

Exploring Machine Learning Algorithms in Soil Management and Precision Agriculture – A Survey

Kurnool. Anusha Devi

1. Department of Computer Science,
SCNR Government Degree College,
Proddatur, YSR Kadapa Dist, India.

2. Research Scholar, Department of Computer Science,
Sri Padmavati Mahila Visvavidyalayam, Tirupati, India.

K.Usha Rani

3. Department of Computer Science,
Sri Padmavati Mahila Visvavidyalayam,
Tirupati, India.

Abstract— Machine learning (ML) is a branch of Artificial Intelligence wherein computers acquire the ability to learn from data, enhancing their performance on a given task without the need for explicit programming. Many industries and fields have benefited from the use of ML, including Agriculture, Bioinformatics, self-driving vehicles, identification of fraudulent credit card activities, filtering unwanted emails and malicious software, healthcare diagnostics, identification of pollutants in urban water systems, analysis of chronological data, processing of human language, recognition of speech, and interpretation of images. A number of Agricultural management tasks are made easier by ML. India's economy is based primarily on Agriculture; it supports industrial as well as international trade in both imports and exports, and its contribution is 18.3% of GDP (gross domestic product) for the fiscal year 2022–2023. During the previous two decades, as a result of significant digitalization in technology, farming has transitioned from the conventional way to Precision Agriculture. Precision Agriculture harmonizes technology and data to elevate farming's efficiency, sustainability, and productivity. Precision farming has numerous uses for ML, including choosing right crop, forecasting yield, classifying soil, predicting soil quality and weather, watering systems, prescribing fertilizer, weed detection, forecasting diseases, figuring out the least expensive support for optimal assured sustainability, maximal productivity, and a safe environment. Soil is a heterogeneous natural resource. Soil's fertile physical and chemical properties play a vital role in the Agriculture Production System. The provision of food, clean water, energy, shelter, and infrastructure to society depends on land and soil. The current study aims to present a review of ML methods applied for various soil management in the Agricultural domain.

Keywords— Machine Learning, ML Algorithms, ML Techniques, Agriculture, Precision Agriculture, Soil Properties, Soil Health Management, Soil Management.

I. INTRODUCTION

Data science employs Machine Learning (ML) for conducting data analysis. Regression and Classification are the two basic operations of ML in Data Science. ML enables the extraction of valuable insights and patterns from data driven to informed decision-making and innovation[1]. In

contemporary times, machine learning finds extensive applications across various domains, including Agriculture, Bioinformatics, self-driving vehicles, identification of fraudulent credit card activities, filtering unwanted emails and malicious software, healthcare diagnostics, identification of pollutants in urban water systems, analysis of chronological data, processing of human language, recognition of speech, and interpretation of images[2], [3], [4], [5], [6], [7], [8], [9] to name just a few. A number of Agricultural management tasks are made easier by ML. ML can be applied to sensor data to improve farm management, offering more robust recommendations and insights to inform subsequent decisions and actions. This will lead to improved Levels of production and the quality of bio-products [10].

Agriculture is backbone of Indian Economy. In the Fiscal Year 2022-23, this industry constituted 18.3% of the country's Gross Domestic Product (GDP). As per the United Nations Department of Economic and Social Affairs, the worldwide population is projected to reach 9.7 billion by 2050, with a substantial portion of the growth occurring in Africa and Asia countries. To feed this growing population food production will need to increase by 50-70% by 2050 with a focus on sustainable and resilient agriculture practices. But Agriculture faces challenges such as pests, weather issues, improper harvesting and inadequate support. These unexpected problems along with excessive chemical use, insufficient subsidies and corruption cause global economic losses. This disconnects farmers causing debt and suicides. Technology like ML, Image Processing, Internet of Things [11], Data Analytics, Cloud Computing, Block Chain Technology can aid the industry and turns towards digitalization of Agriculture known as Precision Agriculture or Digital Agriculture [2], [12] In the realm of Precision Agriculture, ML models are created to help with tasks like choosing the right crops, guessing how much they will produce, categorizing soil, predicting weather, managing irrigation, suggesting fertilizers, foreseeing diseases, and setting the lowest fair price for the crops [2]. The subsequent portions of this systematic review are structured in the following manner. Section II describes ML, its types and Techniques. Section III focuses on the

Agriculture, Precision Agriculture and role of ML in Precision Agriculture. Section IV provides the Soil, with its Properties, and its management. Section V presents the current state of ML models, various ML algorithm applications, its brief explanation, and its best results for Soil Management. Section VI presents the conclusion and future research directions.

II. MACHINE LEARNING

ML is a subset of Artificial Intelligence where computers gain the capability to learn from data, improving their performance on a specific task without the necessity for explicit programming. In this data-driven approach, ML is more effective when more data is used. The data comprises examples characterized by features; machine learning involves learning and testing processes. Features typically form a Feature Vector, it can be Binary, Numeric, Ordinal, or Nominal [13].

In the learning process, the input is represented by the Feature Vector, where the machine learns from training data to achieve task proficiency, leading to model application for classification, clustering, or prediction [14] which is shown in figure 1.

Machine learning is applied in data science across a wide range of domains, including forecasting analytics, Processing of natural language, Recognition of images, Systems for providing recommendations, fraud detection, and medical diagnosis, disease detection, pest control, soil quality assessment, automated farming equipment, irrigation management, and optimizing resource allocations.

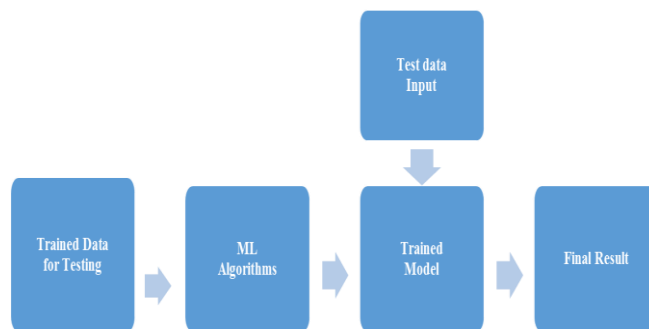


Figure 1: Steps in Machine Learning Process

A. ML Techniques:

The Domain knowledge, Statistical properties is leads vital role when selecting ML methods [15].

While Selecting the ML Algorithms, the scale and quality of the data are critical factors. Certain algorithms require extensive data to achieve strong generalization, while others can perform exceptionally well with smaller datasets effectively avoiding over fitting problems. These factors should be taken into account when selecting the appropriate technique(s) for making predictions of various Soil Properties. Table1 provides a concise overview of well-known machine learning methods along with their respective Advantages and Drawbacks.

Table1: Overview of Machine Learning Techniques, including their Advantages and Drawbacks.

ML Techniques	Advantages	Drawbacks
Linear Regression (LR) [16][17]	<ul style="list-style-type: none"> Simple interpretation and implementation. Effective for linear relationships in data. Quick computations and baseline modeling. 	<ul style="list-style-type: none"> Assumes linearity and is sensitive to outliers. Vulnerable to violations of assumptions. Limited by nonlinearity, multi collinearity, and categorical variables.
Logistic Regression (LGR) [17][18]	<ul style="list-style-type: none"> Simple to understand and interpret. Suitable for binary classification tasks. Provides probabilities for predictions. Can handle feature interactions. 	<ul style="list-style-type: none"> Assumes a linear relationship between features and log-odds. May struggle with non-linear patterns. Sensitive to outliers and multi collinearity. Not well-suited for multiclass classification tasks.
Artificial Neural Networks (ANN) [17][19][20][21]	<ul style="list-style-type: none"> Highly capable in capturing complex patterns and relationships. Suitable for various tasks including image and text analysis. Adaptive and can learn from large datasets. Can model both linear and non-linear relationships. 	<ul style="list-style-type: none"> Complexity can lead to longer training times and overfitting. Prone to black-box behavior, making interpretation difficult. Requires careful tuning of hyper parameters. Large networks may need substantial computational resources.
Support Vector Machines (SVM) [22][23]	<ul style="list-style-type: none"> Effective for both linear and non-linear classification tasks. Robust against overfitting due to regularization. Excels in managing high-dimensional data effectively. Exhibits versatility by employing various kernel functions. 	<ul style="list-style-type: none"> Computationally intensive for large datasets. Choosing the appropriate kernel and tuning parameters can be challenging. Doesn't directly provide probability estimates (usually requires additional steps). Can struggle with noisy or overlapping data.
K-Nearest Neighbors (KNN) [24][25][26]	<ul style="list-style-type: none"> Simple and intuitive algorithm. Can capture complex decision boundaries. Works well for both classification and regression tasks. Adaptively adjusts to data without requiring explicit training. 	<ul style="list-style-type: none"> Computationally expensive during prediction for large datasets. Sensitive to irrelevant features and noisy data. Choice of the right value of k is crucial and can impact results. Not suited for high-dimensional data due to the "curse of dimensionality."
Decision Trees (DT) [27][28]	<ul style="list-style-type: none"> Easy to understand and interpret, providing transparent decisions. Can handle both numerical and categorical data. Automatically selects important features through splits. Suitable for nonlinear relationships and interactions. 	<ul style="list-style-type: none"> Prone to overfitting, especially with deep trees. Can be unstable, sensitive to small changes in data. Limited expressiveness for capturing complex relationships. May create biased trees if one class dominates the data.
Naïve Bayes (NB) [29][30]	<ul style="list-style-type: none"> Simple and efficient algorithm, particularly for text classification. Handles high-dimensional data well. Works with small datasets and requires minimal tuning. Performs surprisingly well in various real-world scenarios. 	<ul style="list-style-type: none"> Assumes independence between features (naive assumption). May not capture complex relationships in the data. Sensitive to irrelevant features. Probability estimates can be unreliable for extreme cases.
Random Forest (RF) [31][32]	<ul style="list-style-type: none"> Efficient handling of high-dimensional data. Reduced overfitting due to ensemble techniques. Provides feature importance rankings for interpretability. 	<ul style="list-style-type: none"> Computationally expensive for large datasets. Reduced transparency due to ensemble complexity. May struggle with capturing intricate relationships in certain datasets.

Gradient Boosting (GB) [33][34]	<ul style="list-style-type: none"> Effectively captures complex patterns and improves model performance. Handles diverse data types without extensive preprocessing. Reduces overfitting through boosting and ensemble. 	<ul style="list-style-type: none"> Can be computationally intensive and time-consuming. Prone to overfitting if not tuned correctly. Requires careful hyper parameter tuning for optimal results.
---------------------------------	--	--

III. PRECISION AGRICULTURE

Agriculture plays a vital role in the Indian economy, where 54 per cent of the population is directly or indirectly involved in agriculture and its allied activities [35]. Increasing global population, changing consumer demands, and not having enough land, water, and energy becomes challenges to the agricultural industry. Technology significantly contributes in mitigating all these pressure on Agriculture Sector. The Indian Council of Agricultural Research (ICAR) launched "Attracting and Retaining Youth in Agriculture (ARYA)" program in 2015-16 to emphasize the significance of rural youth in advancement of agriculture, particularly for food security and still it is going on. By this program Many rural areas in India have embraced digitalization, and applications related to ML and Artificial Intelligence (AI) are gradually gaining prominence.

The program aims to engage rural individuals under 35 years old in agriculture, offering income-generating prospects to empower them in this sector. From the past two decades, due to huge technology evolution mainly Advances in ML, geographic information system (GIS), internet of things, cloud computing, global positioning system (GPS), AI etc. turns the traditional / Conventional way of Agriculture to Digital Agriculture also known as Precision Agriculture (PA). PA begins with the planting of a seed in the ground and continues with soil preparation, seed cleaning, careful monitoring of crop health, accurate water dosage calculation, and culminates as robots employ computer vision methods to gather the ripe harvest [36]. ML has the potential to completely transform economies in developing countries like India, where the primary industry for employment is agriculture. Applications of ML will contribute to raising the yield by making prompt decisions that lower costs and boost profitability[2]. According to study by V Meshram et al. [37], Agriculture tasks categorized into Pre harvesting Tasks, Harvesting Tasks, and Post harvesting Tasks.

A brief explanation of several studies related to ML models used in PA is presented below.

According to the studies [10], [14], [38] PA generally classified as four major categories, those are Crop Management, Soil Management, Water Management, Live Stock Management which are shown in the figure 2. A Crop Management process includes identifying weeds, anticipating crop yield, detecting diseases, recognizing crops, and assessing crop quality. Soil management encompasses various aspects of soil protection and management. Optimal use of water resources is the goal of water management. Animal Welfare and Livestock Production are part of Live Stock Management.

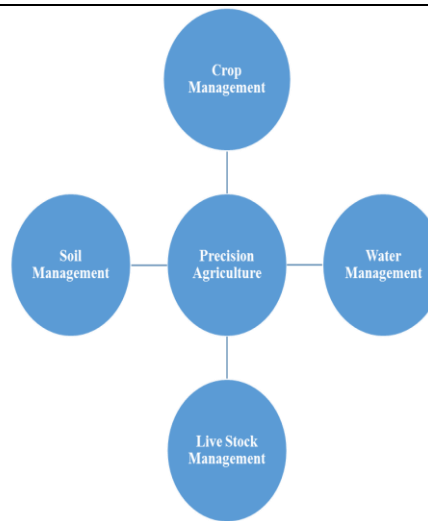


Figure 2. General Classification of Precision Agriculture

The survey conducted by Liakos et al.[10] proposed that various ML models have been applied in soil management, crop management, water management, Weed control, Live Stock Management in Agriculture. The study revealed that ANNs being the most popular model among Eight machine learning models (ANN, SVM, Clustering, DT, Regression, EL, Bayesian Algorithms (BA), Instance Based Algorithm(IBA)) have been implemented. SVMs were the most popular model in livestock management and ANNs were the most popular model in soil, water, weed management. ANNs and SVM models were used most often in crop management.

The study by K. Jhajharia et al. [14]covered the papers that examine how ML algorithms are used in different agricultural domains from the year 2005 to 2019. In Crop Prediction, seven ML algorithms (Decision Tree, Clustering, ANN, DL, Ensemble Algorithm, SVM, IBA) were utilized. Study of the publications revealed given ANN is the prevalent method for Crop Prediction. SVM and regression performed well among various ML algorithms of Regression, SVM, Ensemble Algorithm, ANN, DL were employed in Soil management. In Pest Management, SVM being the prominent choice among the ML algorithms of SVM, BA, DL, ANN, clustering. In weed management, SVM emerged as the predominant option among five ML algorithms of SVM, DL, ANN, BA, DT were implemented. Lastly, for crop disease, four ML algorithms named as SVM, ANN, DL, IBA were applied, with SVM being the major choice. Hence, based on the literature review, it can be inferred that ANN and SVM are the most commonly used algorithms in these agricultural applications.

The survey of Benos et al. [38] furnishes an extensive overview of utilizing ML methodologies in agriculture. Authors explained the potential of a variety of ML models and their families, including ANN, EL, SVM, DL, Regression, IBM, Dimensionality Reduction (DR), BA, and Clustering, to improve crop yield, disease detection, and pest management.

The most frequent and effective ML model is ANNs, comprising 51.8% of the studies. Among ANNs, Convolutional Neural Networks (CNNs) stand out, excelling in all sub-categories due to their efficiency in image-based detection. Recurrent neural networks, including Long short-term memory, constitute around 10% of ANNs, handling sequential data with memory retention. Other ANNs like Multi-layer perceptron, fully convolutional networks, and Radial basis function networks perform well in 3-5% of ANNs. Less common ANNs include adaptive-neuro fuzzy inference system (ANFIS), subtractive clustering fuzzy inference system (SCFIS), back-propagation neural networks (BPNNs), modular artificial neural networks (MANNs), deep belief networks (DBNs), TSagi-Sugeno fuzzy neural networks (TS-FNN), and feed forward neural networks (FFNNs). Second-best among ML models is EL, contributing around 22.2%. EL combines multiple inducers to improve decision-making, particularly in supervised ML tasks. SVM follows at 11.5%, known for accurate pattern learning and classification. DT and Regression models rank at 4.7%, found in all categories. Less prevalent but effective models include Clustering (0.3%), BA (0.9%), DR (1.5%), and IBM (2.7%). Overall, ANN and SVM dominate, followed by EL and other models, depending on their applicability and performance in different agricultural domains.

The implementation of machine learning in Agriculture facilitates improved precision and efficiency in farming, minimizing the need for extensive human labor while ensuring high-quality production.

IV. SOIL MANAGEMENT

Soil is indeed a complex and heterogeneous natural resource. It comprises both organic matter which supports plant growth and inorganic components. Soil exhibits a splendid array of diversity in both its chemical composition and physical attributes[39]. Physical properties of soil include textures, color, depth, structure, porosity, and stone content. Soil's chemical properties include pH level, nutrient content, organic matter, salinity and mineral composition.

The growth of plants is influenced by soil structure, as it impacts the movement of water, air, and nutrients to the plants. These properties play a crucial role in determining soil fertility, plant growth and overall soil health[40]. There are three fundamental soil types: sand, silt, and clay. Sand is characterized by its coarse, gritty texture, consisting of tiny rock fragments[41]. Clay exhibits a sticky or greasy consistency when wet and hardens significantly when dry. Silt falls in between with a texture that lies between that of sand and clay. Loam, considered the optimal soil for most plants, is a blend of sand, silt, and clay, enriched with a substantial amount of organic matter. In nature, nearly all soils exhibit combinations of these three soil types along with varying proportions of organic matter. Consequently, these soils are categorized as loam but their specific characteristics differ based on the relative proportions of clay, silt, sand and organic material they contain[42]. Understanding the different types of soil and their properties is essential for effective soil management and conservation[43].

Soil management deals with issues like soil damage from nature or excessive use of fertilizers. To keep soil healthy, it's important to rotate crops properly to prevent erosion. Soil analysis provides valuable insights for farmers and consumers helping to determine the timing and quantity of fertilizer and farmyard manure needed at various stages of a crop's growth cycle[44]. Accurately predicting the soil's characteristics is a crucial step in determining the "selection of crop, land preparation, selection of seed, crop yield, and selection of fertilizers". The location's climate and geography have an impact on the soil's characteristics. Foreseeing soil nutrients, soil surface humidity, and meteorological conditions throughout the crop's lifecycle are the main components of predicting soil qualities. Traditional soil mapping uses digital elevation models, aerial photos, and Landsat photos which are subsequently verified against actual data. Conventional methods for assessing soil typically involve laboratory analysis and soil sampling, which are usually costly and time-consuming. However, the use of remote sensing and soil mapping sensors offers an economical and effortless approach to studying the spatial variability of soil [43]. Digital soil mapping uses analytical and experimental observational methods paired with spatial and non-spatial soil inference systems to develop and update spatial soil information systems. Challenges arise when dealing with the fusion and management of heterogeneous big data, where conventional analysis methods fall short. ML techniques can provide a reliable and cost-effective solution for addressing these challenges.

V. ML APPROACHES FOR ASSESSING AND PREDICTING SOIL PARAMETERS:

This segment addresses the application of ML in predicting and identifying properties of agricultural soil. This includes estimations of soil fertility, drying conditions, temperature, moisture content, organic compounds, etc. A summary of some literature reviews in the study area is provided in Table 2.

According to the Systematic Survey conducted by Motia and Reddy [45], a review was undertaken to assess how various ML techniques contribute to soil analysis. The classification of soil property assessment and evaluation, including soil property analysis, fertilizer recommendation, soil physio-chemical property prediction, and nutritional conditions, was the main objective of the study. The review work also described various kinds of machine learning methods that are applied to agricultural soil studies in order to perform predictive modeling. Regression-based techniques are the most widely used approach for predictive modeling in soils for farming, according to the findings. Back propagation neural networks (BPNN) and SVM were the preferred methods for estimating soil nutrients. As the top machine learning techniques for forecasting soil parameters for soil health management, SVM and RF have emerged. The most widely used methods for predicting the physio-chemical characteristics of agricultural soils were Ridge Regression (RR), RF, and Least Absolute Shrinkage and Selection Operator (LASSO). RR and RF were found to be the most suitable solutions for fertilizer recommendation applications.

The majority of ML-based analysis applications in soil health management were favored by regression-based models. The best metrics for assessing the effectiveness of machine learning (ML) models used in the analysis of soil were found to be Root Mean Squared Error (RMSE) and coefficient of determination (R^2).

A brief description of literature review related to different ML models in soil management is provided here.

To optimize agricultural practices and minimize environmental harm, the development of soil management zones (MZs) is crucial. The study Maleki et al. [46] employed machine learning methods to delineate MZs based on various soil properties. Two hundred and two (202) soil samples were collected from pomegranate, pistachio and saffron agricultural areas of Bajestan, Iran. Environmental covariates were used to map the properties of the soil using the RF model. The validation of soil properties indicated a range of 'Lin's concordance correlation coefficient' (CCC) values from 0.65 to 0.79, $R^2 > 0.50$, MAE=0.06–46 and RMSE=0.02–67.58 with the maps highlighting deficiencies in total nitrogen, available phosphate, available potassium, and soil organic carbon throughout the region. Using PCA and fuzzy k-means method, it was determined that the ideal quantity of MZs in the research region is four (4). The connections between soil characteristics and environmental covariates were used to identify four distinct MZs. The soil quality map showed that MZ4 had the highest ranking in terms of soil fertility, followed by MZ1, MZ3, and MZ2.

To improve sustainable agricultural landscape management, accurate soil mapping is essential. The study Adeniyi et al. [47] assessed the efficacy of linear and nonlinear machine learning models in predicting various soil properties in the agricultural lowlands of Lombardy, Italy. An ensemble learning model using a stacking approach was employed, but it did not surpass the performance of the individual base learners. While the nonlinear single models, particularly RF demonstrated strong performance. The results of RF are of CCC (Mean, SD), RMSE (Mean, SD) of clay, SOC, PH, Topsoil Depth are (0.76, 0.08), (1.85, 0.53), (0.34, 0.13), (0.73, 0.29), (0.55, 0.06), (0.32, 0.07), (0.60, 0.10), (5.38, 1.28) respectively. The stacking models did not show superior results.

The measurement of soil water and salt contents is frequently done using time domain reflectometry (TDR), but accuracy can be affected by various factors, particularly in salinized soils. The article by Wan, Qi, and Shang[48] explains how to improve the estimation of soil characteristics, such as GWC and VWC, TS, and BD, by using eight machine learning algorithms (MLR, KNN, ANN, SVM, Cubist, RF, GBRT, and XGB) and various model input schemes. Soil particle-size fractions were found to be essential inputs for forecasting all target soil properties in the Hetao Irrigation District in Northwest China. Notably, XGB and GBRT demonstrated strong performance with XGB recommended for accurate GWC and BD estimation of $R^2 = 0.80$ and 0.69 respectively and GBRT for precise VWC and TS estimation of $R^2 = 0.71$ and 0.84 respectively.

The presence of the hazardous element cadmium (Cd) in rice poses a significant concern for human health. It's still difficult to estimate grain Cd content from soil characteristics. The study by Huang et al. [49] covers a comprehensive three-year survey encompassing six hundred and one regional pairs of soil and rice sample pairs. It was observed that the majority of both soil and rice samples exceeded safety limits for Cd. Fermi-Mn oxide-bound Cd, soil pH, field soil moisture content, and the amount of soil reducible manganese were identified as important factors influencing grain Cd concentration by both machine learning and linear regression techniques. Predicting grain Cd concentrations at a regional scale was most successfully accomplished by SVM ($R^2 = 0.87$), followed by RF ($R^2 = 0.67$) and BP-NN models ($R^2 = 0.64$).

For efficient soil management and regulation, accurate mapping of space via remote sensing is required, as soil organic matter is a critical indicator of soil nutritional status. Based on spectral response features in Northeast China, Zhou et al.'s study [50] proposed two new soil indicators: GDVIrededge2 and NLirededge2. The study successfully mapped SOM with high accuracy of $R^2 = 0.91$, MBE = 0.49, RMSE = 0.95, RPIQ=3.25 compared to various ML algorithms of RF, SVR, EBR, and LR by utilizing UAV-based multispectral imagery and the random forest machine learning method. The results highlighted a negative relationship between altitude and the content of SOM, providing valuable insights for agricultural decision-making and UAV-based monitoring of SOM.

Soil organic matter is vital for soil fertility and ecosystem health. The objective of the research by Khalaf and Mustafa [51] was to use RF and extreme gradient boosting (XGBoost) models to map SOM levels in the northern Iraqi Batifa region. Ninety-six soil samples were collected from croplands and soil areas, in addition to Landsat 8 remote sensing data. With respect to accuracy, the XGBoost model outperformed the RF model ($R^2 = 0.79$, MAE = 0.65, RMSE = 0.96,) with values of RMSE = 0.62, $R^2 = 0.92$ and MAE = 0.41.

Soil organic carbon is a crucial element upon which soil quality relies. For the purpose of attaining sustainable soil management, it is therefore essential to understand the geographical distribution and basic variables influencing SOC. The article by Meliho et al. [52] explains about SOC prediction for the Ourika watershed in Morocco was conducted using four ML algorithms: SVM, Cubist, GBM, and RF. Three distinct depths were used to collect a comprehensive set of 420 samples of soil (0–10 cm, 10–20 cm, and 20–30 cm), enabling the measurement of SOC concentration and BD, leading to the determination of SOCS. Eighty-eight variables were included in the modelling data, which integrated factors related to the environment like topography, soil characteristics, climate, and satellite imagery variables used as predictors. The most accurate models for predicting SOC were found to be RF (RMSE = 1.2%, $R^2 = 0.79$) and Cubist (RMSE = 1.2%, $R^2 = 0.77$), whereas none of the models were able to predict BD across the watershed with any degree of quality. The models with the highest predictive capacities

for SOCS were Cubist (RMSE = 11.62 t/ha, $R^2 = 0.86$) and RF (RMSE = 13.26 t/ha, $R^2 = 0.79$).

The temperature of the soil affects how the land and atmosphere interact, which is important for biological, physical, and chemical mechanisms in ecosystems on earth. Four machine learning techniques— CART, ANN, ELM, and GMDH—are compared in the study by Alizamir et al. [53] in order to estimate monthly temperatures of the soil at various depths of 5, 10, 50, and 100 cm. To develop these kinds of models, various combinations of environmental variables are employed as input. The best technique for estimating soil temperatures is found to be ELM, outperforming the other techniques. Additionally, the study notes a decrease in the models' performance as soil depth increased. Interestingly, it is discovered that soil temperatures at depths of 5, 10 and 50 cm can be predicted using solely air temperature data while the inclusion of wind speed and solar radiation data is necessary for estimating soil temperature at the 100 cm depth.

The balance of life on land depends on the temperature of the soil. To understand it better, the study by Feng et al.[54] proposed four computer models: BPNN, ELM, RF and GRNN. On the Chinese Loess Plateau, they aimed to determine whether these models could forecast soil temperature every 30 minutes at depths of 2 cm, 5 cm, 10 cm, and 20 cm. They measured soil temperature and other weather details in the field. They based their models on data such as temperature of the air, vapor pressure, moisture, sunlight, and speed of the wind. In order to predict soil temperature at all depths, they discovered that the RF, BPNN, ELM, and GRNN models performed admirably. Faster than the others and marginally superior was the ELM model.

Farmers can benefit from knowing the soil moisture content ahead of time. This study Prakash, Sharma and Sahun.d.[44] employs various machine learning methods including MLR, SVR, and RN to forecast the soil moisture for the next one, two, and seven days. The analysis involves three distinct datasets sourced from various online repositories. Model performance is assessed using MSE and R^2 . Results indicate that MLR outperforms other methods, displaying R^2 and MSE values for the next seven days as follows: 0.786 and 1.59, 0.939 and 0.353 for two days, and 0.975 and 0.14 for one day. Most cases that were connected to finding out the various soil's properties, RF, XGBoost Methods performing well. To summarize, the choice of most appropriate algorithm should be determined by the specific objectives of the soil management task, the nature of the data available, and the particular challenges facing agricultural practices today. For successful implementation and interpretation of results, domain expertise and data-driven insights must be integrated. For quick reference the summary of all these studies along with ML models used and Best output is presented in the following Table 2.

Table 2: A Summary of ML models used in the Soil Management.

Reference	Characteristics	Data input	Features	Algorithms and Models & Tools	Optimal result
Maleki et al. [46]	Soil Properties and fertility.	202 soil samples.	Developing soil management zones (MZs).	1. RF, PCA, fuzzy k-means clustering. 2. IBM SPSS 22.	1. RF Model: R2 > 0.50, CCC=0.65–0.79 MAE=0.06–46 and RMSE=0.02– 67. 2. Rank wise soil fertility: MZ4 > MZ1 > MZ3 > MZ2.
Adeniyi et al.[47]	Soil Texture, SOC, pH, Topsoil depth.	130 Soil Samples.	Predict and map the spatial distribution of different soil properties.	1. Cubist, GLM, GBM, RF, SVM, Stack_GLM, Stack_GBM. 2. R software.	RF model: Sand: CCC=0.77, RMSE=5.07 Slit: CCC=0.74, RMSE=4.99 Clay: CCC=0.76, RMSE=1.85 SOC: CCC=0.34, RMSE=0.73 pH: CCC=0.55, RMSE=0.32 Topsoil depth: CCC=0.60, RMSE=5.38.
Wan, Qi, and Shang [48]	Soil Properties.	173 Soil Samples.	Spatial distribution soil Properties.	1. Multiple LR, RF, KNN, SVM, ANN, XGB, Cubist, and GBRT. 2. R Software.	1. XGB Model: GWC: $R^2 = 0.80$ BD: $R^2 = 0.69$. 2. GBRT Model: VWC: $R^2 = 0.71$ TS: $R^2 = 0.84$.

Huang et al. [49]	Soil Properties.	601 pairs of rice and soil samples.	Predict Cd concentration of grain from soil properties.	1. BPNN, SVM, RF. 2. R Software.	SVM Model: $R^2 = 0.87$.
Zhou et al. [50]	Soil Organic Matter (SOM).	118 Soil Samples.	Modeling SOM inversion and SOM mapping.	1. RF, SVR, EBR, LR. 2. R Software.	RF Model: RPIQ=3.25, RMSE=0.95, $R^2 = 0.91$ and, MBE=0.90.
Khalaf and Mustafa [51]	SOM.	96 Soil Samples.	Digital Mapping of SOM.	1. XGBoost, RF. 2. R Software.	XGBoost Model: $R^2 = 0.92$, RMSE=0.62 and, MAE=0.1.
Meliho et al. [52]	Soil Organic Component (SOC), BD, SOC Stock (SOCS).	420 Soil Samples.	Spatial Modelling Of SOC.	1. RF, Cubist, SVM, GBM. 2. R Software.	1. RF Model for SOC: $R^2 = 0.79$, RMSE=1.2%. 2. No Model to demonstrated Satisfactory result for BD. 3. Cubist Model for SOCS: RMSE = 11.62 t/ha, $R^2 = 0.86$.
Alizamir et al. [53]	Soil Temperature.	Various types of variable	Monthly temperature of the soil in depths of 5, 10, 50, and 100 cm.	ELM, CART, ANN, and GMDH.	ELM is performed well. At 5 cm: RMSE=6.711, $R^2 = 0.571$ At 10 cm: RMSE=7.110, $R^2 = 0.461$ At 50 cm: RMSE=6.335, $R^2 = 0.390$ At 100 cm: RMSE=3.215, $R^2 = 0.915$.
Feng et al. [54]	Soil Temperature.	Various types of variable	Half Hourly Soil Temperatures 2, 5, 10, 20 cm in So depths.	ELM, GRNN, BPNN, RF.	ELM model MAE=1.37, RMSE=1.74 at 2 cm MAE=1.44, RMSE=1.85 at 5 cm MAE=1.60, RMSE=2.05 at 10 cm MAE=1.91, RMSE=2.47 at 20 cm.
Prakash, Sharma, and Sahu n.d. [44]	Soil Moisture.	Three different data set (569, 4749 92 samples) in different time Periods.	Soil moisture prediction for the first, second, and seventh days.	Multiple LR, RNN and SVR.	MLR model 1 day: MSE=0.14, $R^2 = 0.975$ 2 days: MSE= 0.353, $R^2 = 0.939$ 7 days: MSE= 1.59, $R^2 = 0.786$.

PCA: Principal Component Analysis; CCC: Lin’s Concordance Correlation Coefficient; MAE: Mean Absolute Error; GBM: Gradient Boosting Model; GLM: Generalized Linear Model; Stack_GBM: Stacking Generalization GBM; Stack_GLM: Stacking Generalization GLM; SOC: Soil Organic Component; GWC: Gravimetric water contents; VWC: Volumetric Water Contents; TS: Total Salt content; BD: Bulk Density; EBR: Elastic Bayesian Ridge; MBE: Mean Bias Error; RPIQ: Ratio of Performance to Interquartile distance; XGBoost: Extreme Gradient Boosting; MLR: Multiple Linear Regression;

VI. CONCLUSION

Researchers have focused a lot of attention on ML-based methods to increase agricultural productivity. This review highlights the different ML methods used to predict soil properties, soil fertility, soil temperature, SOM, SOC etc., in Agriculture over the past few years. It provides valuable insights into the current research in this area and offers useful information on predicting soil attributes and its management. According to the findings, machine learning methods are currently the most effective in forecasting soil characteristics. In summary, diverse methods are employed for distinct purposes when forecasting soil conditions. For general soil

parameter prediction, the best machine learning techniques are XGBoost and RF. Also, RMSE and R^2 are widely used to evaluate how well predictive algorithms perform.

REFERENCES

- [1] L. Bheemavarapu and K. U. Rani, "Machine Learning Models Used For Prakriti Identification Using Prasna Pariksha In Ayurveda – A Review," vol. 72, no. 1, pp. 1942–1951, 2023.
- [2] A. Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine Learning Applications for Precision Agriculture: A Comprehensive Review," IEEE Access, vol. 9, pp. 4843–4873, 2021, doi: 10.1109/ACCESS.2020.3048415.
- [3] F. Selection and S. Data, "Chapter 11 Spectrometry Data," pp. 25–48, doi: 10.1007/978-1-60327-194-3.
- [4] H. Fujiyoshi, T. Hirakawa, and T. Yamashita, "Deep learning-based image recognition for autonomous driving," IATSS Res., vol. 43, no. 4, pp. 244–252, 2019, doi: 10.1016/j.iatssr.2019.11.008.
- [5] A. K. Rai and R. K. Dwivedi, "Fraud Detection in Credit Card Data using Unsupervised Machine Learning Based Scheme," Proc. Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2020, no. Icesc, pp. 421–426, 2020, doi: 10.1109/ICESC48915.2020.9155615.
- [6] T. Gangavarapu, C. D. Jaidhar, and B. Chanduka, Applicability of machine learning in spam and phishing email filtering: review and approaches, vol. 53, no. 7. 2020. doi: 10.1007/s10462-020-09814-9.
- [7] Q. Cheng, S. Zhang, S. Bo, D. Chen, and H. Zhang, "Augmented Reality Dynamic Image Recognition Technology Based on Deep Learning Algorithm," IEEE Access, vol. 8, pp. 137370–137384, 2020, doi: 10.1109/ACCESS.2020.3012130.
- [8] L. Kranjč, "Machine-Learning Classification of a Number of Contaminant Sources in an Urban Water Network," pp. 1–15, 2021.
- [9] A. Anagnostis, L. Benos, D. Tsaopoulos, A. Tagarakis, N. Tsolakis, and D. Bochtis, "applied sciences Human Activity Recognition through Recurrent Neural Networks for Human – Robot Interaction in Agriculture," 2021.
- [10] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," Sensors (Switzerland), vol. 18, no. 8, pp. 1–29, 2018, doi: 10.3390/s18082674.
- [11] H. M. Reeve, A. M. Mescher, and A. F. Emery, "Experimental and numerical investigation of polymer preform heating," Am. Soc. Mech. Eng. Heat Transf. Div. HTD, vol. 369, no. 6, pp. 321–332, 2001, doi: 10.1115/imece2001/htd-24365.
- [12] P. S. Baburao, R. B. Kulkarni, P. A. Kharade, and S. S. Patil, Review of Machine Learning Model Applications in Precision Agriculture, no. May. Atlantis Press International BV, 2023. doi: 10.2991/978-94-6463-136-4_81.
- [13] I. Lopez-Arevalo, E. Aldana-Bobadilla, A. Molina-Villegas, H. Galeana-Zapién, V. Muñoz-Sanchez, and S. Gausin-Valle, "A memory-efficient encoding method for processing mixed-type data on machine learning," Entropy, vol. 22, no. 12, pp. 1–21, 2020, doi: 10.3390/e22121391.
- [14] K. Jhajharia and P. Mathur, "A comprehensive review on machine learning in agriculture domain," vol. 11, no. 2, pp. 753–763, 2022, doi: 10.11591/ijai.v11.i2.pp753-763.
- [15] O. Folorunso et al., "Exploring Machine Learning Models for Soil Nutrient Properties Prediction: A Systematic Review," Big Data Cogn. Comput., vol. 7, no. 2, p. 113, 2023, doi: 10.3390/bdcc7020113.
- [16] D. Hsu, "Linear regression Empirical risk minimization," vol. 2, no. Coms 4771, pp. 1–5.
- [17] K. G. Liakos, P. Busato, D. Moshou, and S. Pearson, "Machine Learning in Agriculture : A Review," no. ML, pp. 1–29, doi: 10.3390/s18082674.
- [18] M. P. Lavalley, "Statistical Primer for Cardiovascular Research Logistic Regression," pp. 2395–2399, 2008, doi: 10.1161/CIRCULATIONAHA.106.682658.
- [19] A. D. Dongare, R. R. Kharde, and A. D. Kachare, "Introduction to Artificial Neural Network," vol. 2, no. 1, pp. 189–194, 2012.
- [20] I. Journal, O. Advance, and K. Shiruru, "AN INTRODUCTION TO ARTIFICIAL," no. September 2016, 2017.
- [21] V. Sharma and A. Dev, "A Comprehensive Study of Artificial Neural Networks," vol. 2, no. 10, pp. 278–284, 2012.
- [22] L. Vanneschi and S. Silva, "Support Vector Machines," Nat. Comput. Ser., pp. 271–281, 2023, doi: 10.1007/978-3-031-17922-8_10.
- [23] W. S. Noble, "What is a support vector machine?," Nat. Biotechnol., vol. 24, no. 12, pp. 1565–1567, 2006, doi: 10.1038/nbt1206-1565.
- [24] F. Sets and J. C. Bezdek, "Generalized k-Nearest Neighbor rules," vol. 0114, no. March, 2018, doi: 10.1016/0165-0114(86)90004-7.
- [25] K. Hajebi and Y. Abbasi-yadkori, "Fast Approximate Nearest-Neighbor Search with k -Nearest Neighbor Graph," 2009.
- [26] B. Rajagopalan and U. Lall, "A k-nearest-neighbor simulator for daily precipitation," vol. 35, no. 10, pp. 3089–3101, 1999.
- [27] A. Priyam, R. Gupta, A. Rathee, and S. Srivastava, "Comparative Analysis of Decision Tree Classification Algorithms," pp. 334–337, 2013.
- [28] D. Landgrebe, "iiiiiiiiiiiiiiiiiiiiiiiiiiii /-// -," 1990.
- [29] G. I. Webb, "Naïve Bayes," no. January 2016, 2019, doi: 10.1007/978-1-4899-7502-7.
- [30] D. Berrar, "Bayes ' Theorem and Naive Bayes Classifier," vol. 1, no. 2018, pp. 403–412.
- [31] Y. Qi, "Random Forest for Bioinformatics," pp. 1–18.
- [32] L. E. O. Breiman, "Random Forests," pp. 5–32, 2001.
- [33] C. Bent, A. Cs, and G. Mart, "A Comparative Analysis of XGBoost arXiv : 1911 . 01914v1 [cs . LG] 5 Nov 2019," pp. 1–20.
- [34] J. H. Friedman, "Stochastic Gradient Boosting 1 Gradient Boosting," vol. 1, no. 3, pp. 1–10, 1999.
- [35] S. Sharma and A. Srushtideep, "Precision Agriculture and Its Future," Int. J. Plant Soil Sci., no. December, pp. 200–204, 2022, doi: 10.9734/ijpss/2022/v34i242630.
- [36] V. Meshram, K. Patil, V. Meshram, D. Hanchate, and S. D. Ramkteke, "Machine learning in agriculture domain: A state-of-art survey," Artif. Intell. Life Sci., vol. 1, no. October, p. 100010, 2021, doi: 10.1016/j.aillsci.2021.100010.
- [37] V. Meshram, K. Patil, V. Meshram, D. Hanchate, and S. D. Ramkteke, "Machine learning in agriculture domain: A state-of-art survey," Artif. Intell. Life Sci., vol. 1, no. October, p. 100010, 2021, doi: 10.1016/j.aillsci.2021.100010.
- [38] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine learning in agriculture: A comprehensive updated review," Sensors, vol. 21, no. 11, pp. 1–55, 2021, doi: 10.3390/s21113758.
- [39] P. Taneja, H. Kumar, P. Daggupati, and A. Biswas, "Geoderma Multi-algorithm comparison to predict soil organic matter and soil moisture content from cell phone images," vol. 385, no. December 2020, 2021.
- [40] L. Zhong, X. Guo, Z. Xu, and M. Ding, "Geoderma Soil properties : Their prediction and feature extraction from the LUCAS spectral library using deep convolutional neural networks," vol. 402, no. July, 2021.
- [41] B. Bajat, "Geoderma Soil type classification and estimation of soil properties using support vector machines," vol. 154, pp. 340–347, 2010, doi: 10.1016/j.geoderma.2009.11.005.
- [42] P. S. Baburao, R. B. Kulkarni, P. A. Kharade, and S. S. Patil, Review of Machine Learning Model Applications in Precision Agriculture, no. June. Atlantis Press International BV, 2023. doi: 10.2991/978-94-6463-136-4_81.
- [43] A. Rachmad et al., "Ensemble classifier to support decisions on soil classification Ensemble classifier to support decisions on soil classification", doi: 10.1088/1757-899X/1022/1/012044.
- [44] S. Prakash, A. Sharma, and S. S. Sahu, "SOIL MOISTURE PREDICTION USING".
- [45] S. Motia and S. R. N. Reddy, "Exploration of machine learning methods for prediction and assessment of soil properties for agricultural soil management: A quantitative evaluation," J. Phys. Conf. Ser., vol. 1950, no. 1, 2021, doi: 10.1088/1742-6596/1950/1/012037.
- [46] S. Using and M. Learning, "Scale Using Machine Learning," pp. 1–19, 2023.
- [47] O. D. Adeniyi, A. Brenning, A. Bernini, S. Brenna, and M. Maerker, "Digital Mapping of Soil Properties Using Ensemble Machine Learning Approaches in an Agricultural Lowland Area of Lombardy, Italy," Land, vol. 12, no. 2, 2023, doi: 10.3390/land12020494.
- [48] H. Wan, H. Qi, and S. Shang, "Estimating soil water and salt contents from field measurements with time domain reflectometry using machine learning algorithms," Agric. Water Manag., vol. 285, no. January, p. 108364, 2023, doi: 10.1016/j.agwat.2023.108364.
- [49] B. Y. Huang et al., "Machine learning methods to predict cadmium (Cd) concentration in rice grain and support soil management at a regional scale," Fundam. Res., no. xxxx, 2023, doi: 10.1016/j.fmre.2023.02.016.
- [50] J. Zhou et al., "High-Precision Mapping of Soil Organic Matter Based on UAV Imagery Using Machine Learning Algorithms," Drones, vol. 7, no. 5, pp. 1–20, 2023, doi: 10.3390/drones7050290.
- [51] H. S. Khalaf and Y. T. Mustafa, "applied sciences Digital Mapping of Soil Organic Matter in Northern Iraq : Machine Learning Approach," 2023.
- [52] M. Meliho et al., "Spatial Prediction of Soil Organic Carbon Stock in the Moroccan High Atlas Using Machine Learning," Remote Sens., vol. 15, no. 10, 2023, doi: 10.3390/rs15102494.

- [53] M. Alizamir et al., "Advanced machine learning model for better prediction accuracy of soil temperature at different depths," PLoS One, vol. 15, no. 4, pp. 1–25, 2020, doi: 10.1371/journal.pone.0231055.
- [54] Y. Feng, N. Cui, W. Hao, L. Gao, and D. Gong, "Estimation of soil temperature from meteorological data using different machine learning models," Geoderma, vol. 338, no. June 2018, pp. 67–77, 2019, doi: 10.1016/j.geoderma.2018.11.044.