JOURNAL OF APPLIED SCIENCES TECHNOLOGY AND COMPUTING
e-ISSN: 3036-0250

JASTEC

Vol. 1 No. 1 (2024) 29-38
https://publisher.uthm.edu.my/ojs/index.php/jastec

Machine Learning for Soil Classification: Challenges and Opportunities

Siti Nur Fatin Liyana Mohd Azmin[1], Hamijah Mohd Rahman[2], Nur Nadhirah Mohd Harith Lim[3], Nureize Arbaiy[4]*

[1] *Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, MALAYSIA*

*Corresponding Author: nureize@uthm.edu.my
DOI: https://doi.org/10.30880/jastec.2024.01.01.004

Article Info

Received: 30 October 2023
Accepted: 04 February 2024
Available online: 27 February 2024

Keywords

Soil classification, machine learning, human expert availability, human expert consistency, agriculture, environmental science, land management, crop selection

Abstract

In agriculture and environmental science, soil classification is essential for making well-informed decisions about crop selection, land management, and environmental protection. However conventional methods of classifying soil require a lot of work and time, and they mostly rely on human expertise. This work investigates the possibilities of machine learning (ML) models to automate soil classification utilizing large datasets of soil samples to overcome the shortcomings of existing techniques. In this paper, many machines learning techniques, including support vector machines (SVM), decision trees (DT), random forests (RF), and neural networks (NN), are examined for the classification of soil. There are certain models that work better than others, though, and this is based on the qualities of the soil samples. In addition to that, experiments using Random Forest, Naïve Bayes, and k-Nearest Neighbor (k-NN) were also undertaken. Classification strategies are being chosen to create a classified model using data mining. The algorithm with the highest accuracy is Random Forest (97.23%), followed by Naïve Bayes (96.82%), and k-Nearest Neighbor (k-NN), which has the lowest accuracy (92.92%). The paper highlights the challenges of applying machine learning to soil classification, such as consistency and human specialist availability, to effectively categorize soil samples. The results indicate that, despite these challenges, ML models present a potential substitute for labor-intensive conventional methods in the classification of soil.

1. Introduction

Since long ago, agriculture has been regarded as the primary cultural practice. Understanding the type of soil to utilize for agricultural cultivation is essential for achieving the highest crop yield because soil is such an important aspect of agriculture. The kind of soil can be identified using a variety of techniques, including technological advancements, experience, and conventional procedures [1]. Understanding the soil's classification makes it easier to predict its behavior [2]. Through soil behavior and properties, observation prediction of soil potential can be done for agricultural growth. Soils are categorized and given names based on the chemical and physical characteristics of their horizons. Soil taxonomy classifies soil based on its color, texture, structure, and other characteristics. Soil classification is a way to arrange soil-related knowledge. It's because different soils have different physical, chemical, and biological characteristics [3]. The classification of soil allows land to be grouped or classes with similar properties and behaviors (chemical, physical, and biological), and it can be geo-mapped and referenced [4].

For years, traditional farmers have been deeply acquainted with the soils in their area, and they use this information to choose crops that are compatible with the soil's characteristics [5]. As an illustration, a farmer may be aware that some soil types are more suited for growing rice, while others are better for growing wheat. Depending on the region and the requirements of the farmers, traditional soil classification systems can be more complicated than these examples. Nonetheless, the foundation of every conventional system for classifying soil is an in-depth comprehension of the soil and how it affects plant development [6]. Conventional techniques for classifying soil need a lot of work and time. This is because conventional approaches necessitate the tedious and costly procedure of gathering and analyzing soil samples in a laboratory [7], [8]. Because machine learning (ML) models can be trained on massive datasets of soil samples, they provide a prospective substitute for traditional soil classification techniques in automatically classifying different types of soil. This can increase the accuracy of soil classification while also saving a substantial amount of time and effort.

In several fields, such as environmental science and agriculture, soil classification is crucial because it helps with decision-making about crop selection, land management, and environmental preservation. Traditional methods of classifying soil, however, frequently require a lot of time and labor and mostly rely on human expertise. There are drawbacks to this reliance on human expertise, including possible discrepancies and restricted availability. Given these circumstances, machine learning (ML) shows promise as a means of resolving the problems related to soil classification. Because ML models are trained on large databases of soil samples, they can automate and expedite the classification process [23]. Accurate and effective soil classification is made feasible by utilizing the capabilities of machine learning algorithms.

The effectiveness of various ML models for soil classification is discussed in this work. Few research papers are examined, and the outcomes demonstrated that every model could accurately classify different types of soil. Additionally, an experiment is carried out with three algorithms to classify the soil for crop suggestion: Naïve Bayes, Random Forest, and k-Nearest-Neighbors (k-NN). The acquired soil data is used to train and test these three algorithms for soil classification. Depending on the unique characteristics of the soil samples, some models performed better than others.

The paper also notes the difficulties and possibilities associated with classifying soil using machine learning. A difficulty with machine learning models is their high training data requirements. Another difficulty is that the quality of the training data might have an impact on how well machine learning models perform. The study suggests that ML is a potential approach for soil classification despite these obstacles. A wide range of stakeholders, including farmers, land managers, and environmental scientists, can profit from the development of precise and effective soil classification systems through the application of machine learning algorithms. This study is significant because it provides an inclusive evaluation of different ML models for soil classification. The study also identifies the challenges and opportunities of using ML for soil classification, which can help researchers and practitioners to develop and deploy ML-based soil classification systems more effectively.

This paper contains several important sections. Section 1 gives the introduction; section 2 describes the domain application which is soil application. Section 3 review the machine learning models. Section 4 provides opportunities and challenges in the context of machine learning applications in soil classification. Section 5 ends the discussion.

2. Soil Classification

Most of the Earth's land area is covered in loose surface material, or soil [9]. It is made up of inorganic particles and organic stuff that work together to support life. In addition to providing a medium for plant growth, water, and nutrients, soil also supports agricultural plants structurally [10]. Weathering, leaching, and microbial activity all have an impact on the wide range of characteristics found in soil [11]. For a farming effort, determining the appropriate soil type is crucial to the growth of healthy plants. Soil classification is the process of separating soil groups or classes that may exhibit similar behavior and share characteristics that allow for georeferencing and mapping [4]. The measurement and description of several attributes of representative soil profiles that are indicative of soil formation processes serve as the foundation for the classification of soils [12]. Nevertheless, these attributes can also be utilized separately or in combination to produce soil databases or maps that categorize soils based on the requirements of a potential user.

Soil classification is a method for identifying the kind of soil is [13];[14]. Soil classification is an important task in agriculture, environmental science, and land management. It helps to identify the different types of soil in each area, which can inform decisions about crop selection, land management, and environmental protection. The first official attempt to classify soils appears to have taken place in China about 400 years ago [15]. The oldest known soil classification system in the world is found in the ancient Chinese book Yugong (2500 y.b.p.),

J. of Applied Sciences Technology and Computing Vol. 1 No. 1 (2024) p. 29-38

where soil color, texture, and hydrologic features were used to classify China's soils into three categories and nine classes for land evaluation [16], [17]. However, the assessment of the land classification system can be broken down into three phases. The first soil classification method (Russia, USDA 1938) produced soil profiles by classifying soils into three categories based on the environment and soil-forming variables: zonal, azonal, and intrazonal. Following that, the soil is described using the soil parameters to create a soil classification. A notable example of the letter technique is the French classification system [18];[19]. Concurrently, the publication of the USDA Soil Taxonomy signaled the advent of contemporary soil and classification [20]. It lists the properties of the soil that have been calculated and evaluated using technology [21], [22]. Country extension agents provide maps of the local soil classification based on soil surveys to assist farmers in selecting which crops to plant in which regions,

Traditional soil classification methods require labor and time. These steps include:

1. Sampling: Soil samples are collected from different depths and locations within a given area.
2. Preparation: Soil samples are processed for analysis by drying, grinding, and sieving.
3. Analysis: Soil samples are analyzed to determine their physical and chemical properties, such as texture, pH, organic matter content, and nutrient levels.
4. Classification: Soil samples are then classified based on their physical and chemical properties.

Traditional methods of classifying soils are frequently more comprehensive than contemporary ones, taking into consideration a variety of elements like the soil's function in the local ecosystem and its relationship to plant growth. Because traditional approaches don't require specialized knowledge or equipment, they are also frequently easier for farmers and other land users to implement. However, compared to modern techniques, traditional approaches for classifying soil may be less accurate and may not always be able to predict the behavior of the soil. The traditional method of determining soil type takes a long time and mostly depends on the knowledge of agricultural specialists. Results from this process can be inconsistent and unreliable due to human error and procedural delays. It can also be more challenging to standardize ancient procedures, which can hinder the exchange of soil data between other nations or areas. The comparison of conventional and contemporary approaches for classifying soil is shown in **Table 1**.

Table 1 *Traditional vs modern soil classification system*

Characteristic	Traditional	Modern
Precision	Less precise	More precise
Standardization	Less standardized	More standardized
Range of soil properties considered	holistic, considering connections between ecosystems and plant growth	Numerous factors related to soil are considered, such as its chemistry, biology, and mineralogy.
Accessibility	More accessible to farmers and other land users	Less accessible to farmers and other land users
Appropriateness for all regions	More appropriate for all regions	Less appropriate for all regions

3. Machine Learning Models

Artificial intelligence (AI) in the form of machine learning (ML) enables software programs to improve their prediction accuracy without having to be specifically designed to do so [23] – [25]. The foundation of machine learning is the notion that computers are capable of learning from data and gradually becoming more efficient. Algorithms for machine learning are trained on large datasets of labelled data, or data that has been given the appropriate output values. The algorithm can be taught and used to forecast the output values of fresh data [26]. While machine learning algorithms come in a wide variety of forms, they may be roughly categorized into three groups: reinforcement learning, unsupervised learning, and supervised learning. Machine learning algorithms forecast new values by utilizing past data as input.

Classification constitutes a supervised machine learning technique wherein the algorithm gains knowledge from the input data and applies that knowledge to categorize subsequent observations [26]. This approach yields a set of sample data with its classes and determines what data needs to be recognized. There are two stages to it: training and testing. During the training phase, the training set must decide which parameter to focus on and how to combine the various data kinds into a single form. Testing will involve applying the set to test data with a predetermined goal and comparing it with selected data from the training set [27], [28], [36]. To provide further details, the testing set will produce a result that shows how long it takes to precisely analyse each item of data and assess whether it has a high degree of accuracy.

Classifying soil types according to their suitability for various uses, including construction, agriculture, or environmental cleanup, can also be done with ML models. The classification of soil could be completely changed using ML. To categorize soil types automatically, machine learning models can be trained on massive datasets of soil samples. In addition to saving a great deal of time and effort, this can increase the classification accuracy of soil. A variety of machine learning models include Support vector machines (SVM), Decision trees (DT), Random forests (RF) and Neural networks (NN).

Support vector machines (SVMs) [37]: Supervised learning models, or SVMs, are a kind of model that may be applied to tasks involving both regression and classification. For SVMs to function, a hyperplane with the greatest feasible margin that divides the data into two classes must be found. SVMs have demonstrated remarkable efficacy in soil classification tasks, attaining excellent accuracy across multiple datasets.

Decision trees (DTs) [38]: Another kind of supervised learning model that may be applied to regression and classification applications is the deep learning model (DT). DTs function by constructing a tree structure that illustrates the connections between the various data aspects. After that, fresh data points are classified using the tree by going through it and adhering to the decision criteria at each node. DTs can be used to create classification criteria and are generally easy to read.

Random forests (RFs) [39]: An ensemble learning model called an RF combines the predictions of several different DTs to provide a forecast that is more accurate. For soil classification tasks, RFs have proven to be quite effective, delivering high accuracy across a range of datasets.

Neural networks (NNs) [40]: A class of machine learning models called neural networks (NNs) is modelled after the composition and operations of the human brain. Neural networks (NNs) consist of a network of interconnected nodes that analyse and interpret data. For soil classification tasks, NNs have proven to be quite effective, reaching high accuracy on a range of datasets. NNs, however, can be complicated and challenging to train.

Other machine learning models that have been used for soil classification include:

- K-nearest neighbors (KNN)
- Naïve Bayes classifiers
- Logistic regression
- Boosting
- Stacking

The optimal machine learning model for a given soil classification task will vary depending on the details of the data, including the kind of soil classes, number of features, and dataset size. For soil classification tasks, it has been demonstrated that SVMs, RFs, and NNs are the most efficient machine learning models [29] – [32]. Even with a very limited number of training samples, these models can attain excellent accuracy on a range of datasets. These models are appropriate for large-scale soil classification projects because they are computationally efficient and have demonstrated good accuracy in a range of soil classification applications.

It is crucial to remember that the quality of the training data will determine how well any machine learning model performs. The model's ability to generalize to new data will be hindered if the training data is noisy, incomplete, or not representative of the target population. Thus, before training a machine learning model for soil classification, it is crucial to carefully choose and prepare the training data.

4. Challenges and Opportunities

4.1 Challenges

Challenges of using machine learning for soil classification are summarized in the following:

(a) Data Quality and Availability:

Availability and quality of data provide major obstacles to using machine learning models for soil classification. A large and varied training dataset is one of the main needs for machine learning models to learn efficiently. On the other hand, gathering high-quality soil data can be expensive and time-consuming. Careful planning and execution are needed to guarantee that the soil samples are representative of the target area and appropriately labelled.

Furthermore, soil data's temporal and geographical variability adds even another level of complication. The properties of soil can range greatly between locations and evolve throughout time as a result of both natural and man-made activities. In such cases, it becomes difficult to train a model that can generalize to new and unknown data. To capture the intrinsic variability and guarantee the robustness of the model, a wide range of soil samples from different places and times must be included.

Proper collection and processing efforts are needed to address these difficulties. To reduce mistakes and discrepancies in the soil data that has been gathered, strict quality control procedures must be put in place. Creating thorough and representative datasets that capture the temporal and geographical variability in soil properties is also essential for training well-generalizable models.

Advances in data collection methods, remote sensing technologies and collaborative efforts across academia and institutions can help address challenges related to data availability and quality. Better decision-making in land management, environmental science and agriculture can be achieved by improving the effectiveness and reliability of machine learning models for soil classification through improved quality and accessibility of soil data.

(b) Interpretability of the Model

A major obstacle in the context of classifying soils is the interpretability of machine learning models. Understanding the inner workings of a machine learning model and interpreting its predictions becomes more challenging as the model grows more complex and advanced. We may be less able to understand how the model makes predictions and determine the important factors that influence its output as a result of its interpretability issues.

Interpretability is important in the field of land classification. Determining which soil properties or qualities most strongly influence classification decisions is important, as is understanding the reasoning behind model predictions. Decisions related to agriculture, environmental science and soil management can be made with greater understanding thanks to the ability of domain experts to see underlying patterns and relationships in soil data.

The interpretability of machine learning models is being improved. In order to obtain insights into the decision-making processes of complicated models, researchers are creating methods and procedures for doing so. One aspect of this is the creation of post-hoc interpretability techniques, which offer justifications for the model's predictions after they have been produced. These techniques assist in comprehending and verifying the model's predictions by displaying the internal workings of the model or highlighting its most important characteristics.

To build trust and confidence in the outputs of machine learning models for soil classification, interpretability must be promoted. It enables interested parties to evaluate the model's dependability and comprehend how well it fits in with the body of current domain knowledge. In order to guarantee that the insights obtained from these models are significant and useful, it is imperative to strike a balance between model complexity and interpretability.

(c) Computational Resources:

One of the challenges in applying and training machine learning models for land classification, especially for small firms, is the availability of computational resources. To process and analyze large data sets and train complex models, machine learning algorithms often demand large amounts of processing power. Accessing and deploying the infrastructure required to successfully apply machine learning to land classification may prove challenging for small enterprises with low computational resources.

Small firms may have lengthier processing times and decreased productivity as a result of the computational demands of machine learning models. Furthermore, for businesses with tight budgets, the expense of purchasing and maintaining the hardware or cloud-based resources required to carry out resource-intensive machine learning operations may be a hurdle.

To solve this problem, other approaches must be investigated, such as making use of cloud-based platforms or distributed computing frameworks that facilitate the effective use of existing resources. Smaller enterprises can potentially get over these resource constraints through joint ventures and collaborations with larger organizations or research institutes that have the necessary computational capacity.

4.2 Opportunities

Opportunities of using machine learning for soil classification are given as follows:

Precision agriculture, environmental monitoring, land management, and other fields greatly benefit from machine learning models since they are more accurate and efficient than traditional soil classification methods. Machine learning models have the potential to reduce the time and effort necessary for soil classification by processing and analyzing data more quickly. This is achieved by utilizing huge datasets and sophisticated algorithms. Furthermore, these models may be able to attain higher accuracy levels by identifying intricate linkages and patterns in the data that might be difficult for human specialists to recognize. In fields linked to land management, environmental protection, and precision agriculture, better decision-making and resource allocation are made possible by machine learning's increased efficiency and accuracy in classifying soil, which eventually results in improved outcomes and sustainable practices.

Through the analysis of massive amounts of soil data, machine learning algorithms have the potential to provide new understandings and insights about soil. These algorithms have the capacity to uncover previously undiscovered patterns and relationships, which can yield important information. This newfound comprehension advances our understanding of the patterns and functions of soil, resulting in the creation of better soil management strategies. We may get important knowledge about the properties, composition, and behavior of soil by utilizing machine learning. This knowledge will help us make better informed decisions and put more successful plans for sustainable soil management methods into action.

Increased automation of soil classification procedures through machine learning models could result in improved efficiency and resource management. Automated soil type classification can be achieved by training machine learning models on large datasets of soil samples, hence decreasing the need for labor-intensive human procedures. In addition to speeding up the process of classifying soil, automation frees up labor resources, enabling specialists to concentrate on other important assignments and initiatives. Organizations can obtain better overall results by optimizing operations, increasing productivity, and more wisely allocating human resources by utilizing machine learning for soil classification.

The benefits of applying machine learning to soil classification generally exceed the drawbacks. It's crucial to recognize the difficulties and take action to lessen them, though. For instance, it's crucial to have access to the required computational resources, to apply interpretable machine learning models, and to properly choose and prepare the training data. The following are some instances of the potential for soil classification that machine learning can present:

- Precision agriculture: By using machine learning, farmers may minimize their environmental effect and maximize crop yields through the development of precision agriculture systems. Machine learning algorithms, for instance, can be used to forecast soil fertility and suggest the best rates for applying fertilizer.
- Environmental monitoring: Systems for tracking changes in soil conditions over time can be developed using machine learning. Areas in danger of pollution or soil erosion can be found using this information.
- Land management: Systems for managing land that preserve and enhance soil health can be created with the use of machine learning. Machine learning algorithms, for instance, can be used to determine which regions are best suited for growing kinds of crops and to suggest sustainable land management techniques.

Traditional methods for land classification and management have already been revolutionized by machine learning, demonstrating its transformative potential. It has created new opportunities for more accurate and effective land management and classification techniques. To optimize the advantages that machine learning may bring to the world, continuous research is focused on improving these techniques. To ensure that machine learning plays an important role in improving land classification and management procedures and ultimately produces favorable consequences worldwide, this growing field of study aims to unlock even greater improvements.

5. Applications

Three relevant research papers were reviewed and summarized as follows:

(i) Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis [33]

This study forecasts mustard crop output using machine learning based on soil analysis [17]. This study used five supervised machine learning techniques—k-Nearest Neighbor (k-NN), Naïve Bayes, Multinomial Logistic Regression, Artificial Neural Network (ANN), and Random Forest—to estimate Mustard Crop production from soil data. They concluded that machine learning techniques might be successfully used for yield prediction based on the experiment's outcomes. The results of this study's experiment showed that k-NN and Random Forest predicted the highest accuracy (88.67 % and 94.13 %, respectively), while Naïve Bayes predicted the lowest accuracy (72.33 %), ANN predicted 76.86 %, and Multinomial Logistic Regression predicted 80.24 %.

(ii) Data Mining Classification Algorithms for Analyzing Soil Data [34]

Researchers employed Naïve Bayes, Random Forest, Decision Tree, and k-Nearest Neighbor (k-NN) to determine soil type [18]. Using classifier approaches, these algorithms extract information from soil data. Finding the optimal machine learning for soil classification is the primary goal of these four classifiers. In the testing data (53.85%), k-NN has the highest accuracy of 84% when compared to Naïve Bayes (69.23%), Decision Tree, and Random Forest. As a result, it performs better than other classification methods. Based on the data, it appears that k-NN could be useful for classifying agricultural soil types.

(iii) Ensemble Classifier to Support Decisions of Soil Classification [35]

Three popular classification models—k-Nearest-Neighbor (k-NN), Naïve Bayes, and Decision Tree—were tested by [24] using a publicly accessible agricultural soil dataset. The investigation aids in obtaining machine learning methodologies and domain knowledge in soil science to address a variety of soil research issues. For calculating accuracy and assessing performance, every classifier model is applied and assessed using the same dataset. The Naïve Bayes algorithm yielded the lowest accuracy of 72.90% in the testing, while the k-NN approach achieved an accuracy of 73.56%. At 80.84%, Decision Tree demonstrated the highest accuracy. Using the same dataset, the fused or recommended approach outperforms the other three classifiers with an accuracy of 84.14%. The study's conclusions indicated that, in terms of accuracy, the recommended ensemble classifier performed better than the widely used three classifiers.

The effectiveness of the algorithms used in three research publications that used the soil data set to identify the soil for crop suggestions is examined in **Table 2**.

Table 2 *Algorithms outcome comparison*

Paper	Algorithm (s)	Accuracy (%)
Pandith et al. [33]	Naïve Bayes	72.33
	k-Nearest-Neighbor (k-NN)	88.67
	Multinomial Logistic Regression	80.24
	Random Forest	94.13
	Artificial Neural Network (ANN)	76.86
Taher et al. [34]	Naïve Bayes	69.23
	Decision Tree	53.84
	Random Forest	53.84
	k-Nearest-Neighbor (k-NN)	84.61
Motia and Reddy [35]	Decision Tree	80.84
	k-Nearest-Neighbor (k-NN)	73.56
	Naïve Bayes	72.90
	Ensemble Classifier	84.14
This study	Random Forest	97.23
	Naïve Bayes	96.82
	k-Nearest-Neighbor (k-NN)	92.92

Additionally, an experiment has been carried out in this study. As a training and testing model, Random Forest, Naive Bayes, and k-NN were used. In this experiment, the two models for each approach must be generated using the training and testing models. Because each algorithm cannot be applied directly due to data and algorithm incompatibility, each model will be used in the training model to ensure that the algorithm can be trained using the crop data set.

The main measure of performance is evaluated in terms of accuracy and kappa from the confusion matrix of classification. The measures and experimental errors are computed by using equations that are described in the following:

1. Accuracy:

$$Accuracy = \frac{TN+TN}{TP+TN+FN+FP} = \frac{TP+TN}{P+N} \tag{1}$$

2. kappa:

$$\mathcal{K} = \frac{p_o - p_e}{1-p_e} \tag{2}$$

3. Errors:

$$Errors = 1 - accuracy \tag{3}$$

The tests aimed to examine the effectiveness of three algorithms on crop recommendation datasets: Random Forest, Naïve Bayes, and k-Nearest Neighbor (k-NN). The accuracy and kappa performance measurements were used to assess the study's results. Four distinct cross-validation strategies were used to guarantee robustness: 50-50, 60-40, 70-30, and 80-20. Table 2 provides a summary of the experiment's findings, including those from the three algorithms, the performance measures-based evaluation, and the experimental error analysis.

Among Naïve Bayes and k-Nearest Neighbor (96.82 % and 92.92 %), Random Forest has the best accuracy of 97.23%, according to this study, which employed 2570 data size. A comparison with the findings of a study from [33] that employed 5000 data sizes further demonstrates that Random Forest has an accuracy of 94.13% and is proved to perform better than other algorithms. However, the experimental results in this study demonstrate that Naïve Bayes has a lower accuracy of 72.33%. With an accuracy of 88.67%, k-NN is in second place in the meanwhile. In the study, Random Forest performed the worst with 53.84 % accuracy, followed by Naïve Bayes with 69.23 % [34]. Compared to these two approaches, k-NN performs better with an accuracy of 84.61.

Nevertheless, compared to this study and [33], the data set utilized by [34] is smaller. Based on the comparing results, Random Forest is utilized in many applications to categorize large datasets and offers better accuracy when used with larger data sets. When the results and performance were evaluated, Random Forest outperformed the other two classifiers, Naïve Bayes and k-NN, in terms of accuracy. Given its capabilities, Random Forest might be the ideal option for studies pertaining to agricultural soils.

6. Conclusion

Each type of soil classification technique, traditional and modern, has benefits and drawbacks of its own. Traditional procedures are more extensive and accessible, but modern technologies offer higher precision and standardization. The kind of soil information to be used, the degree of accuracy needed, and the resources at hand all play a role in the soil classification technique selection. A hybrid system that blends aspects of traditional and contemporary methods may prove advantageous in specific scenarios. Machine learning-based techniques for classifying soil are predicted to become more and more common as technology advances. The benefits of both conventional and contemporary methods can be combined to produce accurate and readily available soil information by utilizing machine learning models. Among the many benefits of machine learning models are the reduction of dependency on human specialists and the resolution of availability and consistency concerns. These algorithms can swiftly handle vast amounts of soil data, which produces classification results that are more reliable and appropriate. Furthermore, complicated patterns and interactions in soil records can be handled by machine learning algorithms, which can provide insights that human specialists might miss. In the end, this improves the precision and thoroughness of soil classification.

Acknowledgement

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 (Vot Q507).

References

[1] Lombardozzi, D. L., Lu, Y., Lawrence, P. J., Lawrence, D. M., Swenson, S., Oleson, K. W., ... & Ainsworth, E. A. (2020). Simulating agriculture in the Community Land Model version 5. Journal of Geophysical Research: Bio geosciences, 125(8), e2019JG005529.

[2] Olivares, B. O., Araya-Alman, M., Acevedo-Opazo, C., Rey, J. C., Cañete-Salinas, P., Kurina, F. G., ... & Gómez, J. A. (2020). Relationship between soil properties and banana productivity in the two main cultivation areas in Venezuela. *Journal of soil science and plant nutrition*, *20*, 2512-2524.

[3] Arnold, R.W (2004) Classification of Soils. In D. Hillel (Ed.), *Encyclopedia of Soils in the Environment* (pp. 204-210). ScienceDirect.

[4] Food and Agriculture Organization of the United Nations. (n.d.) Soil classification. Retrieved August 17, 2021, from http://www.fao.org/soils-portal/data-hub/soil-classification/en/

[5] Zeraatpisheh, M., Jafari, A., Bodaghabadi, M. B., Ayoubi, S., Taghizadeh-Mehrjardi, R., Toomanian, N., ... & Xu, M. (2020). Conventional and digital soil mapping in Iran: Past, present, and future. *Catena*, *188*, 104424.

[6] Kappali, H. R., Sadyojatha, K. M., & Prashanthi, S. K. (2023). Computer Vision and Machine Learning in Paddy Diseases Identification and Classification: A Review. *Indian Journal Of Agricultural Research*.

[7] Sharu, M. B., Yakubu, M., Noma, S. S., & Tsafe, A. I. (2013). Characterization and classification of soils on an agricultural landscape in Dingyadi District, Sokoto State, Nigeria. *Nigerian Journal of Basic and Applied Sciences*, *21*(2), 137-147.

[8] Sharu, M. B., Yakubu, M., Noma, S. S., & Tsafe, A. I. (2013). Characterization and classification of soils on an agricultural landscape in Dingyadi District, Sokoto State, Nigeria. Nigerian Journal of Basic and Applied Sciences, 21(2), 137-147.

[9] Bronick, C. J., & Lal, R. (2005). Soil structure and management: a review.

[10] Lin, B. B. (2011). Resilience in agriculture through crop diversification: adaptive management for environmental change. BioScience, 61(3), 183-193.

J. of Applied Sciences Technology and Computing Vol. 1 No. 1 (2024) p. 29-38

[11] Mitchell, J. K., & Soga, K. (2005). Fundamentals of soil behavior (Vol. 3). New York: John Wiley & Sons.

[12] Butler B. E. and B. Butler, Soil Classification for Soil Survey. Oxford: Clezedon Press.

[13] Singh, R. K., Sinha, V. S. P., Joshi, P. K., & Kumar, M. (2020). Modelling Agriculture, Forestry and Other Land Use (AFOLU) in response to climate change scenarios for the SAARC nations. *Environmental Monitoring and Assessment*, *192*, 1-18.

[14] A. John, D. Leslie and S. Sarah, 2021. Available: https://environment.uwe.ac.uk/geocal/SoilMech/classification/default.htm.

[15] Robert J. A, USDA, NRCS; Thomas J. Rice, Jr., Calif. Poly. State Univ., SLO & Hari Eswaran, USDA, NRCS. (2002). National Cooperative Soil Survey Newsletter. (19), 1.

[16] Krasilnikov, P., Arnold, R. W., & Ibáñez, J.-J. (2010). Soil classifications: Their origin, the state-of-the-art and perspectives. 2010 19 World Congress of Soil Science, Soil Solutions for a Changing World. https://www.researchgate.net/publication/255964346_Soil_classifications_Their_origin_the_state-of-the-art_and_perspectives

[17] Gong Z. (Ed) (1994) 'Chinese soil taxonomic classification (First proposal)'. (Institute of Soil Science, Academia Sinica: Nanjing)

[18] Ahrens, R. J., Rice, T. J., & Eswaran, H. (2002). Soil classification: past and present. In Soil Classification (pp. 19-26). CRC Press.

[19] Ikazaki, K., Nagumo, F., Simporé, S., & Barro, A. (2018). Soil toposequence, productivity, and a simple technique to detect petroplinthites using ground-penetrating radar in the Sudan Savanna. *Soil Science and Plant Nutrition*, *64*(5), 623-631.

[20] Hughes, P., McBratney, A. B., Minasny, B., Huang, J., Michéli, E., Hempel, J., & Jones, E. (2018). Comparisons between USDA soil taxonomy and the Australian Soil Classification system II: Comparison of order, suborder and great group taxa. Geoderma, 322, 48-55.

[21] Salehi, M. H. (2018). Challenges of Soil Taxonomy and WRB in classifying soils: Some examples from Iranian soils. Phys. Geogr. Ser, 14, 63-70.

[22] Paramananthan, S. (2000). Soils of Malaysia: their characteristics and identification, Volume 1. Academy of Sciences Malaysia.

[23] Kubat, M., Bratko, I., & Michalski, R. S. (1998). A review of machine learning methods. *Machine Learning and Data Mining: Methods and Applications*, 3-69.

[24] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.

[25] Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). *Automated machine learning: methods, systems, challenges* (p. 219). Springer Nature.

[26] Frank, M., Drikakis, D., & Charissis, V. (2020). Machine-learning methods for computational science and engineering. *Computation*, *8*(1), 15.

[27] Rahman, S. A. Z., Mitra, K. C., & Islam, S. M. (2018, December). Soil classification using machine learning methods and crop suggestion based on soil series. In *2018 21st International Conference of Computer and Information Technology (ICCIT)* (pp. 1-4). IEEE.

[28] Nguyen, M. D., Costache, R., Sy, A. H., Ahmadzadeh, H., Van Le, H., Prakash, I., & Pham, B. T. (2022). Novel approach for soil classification using machine learning methods. *Bulletin of Engineering Geology and the Environment*, *81*(11), 468.

[29] Belgiu, M., & Drăguţ, L. (2011). Random forests for soil classification. In 2011 IEEE International Geoscience and Remote Sensing Symposium (pp. 3524-3527). IEEE.

[30] Wu, W., Li, A. D., He, X. H., Ma, R., Liu, H. B., & Lv, J. K. (2018). A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. Computers and Electronics in Agriculture, 144, 86-93.

[31] Beucher, A., Møller, A. B., & Greve, M. H. (2019). Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark. Geoderma, 352, 351-359.

[32] Inazumi, S., Intui, S., Jotisankasa, A., Chaiprakaikeow, S., & Kojima, K. (2020). Artificial intelligence system for supporting soil classification. Results in Engineering, 8, 100188.

J. of Applied Sciences Technology and Computing Vol. 1 No. 1 (2024) p. 29-38

[33] Pandith, V., Kour, H., Singh, S., Manhas, J., & Sharma, V. (2020). Performance evaluation of machine learning techniques for mustard crop yield prediction from soil analysis. *Journal of scientific research*, *64*(2), 394-398.

[34] Taher, K. I., Abdulazeez, A. M., & Zebari, D. A. (2021). Data mining classification algorithms for analyzing soil data. *Asian Journal of Research in Computer Science*, *8*(2), 17-28.

[35] Motia, S., & Reddy, S. R. N. (2021). Ensemble classifier to support decisions on soil classification. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1022, No. 1, p. 012044). IOP Publishing.

[36] Azmin, S. N. F. L. M., & Arbaiy, N. (2022). Soil Classification based on Machine Learning for Crop Suggestion. *Journal of Soft Computing and Data Mining*, *3*(2), 79-91.

[37] Steinwart, I., & Christmann, A. (2008). Support vector machines. Springer Science & Business Media.

[38] Navada, A., Ansari, A. N., Patil, S., & Sonkamble, B. A. (2011, June). Overview of use of decision tree algorithms in machine learning. In 2011 IEEE control and system graduate research colloquium (pp. 37-42). IEEE.

[39] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

[40] Müller, B., Reinhardt, J., & Strickland, M. T. (1995). Neural networks: an introduction. Springer Science & Business Media.