International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(22s) 2026 pp. 809-821
https://doi.org/10.70917/ijcisim-2026-5478
Article

Development of Accurate Soil Profile Prediction for the Agriculture Sector using Machine Learning

Dr. P. Soubhagyalakshmi[1], Savitha H. P[2], Sandhya S[3], Shriya Sinha[4]

[1]Associate Professor, Department of AIML, The Oxford College of Engineering, Bangalore-560068, KA, India.
[2,3,4]Assistant Professor, Department of AIML, The Oxford College of Engineering, Bangalore-560068, KA, India.
*Corresponding Email: Slakshmi.P@Gmail.Com

Abstract: India is one of the largest agricultural countries in the world. It contributes major production of pulses and spices and is considered one of the powerhouses of agriculture. However, the agriculture sector is facing the challenges such as soil degradation, low productivity, variations in the monsoons, lack of agriculture infrastructure, water scarcity and irrigation. There are several methods to find soil profile prediction such as field observation, soil sampling analysis, Pedogenesis model, remote sensing, and machine learning algorithms. Among these model, machine learning algorithm is one of the emerging trends because of scalability, adaptability, versatility, robustness and continuous improvement. In this regard, an Improved Regression by Discretization Algorithm (IRDA) is used to predict the soil profile of an agriculture unit for the better cultivation. The accurate value of PH, presence of essential nutrients , and the nature of soil are determined using IRDA algorithms. The performance of IRDA algorithm is compared with other popular machine learning algorithms using cover coefficient, mean absolute error, and root relative square errors.
Keywords: Cover coefficient, IRDA algorithm, Machine Learning, Mean Absolute Error and Root Relative Square Error

1. Introduction

India's agricultural sector boasts a rich and storied history, tracing its roots back an impressive ten thousand years. Today, it stands as the globe's second-largest producer of agricultural goods, playing a pivotal role in the nation's GDP and providing employment for nearly half of its workforce. Although its contribution to the GDP has experienced a gradual decline over time, agriculture remains the cornerstone of India's socioeconomic progress, underscoring its enduring importance in the country's developmental landscape. With projections indicating that India's population will exceed 1.5 billion by 2030 [1], the demand for resources, particularly food, has surged substantially. Despite over half of India's populace being engaged in agricultural activities, the sector confronts numerous hurdles. These challenges range from unsustainable farming practices and inadequate adoption of technology to constrained market accessibility and the looming threat of insufficient food production to meet the needs of the expanding population. Soil profile prediction is of utmost importance for the agricultural sector as it enables farmers to make informed decisions and implement sustainable farming practices. By studying the composition and characteristics of soil at various depths, farmers can effectively optimize their crop selection, irrigation methods, fertilization strategies, and soil management techniques. Soil profile prediction is crucial in agriculture. It helps farmers choose the right crops, manage water, maintain soil fertility, prevent degradation, and implement precise farming techniques. Effective soil profile prediction supports sustainable farming practices by promoting soil health and biodiversity.

Predicting soil profiles involves looking at soil in the field, testing it in labs, and sometimes using models. Soil scientists dig holes to see soil layers and note details like color, texture, and depth. This helps them understand soil better. Soil can also be put into categories based on its properties, like how the USDA does it. Scientists take soil samples from different depths and check them for things like texture and pH. This gives them more clues about what the soil profile might look like. Models can also help, like ones that show how soil forms over time. Tools such as ground-penetrating radar let scientists see below the surface. Satellite images help map soil properties over large areas. Geographic information systems (GIS) [2] help bring together different kinds of data for better predictions. Experts use their experience and what they see in the field to add to predictions. Advanced methods like Machine Learning (ML) [3] analyze big sets of data to predict soil profiles better. The ML techniques used for soil profile prediction is shown in Fig. 1. Machine learning algorithms analyze soil data to predict profiles more accurately, allowing farmers to make better decisions about crop selection, irrigation, and fertilization. This accuracy helps optimize crop management, reduces input waste, and improves productivity. Machine learning predictions also aid in identifying areas at risk of degradation, erosion, or nutrient depletion, enabling farmers to mitigate these risks. Additionally, machine learning algorithms provide real-time monitoring of soil conditions, empowering farmers to make timely adjustments to their agricultural practices, resulting in better outcomes.

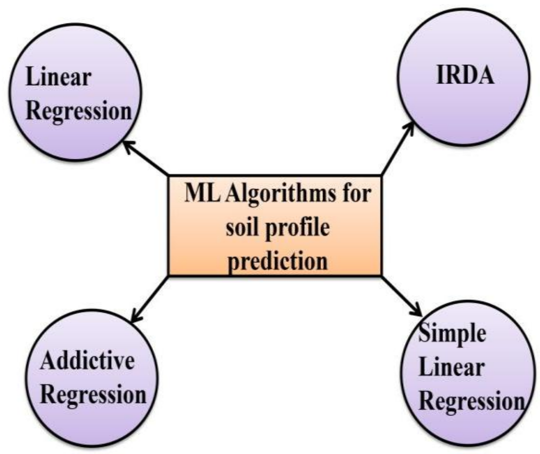

Fig. 1.ML algorithms for soil profile prediction

There are several ML algorithms such as linear regression, Simple linear regression, additive regression, and IRDA are used for the agricultural application like soil profile prediction. In this research, the implementation of IRDA in the soil profile prediction is illustrated with justification. The entire research paper is divided as follows: Section II describes extensive literature review; Section III elaborates detailed methodology, Section IV presents the results and discussion, and Section V concludes the paper.

2. Literature Review

2.1 Artificial Neural Network based soil profile prediction

S. Selvalakshmi *et al*, utilized Artificial Neural Networks (ANNs) [4] to forecast Soil Profiles across five regions of Sudan, drawing upon information gathered from 417 sites. Authors developed ten models for each zone, integrating depth and coordinate data as inputs, while soil classification and parameters served as outputs. Their investigation affirmed that ANNs serve as valuable decision support mechanisms for soil profiles, particularly in localized applications.

In their research, investigated the efficacy of Artificial Neural Network in discerning soil nutrient levels and pH with remarkable efficiency. To acquire essential data, authors utilized the Soil Test Kit (STK) and Rapid Soil

Testing (RST) provided by the Bureau of Soils and Water Management. Employing an artificial neural network approach, the researchers expedited the process of delivering precise outcomes.

Y. Liu *et al.* demonstrated ingenuity in employing empirical methods to evaluate historical rainfall data and its correlation with diverse atmospheric variables across different regions of the nation. Commonly employed techniques for weather forecasting encompass regression, artificial neural networks, fuzzy logic, and group statistics [5]. Furthermore, the authors harnessed advanced data mining techniques such as clustering and classification to enhance rainfall prediction.

2.2 Soil prediction using Decision Tree techniques

J. S. Parra *et al*, have innovated a system designed to aid farmers by furnishing them with pertinent information regarding their soil conditions. Understanding the soil's nutrient composition holds paramount importance in optimizing crop yield. Soil toxicity poses a significant threat to soil nutrients and crop vitality. The system anticipates soil toxicity levels and duly notifies farmers. Given the substantial reliance of many farmers on rainfall for agricultural activities, the system further offers recommendations regarding suitable crops, soil fertility levels, and water supply based on prevailing soil conditions. The precision of both the sensor and the classification algorithm [6] employed in the recommendation system is crucial. To circumvent challenges associated with power supply, the integration of a solar panel system stands as a viable solution.

Y. Wang *et al.*, employ data mining [7] classification techniques to forecast soil types. Various algorithms like JRip, J48, and Naive Bayes are utilized for this purpose. These classifier algorithms are employed to glean insights from soil data, with a focus on two main types: Red and Black soils. Among these algorithms, JRip stands out for its superior performance in accurately classifying a maximum number of instances within the soil dataset. Consequently, the authors advocate for the adoption of JRip as the preferred method for predicting soil types.

2.3 Additioanl methods to predict soil profile

J. Bu *et al.*, delved into an irrigation technique reliant on soil moisture content, incorporating a flow sensor. This system encompasses a closed-loop automatic irrigation mechanism [8], coupled with temperature and water usage monitoring. Users can conveniently establish desired moisture thresholds and receive real-time updates on all parameter values via an LCD display. P. Kotak *et al.*, employed data mining methodologies within the domain of soil science to uncover valuable correlations among different soil types. They meticulously curated a subset of soils from the World Soil Science Database for their investigation. Their methodology entailed the application of a clustering technique utilizing the WEKA tool [9] to categorize soils according to their salinity levels. Through this process, they garnered insights into the varying degrees of salinity across different soil types.

A. Tabatabaeenejad *et al.*, introduced a novel retrieval algorithm aimed at estimating Root Zone Soil Moisture (RZSM) [10]. This algorithm leverages reflectometry signals at L-band, P-band, and their combinations to achieve precise RZSM estimates. Initially, they employ an RZSM profile scattering model incorporating both coherent and incoherent scattering components. To address this, their team previously devised numerical and approximate analytical techniques. Subsequently, they establish a link between polarimetric scattering cross-sections and circularly polarized DDM observations [11]. The subsequent phase involves solving the inverse problem for surface soil moisture and RZSM utilizing single frequencies (L- and P-band) and their combinations within a reasonable observation timeframe. To tackle this, they utilize a robust inversion method developed within their research group, employing a hybrid of global and local optimization approaches. Following this, they conduct an exhaustive sensitivity analysis to assess the efficacy of individual frequencies and their combinations for soil moisture inversion at both surface and root zone levels. They then demonstrate the application of this methodology to actual Signals-of-Opportunity (SoOp) data [12], to the extent feasible. Furthermore, they advocate for the combined utilization of GNSS and MUOS for RZSM retrieval based on their findings.

3. Methodology

Precision agriculture employs autonomous processes to gather and analyze data, aiming to enhance farming practices. This includes employing data mining techniques to extract novel insights. While soil sampling offers valuable insights for agronomic decisions, it can prove costly and time-intensive. In instances where field conditions mirror those of neighboring fields, conducting soil tests for all fields may be redundant. Utilizing data mining techniques allows for the examination of various soil profiles and the monitoring of factors impacting crop yield. These methods prove valuable in classifying soil data, forecasting soil maps, and identifying soil salinity. Leveraging data from neighboring fields enables the prediction of soil characteristics and the completion of missing values in soil profiles, thereby mitigating soil sampling expenses. The pH value of the soil plays a major role in selecting crops for cultivation. Some crops, such as pineapple, blueberries, and potatoes, thrive in acidic soil, while others like cabbage, garlic, spinach, and lettuce are best suited to basic soil. Soil testing is carried out by a private laboratory located in Palakkad, Kerala. The analysis includes important elements such as Nitrogen, Phosphorus, Potassium, Iron, Zinc, organic carbon, Manganese, copper, electrical conductivity, and PH values. These attributes are taken into consideration during the data collection process. The IRDA technique utilizes historical data collected from soil testing centers to identify soil profile features.

3.1 IRDA Algorithm

IRDA comes under regression method which finds the relation between dependent variable and independent variable. The main setbacks of conventional regression methods are multicolinearity, limited flexibility, over lifting or under lifting, difficulty with missing data, and difficulty in deal with large scale data. To overcome those difficulties, IRDA algorithm is proposed for the application of soil profile prediction. The steps involved in the IRDA algorithm is shown in Fig.2

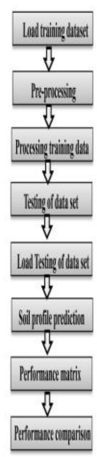

Fig. 2 Flow diagram of IRDA algorithm

The application of the IRDA algorithm on the Palakkad soil training dataset facilitates soil profile prediction. Initially, the dataset undergoes pre-processing to address any missing values. Following this, the algorithm enters the training phase, where it formulates rules for soil profile prediction. Once these rules are established, users can input values for a single record to predict the soil profile. Subsequently, the testing phase is implemented, utilizing the training rules to predict the soil profile for the entered record. Moreover, users have the option to input a testing dataset for Soil Profile Prediction, which undergoes the IRDA Testing phase. This phase utilizes the training rules to predict the soil profile for the entered testing dataset, enhancing the algorithm's predictive capabilities.

The algorithm for predicting soil profile begins by loading the training dataset (TrD) and pre-processing it to eliminate any missing values, which results in a refined dataset (PPTrD). In the training phase (TR), the algorithm employs the modified regression by discretization (MRD) to generate training rules (TR). Users can input single record values (SR) for soil profile prediction once the rules are established. These values undergo prediction using the training rules, resulting in soil profile prediction for a single record (SPP_SR).

The algorithm also allows users to load a testing dataset (TeD) for soil profile prediction [13]. The testing dataset is subjected to the MRD testing phase using the established training rules, yielding soil profile predictions for the entire dataset (SPP_TeD). The provided predictions include categories such as low, medium, and high, offering comprehensive insights into soil profile characteristics.

3.2 Soil profile prediction using IRDA algorithm

The Regression By IRDA is employed in soil profile prediction, leveraging a robust training dataset comprising 30,000 records. In its predictive process, the algorithm first loads the training dataset, conducting pre-processing to eliminate any instances of missing data. This meticulous preparation ensures the integrity and reliability of the subsequent model. Once the model is established, it undergoes rigorous testing using multiple records representing diverse soil types. Through this comprehensive testing, the algorithm validates its predictive capabilities across various soil profiles, demonstrating its efficacy and versatility in soil analysis [14].

Sample Data Set

CLS	B	Mn	Cu	Fe	Zn	S	K	P	N	OC	EC	PH
2	0.1	3.61	0.94	6.75	1.08	12.10	282.50	25	101.5	0.10	0.28	7.11
1	0.1	3.26	0.93	5.89	0.82	12.10	292	10	94.5	0.21	0.21	7.10
2	0.1	3.85	1.12	6.13	0.92	12.47	282.50	38.75	115.50	0.25	0.25	7.00
2	0.1	3.61	1.16	6.91	1.02	12.59	285.0	25	112	0.18	0.27	7.00
1	0.10	3.73	1.09	4.15	1.12	12.59	297.5	10	94.50	0.08	0.23	7.13

This meticulous preparation ensures the integrity and reliability of the subsequent model. Once the model is established, it undergoes rigorous testing using multiple records representing diverse soil types. Through this comprehensive testing, the algorithm validates its predictive capabilities across various soil profiles, demonstrating its efficacy and versatility in soil analysis. The sample data set is shown in the Table 1. Table 1 displays a concise dataset featuring data and attribute values obtained from a soil testing laboratory, specifically focusing on alkali soils. Soil attributes play a pivotal role in facilitating plant growth, with properties like soil pH, texture, moisture level, calcium carbonate content, and organic matter content exerting substantial influence on plant development. The details of after applying discretization are presented in the Table 2.

Data set after applying discretization

Sl.No	Data after applying discretization
	(63.92-71.86]
	(55.98-63.92]
	(48.04-55.98]
	(40.1-48.04]
	(32.16-40.1]

	(24.22-32.16]
	(16.28-24.22]
	(8.34-16.28]
	(-inf-8.34]

Discretization [15] is the process of converting continuous data into a finite set of intervals and specific data values. It simplifies the management and evaluation of data for machine learning algorithms that handle only discrete data. The table 3 depicts data conversion.

Data Conversion

(165.53-329.36]	(-inf-49.05]	(90.8-116.4]	(-inf-5.045]	(-inf-0.468]	(-inf-8.34]
'(-inf-0.816]	(2.96-4.39]	(0.702-1.038]	(4.852-7.228]'	(-inf-7.818]	(-inf-25.468]
(165.53-329.36]	(-inf-49.05]	(90.8-116.4]	(-inf-5.045]	(-inf-0.468]	(-inf-8.34]
'(-inf-0.816]	(2.96-4.39]	'(0.702-1.038]	(4.852-7.228]	(-inf-7.818]	(-inf-25.468]
(165.53-329.36]	(-inf-49.05]	(90.8-116.4]	(-inf-5.045]	(-inf-0.468]	(-inf-8.34]

The table 3 gives a clear picture on how the PH range changes. Once the data conversion is taken place, IRDA algorithm is applied. The results are obtained in terms of three parameters such as high, medium and low. The results are presented in Table IV. Table IV was updated by specifying the intervals as low, medium, and high instead of using infinite intervals. This categorization helped predict and compare results with other classifiers, leading to a detailed discussion of the outcomes.

Regression by IRDA

Sl.No	Parameters	Attributes
	>7.5 Alkaline	High
	6.5 to 7.5 Neutral	Medium
	<6.5 Acidic	Low

3.3 Performance validation of IRDA algorithm

In order to compare the performance of IRDA algorithm quantitatively, the parameters such as Mean Absolute Error (MAE), Cover Co-efficient (CC), Root Mean Square Error (RMSE), Root Absolute Error (RAE), and Root Relative Square Error (RRSE) are calculated for each algorithm [16]-[20]. The MAE is used to evaluate the performance of predictive model which can be evaluated using (1)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \quad |y_i' - y_i| \qquad (1)$$

In the equation (1), n is the number of observations, y' is the predicted value of observation i, yi is the actual value of observation i. MAE must be minimum for a robust algorithm. The Coverage Coefficient measures data dispersion around the mean. The CC is calculated using the equation (2)

$$CC = \frac{S\tan dard\ Deaviation}{Mean} \times 100 \qquad (2)$$

RMSE is a common metric in statistics and machine learning to evaluate a predictive model's accuracy. It calculates the average difference between predicted and actual values. RMSE is evaluated using (3)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y'\right)^2} \quad (3)$$

Root Absolute Error (RAE) is a metric to evaluate prediction accuracy. It measures the average difference between predicted and actual values, irrespective of error direction. RAE is calculated using the equation (4)

$$RAE = \sqrt{\frac{\sum_{i=1}^{n}\left|y_i - y'\right|}{\sum_{i=1}^{n}\left|y_i - \bar{y}\right|}} \quad (4)$$

Here, yi denotes actual value, y' represents predicted value, represents the mean of actual values, and n is the number of observations. RRSE is a metric to evaluate a model's prediction accuracy by comparing the RMSE of predicted values with that of a simple baseline model. RRSE is calculated using the equation (5)

$$RRSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y'\right)^2}{n}}}{\sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}{n}}} \quad (5)$$

4. Results and Discussion

The IRDA is implemented and validated using Waikato Environment for Knowledge Analysis (WEKA) and Net Beans [21]. The system configuration is Intel core i-5, 8 GB RAM, and 1.90 GHz . First, the data was divided into a training dataset comprising 80% of the data and a testing dataset containing the remaining 20%. Next, k-fold cross-validation was applied to the training dataset to ensure robust model evaluation. Hyperparameters were then selected based on their performance across the k validation sets. The best-performing model was trained using the entire training dataset. Subsequently, an unbiased performance estimate was obtained by evaluating the model on the testing dataset. To further refine the model, it was trained again using the complete dataset. Finally, the finalized model was utilized to make predictions on future data.

First, the dataset is checked for missing values, noisy data, and other inconsistencies to prepare it for the algorithm. This step is crucial for creating a machine learning model. Once the data is pre-processed, it's used to train and make predictions. The trained data set taken from the model is presented in the Fig.3

pH	EC	OC	N	P	K	S	Zn	Fe	Cu	Mn	B	Cls
7.11	0.28	0.10	101.50	25.00	282.50	12.10	1.08	6.75	0.94	3.61	0.10	2
7.10	0.21	0.21	94.50	10.00	290.00	12.10	0.82	5.89	0.93	3.26	0.10	1
7.00	0.25	0.25	115.50	38.75	282.50	12.47	0.92	6.13	1.12	3.85	0.10	2
7.00	0.27	0.18	112.00	[illegible]	285.00	12.59	1.02	6.91	1.16	3.61	0.10	2
7.13	0.23	0.08	94.50	10.00	[illegible]	12.59	1.12	4.15	1.09	3.73	0.10	1
7.10	0.20	0.06	223.00	[illegible]	[illegible]	12.59	1.02	5.49	1.14	2.60	0.10	1
7.15	0.21	0.06	108.50	[illegible]	[illegible]	12.47	1.07	6.44	1.11	3.92	0.10	2
7.17	0.18	0.15	140.00	10.00	260.00	12.34	1.07	6.98	1.08	4.07	0.10	1
6.83	0.21	0.33	133.00	5.00	285.00	12.34	1.15	5.66	1.01	2.46	0.10	1
6.67	0.20	0.36	129.50	5.00	277.50	12.47	1.06	6.66	1.14	4.28	0.10	1
6.89	0.23	0.08	105.00	5.00	275.00	12.34	1.14	5.39	0.96	3.15	0.10	1
6.90	0.22	0.12	105.00	10.00	265.00	12.22	1.14	4.68	1.13	3.59	0.10	1
6.75	0.25	0.09	98.00	5.00	247.50	12.22	1.01	5.40	0.93	3.75	0.10	1
6.86	0.25	0.09	112.00	5.00	275.00	12.22	0.87	6.23	1.11	3.57	0.10	1
7.00	0.23	0.42	112.00	10.00	280.00	12.34	0.89	5.82	0.95	3.70	0.10	1

. Fig.3 Trained data set

In the next phase, IRDA is applied accordingly and predicted the soil profile. The attributes used for the soil profile prediction is high, medium and low. The soil profile prediction is presented in the Fig.4

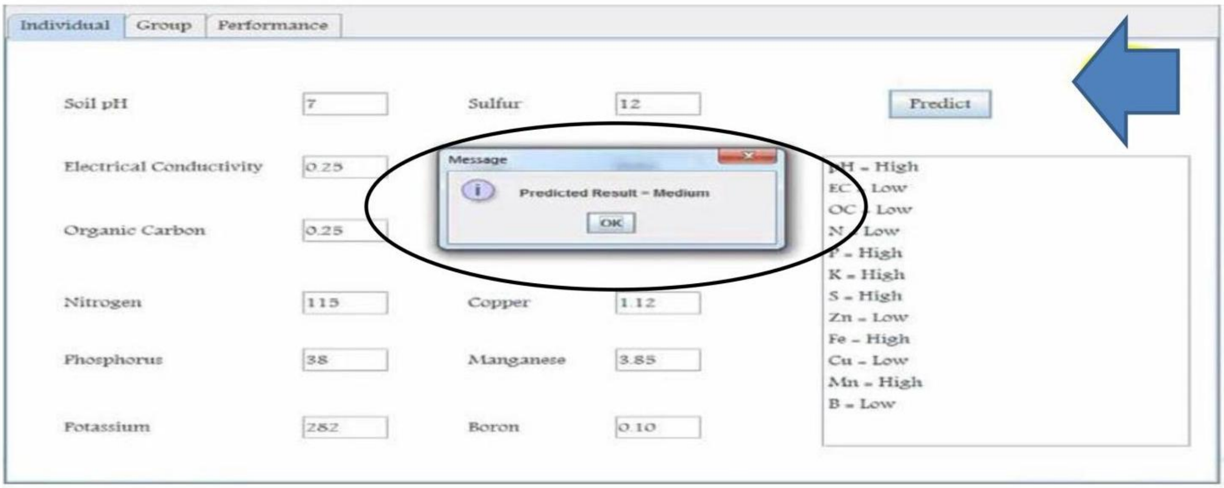

Fig.4 Soil profile prediction result

As observed in the Fig.4, only one type of soil is analysed in the individual tab, whereas more number of soil is compared in the group tab. The performance tab is used for comparing the performance of algorithms other than IRDA. The result obtained during the testing phase is shown in Fig.5. The IRDA algorithm carries out the IRDA Testing phase, predicting soil profiles for the testing dataset using training rules. The performance of the proposed system is then evaluated by comparing it with other ML algorithms such as Simple Linear Regression, Regression by Discretization, Linear Regression, and Additive Regression [22]-[25] to determine the best results.

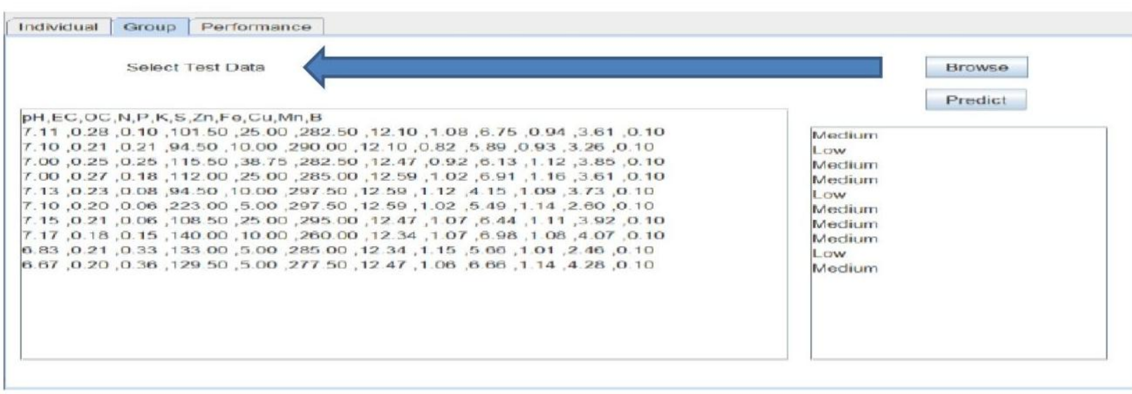

Fig.5 Testing phase of soil profile prediction

There are 30000 records in the data sets, which covers 80% Palakkad district, Kerala, India. The data is collected by government and non-governmental organizations across districts. The model is implemented using Waikato Environment for Knowledge Analysis. Initially, the performance of IRDA algorithm is verified in the environment. The results obtained using IRDA algorithm is compared with Linear Regression (LR), Simple Linear Regression (SLR), Regression by Discretization (RD), and Additive Regression (AR). It is desirable to compare the performance of selected ML algorithms using for accurate soil profile prediction.In this regard, performance of ML algorithms are compared using the quantitative techniques such as Mean Absolute Error (MAE), Cover Co-efficient (CC), Root Mean Square Error (RMSE), Root Absolute Error (RAE), and Root Relative Square Error (RRSE) [26-29]. The calculations are performed using the mathematical equations illustrated in the equations from (1) to (5). The comparative result using MAE is shown in Fig.6

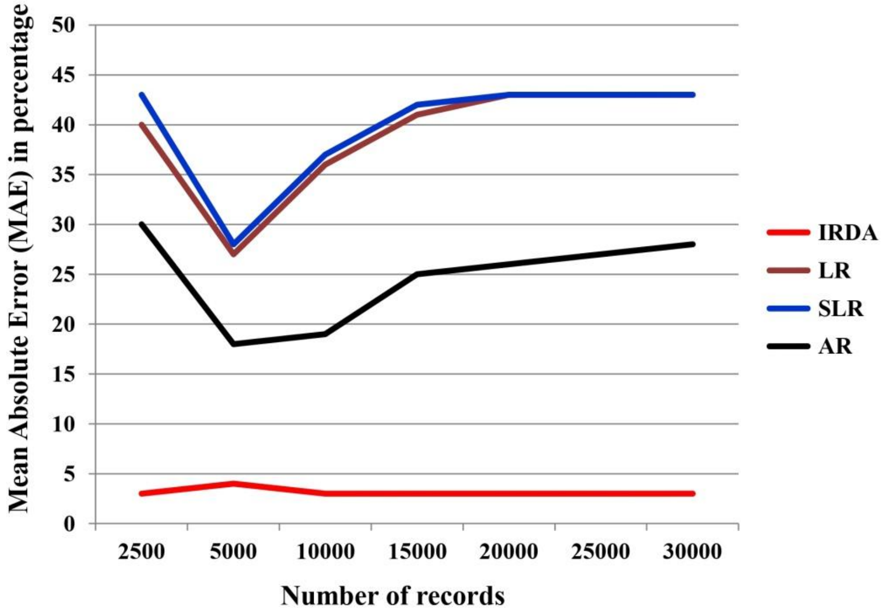

Fig.6 Comparison of ML algorithms using MAE

As compared to other ML algorithms, MAE of IRDA algorithm is less which enhance the superiority in accuracy. Both the LR and SLR shows MAE comparatively high where AR shows moderate MAE in their performance. The performance analysis of the ML algorithms using Root Mean Square Error (RMSE) is illustrated in Fig.7

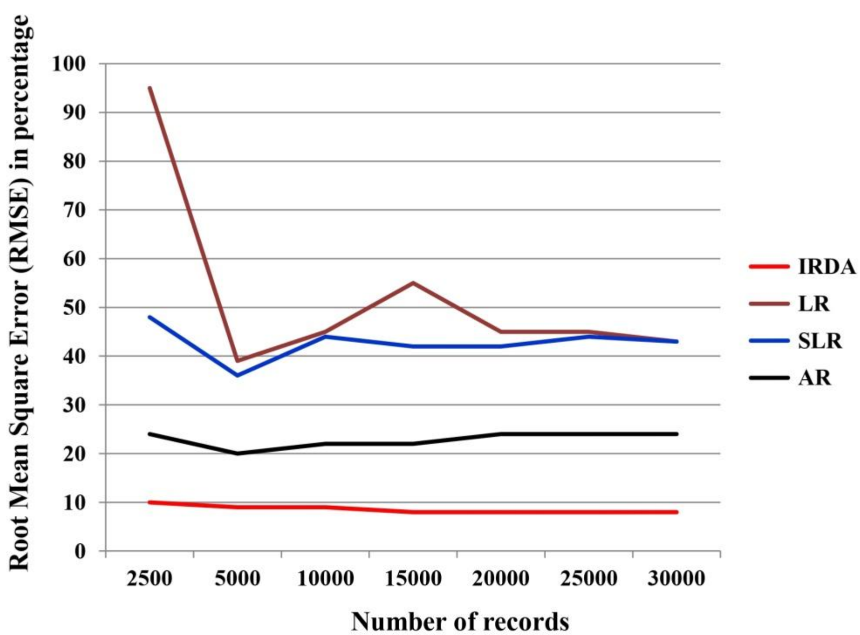

Fig.7 Comparison of ML algorithms using RMSE

It is observed that RMSE is highest in the case of LR algorithm whereas IRDA shows minimal value of RMSE.RMSE must be at the minimum for a convenient ML algorithm. The Root Absolute Error (RAE) is another important parameter for the performance comparison of ML algorithm. RAE should be as minimum as possible for a desirable algorithm. The RAE of ML algorithms is depicted in Fig.8

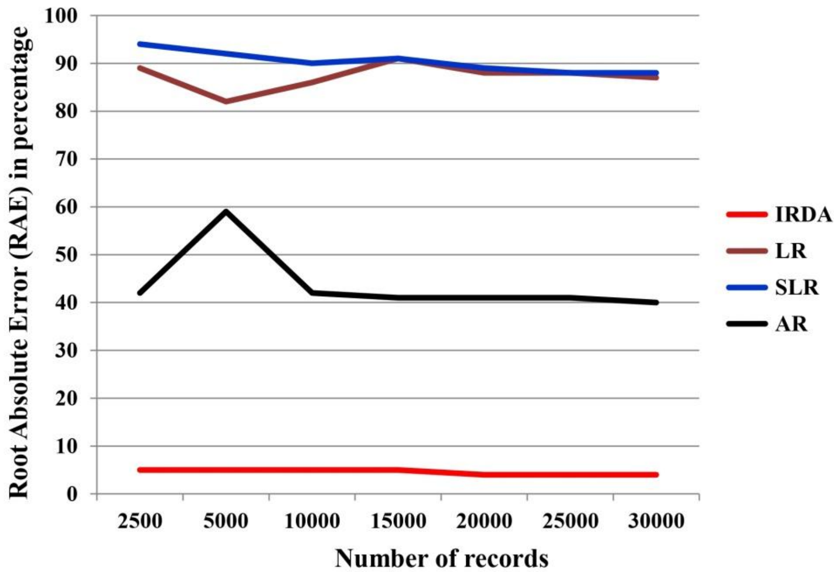

Fig.8 Comparison of ML algorithms using RAE

The Root Relative Squared Error (RRSE) is another important parameter for the performance comparison of ML algorithms used for specific application like soil profile prediction system. The RRSE of ML algorithms are illustrated in the Fig.9

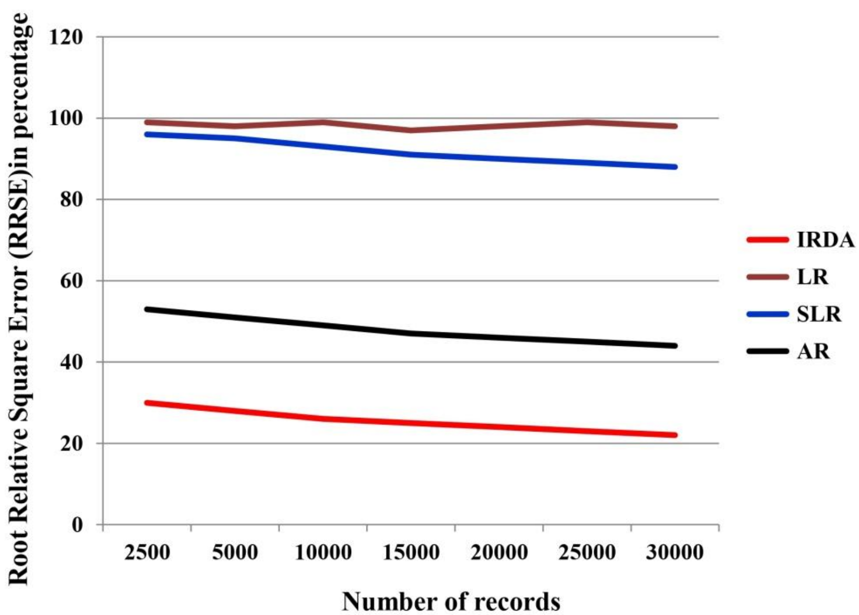

Fig.9 Comparison of ML algorithms using RRSE

It is found that RRSE is minimum in IRDA and highest in LR algorithm. Another important parameter is Cover Co-efficient (CC) which indicating the strength and direction of relationship with the target variable. The higher value of CC indicates deterministic output which is desirable for an efficient ML algorithm. The CC of ML algorithms selected for this research is shown in Fig.10

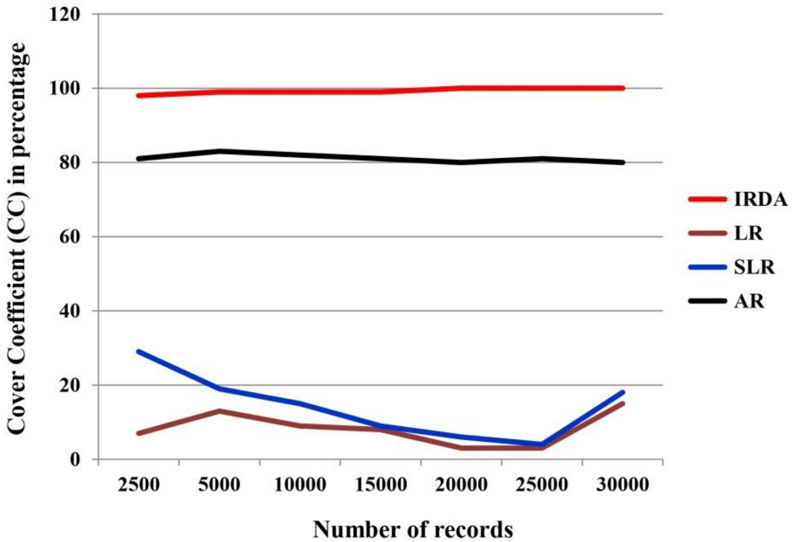

Fig.10 Comparison of ML algorithms using CC

The CC must be highest for a desirable ML algorithm. As per the analysis, CC is highest for IRDA algorithm which helps for accurate prediction and decision making. While considering LR and SLR, the value of CC is consistently poor and not admissible to all the circumstances. However, AR shows moderate performance in CC determination. It is found that IRDA is most convenient ML algorithm for accurate soil profile prediction.

5. Conclusions

The agriculture sector plays major role in the economy of India. The ML algorithms help to overcome a few challenges faced by the agriculture sector. Improved Regression by Discretization Algorithm (IRDA) is a powerful ML algorithm helps to predict the soil profile of an agriculture unit for the better cultivation. In this research work, IRDA is implemented using Waikato Environment for Knowledge Analysis (WEKA) and Net Beans by using 30,000 records in the data sets, which covers 80% Palakkad district, Kerala, India.The performance of IRDA algorithm is validated by comparing the performance of other ML algorithms such as Linear Regression (LR), Simple Linear Regression (SLR), and Additive Regression (AR). The important statistical parameters such as Mean Absolute Error (MAE), Cover Co-efficient (CC), Root Mean Square Error (RMSE), Root Absolute Error (RAE), and Root Relative Square Error (RRSE) are evaluated for each ML algorithm. It is found that IRDA algorithm has superior performance over other ML algorithm. The Cover Coefficient of IRDA is highest compared to other algorithm so that soil profile prediction becomes more accurate. Waikato Environment for Knowledge Analysis (WEKA) and Net Beans are the helpful tool to compare the performance of various ML algorithms. The accurate value of PH, presence of essential nutrients , and the nature of soil can be accurately determined using IRDA algorithm.

Acknowledgment

I express my sincere gratitude to my research guide, ---------, the management and Principal of -------, and ------ University for their support in completing this research work successfully.

REFERENCES

1. W. Purcell, T. Neubauer, and K. Mallinger, "Digital Twins in agriculture: challenges and opportunities for environmental sustainability," Current Opinion in Environmental Sustainability, vol. 61, pp. 101252, 2023. DOI: 10.1016/j.cosust.2022.101252.
2. M. Maurya, S. Dixit, N. Zaidi and A. Maurya, "Geographic Information Systems for an Intelligent Geo-Location Based Value Delivery Model," 2021 International Conference on Technological Advancements and Innovations (ICTAI), Tashkent, Uzbekistan, 2021, pp. 63-68, doi: 10.1109/ICTAI53825.2021.9673488.

3. K. S. Bohnsack, M. Kaden, J. Abel and T. Villmann, "Alignment-Free Sequence Comparison: A Systematic Survey From a Machine Learning Perspective," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 20, no. 1, pp. 119-135, 1 Jan.-Feb. 2023, doi: 10.1109/TCBB.2022.3140873.
4. S. Selvalakshmi, R. Immanual, B. Priyadharshini, and J. Sathya, "Artificial neural network (ANN) modelling for the thermal performance of bio fluids," in Materials Today: Proceedings, vol. 66, pt. 3, pp. 1289-1294, 2022.
5. Y. Liu, Q. Zhang, Z. -P. Fan, T. -H. You and L. -X. Wang, "Maintenance Spare Parts Demand Forecasting for Automobile 4S Shop Considering Weather Data," in IEEE Transactions on Fuzzy Systems, vol. 27, no. 5, pp. 943-955, May 2019, doi: 10.1109/TFUZZ.2018.2831637.
6. J. S. Parra, Z. M. de S. Souza, S. R. de M. Oliveira, C. V. V. Farhate, J. Marques, and D. Siqueira, "Phosphorus adsorption prediction through Decision Tree Algorithm under different topographic conditions in sugarcane fields," CATENA, vol. 213, p. 106114, 2022.
7. Y. Wang et al., "Hyperspectral Estimation of Soil Copper Concentration Based on Improved TabNet Model in the Eastern Junggar Coalfield," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-20, 2022, Art no. 5534020, doi: 10.1109/TGRS.2022.3190310.
8. J. Bu, K. Yu, S. Han, N. Qian, Y. Lin and J. Wang, "Retrieval of Sea Surface Rainfall Intensity Using Spaceborne GNSS-R Data," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-16, 2022, Art no. 5803116, doi: 10.1109/TGRS.2022.3168928.
9. P. Kotak and H. Modi, "Enhancing the Data Mining Tool WEKA," 2020 5th International Conference on Computing, Communication and Security (ICCCS), Patna, India, 2020, pp. 1-6, doi: 10.1109/ICCCS49678.2020.9276870.
10. A. Tabatabaeenejad et al., "Assessment and Validation of AirMOSS P-Band Root-Zone Soil Moisture Products," in IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 9, pp. 6181-6196, Sept. 2020, doi: 10.1109/TGRS.2020.2974976.
11. T. Beltramonte et al., "Simulation-Based Feasibility Analysis of Ship Detection Using GNSS-R Delay-Doppler Maps," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 1385-1399, 2020, doi: 10.1109/JSTARS.2020.2970221.
12. B. Hu et al., "DOA Robust Estimation of Echo Signals Based on Deep Learning Networks With Multiple Type Illuminators of Opportunity," in IEEE Access, vol. 8, pp. 14809-14819, 2020, doi: 10.1109/ACCESS.2020.2966653.
13. V. Nguyen, J. Li, D. Mishra and A. Seneviratne, "Frequency-Sensitive Soil Moisture Profiling Using WiFi Sensing," 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Gothenburg, Sweden, 2023, pp. 641-646, doi: 10.1109/EuCNC/6GSummit58263.2023.10188284.
14. K. Huang et al., "A Novel Accurate Insecticidal Counting Method Based on Solar Insecticidal Lamp Using Machine Learning," in IEEE Transactions on Consumer Electronics, vol. 69, no. 4, pp. 1045-1054, Nov. 2023, doi: 10.1109/TCE.2023.3297723.
15. J. Liu et al., "Estimating the Forage Neutral Detergent Fiber Content of Alpine Grassland in the Tibetan Plateau Using Hyperspectral Data and Machine Learning Algorithms," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-17, 2022, Art no. 4405017, doi: 10.1109/TGRS.2021.3105482.
16. D. Vimalajeewa, C. Kulatunga, D. P. Berry and S. Balasubramaniam, "A Service-Based Joint Model Used for Distributed Learning: Application for Smart Agriculture," in IEEE Transactions on Emerging Topics in Computing, vol. 10, no. 2, pp. 838-854, 1 April-June 2022, doi: 10.1109/TETC.2020.3048671.
17. N. Rezaei, M. N. Uddin, I. K. Amin, M. L. Othman, M. B. Marsadek and M. M. Hasan, "A Novel Hybrid Machine Learning Classifier-Based Digital Differential Protection Scheme for Intertie Zone of Large-Scale Centralized DFIG-Based Wind Farms," in IEEE Transactions on Industry Applications, vol. 56, no. 4, pp. 3453-3465, July-Aug. 2020, doi: 10.1109/TIA.2020.2990584.
18. T. Pamuklu, A. C. Nguyen, A. Syed, W. S. Kennedy and M. Erol-Kantarci, "IoT-Aerial Base Station Task Offloading With Risk-Sensitive Reinforcement Learning for Smart Agriculture," in IEEE Transactions on Green Communications and Networking, vol. 7, no. 1, pp. 171-182, March 2023, doi: 10.1109/TGCN.2022.3205330.
19. M. Akbari, A. Syed, W. S. Kennedy and M. Erol-Kantarci, "Constrained Federated Learning for AoI-Limited SFC in UAV-Aided MEC for Smart Agriculture," in IEEE Transactions on Machine Learning in Communications and Networking, vol. 1, pp. 277-295, 2023, doi: 10.1109/TMLCN.2023.3311749.
20. Y. Zhang et al., "Improving EEG Decoding via Clustering-Based Multitask Feature Learning," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 8, pp. 3587-3597, Aug. 2022, doi: 10.1109/TNNLS.2021.3053576.
21. A. Trojan Fenerich, M. T. Arns Steiner, J. C. Nievola, K. Borges Mendes, D. P. Tsutsumi and B. Samways dos Santos, "Diagnosis of Headaches Types Using Artificial Neural Networks and Bayesian Networks," in IEEE Latin America Transactions, vol. 18, no. 01, pp. 59-66, January 2020, doi: 10.1109/TLA.2020.9049462.
22. H. Zhang, P. Gao, J. Yu, J. Lin and N. N. Xiong, "Machine Learning on Cloud With Blockchain: A Secure, Verifiable and Fair Approach to Outsource the Linear Regression," in IEEE Transactions on Network Science and Engineering, vol. 9, no. 6, pp. 3956-3967, 1 Nov.-Dec. 2022, doi: 10.1109/TNSE.2021.3110101.
23. H. Wang, W. Wang, Y. Liu and B. Alidaee, "Integrating Machine Learning Algorithms With Quantum Annealing Solvers for Online Fraud Detection," in IEEE Access, vol. 10, pp. 75908-75917, 2022, doi: 10.1109/ACCESS.2022.3190897.
24. J. Nokkala, R. Martínez-Peña, R. Zambrini and M. C. Soriano, "High-Performance Reservoir Computing With Fluctuations in Linear Networks," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 6, pp. 2664-2675, June 2022, doi: 10.1109/TNNLS.2021.3105695.

25. Y. Yoshikawa and T. Iwata, "Gaussian Process Regression With Interpretable Sample-Wise Feature Weights," in IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 9, pp. 5789-5803, Sept. 2023, doi: 10.1109/TNNLS.2021.3131234.
26. Huang, Y. et al. Data-driven soil analysis and evaluation for smart farming using machine learning approaches. Agriculture 13(9), 1777. https://doi.org/10.3390/agriculture13091777 (2023).
27. Kiran Pandiri, D. N., Murugan, R. & Goel, T. Smart soil image classification system using lightweight convolutional neural network. Expert Syst. Appl. 238, 122185. https://doi.org/10.1016/j.eswa.2023.122185 (2024).
28. Pal, A., Dubey, S. K., Goel, S. & Kalita, P. K. Portable sensors in precision agriculture: assessing advances and challenges in soil nutrient determination. TRAC Trends Anal. Chem. 180, 117981. https://doi.org/10.1016/j.trac.2024.117981 (2024).
29. Shahab, H., Naeem, M., Iqbal, M., Aqeel, M. & Ullah, S. S. IoT-driven smart agricultural technology for real-time soil and crop optimization. Smart Agricultural Technol. 10, 100847. https://doi.org/10.1016/j.atech.2025.100847 (2025).