Heliyon 10 (2024) e28626

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Enhancing agricultural automation through weather invariant soil parameter prediction using machine learning

Monisha Mushtary Uttsha [a], A.K.M. Nadimul Haque [a], Tahsin Tariq Banna [a], Shamim Ahmed Deowan [a,*], Md. Ariful Islam [a], Hafiz Md. Hasan Babu [b]

[a] *Department of Robotics and Mechatronics Engineering, University of Dhaka, Dhaka, Bangladesh*
[b] *Department of Computer Science and Engineering, University of Dhaka, Dhaka, Bangladesh*

ARTICLE INFO

Dataset link: https://github.com/Nadimulhaque0403/Soil_parameter_prediction_dataset

Keywords:
Agricultural automation
Machine learning
Artificial intelligence
Soil parameter prediction

ABSTRACT

Soil parameters are crucial aspects in increasing agricultural production. Even though Bangladesh is heavily dependent on agriculture, little research has been done regarding its automation. And a vital aspect of agricultural automation is predicting soil parameters. Generally, sensors relating to soil parameters are quite expensive and are often done in a controlled environment such as a greenhouse. However, a large scale implementation of such expensive sensors is not very feasible. This work tries to find an inexpensive solution towards predicting soil parameters such as soil moisture and temperature, both of which are crucial to the growth of crops. We focus on finding a robust relation between the above mentioned soil parameters with the nearby weather parameters such as humidity and temperature, irrespective of the weather. We apply different machine learning models like multilayer perceptron (MLP), random forest, etc. to predict the soil parameters, given the humidity and temperature of the surrounding environment. For all the experiments we have used a custom made dataset, which contains around 9000 datapoints of soil moisture & temperature, ambient humidity & temperature. The data has been collected in an uncontrolled agriculture bed via inexpensive sensors. Our results show that XGBoost regressor achieves the best results with an R^2 score of 0.93 and 0.99 for soil moisture and soil temperature data respectively. This suggests very high correlation between the weather parameters and soil parameters. The model also portrayed a very low root mean squared error and mean absolute error of 0.037 & 0.015 for soil moisture and 0.001 & 0.0008 for soil temperature. Our results show that it is indeed possible to find the soil parameters from the corresponding weather, which will have great impact on mass agricultural automation. The dataset has been made publicly available at https://github.com/Nadimulhaque0403/Soil_parameter_prediction_dataset.

1. Introduction

With the profound expansion of humanity and the escalating food requirements, agriculture has persistently stood as one of the most coveted domains throughout history. Given the prevalence of hunger, famine, and drought worldwide, the pressing necessity for agricultural advancements remains unabated. Moreover, these advancements have invariably walked hand in hand with the progress of science and technology. In this era characterized by automation, a dire need arises for the complete automation of the agricultural

* Corresponding author.
E-mail address: shamimdeowan.rme@du.ac.bd (S.A. Deowan).

https://doi.org/10.1016/j.heliyon.2024.e28626
Received 4 August 2023; Received in revised form 21 March 2024; Accepted 21 March 2024

Available online 27 March 2024
2405-8440/

realm. It can be confidently asserted that a fully automated agricultural system stands as the sole means to satisfy the escalating demands not only for food but also for adequate nutrition of our vast population. While diverse approaches have been undertaken to automate the agricultural sector, significant strides towards a fully automated agricultural system, encompassing soil preparation to crop harvesting without any human intervention, have been scarce. This challenge predominantly arises from persistent hurdles in accurately sensing and monitoring soil parameters [1–3].

Ensuring optimal soil quality stands as a fundamental aspect in achieving a bountiful harvest, serving as the cornerstone of agricultural production [4]. To uphold this quality, continuous monitoring of crucial soil parameters is imperative. Among these parameters, soil moisture and soil temperature play a pivotal role in crop growth. However, the sensors employed for monitoring such parameters tend to be costly, rendering comprehensive soil parameter monitoring in vast agricultural expanses a formidable challenge. As a result, a more viable solution emerges in the form of externally controlling soil parameters within controlled environments like greenhouses. Nevertheless, this approach proves impractical when considering large-scale agricultural automation. The financial burden alone would be substantial, and the consequential global climate impact would be profound [5,6]. While rooftop greenhouses exhibit fewer environmental drawbacks compared to traditional counterparts [7], their applicability remains limited. With environmental conditions progressively deteriorating, the pursuit of sustainable agriculture becomes increasingly arduous. Hence, it becomes crucial to explore alternative solutions, particularly for mass agriculture. One potential avenue lies in automating the agricultural process, even in open, uncontrolled fields.

As delineated earlier, the primary hurdle in achieving an automated agricultural system lies in effectively monitoring and controlling soil parameters. Traditionally, farmers have relied on manual methods, often relying on intuition and weather conditions. While this approach has its limitations, the prevailing alternative of greenhouses proves detrimental and unsustainable. Thus, integrating artificial intelligence to leverage weather data for accurate and swift predictions would constitute a significant and crucial step towards automating agriculture in open fields.

To emulate human farmers, it becomes essential to amass a vast database encompassing weather patterns and subsequent changes in soil parameters, compensating for decades of experiential knowledge. While assembling a dataset of comparable magnitude poses a formidable challenge, testing the hypothesis on a smaller scale with minimal data offers a viable approach. Accordingly, this work adopts a preferred methodology centered around a small agricultural bed, where extensive data collection occurs, encompassing soil temperature, moisture, weather temperature, and humidity.

This research aspires to contribute to the augmentation of agricultural production in Bangladesh, offering an accessible and affordable approach to predicting soil parameters. The central focus of this research is to:

1. Use inexpensive sensors to collect weather data to make predicting soil parameters affordable and accessible to farmers in regions with limited resources.
2. Utilize soil parameters such as moisture and temperature using nearby weather parameters such as humidity and temperature. This can help farmers make informed decisions about irrigation and fertilization, which can lead to increased crop yield.

2. Related works

In the domain of soil parameter prediction quite a lot of studies have been conducted. We categorize them into three different categories and detail how the following key studies have made significant contributions:

Soil Strength and Composition Analysis: Keller et al. [8] presented a mathematical model for describing the stress-strain relationship in soils, which is instrumental in comprehending soil degradation. Although their work did not yield an automatic predictor, it laid the groundwork for the development of such predictive models.

Sirsat et al. [9] conducted research aimed at classifying soil fertility indices. They utilized parameters like nitrogen dioxide (N_2O), pH, and elements such as phosphorus pentoxide (P_2O_5) and iron (Fe). Their approach, employing the Random Forest (RF) algorithm, achieved an impressive accuracy rate of 90.65%. This study underscores the potential of machine learning in the analysis of soil fertility.

Zhang et al. [10,11] employed a genetic algorithm to predict the compression module of soft clays based on a dataset consisting of 221 samples from 65 surveys. Their model exhibited a remarkably low mean squared error (MSE) of 0.13 when trained with 10-fold cross-validation. However, it is essential to note that their model's applicability on-site was limited due to site-specific considerations.

Pham et al. [12] harnessed neural networks to predict the consolidation coefficient, leveraging a dataset comprising 188 tests. Their model displayed exceptional performance, with an R^2 of 0.9973, RMSE of 0.0614, and MAE of 0.0415. This study highlights the potential of machine learning for the precise prediction of geotechnical parameters.

Kiran et al. [13] employed probabilistic neural networks to predict shear strength. Their model consistently provided accurate predictions, typically within the range of 7% to 14% deviation from actual values. This research contributes to the understanding of soil mechanics and the prediction of shear strength.

Morellos et al. [14] sought to estimate total nitrogen, organic carbon, and moisture content in soil. They investigated various algorithms, including Principal Component Regression (PCR), Partial Least Square Regression (PLSR), Least Square Support Vector Machines (LS-SVM), and Cubist. Their study, based on a dataset of wet soil samples collected from a German village, emphasized the superiority of LS-SVM and Cubist methods for accurate soil composition estimation.

Soil Moisture and Nutrient Levels Prediction: Suchithra et al. [15–18] explored the prediction of available phosphorus, potassium, boron, organic carbon, and soil reaction. Leveraging extreme learning machines, they achieved accuracy exceeding 80%, with their best-performing model utilizing a radial basis function (RBF) kernel reaching nearly 90% accuracy. This research underscores

Table 1
Property of the soil used in the experiments.

Property	Description
Particle Composition	Approximately 40% sand, 40% silt, 20% clay
Texture	Crumbly and easy to work with
Drainage	Good drainage and water infiltration
Water Retention	Maintains moisture while preventing waterlogging
Nutrient-Holding Capacity	Excellent ability to retain and provide nutrients
Fertility	High organic matter content for enhanced fertility
Aeration	Good aeration, supporting healthy root growth
Agricultural Use	Ideal for farming and gardening

the potential of machine learning in agricultural and environmental applications.

Li et al. [19,20] employed extreme learning machines to predict soil moisture, achieving accuracy rates of over 80%. This aligns with similar findings by Liu et al. (2014), who also utilized extreme learning machines to predict soil moisture and outperformed established algorithms like Support Vector Machines (SVM).

Tizpa et al. [21] proposed a neural network-based approach to predict several soil parameters, including permeability, compaction, and shear strength. Their comprehensive dataset, which included grain size distribution, permeability, Atterberg limits, and compaction data, yielded accurate predictions within a 95% confidence interval.

Smartphone-Based Soil Analysis: Yang et al. [1,22] introduced a cost-effective approach for identifying soil organic matter using smartphone photos. By preprocessing the data and employing step-wise multiple linear regression and partial least squares regression, their results demonstrated the potential of smartphone photos for accurate soil analysis. However, they noted that the reliability of smartphone capabilities remains a concern.

While these studies have significantly contributed to soil parameter prediction, it is worth noting that none have explored the correlation between weather parameters and soil conditions. This research seeks to bridge this gap by constructing a comprehensive database that integrates weather data, specifically temperature and humidity, along with corresponding soil temperature and moisture measurements. Subsequently, a neural network model has been developed and trained on this integrated dataset to predict soil parameters by leveraging the weather parameters. This novel approach is elaborated upon in the subsequent section of this paper.

3. Methodology

We present our methodologies in four parts. First, we discuss the details of the agriculture bed, the dimensions of it and soil properties. We then move onto the details of dataset creation, preprocessing and the machine learning models used to regress the soil parameters.

3.1. Agriculture bed

To create a small agriculture bed, a square shaped, steel box was designed with a barrier in between. The bed was approximately $1.2\ \text{m} \times 1.2\ \text{m} \times 0.35\ \text{m}$ in dimensions. It was filled with soil and eggplants, tomatoes and chillies were planted in it phase by phase. The bed was put in a position where it would obtain abundant natural sunlight. The soil was collected from the Rangpur district, from the northern region of Bangladesh. The soil collected is locally known as "Doash" or "Bele-Doash" soil, which is more commonly known as Loam soil [23,24]. It is a common soil type found within the broader category of Cambisols [25]. In terms of general category soil, these soils fall under the floodplain category, which is very commonly found in Bangladesh and covers almost 78.96% of the total land area [24]. Cambisols are mineral soils that have experienced some degree of soil horizon development, often characterized by the presence of a horizon (a layer of soil) that has undergone weathering processes. Loam soil is one of the most desirable types of soil for agricultural purposes due to its balanced mixture of sand, silt, and clay particles. Loam soil is often referred to as the "ideal" soil type because it provides a great balance of drainage, water retention, and nutrient-holding capacity [26].

Loam soil has specific properties that make it highly productive for farming and gardening. It consists of approximately 40% sand, 40% silt, and 20% clay, giving it a near-even distribution of particle sizes. This balance ensures good water infiltration and drainage, preventing both waterlogging and excessive drying out. Additionally, loam soil has excellent nutrient-holding capacity, allowing it to provide essential minerals and nutrients to plants. Its crumbly texture and good aeration support root growth, and it is generally easy to work with. Loam soil is also known for its high organic matter content, which enhances its fertility and overall productivity. It is highly versatile and suitable for a wide range of crops and plants, making it a preferred choice for many agricultural and horticultural applications. Some of the soil properties are shown in Table 1.

3.2. Creating the dataset

A dataset has been created by collecting both weather data and soil data [27,28] from the agricultural bed, with the help of multiple sensors (Fig. 1) connected to a PC (Fig. 1A) through an Integrated Controller Circuit (Fig. 1B) with additional power supply (Fig. 1C) for the controller. 4 sensors had been used, weather temperature & humidity sensor (Fig. 1E), soil temperature sensor (Fig. 1F) and a capacitive soil moisture sensor (Fig. 1D). The sensors for the weather data collection were placed in close proximity

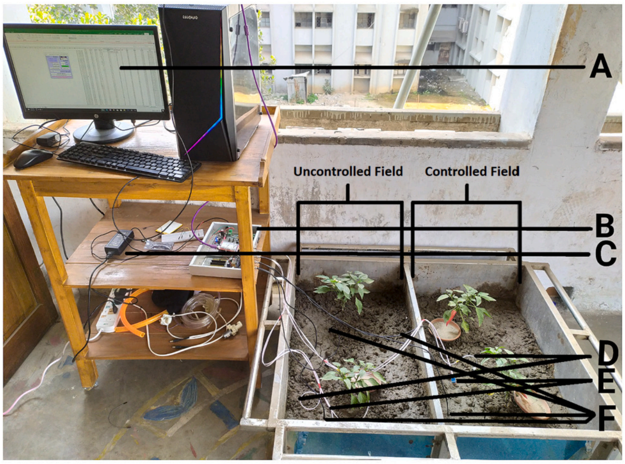


Fig. 1. Experimental setup for collecting soil parameter data. **A**: Control Interface, **B**: Integrated Controller Circuit, **C**: Power Supply **D**: Soil Moisture Sensor, **E**: Temperature and Humidity Sensor and **F**: Soil Temperature Sensor.

to the agricultural bed [29–31]. The soil moisture sensors were approximately 10 cm long and were inserted 8 cm into the soil itself. The soil temperature sensors were 4 cm long are were inserted 3.5 cm in the soil. The soil parameter sensors were placed about 5-10 cm from the root of the plants and the environmental parameter sensors were placed next to them.

The data has been collected and compiled on an excel file along with the timestamp of the data collection. Data has been collected in three different seasons, summer, winter and rainy. The distribution of sensor data according to the seasons are shown in Fig. 2 and 3. More than 9,000 data points have been compiled. The dataset contains the soil temperature value, the raw voltage of the capacitive moisture sensor, the weather temperature and humidity along with the season of the data collection.

3.3. Data preprocessing and cleanup

To gain insights into the collected data [32,33], a merging process was initially undertaken, followed by visualization techniques to comprehend the interrelationships and characteristics among the various collected fields. To optimize resource utilization and save time, the models were trained on a subset of the dataset. Notably, it became evident that the correlations between the fields were neither linear nor polynomial, rendering attempts to implement simple linear or polynomial regression models inadequate. These models exhibited poor fit and yielded substantial errors.

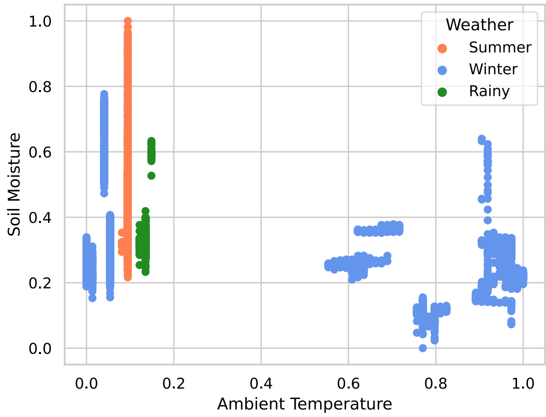
During the model training process, discrepancies were discovered in certain data points, attributed to sensor connectivity issues resulting in negative values where only positive values were expected. Consequently, the non-positive values were cleansed from the dataset. Additionally, normalization was applied to the data, as depicted in Figs. 2 and 3. Normalization plays a vital role in achieving improved results, given that different features within the dataset may possess distinct scales. By normalizing the data, features are rescaled to a common range, facilitating the model's ability to discern patterns within the data. This normalization process enhances model score while reducing the likelihood of overfitting.

3.4. Using different training models

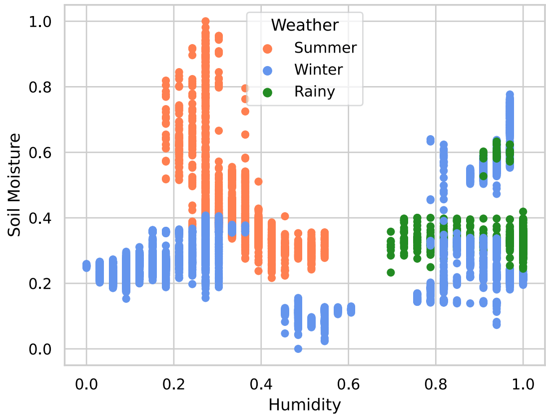
In this study, we employed a variety of machine learning models to analyze and predict soil parameters in an agricultural setting [34–37]. These models were selected to explore their performance in capturing the complex relationships within the dataset. We initiated our model selection with simpler linear and polynomial regression models. However, upon initial evaluation, it became evident that these models did not adequately represent the dataset, as observed from the normalized data plots. Their limited capacity to capture the nonlinear and intricate relationships in the data led us to seek more robust alternatives.

Support Vector Regressor (SVR): As the limitations of linear and polynomial regression models became apparent, we turned to the SVR [38–40]. SVR displayed a notable improvement in capturing the underlying patterns within the dataset, but we recognized that further enhancements were still possible. The initial SVR model achieved an R^2 score of approximately 0.77 for Soil Moisture and 0.96 for Soil Temperature. It was a significant step forward, but we remained focused on refining our approach.

Multi-Layered Perceptron Regressor (MLPRegressor): To pursue better results and harness the potential of non-linear relationships within the data, we employed an MLPregressor, a type of neural network (as shown in Fig. 4). Neural networks are adept at capturing complex, non-linear patterns within data. The initial neural network model exhibited remarkable promise, achieving an R^2 score of approximately 0.88 for Soil Moisture and 0.99 for Soil Temperature. To improve the performance further, hyperparameter tuning and data normalization were explored, as the dataset contained fields with varying value ranges. The neural network performed optimally when the data was normalized, allowing for a more uniform range of data to be processed.



(a) Soil Moisture vs Ambient Temperature



(b) Soil Moisture vs Humidity

Fig. 2. Normalized weather temperature (a) and humidity (b) values with respect to soil moisture.

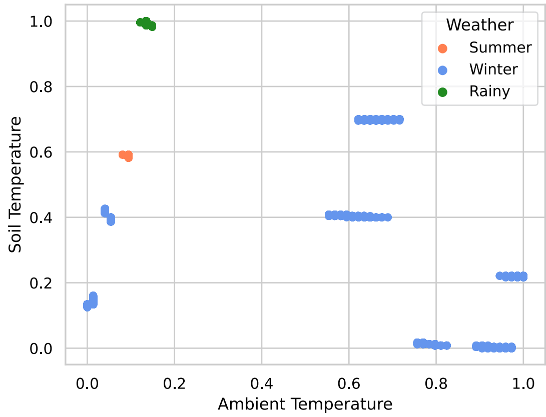
Random Forest (RF) Regressor: RF [41] works by creating an ensemble of decision trees, where each tree independently makes predictions. The final prediction is the average (or majority vote) of these individual tree predictions. RF is robust and generally performs well without the need for extensive hyperparameter tuning. While RF Regressor did not surpass XGBRegressor in predictive score, it still yielded competitive results, making it a viable alternative for specific applications.

Gradient Boosting Regressor (GBRegressor): GBregressor constructs an ensemble of decision trees in a sequential manner. Each new tree is trained to correct the errors made by the previous trees, which often leads to highly accurate predictions. Gradient boosting [42,43] is widely used in various regression problems and can capture complex patterns in data. The GBregressor showed significant promise in our experiments, achieving high correlations among the datasets and capturing intricate patterns. It demonstrated the ability to adapt to non-linear relationships and exhibited notable predictive power.

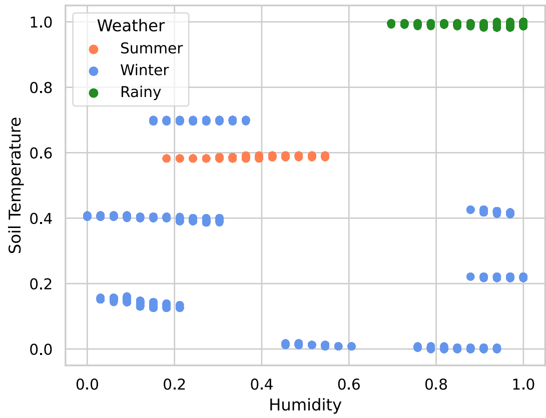
LGBMRegressor: LGBMRegressor is a regression model based on Light Gradient Boosting Method [44], another ensemble learning algorithm, designed for efficiency and handling large datasets. It uses a histogram-based approach to build decision trees, which speeds up the training process. LightGBM is known for its ability to handle big data and is particularly efficient when it comes to categorical features.

XGBRegressor: XGBoost [45] is an ensemble learning algorithm known for its high performance. XGBRegressor is a regression model based on XGBoost. It combines multiple decision trees to make predictions. These decision trees are trained sequentially to correct the errors made by the previous ones. XGBoost is effective at handling complex relationships in data and is a popular choice for various regression tasks. Among the ensemble models, the XGBRegressor emerged as the top performer, consistently delivering the highest results. Its predictive score outperformed the other models, making it the preferred choice for our predictive modeling tasks.

By incorporating a diverse set of machine learning models in our analysis, we aimed to ensure a comprehensive exploration of the dataset's intricacies and to develop a reliable predictive model for soil parameters in agricultural contexts. In the subsequent sections, we present the results of these models, which highlight their respective performances and insights gained from their application.



(a) Soil Moisture vs Ambient Temperature



(b) Soil Temperature vs Humidity

Fig. 3. Normalized weather temperature (a) and humidity (b) values with respect to soil temperature.

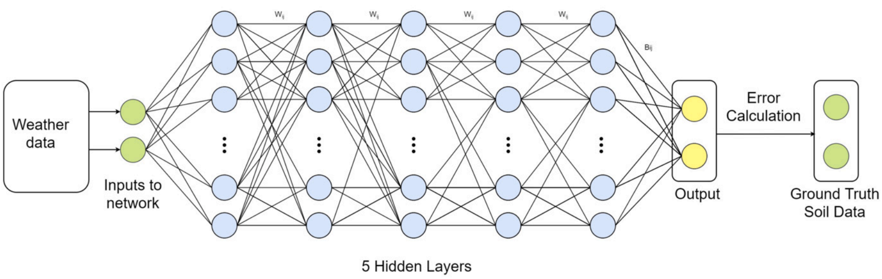


Fig. 4. MLP regressor model used to train the data.

4. Experiments

4.1. Evaluation metrics

Predicting soil parameters fall under supervised regression tasks. Usually these models are evaluated using regular regression metrics such as Root Mean Square Error (RMSE), Coefficient of determination (R^2) and Mean Absolute Error (MAE) values.

RMSE: The RMSE is a frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed. The RMSE of predicted values $\hat{y}_t$ for times t of a regression's dependent variable y_t, with variables observed over T times, is computed for T different predictions as the square root of the mean of the squares of the deviations using the equation (1).

Table 2
Results on the soil moisture and soil temperature data. The best scores are shown in **bold**.

Model Name	Soil Moisture			Soil Temperature		
	R^2	RMSE	MAE	R^2	RMSE	MAE
SVR	0.772083	0.066724	0.051517	0.955363	0.058583	0.049610
RFRegressor	0.923326	0.038701	0.015793	0.999971	0.001506	**0.000807**
GBRegressor	0.904779	0.043128	0.021427	0.999972	0.001476	0.001005
LGBMRegressor	0.923700	0.038606	0.015898	**0.999975**	**0.001393**	0.000885
XGBRegressor	**0.926852**	**0.037800**	**0.015269**	**0.999975**	0.001398	0.000875
MLPRegressor	0.881328	0.048147	0.028521	0.999947	0.002023	0.001572

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}} \tag{1}$$

$\mathbf{R^2}$: The coefficient of determination (R^2) is the proportion of the variation in the dependent variable that is predictable from the independent variables. If SS_{res} is the sum of squares of residuals and SS_{tot} is the total sum of squares, the coefficient of determination can be expressed as the equation (2):

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{2}$$

MAE: The mean absolute error (MAE) is a measure of errors between paired observations expressing the same phenomenon. It is calculated as the sum of absolute errors divided by the sample size using equation (3):

$$\text{MAE} = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \tag{3}$$

4.2. Results

Fig. 5 shows the changes the evaluation metrics, R^2, RMSE and MAE with the change in training size. The plots show a general upward rise for the coefficient of determination for both soil moisture and soil temperature. As a result, the RMSE and MAE values decrease as the models get better and better and we see a decrease with training size. The mean R^2, RMSE and MAE values for soil moisture are shown in Fig. 5. This gives us the best model, XGBRegressor. While the SVR model struggles to learn well, other models achieve better results but fall short to the 0.93 coefficient of determination of the XGBRegressor model. This is also mirrored in the MAE and RMSE errors where the model achieves the lowest error.

In contrast to soil moisture, which frequently demonstrates notable spatial and temporal fluctuations, soil temperature generally exhibits a more stable profile and is predominantly shaped by diurnal and seasonal rhythms. Consequently, the data attributes pertaining to soil temperature were comparatively straightforward [46–49]. This is reflected in the evaluation metrics in Fig. 5 as well as the mean scores in Table 2. Other than SVR, all other model fits the dataset perfectly and achieves a perfect correlation. The errors are also near zero as a result.

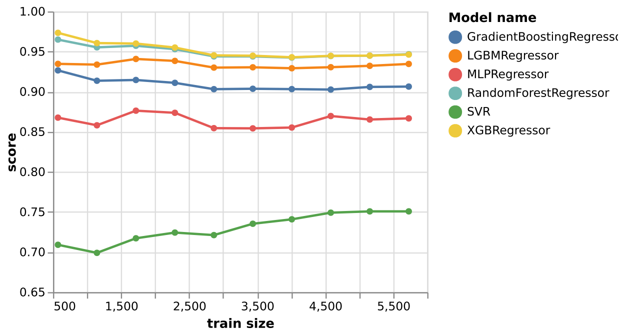
The results show a definite correlation between the weather parameters and soil parameters. This indicates that it is indeed possible to predict soil parameters such as moisture and temperature from weather humidity and temperature.

4.3. Tuning hyperparameters

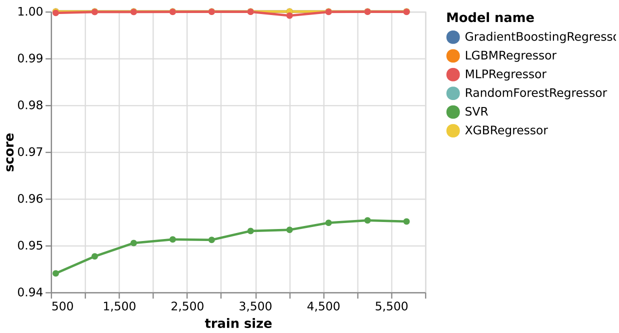
After the model was selected, the default hyperparameters of the model were tweaked ever so slightly to increase the score. The maximum iteration for convergence was increased to a value of 200000 from the default 200 as the prediction score of the model did not converge with small number of iterations. The relu activation function was chosen as it yielded in better performance. The adam optimizer [50] had been chosen through trial and error as others did not improve the overall score. The internal hidden layer size and the number of neurons in each of the layers were also changed to be over 350 in order to better fit the data. The learning rate was reduced 10-folds and some other hyperparameters were also changed slightly. For the regressor based models, learning rate, gamma, max depth and regularization terms were tuned.

5. Discussions and future work

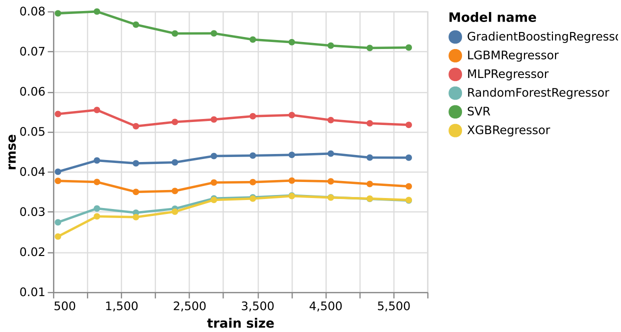
The results clearly indicate that there exists a strong correlation between soil and weather parameters. Although all the models produce strong and consistent results, XGBRegressor produced the best results. One of the main limitations of this work is that there exists a relatively high data imbalance, with more datapoints in winter, compared to the other two. Although data imbalance is not as crucial in regression as it is in classification, it can still have a negative effect, especially in neural networks. This could be the reason why MLP regressor did not perform up to the mark. Also, we tried to keep the neural network shallow in order for it to be usable an automated system operating in real time, which could have limited its potential as well. XGBRegressors however, are fast and they excel with imbalanced data to begin with. Ensemble learning methods also performed well due to their ability to



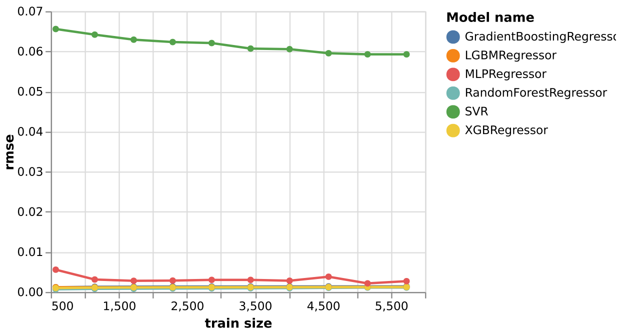
(a) R^2 score on soil moisture



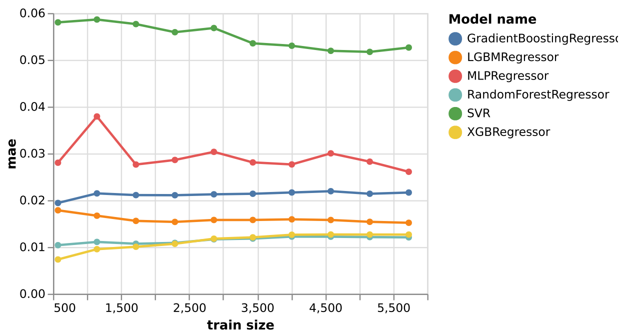
(b) R^2 score on soil temperature



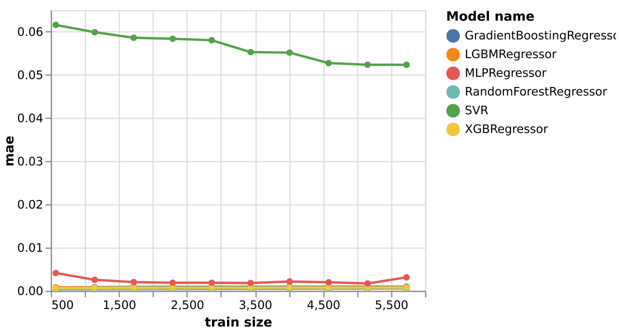
(c) RMSE score on soil moisture



(d) RMSE score on soil temperature



(e) MAE score on soil moisture



(f) MAE score on soil temperature

Fig. 5. (a–b) shows increase in R^2 values, (c–d) shows decrease of root mean squared error (RMSE) and (e–f) shows decrease in mean absolute error (MAE) with training size.

derive complex relationships between input and output data. Despite the R^2 values, all the algorithms exhibit minimal error (Both in RMSE and MAE), and represent how accurately we can predict soil parameters with inexpensive ambient temperature and humidity sensors.

In future, this experiment will be replicated in a larger, open, agricultural field to ensure the reliability of the model's performance. Some future work based on the limitations of what can be done in this area are as follows:

1. Other factors that affect soil parameters, such as soil type, topography, wind speed and solar radiation can be used to conduct a more robust experiment.
2. Our experiments focused on a single test bed. It can be extended to different agricultural beds and regions to test the generalizability.
3. This experiment can be extended to predict other soil parameters such as pH and nutrient levels.
4. This work can be integrated with an automated irrigation and fertilization system to provide real-time feedback and recommendations to farmers.
5. This can be used to develop a mobile application that can provide farmers with real-time information about soil parameters and weather conditions based on local regions.

6. This work can facilitate a fully automated decision support system that can help farmers make informed decisions about irrigation and fertilization.

6. Conclusion

The present study endeavors to establish a substantive correlation between critical soil parameters, specifically soil moisture and temperature, and ambient factors like temperature and humidity. The findings underscore a robust relationship between soil and ambient parameters, indicating a significant association. Employing various machine learning models—XGBRegressor, RFRegressor, GBRegressor, LGBMRegressor, and a customized MLP regressor—the research demonstrates consistently low RMSE errors across distinct seasonal variations. This capacity to generate precise predictions despite considerable fluctuations in data due to seasonal changes holds considerable promise for the advancement of agricultural automation in open-field settings.

The performance of the models, especially amid varying seasonal conditions, signals a potential breakthrough in the realm of agricultural automation. This insight offers a departure from conventional methods that involve controlling entire environmental conditions, a practice that often exacerbates climate challenges. Instead, the findings present an avenue for sustainable agricultural automation, thereby fostering a pathway toward a more promising and environmentally conscious future for agricultural practices. This empirical evidence supports the feasibility of leveraging weather-related data for accurate soil parameter predictions, marking a pivotal stride towards advancing agricultural automation at scale in uncontrolled agricultural settings.

CRediT authorship contribution statement

Monisha Mushtary Uttsha: Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation. **A.K.M. Nadimul Haque:** Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation. **Tahsin Tariq Banna:** Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Shamim Ahmed Deowan:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Md. Ariful Islam:** Writing – review & editing, Supervision, Project administration, Investigation. **Hafiz Md. Hasan Babu:** Supervision, Project administration, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset used in this study, comprising weather parameters, soil parameters, and the corresponding experimental details, is openly available and can be accessed on the following repository: https://github.com/Nadimulhaque0403/Soil_parameter_prediction_dataset.

Acknowledgement

We would like to thank the Ministry of Science and Technology of the People's Republic of Bangladesh, under the project title "Artificial Intelligent Supported Climate-Smart Agriculture (AIAgri)" Financial Year: 2018-19 (Phase-1), 2019-20 (Phase-2) & 2020-21 (Phase-3), for funding this work. We also thank the Department of Robotics and Mechatronics Engineering, University of Dhaka and everyone involved in the department for their continuous support in this work.

References

[1] H.-Q. Yang, L. Zhang, Q. Pan, K.-K. Phoon, Z. Shen, Bayesian estimation of spatially varying soil parameters with spatiotemporal monitoring data, Acta Geotech. 16 (2021) 263–278.
[2] P. Placidi, R. Morbidelli, D. Fortunati, N. Papini, F. Gobbi, A. Scorzoni, Monitoring soil and ambient parameters in the iot precision agriculture scenario: an original modeling approach dedicated to low-cost soil water content sensors, Sensors 21 (15) (2021) 5110.
[3] B.B. Bhanu, M.A. Hussain, P. Ande, Monitoring of soil parameters for effective irrigation using wireless sensor networks, in: 2014 Sixth International Conference on Advanced Computing (ICoAC), IEEE, 2014, pp. 211–215.
[4] R.I. Papendick, J.F. Parr, Soil quality—the key to a sustainable agriculture, Am. J. Altern. Agric. 7 (1–2) (1992) 2–3.
[5] T. Boulard, C. Raeppel, R. Brun, F. Lecompte, F. Hayer, G. Carmassi, G. Gaillard, Environmental impact of greenhouse tomato production in France, Agron. Sustain. Dev. 31 (4) (2011) 757.
[6] I. Stavi, R. Lal, Agriculture and greenhouse gases, a common tragedy. A review, Agron. Sustain. Dev. 33 (2) (2013) 275–289.
[7] I.T. Pineda, J.H. Cho, D. Lee, S.M. Lee, S. Yu, Y.D. Lee, Environmental impact of fresh tomato production in an urban rooftop greenhouse in a humid continental climate in South Korea, Sustainability 12 (21) (2020) 9029.
[8] T. Keller, P. Défossez, P. Weisskopf, J. Arvidsson, G. Richard, SoilFlex: a model for prediction of soil stresses and soil compaction due to agricultural field traffic including a synthesis of analytical approaches, Soil Tillage Res. 93 (2) (2007) 391–411.
[9] M.S. Sirsat, E. Cernadas, M. Fernández-Delgado, R. Khan, Classification of agricultural soil parameters in India, Comput. Electron. Agric. 135 (2017) 269–279.
[10] D.-m. Zhang, J.-z. Zhang, H.-w. Huang, C.-c. Qi, C.-y. Chang, Machine learning-based prediction of soil compression modulus with application of 1D settlement, J. Zhejiang Univ. Sci. A 21 (6) (2020) 430–444.

[11] S. Mirjalili, Genetic algorithm, in: Evolutionary Algorithms and Neural Networks, Springer, 2019, pp. 43–55.
[12] B.T. Pham, S.K. Singhº, H.-B. Ly, Using Artificial Neural Network (ANN) for prediction of soil, Vietnam J. Earth Sci. 42 (4) (2020) 311–319.
[13] S. Kiran, B. Lal, S. Tripathy, Shear strength prediction of soil based on probabilistic neural network, Indian J. Sci. Technol. 9 (41) (2016).
[14] A. Morellos, X.-E. Pantazi, D. Moshou, T. Alexandridis, R. Whetton, G. Tziotzios, J. Wiebensohn, R. Bill, A.M. Mouazen, Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy, Biosyst. Eng. 152 (2016) 104–116.
[15] M. Suchithra, M.L. Pai, Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters, Inf. Process. Agric. 7 (1) (2020) 72–82.
[16] H. Abdi, Partial least square regression (PLS regression), Encycl. Res. Methods Soc. Sci. 6 (4) (2003) 792–795.
[17] J.A. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J.P. Vandewalle, Least Squares Support Vector Machines, World Scientific, 2002.
[18] M. Kuhn, S. Weston, C. Keefer, N. Coulter, Cubist models for regression, in: R Package Vignette R Package Version 0.0, vol. 18, 2012.
[19] M.-B. Li, G.-B. Huang, P. Saratchandran, N. Sundararajan, Fully complex extreme learning machine, Neurocomputing 68 (2005) 306–314.
[20] Y. Liu, L. Mei, S.K. Ooi, Prediction of soil moisture based on extreme learning machine for an apple orchard, in: 2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems, IEEE, 2014, pp. 400–404.
[21] P. Tizpa, R.J. Chenari, M.K. Fard, S.L. Machado, ANN prediction of some geotechnical properties of soil from their index parameters, Arab. J. Geosci. 8 (5) (2015) 2911–2920.
[22] J. Yang, F. Shen, T. Wang, M. Luo, N. Li, S. Que, Effect of smart phone cameras on color-based prediction of soil organic matter content, Geoderma 402 (2021) 115365.
[23] M.A. Islam, M.A. Hasan, M.A. Farukh, Application of gis in general soil mapping of Bangladesh, Int. J. Geogr. Inf. Syst. 9 (5) (2017) 604–621.
[24] Banglapedia, 17 June 2021, Bangladesh soil, 18 October 2023 [Online]. Available: https://en.banglapedia.org/index.php/Bangladesh_Soil.
[25] H. Brammer, J. Antoine, A. Kassam, H. Van Velthuizen, Land resources appraisal of Bangladesh for agricultural development, Report-2 (BGD/81/035), vol. 20, FAO of United Nations, Rome, 1988.
[26] M. Quan, J. Liang, The influences of four types of soil on the growth, physiological and biochemical characteristics of Lycoris aurea (L'Her.) Herb, Sci. Rep. 7 (1) (2017) 43284.
[27] Y. Yang, P. Gao, Z. Sun, H. Wang, M. Lu, Y. Liu, J. Hu, Multistep ahead prediction of temperature and humidity in solar greenhouse based on FAM-LSTM model, Comput. Electron. Agric. 213 (2023) 108261.
[28] M.K. Gill, T. Asefa, M.W. Kemblowski, M. McKee, Soil moisture prediction using support vector machines 1, J. Am. Water Resour. Assoc. 42 (4) (2006) 1033–1046.
[29] K.I. Paul, P.J. Polglase, P.J. Smethurst, A.M. O'Connell, C.J. Carlyle, P.K. Khanna, Soil temperature under forests: a simple model for predicting soil temperature under a range of forest types, Agric. For. Meteorol. 121 (3–4) (2004) 167–182.
[30] G. Patrizi, A. Bartolini, L. Ciani, V. Gallo, P. Sommella, M. Carratù, A virtual soil moisture sensor for smart farming using deep learning, IEEE Trans. Instrum. Meas. 71 (2022) 1–11.
[31] Z. Hong, Z. Kalbarczyk, R.K. Iyer, A data-driven approach to soil moisture collection and prediction, in: 2016 IEEE International Conference on Smart Computing (SMARTCOMP), IEEE, 2016, pp. 1–6.
[32] L.D. Rodić, T. Županović, T. Perković, P. Šolić, J.J. Rodrigues, Machine learning and soil humidity sensing: signal strength approach, ACM Trans. Internet Technol. 22 (2) (2021) 1–21.
[33] H. Yin, Y. Cao, B. Marelli, X. Zeng, A.J. Mason, C. Cao, Soil sensors and plant wearables for smart and precision agriculture, Adv. Mater. 33 (20) (2021) 2007764.
[34] O. Ozgener, L. Ozgener, J.W. Tester, A practical approach to predict soil temperature variations for geothermal (ground) heat exchangers applications, Int. J. Heat Mass Transf. 62 (2013) 473–480.
[35] S. Kang, S. Kim, S. Oh, D. Lee, Predicting spatial and temporal patterns of soil temperature based on topography, surface cover and air temperature, For. Ecol. Manag. 136 (1–3) (2000) 173–184.
[36] X. Li, J. Tang, C. Yin, Sequence-to-sequence learning for prediction of soil temperature and moisture, IEEE Geosci. Remote Sens. Lett. 19 (2022) 1–5.
[37] G. Singh, D. Sharma, A. Goap, S. Sehgal, A. Shukla, S. Kumar, Machine learning based soil moisture prediction for Internet of things based smart irrigation system, in: 2019 5th International Conference on Signal Processing, Computing and Control (ISPCC), IEEE, 2019, pp. 175–180.
[38] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[39] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, IEEE Intell. Syst. Appl. 13 (4) (1998) 18–28.
[40] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222.
[41] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[42] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. (2001) 1189–1232.
[43] Y. Shi, G. Ke, Z. Chen, S. Zheng, T.-Y. Liu, Quantized training of gradient boosting decision trees, Adv. Neural Inf. Process. Syst. 35 (2022) 822–18 833.
[44] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, Lightgbm: a highly efficient gradient boosting decision tree, Adv. Neural Inf. Process. Syst. 30 (2017).
[45] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[46] M. Bilgili, Prediction of soil temperature using regression and artificial neural network models, Meteorol. Atmos. Phys. 110 (2010) 59–70.
[47] L. Xing, L. Li, J. Gong, C. Ren, J. Liu, H. Chen, Daily soil temperatures predictions for various climates in United States using data-driven model, Energy 160 (2018) 430–440.
[48] H. Tabari, A.-A. Sabziparvar, M. Ahmadi, Comparison of artificial neural network and multivariate linear regression methods for estimation of daily soil temperature in an arid region, Meteorol. Atmos. Phys. 110 (2011) 135–142.
[49] Y. Feng, N. Cui, W. Hao, L. Gao, D. Gong, Estimation of soil temperature from meteorological data using different machine learning models, Geoderma 338 (2019) 67–77.
[50] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint, arXiv:1412.6980, 2014.