Exploring Machine Learning Models for Soil Nutrient Properties Prediction: A Systematic Review

Folorunso O.
Department of Computer Science
Federal University of Agriculture Abeokuta
Ogun State, Nigeria
folorunsoo@funaab.edu.ng

Ojo O.
Department of Computer Science
Federal University of Agriculture Abeokuta
Ogun State, Nigeria
ojoeo@funaab.edu.ng

Busari M.
Department of Soil Science and Land Management
Federal University of Agriculture Abeokuta
Ogun State, Nigeria
busarima@funaab.edu.ng

Adebayo M.
Safefood Africa Agroenterprise, Abeokuta
Nigeria
adebayomadebayo@gmail.com

Joshua A.
Department of Computer Science
Federal University of Agriculture Abeokuta
Ogun State, Nigeria
aesther661@gmail.com

Folorunso D.
Department of Computer Science
Federal University of Agriculture Abeokuta
Ogun State, Nigeria
folorunsodaniel5@gmail.com

Ugwunna C.O.
Department of Computer Science
Federal University of Agriculture Abeokuta
Ogun State, Nigeria
ugwunnaco@funaab.edu.ng

Olabanjo O.
Department of Computer Science
Federal University of Agriculture Abeokuta
Ogun State, Nigeria
femstarola11@gmail.com

Olabanjo O.
Department of Computer Science
Federal University of Agriculture Abeokuta
Ogun State, Nigeria
olabanjoolusola@yahoo.com

Abstract

Agriculture is essential to a flourishing economy. Although soil is essential for sustainable food production, its quality can decline as cultivation becomes more intensive and demand increases. The importance of healthy soil cannot be overstated, as a lack of nutrients can significantly lower crop yield. Smart soil prediction and digital soil mapping offer accurate data on soil nutrient distribution needed for precision agriculture. Machine learning techniques are now driving intelligent soil prediction systems. This article provides a comprehensive analysis of the use of machine learning in predicting soil qualities. The components and qualities of soil, the prediction of soil parameters, the existing soil dataset, the soil map, the effect of soil nutrients on crop growth, are the key subjects under inquiry. Smart agriculture, as exemplified by this study, can improve food quality and productivity.

5th Deep Learning Indaba Conference (DLI 2023).

1 Soil Components and Properties

Sustainable agricultural growth and enhanced crop yields are both feasible consequences of land reclamation and productive resource management. Increased yields can be obtained in intensive cropping by using adequate nutrition sources and application rates [1]. Soil quality fundamentally means "the ability of a soil to function"; this ability can be indicated by the estimated soil's physical, chemical, and biological qualities, often known as soil quality indicators (SQI) [2]. Several soil investigations may be envisaged to adequately quantify the soil framework, and science-based indices on SQI provide valuable data to farm managers for decision making. These indices incorporate important soil attributes, including supplying suitable amounts of water and nutrients, resisting and recovering from physical degradation, and supporting plant growth with the right management [3]. Sustainable farmland management requires an in-depth familiarity with the relationships between soil physical qualities and many agronomic and environmental factors [4]. The availability of nutrients is influenced by the soil's chemical and physical properties, such as its parent material and naturally occurring minerals, organic matter, depth to bedrock, sand, or gravel, permeability, water-holding capacity, and drainage. The distribution of nutrients is also determined by plant and atmospheric conditions [5]. The nutrient concentration in the soil solution is influenced by soil water content, depth, pH, cation-exchange capacity, redox potential, soil organic matter, microbial activity, season, and fertilizer application [6]. It is typically time-consuming and costly to estimate and evaluate soil components and qualities. Predictive soil mapping is a common modeling approach used to estimate the spatial distribution of soil components when actual data from samples are unavailable. Many of these approaches rely on predictive maps or the estimation of soil-related variables at unmeasured locations based on field data using mathematical or statistical models of relationships between soil and other environmental elements [7].

1.1 Research Justification

The ability of ML-based methods to accurately forecast soil characteristics, crop growth, and soil fertility has attracted a lot of attention in recent years. Texture, organic matter, pH, nutrient content, soil moisture, and soil structure are just a few of the many soil variables that may be analysed with the ML approach. ML techniques are superior to traditional statistical methods because of their capacity to process massive amounts of complex data and reveal hidden patterns. Several studies have focused on developing ways for applying machine learning to predict soil parameters [8,9,10], crop growth [11,12,13], and soil fertility [14,15]. Recently, a systematic literature review that highlights the research gaps in certain applications of deep learning techniques and evaluates the influence of vegetation indicators and environmental factors on agricultural productivity was published in [16]. The authors examined prior studies from 2012 to 2022 from various databases. The article focuses on the benefits of employing deep learning in agricultural yield prediction, the best remote sensing technology depending on data collection requirements, and the numerous factors that influence crop yield prediction. In general, several studies have demonstrated the efficacy of machine learning algorithms in predicting soil properties, soil fertility, and crop yields. It is vital to keep in mind, though, that ML models' accuracy is extremely sensitive to the quantity and quality of data used in training, in addition to the algorithms and parameters with which they are implemented. Further research is needed to investigate how to construct and refine ML models for predicting soil parameters and evaluate how well they function in different environmental and soil circumstances. Farmers, policymakers, plant breeders, and other professionals in the agricultural sector can all benefit from ML recommendations.

1.2 Soil Dataset

To determine the nutrient level, composition, and other properties of a soil sample, scientists conduct a soil test. Soil testing can involve a variety of techniques and fertilizer recommendations to determine the soil's fertility and pinpoint any deficiencies that need to be addressed. Soil analysis provides information useful to farmers and consumers in deciding when and how much fertilizer and farmyard manure should be administered during a crop's growth cycle [17]. Soil datasets entail information on land suitability for agricultural production, soil maturity, soil texture, meteorological data, moisture content, soil classes, soil colour, covariate data, soil nutrients, and trace elements.

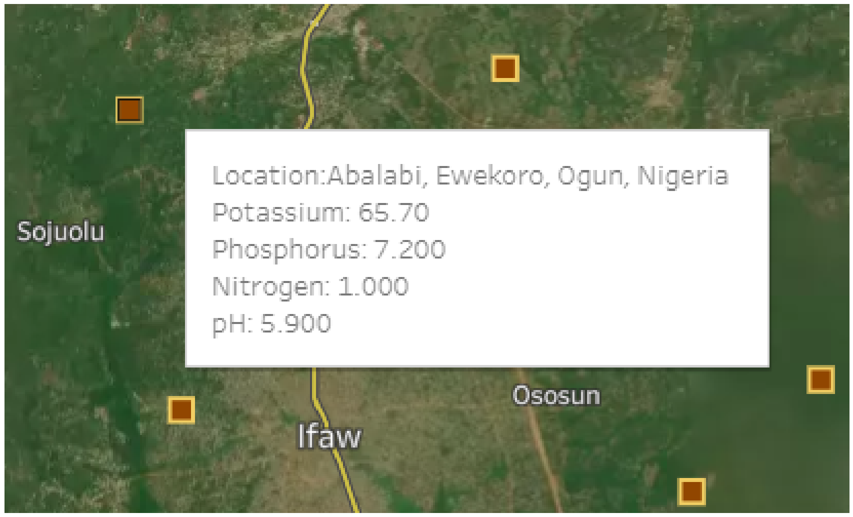

Figure 1: Digital soil map depicting the soil's nutrients for a location in South-West Nigeria.

The utilisation of covariate environmental data facilitates the establishment of associations between soil properties and various environmental factors. The process of soil formation and its characteristics are impacted by several factors, including but not limited to climatic conditions, topographical features, vegetation cover, land utilisation, and the nature of the parent material. The integration of covariate data can enhance the efficacy of soil prediction models by enabling a more comprehensive understanding of the intricate interplay between soil and its surrounding ecosystem. The inclusion of covariate environmental data is imperative in soil prediction due to its ability to augment our comprehension of soil-environment associations, capture spatial heterogeneity, offer insights into fundamental mechanisms, enable data amalgamation, and facilitate informed decisions regarding land management. The integration of covariate data into soil prediction models enhances their precision and usefulness in diverse domains, such as agriculture, environmental governance, and land use management [18,19].

1.3 Soil Map

Environmental elements pertaining to geology, landforms, or vegetation are identified through the use of aerial photographs, Landsat images, and digital elevation models (DEMs) in traditional digital soil mapping. The method is then checked against real-world data [20]. The final outcome is a map labeled with soil classifications, which can be confusing to users. Furthermore, there are other issues caused by mapping's subjective character [21]. In traditional soil surveys, the soil is mapped according to the surveyor's preconceived notions [22]. Classical mapping's conceptual framework was established using quantitative and statistical methods. The method of developing and updating spatial soil information systems via analytical and experimental observational methods paired with spatial and non-spatial soil inference systems is generally known as digital soil mapping [23]. The digital soil map depicted in Figure 1 presents an illustration of the soil nutrient distribution in a specific area located in Ogun State, situated in the south-west region of Nigeria. In prior studies, a digital soil map was considered a digitized conventional soil map in the form of polygons [24]. However, because the map was not created using statistical inference, it cannot be construed as a digital soil map, but rather a digitized soil map. SCORPAN is a mnemonic for an empirical quantitative description of relationships between soil and environmental factors with a view to using these as soil spatial prediction functions for the purpose of digital soil mapping where each letter stands for the following: S = soil classes or attributes, f = function, s = soil, other or previously measured properties of the soil at a point, c = climate, climatic properties of the environment at a point, o = organisms, including land cover and natural vegetation or fauna or human activity, r = relief, topography, landscape attributes, p = parent material, lithology, a = age, the time factor, n = spatial or geographic position.

The initial development of the SCORPAN framework for use in digital soil mapping was accomplished by [25]. Spatial soil prediction functions with an auto-correlated error are often used to forecast soil class or soil attributes from so-called SCORPAN factors [25].

$$S_c = f(s, c, o, r, p, a, n) + e, or S_a = f(s, c, o, r, p, a, n) + e$$

'e' stands for spatially correlated residuals, where S_c and S_a are soil classes and soil properties as a function of soil, climate, organisms, relief, parent material, age, and geographical position [26]. For the quantitative prediction of soil groups or dynamic soil properties based on empirical observations, the SCORPAN model is employed.The majority of effort in digital soil mapping is based on developing a mathematical model that connects field soil data and SCORPAN variables [27,28]. Afterwards, the model is used with extensive spatial environmental data. To extrapolate, update, or disaggregate soil maps, digital soil mapping can also employ conventional soil maps as input [29,30]. The underlying principle is to employ machine learning (ML) techniques to find the knowledge inherent in completed surveys or to reverse engineer the surveyor's soil-landscape mental model [31].

2 Artificial Intelligence (AI) Models for Soil Properties Prediction

AI models have been widely employed in predicting soil attributes. Ref. [32] offered a computerized soil mapping method for preventing gully erosion by advising landowners on preventative steps. Using R-Squared, KC, and RMSE as accuracy metrics, a multiple nonlinear regression model was built with 68% precision. Nonetheless, the low accuracy is understandable given that the soil depth map is not a fair depiction of the sample in reality, making it difficult to conduct research. The use of machine learning algorithms for estimating soil depth has been explored further in [33]. Quality Reference Framework(QRF) models were utilized, and with RMSE as the measure of evaluation, they were able to reach an accuracy of 30%. It can be inferred from the accuracy percentage that soil depth in digital soil mapping is still a discoverable topic. An evaluation of soil fertility using DSM and machine learning techniques was proposed in [34]. Using QRF, great accuracy was attained by utilizing the evaluation metrics RMSE and MAE. However, the model's precision was constrained for some soil characteristics, such as nitrogen (N) and potassium (K). Soil maps for a variety of soil qualities, for which QRF was able to provide the best accuracy, is another issue that was addressed. Self-organizing maps (SOMs) were also employed as a machine learning model [35]. Supervised maps are used to forecast soil moisture using SOM and random forest (RF) models; when tested on a dataset including both soil moisture and land cover, SOM showed greater model accuracy than RF when evaluated with respect to R2 and KC. Multi-sensor data and ML algorithms, including RF, XGBoost, and SVM (supervised vector machine), were also used to make predictions about soil moisture, with an accuracy of 87.5% [36]. Many deep learning methods, such as deep neural networks (DNN) and artificial neural networks (ANN), have been used to predict soil attributes in space. With an AUC of 89.8%, DNN achieved the highest accuracy. Due to the lack of high-quality artificial intelligence solutions for digital soil mapping, researchers from all over the world are paying close attention to the field.

3 Findings and Discussion

Majority of published works (67.3%) dealt with issues of soil nutrient characteristics; 17.3% handled DSM; 11.1% addressed soil erosion; and 5.5% dealt with soil fertility. For soil prediction, RF and neural networks outperform conventional machine learning methods.

4 Conclusion

The study reviews machine learning methods for predicting soil properties, highlighting gaps in research and their findings. Challenges include inaccurate data, regional variations, and feature selection. Collaboration, model monitoring, and adaptation are crucial for refining machine learning techniques. Implementing these techniques in less developed nations faces challenges such as language barriers, limited resources, and data availability. Solutions include investing in data collection, infrastructure, education, and partnerships with foreign organizations to improve soil management and agricultural productivity.

Disclosure of Funding

This research was funded by European Union Under the Accelerating Inclusive Green Growth through Agri-Based Digital Innovation in West Africa (AGriDI), which is an ACP Innovation Fund (grant ref: FED/2020/420-491) led by the International Centre for Insect Physiology and Ecology (ICIPE), Nairobi Kenya.

References

[1] Murmu, K., Swain, D. K., & Ghosh, B. C. (2013). Comparative assessment of conventional and organic nutrient management on crop growth and yield and soil fertility in tomato-sweet corn production system. Australian Journal of Crop Science, 7(11), 1617-1626.

[2] Shukla, M. K., Lal, R., & Ebinger, M. (2006). Determining soil quality indicators by factor analysis. Soil and Tillage Research, 87(2), 194-204.

[3] Rezaei, S. A., Gilkes, R. J., & Andrews, S. S. (2006). A minimum data set for assessing soil quality in rangelands. Geoderma, 136(1-2), 229-234.

[4] Li, Y., Li, Z., Cui, S., & Zhang, Q. (2020). Trade-off between soil pH, bulk density and other soil physical properties under global no-tillage agriculture. Geoderma, 361, 114099.

[5] Fernández, F. G., & Hoeft, R. G. (2009). Managing soil pH and crop nutrients. Illinois agronomy handbook, 24, 91-112.

[6] Marschner, P., & Rengel, Z. (2023). Nutrient availability in soils. In Marschner's Mineral Nutrition of Plants (pp. 499-522). Academic press.

[7] Kovačević, M., Bajat, B., & Gajić, B. (2010). Soil type classification and estimation of soil properties using support vector machines. Geoderma, 154(3-4), 340-347.

[8] Martinelli, G., & Gasser, M. O. (2022). Machine learning models for predicting soil particle size fractions from routine soil analyses in Quebec. Soil Science Society of America Journal, 86(6), 1509-1522.

[9] Payen, F. T., Sykes, A., Aitkenhead, M., Alexander, P., Moran, D., & MacLeod, M. (2021). Predicting the abatement rates of soil organic carbon sequestration management in Western European vineyards using random forest regression. Cleaner Environmental Systems, 2, 100024.

[10] Liu, D., Liu, C., Tang, Y., & Gong, C. (2022). A GA-BP neural network regression model for predicting soil moisture in slope ecological protection. Sustainability, 14(3), 1386.

[11] Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., & Zhang, J. (2020). Prediction of winter wheat yield based on multi-source data and machine learning in China. Remote Sensing, 12(2), 236.

[12] Kuwata, K., & Shibasaki, R. (2015, July). Estimating crop yields with deep learning and remotely sensed data. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 858-861). IEEE.

[13] Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., & Fritschi, F. B. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote sensing of environment, 237, 111599.

[14] Keerthan Kumar, T. G., Shubha, C. A., & Sushma, S. A. (2019). Random forest algorithm for soil fertility prediction and grading using machine learning. Int J Innov Technol Explor Eng, 9(1), 1301-1304.

[15] Benedet, L., Acuña-Guzman, S. F., Faria, W. M., Silva, S. H. G., Mancini, M., dos Santos Teixeira, A. F., ... & Curi, N. (2021). Rapid soil fertility prediction using X-ray fluorescence data and machine learning algorithms. Catena, 197, 105003.

[16] Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H., & Islam, N. (2022). A systematic literature review on crop yield prediction with deep learning and remote sensing. Remote Sensing, 14(9), 1990.

[17] Baskar, S. S., Arockiam, L., & Charles, S. (2013). Applying data mining techniques on soil fertility prediction. Int. J. Comput. Appl. Technol. Res, 2(6), 660-662.

[18] Tziachris, P., Aschonitis, V., Chatzistathis, T., Papadopoulou, M., & Doukas, I. J. D. (2020). Comparing machine learning models and hybrid geostatistical methods using environmental and soil covariates for soil pH prediction. ISPRS International Journal of Geo-Information, 9(4), 276.

[19] Zeraatpisheh, M., Garosi, Y., Owliaie, H. R., Ayoubi, S., Taghizadeh-Mehrjardi, R., Scholten, T., & Xu, M. (2022). Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. Catena, 208, 105723.

[20] Legros, J. P. (2006). Mapping of the Soil. Science Publishers.

[21] Ryan, P. J., McKenzie, N. J., O'Connell, D., Loughhead, A. N., Leppert, P. M., Jacquier, D., & Ashton, L. (2000). Integrating forest soils information across scales: spatial prediction of soil properties under Australian forests. Forest Ecology and Management, 138(1-3), 139-157.

[22] Hudson, B. D. (1992). The soil survey as paradigm-based science. Soil Science Society of America Journal, 56(3), 836-841.

[23] Lagacherie, P., & McBratney, A. B. (2006). Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. Developments in soil science, 31, 3-22.

[24] Tomlinson, R. F. (1978). Design considerations for digital soil map systems

[25] McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. Geoderma, 117(1-2), 3-52.

[26] Florinsky, I. V. (2012). The Dokuchaev hypothesis as a basis for predictive digital soil mapping (on the 125th anniversary of its publication). Eurasian soil science, 45, 445-451.

[27] Kheir, R. B., Greve, M. H., Bøcher, P. K., Greve, M. B., Larsen, R., & McCloy, K. (2010). Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: The case study of Denmark. Journal of Environmental Management, 91(5), 1150-1160.

[28] Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., & Edwards Jr, T. C. (2015). Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma, 239, 68-83.

[29] Subburayalu, S. K., & Slater, B. K. (2013). Soil series mapping by knowledge discovery from an Ohio county soil map. Soil Science Society of America Journal, 77(4), 1254-1268.

[30] Odgers, N. P., Sun, W., McBratney, A. B., Minasny, B., & Clifford, D. (2014). Disaggregating and harmonising soil map units through resampled classification trees. Geoderma, 214, 91-100.

[31] Moran, C. J., & Bui, E. N. (2002). Spatial data mining for enhanced soil map modelling. International Journal of Geographical Information Science, 16(6), 533-549.

[32] Du Plessis, C.; Van Zijl, G.; Van Tol, J.; Manyevere, A. Machine learning digital soil mapping to inform gully erosion mitigation measures in the Eastern Cape, South Africa. Geoderma 2020, 368, 114287.

[33] Baltensweiler, A.; Walthert, L.; Hanewinkel, M.; Zimmermann, S.; Nussbaum, M. Machine learning based soil maps for a wide range of soil properties for the forested area of Switzerland. Geoderma Reg. 2021, 27, e00437.

[34] Hounkpatin, K.O.; Bossa, A.Y.; Yira, Y.; Igue, M.A.; Sinsin, B.A. Assessment of the soil fertility status in Benin (West Africa)—Digital soil mapping using machine learning. Geoderma Reg. 2022, 28, e00444.

[35] Riese, F.M.; Keller, S.; Hinz, S. Supervised and semi-supervised self-organizing maps for regression and classification focusing on hyperspectral data. Remote. Sens. 2019, 12, 7.

[36] Nguyen, T.T.; Ngo, H.H.; Guo, W.; Chang, S.W.; Nguyen, D.D.; Nguyen, C.T.; Zhang, J.; Liang, S.; Bui, X.T.; Hoang, N.B. A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. Sci. Total. Environ. 2022, 833, 155066.