovative research, teaching endeavors, and mentorship contributions. He is currently an esteemed Research Scientist of artificial intelligence-based image signal processing. His contributions to the scientific community are significant, with more than 50 publications featured in renowned international journals and conference proceedings. Additionally, he has authored a printed book titled *Robust and Fragile Watermarking Techniques*, published by LAMBERT Academic Publishing. His research encompasses a broad spectrum, focusing on the development of groundbreaking algorithms and models for image processing, computer vision, and machine learning. His work finds particular applications in the domains of medical imaging and remote sensing. Dedicated to fostering progress and collaboration, he actively participates in the research community. He serves on program committees for various conferences in the fields of computer vision and machine learning. He holds membership in esteemed professional organizations, including the Institute of Electrical and Electronic Engineers (IEEE), Institution of Engineers (India) (MIE), International Association of Engineers (IAENG), International Association of Computer and Information Technology (IACSIT), and the Institute of Research Engineers and Doctors (IRED).