Rev. Peru. Investig. Agropecu. 3(2), e63
ISSN: 2955-8530
e-ISSN: 2955-831X
DOI: 10.56926/repia.v3i2.63
Universidad Nacional Autónoma de Alto Amazonas

Original article / Artículo original

Prediction of soil fertility using machine learning in Alto Amazonas province, Peru

Predicción de la fertilidad del suelo mediante aprendizaje automático en la provincia de Alto Amazonas, Perú

César Oswaldo Arévalo-Hernández [1*]; Enrique Arévalo-Gardini [1]; Luis Alberto Arévalo-López [1]; Oscar Tuesta-Hidalgo [1]; Dayani Shirley Romero-Vela [1]; Claudia Elizabeth Ruiz-Camus [1]

[3]Facultad de Ciencias, Universidad Nacional Autónoma de Alto Amazonas, Yurimaguas, Perú

ABSTRACT
The objective of the work was to predict soil fertility in the province of Alto Amazonas with the use of satellite images and machine learning techniques. The study was in the province of Alto Amazonas in Peru. Soil sampling was carried out in all the provinces, totalling 100 samples. Afterwards, soil physical (texture) and chemical analyses were performed. Satellite images were obtained from USGS, and vegetation indexes were calculated based on these images. Finally, descriptive analysis and machine learning modelling using 06 algorithms (GLM, CUBIST, KKNN, SVM, Random Forest and NN) were used and selected based on their R2 and rmse. In this work, we observed that most soils in the province have low pH, P, Mg, K and high acidity. We also managed to achieve good predictions for pH, Ca, Mg and CEC, and we observed that the most successful algorithm was Random Forest. Nevertheless, for Al, CUBIST performed better. This is one of the first works using machine learning to predict soil fertility in the Peruvian Amazon, and we hope it may serve as a base for future projects.

Keywords: Random Forest; amazon soils; soil modelling; acid soils

RESUMEN
El objetivo del trabajo fue predecir la fertilidad del suelo en la provincia de Alto Amazonas con el uso de imágenes satelitales y técnicas de aprendizaje automático. El estudio se ubicó en la provincia de Alto Amazonas en Perú. Se realizaron muestreos de suelos en toda la provincia, totalizando 100 muestras. Posteriormente se realizaron análisis físicos (textura) y químicos del suelo. Las imágenes satelitales se obtuvieron del USGS y los índices de vegetación se calcularon con base en estas imágenes. Finalmente, se utilizó análisis descriptivo y modelado de aprendizaje automático utilizando 06 algoritmos (GLM, CUBIST, KKNN, SVM, Random Forest y NN) que se seleccionaron en función de su R2 y RMSE. En este trabajo observamos que la mayoría de los suelos de la provincia tienen bajos pH, P, Mg, K y alta acidez. También se lograron obtener buenas predicciones para pH, Ca, Mg y CIC y se observó que el algoritmo más exitoso fue Random Forest. Sin embargo, para Al, Cubist tuvo mejores resultados. Este es uno de los primeros trabajos que utiliza aprendizaje automático para predecir la fertilidad del suelo en la Amazonía peruana y se espera que pueda servir como base para futuros proyectos.

Palabras clave: Random Forest; suelos amazónicos; modelamiento de suelos; suelos ácidos

Citation / Cómo citar: Arévalo-Hernández, C.O., Arévalo-Gardini, E., Arévalo-López, L. A., Tuesta-Hidalgo, O., Romero-Vela, D.S., & Ruiz-Camus, C.E. (2024). Prediction of soil fertility using machine learning in Alto Amazonas province, Peru. *Revista Peruana de Investigación Agropecuaria*, *3*(2), e63. https://doi.org/10.56926/repia.v3i2.63

Editor: Dr. Fred William Chu Koo

Received: 15/06/2024
Accepted: 29/09/2024
Published: 10/10/2024

* carevalo@unaaa.edu.pe (corresponding author)

1. INTRODUCTION

Soil is one of the main natural components in the production of agricultural and forestry crops. To evaluate their physical, chemical or biological characteristics, random sampling is carried out to represent the areas of interest; however, this type of analysis is often expensive and spatially unrepresentative (Watt et al., 2019). However, this has been the classic way of analyzing and diagnosing the characteristics related to soil fertility and its productive capacity in agricultural and forestry crops (Delgado-Caballero et al., 2009).

With the improvement of technology and the use of geographic information systems, it is possible to carry out sampling considering variations in climate, topography, mineralogy and others, allowing mapping areas to be better identified (Campos et al., 2019). In general, the sampling scheme and sampling design are fundamental in digital soil mapping since they allow for obtaining the greatest possible representativeness of an area with the smallest number of samples (Brus, 2019). Thus, technological advances and the possibility of obtaining covariates such as those related to the DEM (digital elevation model) and climate, among others, have allowed the use of new sampling and point selection techniques such as the Latin hypercube. conditional, which also allows sampling points to be selected at the lowest cost (Yang et al., 2020), is currently one of the preferred design methods.

On the other hand, soil fertility is important for agricultural and forestry production and plays a fundamental role in food security. Soil fertility is divided into three main components: Physical, Chemical and Biological, the interaction between the three being essential to ensure the quality and sustainability of the crops (Bünemann et al., 2018). Of all these attributes, the easiest to diagnose by traditional methods are the physical and chemical ones, the former being the least variable in time and space and commonly used in many cases as covariates for predicting chemical attributes (Di Raimo et al., 2022).

Chemical attributes are highly variable in space, especially micronutrients or heavy elements, which increases the probability of misdiagnosis in an area with high variability, as is the case of tropical soils (Macedo Neto et al., 2020). In this way, it is possible to observe nutritional deficiencies in areas where fertilizers or amendments were applied due to the low dose suggested using traditional soil fertility diagnostic methods.

In this way, computer systems have evolved quite a bit, allowing the use of powerful statistical techniques such as machine learning, which base and adjust their predictions from models based on experience or, in this case, from data, one of its applications being agricultural sciences and soil science (Wadoux et al., 2020). Machine learning algorithms are diverse, and to date, more than 300 different models have been registered that can be adjusted and/or adapted to agricultural sciences to be used for the prediction of relationships between different climatic conditions, topography, management and soil fertility, with quite promising results (Dharumarajan et al., 2022; Wadoux et al., 2020).

Some cases that use this type of technique for predicting soil fertility can be found in India (Dharumarajan et al., 2022), Europe (Lu et al., 2023), Africa (Hounkpatin et al., 2022), Brazil (Vieira et al., 2021) and even for the prediction of textural classes in Antarctic soils (Siqueira et al., 2023) in all cases with high prediction indices, which allowed having an idea of the spatial variability of soil characteristics and how these are related to the landscape and environment. In this way, the present research aimed to predict soil fertility in the province of Alto Amazonas with the use of satellite images and machine learning techniques.

2. MATERIALS AND METHODS

2.1. Localization

The research will be carried out in the Province of Alto Amazonas, which has an Am type climate (Köppen, 1931). This region of the country has average annual maximums and minimums of 31.7°C and 21.8°C, respectively. The average annual accumulated precipitation is 2086.2 mm.

2.2. Imaging processing

Images will be obtained freely from the United States Geological Service (USGS), considering the date of collection of soil samples. Images from satellite Sentinel-2 will be selected, taking into consideration the cartographic base available from the company that was visualized in Google Earth. Sentinel-2 has a regular multispectral camera with 13 bands in the spectrum's visible, near-infrared and short-wave infrared parts with main applications such as agriculture, land ecosystems, forest management, and others. To improve values obtained from the satellite, calibrations and conversion will be performed to suppress the effect of atmospheric gases.

2.3. Vegetation Indexes

Proposed Vegetation Indexes (VI) are presented in Table 1 and are based on previous work executed for oil palm (Oliveira Teixeira, 2022). These VI will be calculated to predict their relationship with Alto Amazonas – Loreto soil characteristics.

Table 1.
Proposed VI for predicting soil fertility in Alto Amazonas - Loreto

Index	Equation
Atmospherically Resistant Vegetation Index (ARVI) (Kaufman & Tanré, 1992)	$\frac{NIR - (2 * R - B)}{NIR + (2 * R - B)}$
Difference Vegetation Index (DVI) (Tucker, 1980)	$NIR - R$
Green Chlorophyll Index (GCI) (Gitelson et al., 2003)	$\frac{NIR}{G} - 1$
Green Difference Vegetation Index (GDVI) (Sripada et al., 2006)	$NIR - G$
Leaf Area Index (LAI) (Boegh et al., 2002)	$3.618 * \frac{2.5 * (NIR - R)}{NIR + 6 * R - 7.5 * B + 1} - 0.118$
Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974)	$\frac{NIR - R}{NIR + R}$
Optimized Soil Adjusted Vegetation Index (OSAVI) (Rondeaux et al., 1996)	$\frac{NIR - R}{NIR + R + 0.5}$
Soil Adjusted Vegetation Index (SAVI) (Huete, 1988)	$1.5 * \frac{NIR - R}{NIR + R + 0.5}$

Specific Leaf Area Vegetation Index (SLAVI) (Lymburner et al., 2000)	$\frac{NIR}{R + SWIR}$
Simple Ratio or Ration Vegetation Index (SR) (Jordan, 1969)	$\frac{NIR}{R}$
Triangular Vegetation Index (TVI) (Broge & Leblanc, 2001)	$0.5 * (120 * (NIR - G) - 200 * (R - G))$
Visble Atmospherically Resistant Index (VARI) (Gitelson et al., 2003)	$\frac{G - R}{G + R - B}$
Vegetativen (VEG)	$\frac{G}{R^{0.667} * B^{0.333}}$

2.4. Soil sampling and analysis

Soil sampling was performed at 0-20 cm depth in different regions of Alto Amazonas province to obtain the maximum information from each site, totalling 100 soil samples.

The physical and chemical properties of the soil, such as pH, E.C., organic matter, textural fractions (sand, clay, and lime), exchangeable bases (Ca, Mg, Na, and K), exchangeable acidity, available P, and CEC were determined before and after liming the soil. The chemical methods used for assessing soil characteristics are reported in previous publications (Arévalo-Hernández et al., 2022). Soil texture analysis was performed with the Bouyoucos method, using 1 M $L-1$ of NaOH as a dispersant. Soil pH (1:2.5 H2O) was measured with a potentiometer, electrical conductivity (EC) with a conductivity meter, and organic matter (OM) concentration with the Walkey and Black method by titration. CEC and base cations (Ca2+, Mg2+, Na+, K+) were determined using extraction with 1 M NH4OAc and, after, determined in flame atomic absorption spectrophotometer—FAAS. Yuan's method was used to determine exchangeable acidity (Al3+, H+). Available P was extracted with the Olsen method (0.5 M NaHCO3 pH 8.5) and determined in a UV-VIS spectrophotometer. The selected microelements (Cu, Fe, Mn, Zn) were extracted by DTPA and then analyzed by AAS.

2.5. Modelling and statistical analyses

For modelling, data will be divided into training and testing data. For the training phase, 75% of data from farms will be used, while the remaining 25% will be for prediction assessment (testing data). Descriptive statistics (minimum, quartile-1, mean, median, quartile-3, standard deviation, maximum, interquartile range and coefficient of variation) will be performed for all soil characteristics and vegetation indexes (VI). The data and descriptive statistics will be used for modelling procedures; however, to avoid high correlated variables (r>0.90), Spearman correlation (5% confidence) will be performed, and only low correlated variables will be stored.

Afterwards, data will be submitted to six models as follows: Cubist (C), General Linear Model (GLM), Random Forest (RF), Weighted K-Nearest Neighbor Classifier (KKNN), Support Vector Machine (SVM) and Neural Network (NN). The root mean square error (RMSE) and the coefficient of determination (R2) will be calculated for training and validation to compare and select the best algorithm with the higher R2 and lower RMSE. From the results of the models, an important analysis will be performed to select the most influential variables to produce the final prediction model. All modelling procedures and statistical analyses will be performed in R, version 4.1.2 (R Core Team, 2021).

3. RESULTS AND DISCUSSION

3.1. Soil physical and chemical characteristics

Table 2 presents the descriptive statistics of soil physical and chemical characteristics. In the case of physical characteristics, texture indicated that mean values of sand were higher in comparison to silt or clay, indicating a predominance of clay loam to sandy loam soils, representing 74.7 % of the sampled soils.

Table 2.
Soil chemical and physical attributes mean, median, Interquartile Range and Range in Alto Amazonas province

Soil variables	Mean (SD)	Median (IQR)	Range
pH	4.8 (0.9)	4.6 (1.0)	3.2 - 7.9
CE dS/cm	0.2 (0.4)	0.1 (0.1)	0.0 - 2.0
CaCO3 %	0.0 (0.3)	0.0 (0.0)	0.0 - 2.6
Organic Matter %	2.7 (4.3)	1.8 (2.0)	0.0 - 31.9
P mg/kg	5.8 (7.6)	3.7 (4.0)	0.0 - 54.9
Sand %	48.3 (20.4)	47.7 (31.6)	7.0 - 88.0
Silt %	25.9 (11.2)	25.5 (16.5)	5.3 - 52.0
Clay %	25.9 (13.6)	25.4 (17.2)	2.0 - 57.0
CEC cmol+/kg	10.3 (9.5)	7.5 (9.3)	1.1 - 42.8
Ca cmol+/kg	5.2 (8.0)	1.0 (5.8)	0.0 - 38.0
Mg cmol+/kg	0.7 (0.9)	0.3 (1.0)	0.0 - 3.5
K cmol+/kg	0.1 (0.1)	0.1 (0.1)	0.0 - 0.5
Na cmol+/kg	0.1 (0.1)	0.1 (0.0)	0.1 - 0.4
Al cmol+/kg	2.8 (2.9)	2.0 (3.4)	0.0 - 15.2

For chemical characteristics, the mean and median value of pH in the soil (4.8) was acidic. However, some places showed high pH values (7.9), indicating that a great region area may require lime amendments to achieve better yields. In the case of CE (dS/cm), all the values remain low, with no saline soils observed in this region. In the case of Carbonates, low to zero values were observed, being the mean and median near zero. For organic matter (%), mean values were observed in the medium range and the mean slightly below the critical limit (2%); however, due to wetlands, higher values were also observed in the magnitude of 31.9%. In the case of nutrients, P mean and median values were low (<7 mg kg-1), indicating the high need for P fertilizers for crop production.

For exchangeable bases (Ca, Mg, K and Na), Ca had mean and median with very different values indicating a non-normal distribution of data, while mean values were in the range considered as medium (3-6 cmol+/kg). For Mg, mean values were low (<1.0 cmol+/kg), indicating the need to apply high Mg amendments. Finally, for K and Na, mean values were very low (<0.1 cmol+/kg), indicating the need for high K fertilizers.

Finally, in the case of exchangeable acidity (Al), mean values were high (>2.5 cmol+/kg), indicating the need for the use of acidic tolerant species or lime amendments to reduce Al toxicity in crop production in this province.

3.2. Prediction models

The results of the different machine learning models applied to the study for the prediction of soil fertility in Alto Amazonas province in the Loreto region are presented in Table 3.

Table 3.
Prediction models (GLM, CUBIST, KKNN, SVM, RF and NN) R2, RMSE and MAE for main soil physical and chemical characteristics in Alto Amazonas province

Soil variables	GLM*			CUBIST			KKNN			SVM			RF			NN		
	R2	rmse	MAE	R2	rmse	MAE	R2	rmse	MAE	R2	rmse	MAE	R2	rmse	MAE	R2	rmse	MAE
pH	0.21	0.99	0.66	0.60	0.73	0.49	0.18	0.96	0.64	0.01	1.11	0.74	0.71	0.73	0.49	0.01	3.82	2.55
Organic Matter	0.12	2.54	1.69	0.23	1.71	1.14	<0.01	2.08	1.39	0.12	3.06	2.04	0.19	2.83	1.89	0.20	1.77	1.18
P mg/kg	0.05	9.51	6.34	0.01	6.20	4.13	0.02	4.56	3.04	0.07	5.67	3.78	0.03	4.49	2.99	0.09	4.91	3.27
Ca cmol+/kg	0.46	6.12	4.08	0.06	9.31	6.21	0.52	4.91	3.27	0.42	6.61	4.41	0.75	3.53	2.35	0.28	8.20	5.47
Mg cmol+/kg	0.35	0.62	0.41	0.48	0.52	0.35	0.42	0.56	0.37	0.37	0.61	0.41	0.57	0.48	0.32	0.47	0.56	0.37
K cmol+/kg	0.27	35.76	23.84	0.15	33.37	22.25	0.07	32.76	21.84	0.21	38.71	25.81	0.29	27.94	18.63	0.10	60.10	40.07
Al cmol+/kg	0.47	3.36	2.24	0.66	2.26	1.51	0.09	3.45	2.30	0.44	3.11	2.07	0.34	3.02	2.01	0.05	4.20	2.80
CEC cmol+/kg	0.91	1.88	1.25	0.65	4.29	2.86	0.78	3.15	2.10	0.89	2.19	1.46	0.92	2.32	1.55	0.01	10.49	6.99

* Cubist (C), General Linear Model (GLM), Random Forest (RF), Weighted K-Nearest Neighbor Classifier (KKNN), Support Vector Machine (SVM) and Neural Network (NN)

It was possible to observe that, in general, to predict soil variables, the Random Forest algorithm was satisfactory compared to other models, obtaining the higher R2 and lower RMSE in the Alto Amazonas province. However, in the case of organic matter and Aluminum, the CUBIST algorithm was slightly superior.

Even though the use of these algorithms in the prediction of soil variables is not new, some algorithms have performed better than others, such as random forest, since it has great capacity in the use of nonlinear data and is robust against possible errors (Breinman, 2001). Also, Smith et al. (2020) have obtained better performance with random forest in soil texture prediction in agricultural soils. In the case of Organic Matter, Mosaid et al. (2024) showed that the use of random forest performed better, as shown in the present study.

Even though the perspectives of using machine learning have been discussed elsewhere by Sujatha et al. (2023), it remains an interesting tool to have results in areas where the logistics and costs are often expensive, improving the information and decision-making.

CONCLUSIONS

Alto Amazonas province has high climatic and geologic differences, generating different types of soils and formations; also, access to many areas is limited. The use of machine learning algorithms to predict poses as an alternative to improve information on soil fertility. In this work, we observed that most soils in the province have low pH, P, Mg, K and high acidity. We also managed to achieve good predictions for pH, Ca, Mg and CEC, and we observed that the most successful algorithm was Random Forest. Nevertheless, for Al, Cubist performed better. This is one of the first works using machine learning to predict soil fertility in the Peruvian Amazon, and we hope it may serve as a base for future projects.

FUNDING

The authors thank the Universidad Nacional Autónoma de Alto Amazonas for the funding within the scope of the call for Research Projects 2023 with Presidential Resolution N°245-UNAAA/P. We also appreciate the technical and logistical facilities for field research and laboratory analytical facilities.

CONFLICT OF INTEREST

Los autores declaran que no existe ningún tipo de conflicto de intereses.

AUTHORSHIP CONTRIBUTION

Conceptualization: Arévalo-Hernández, C.O. y Arévalo-Gardini, E.

Formal analysis: Arévalo-López, L. A. y Tuesta-Hidalgo, O.

Research: All authors

Methodology: Arévalo-Hernández, C.O., Arévalo-Gardini, E. y Arévalo-López, L. A.

Project management: Romero-Vela, D.S.

Supervision: Arévalo-Gardini, E. y Ruiz-Camus, C.E.

Validation: Arévalo-López, L. A. y Tuesta-Hidalgo, O.

Visualization: Romero-Vela, D.S.

Writing - original draft: All authors

Writing - proofreading and editing: All authors

REFERENCES

Arévalo-Gardini, E., Canto, M., Alegre, J., Loli, O., Julca, A., & Baligar, V. (2015). Changes in soil physical and chemical properties in long term improved natural and traditional agroforestry management systems of cacao genotypes in Peruvian Amazon. *PLoS ONE*, 10(7). https://doi.org/10.1371/journal.pone.0132147

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M. L., & Bareth, G. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation*, 39, 79–87. https://doi.org/10.1016/j.jag.2015.02.012

Brus, D. J. (2019). Sampling for digital soil mapping: A tutorial supported by R scripts. *Geoderma*, 338, 464–480. https://doi.org/10.1016/J.GEODERMA.2018.07.036

Bünemann, E. K., Bongiorno, G., Bai, Z., Creamer, R. E., De Deyn, G., de Goede, R., Fleskens, L., Geissen, V., Kuyper, T. W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J. W., & Brussaard, L. (2018). Soil quality – A critical review. In Soil Biology and Biochemistry (Vol. 120, pp. 105–125). https://doi.org/10.1016/j.soilbio.2018.01.030

Canero, F. M., Rodriguez-Galiano, V., & Aragones, D. (2024). Machine Learning and Feature Selection for soil spectroscopy. An evaluation of Random Forest wrappers to predict soil organic matter, clay, and carbonates. *Heliyon*, *10*(9). https://doi.org/10.1016/j.heliyon.2024.e30228

Campos, A. R., Giasson, E., Costa, J. J. F., Machado, I. R., Silva, E. B. da, & Bonfatti, B. R. (2019). Selection of Environmental Covariates for Classifier Training Applied in Digital Soil Mapping. *Revista Brasileira de Ciência Do Solo*, 42, e0170414. https://doi.org/10.1590/18069657RBCS20170414

Chen, J. M. (1996). Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing*, 22(3), 229–242. https://doi.org/10.1080/07038992.1996.10855178

Chuvieco, E., Martín, M. P., & Palacios, A. (2002). Assessment of different spectral indices in the red-near-infrared spectral domain for burned land discrimination. *International Journal of Remote Sensing*, 23(23), 5103–5110. https://doi.org/10.1080/01431160210153129

Dash, J., & Curran, P. J. (2007). Evaluation of the MERIS terrestrial chlorophyll index (MTCI). Advances in Space Research, 39(1), 100–104. https://doi.org/10.1016/j.asr.2006.02.034

Delgado-Caballero, C. E., Gómez-Guerrero, A., Valdez-Lazalde, J. R., De los Santos- Posadas, H., Fierros-González, A. M., & Horwath, W. R. (2009). Site index and soil properties in young plantations of Eucalyptus grandis and E. urophylla in southeastern México. *Agrociencia*, 43(1). https://www.agrociencia-colpos.org/index.php/agrociencia/article/view/697

Dharumarajan, S., Lalitha, M., Niranjana, K., & Hegde, R. (2022). Evaluation of digital soil mapping approach for predicting soil fertility parameters—a case study from Karnataka Plateau, India. *Arabian Journal of Geosciences*, 15(5), 1–21. https://doi.org/10.1007/S12517-022-09629-8

Di Raimo, L. A. D. L., Couto, E. G., de Mello, D. C., Demattê, J. A. M., Amorim, R. S. S., Torres, G. N., Bocuti, E. D., Veloso, G. V., Poppiel, R. R., Francelino, M. R., & Fernandes-Filho, E. I. (2022). Characterizing and Modeling Tropical Sandy Soils through VisNIR-SWIR, MIR Spectroscopy, and X-ray Fluorescence. Remote Sensing 2022,14(19), 4823. https://doi.org/10.3390/RS14194823

EMBRAPA. (2009). Manual de análises químicas de solos, plantas e fertilizantes (2nd ed.). Embrapa. www.sct.embrapa.br/liv

Frampton, W. J., Dash, J., Watmough, G., & Milton, E. J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007

Gamon, J. A., & Surfus, J. S. (1999). Assessing leaf pigment content and activity with a reflectometer. New Phytologist, 143(1), 105–117. https://doi.org/10.1046/j.1469-8137.1999.00424.x

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. Journal of Plant Physiology, 160(3), 271–282. https://doi.org/10.1078/0176-1617-00887

Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2001). Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. *Photochemistry and Photobiology*, 74(1), 38–45. https://doi.org/10.1562/0031-8655(2001)0740038opaneo2.0.co2

Hounkpatin, K. O. L., Bossa, A. Y., Yira, Y., Igue, M. A., & Sinsin, B. A. (2022). Assessment of the soil fertility status in Benin (West Africa) – Digital soil mapping using machine learning. *Geoderma Regional*, 28, e00444. https://doi.org/10.1016/J.GEODRS.2021.E00444

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X INEI. (2012). IV Censo Nacional Agropecuario . In Resultados Definitivos. IV Censo Nacional Agropecuario. http://proyectos.inei.gob.pe/web/documentospublicos/resultadosfinalesivcenagro.pdf

Jordan, C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology*, 50(4), 663–666. https://doi.org/10.2307/1936256

Köppen, W. (1931). Grundriß der Klimakunde. In Grundriß der Klimakunde. De Gruyter. https://doi.org/10.1515/9783111667751

Lang, P., Qin, C. zhi, Zhu, A. xing, Hou, Z. wei, Fan, N. qing, & Wang, Y. jie. (2020). A case- based method of selecting covariates for digital soil mapping. *Journal of Integrative Agriculture*, 19(8), 2127–2136. https://doi.org/10.1016/S2095-3119(19)62857-1

Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, 16(1), 65–70. https://doi.org/10.1080/10106040108542184

Lu, Q., Tian, S., & Wei, L. (2023). Digital mapping of soil pH and carbonates at the European scale using environmental variables and machine learning. *Science of The Total Environment*, 856, 159171. https://doi.org/10.1016/J.SCITOTENV.2022.159171

Lymburner, L., Beggs, P. J., & Jacobson, C. R. (2000). Estimation of canopy-average surface-specific leaf area using Landsat TM data. In Photogrammetric Engineering and Remote Sensing (Vol. 66, Issue 2, pp. 183–191).

Macedo Neto, A. A. L.; Farias, P. R. S.; Matos, G. S. B.; Silva, G. B.; Santos, A. V. F. Anhê, B. B. (2020). Diagnosis and spatial variability of soil fertility and crop production in a teak area in Eastern Pará State. *CERNE*, 26(1), 37-47. https://doi.org/10.1590/01047760202026012683

Mosaid, H., Barakat, A., John, K., Faouzi, E.H., Bustillo, V., El Garnaoui, M., & Heung, B. (2024). Improved soil carbon stock spatial prediction in a Mediterranean soil erosion site through robust machine learning techniques. *Environmental monitoring and assessment, 196 2*, 130. https://doi.org/10.1007/s10661-024-12294-x

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

R Core Team. (2021). R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL Http://Www.R-Project.Org/

Richardson, A. J., & Wiegand, C. L. (1977). Distinguishing Vegetation from Soil Background Information. A gray mapping technique allows delineation of any Landsat scene into vegetative cover stages, degrees of soil brightness, and water. *Photogrammetric Engineering and Remote Sensing*, 43(13), 1541-1552. https://www.asprs.org/wp-content/uploads/pers/1977journal/dec/1977_dec_1541-1552.pdf

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55(2), 95–107. https://doi.org/10.1016/0034-4257(95)00186-7

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the great plains with ERTS. NASA Goddard Space Flight Center 3d ERTS-1 Symposium, 1, 309–317. https://doi.org/10.1021/jf60203a024

Roy, D. P., Boschetti, L., & Trigg, S. N. (2006). Remote sensing of fire severity: Assessing the performance of the normalized burn ratio. *IEEE Geoscience and Remote Sensing Letters*, 3(1), 112–116. https://doi.org/10.1109/LGRS.2005.858485

Siqueira, R. G., Moquedace, C. M., Francelino, M. R., Schaefer, C. E. G. R., & Fernandes- Filho, E. I. (2023). Machine learning applied for Antarctic soil mapping: Spatial prediction of soil texture for Maritime Antarctica and Northern Antarctic Peninsula. *Geoderma*, 432, 116405. https://doi.org/10.1016/J.GEODERMA.2023.116405

Sripada, R. P., Heiniger, R. W., White, J. G., & Meijer, A. D. (2006). Aerial color infrared photography for determining early in-season nitrogen requirements in corn. *Agronomy Journal*, 98(4), 968–977. https://doi.org/10.2134/agronj2005.0200

Sujatha, M., & Jaidhar, C. D. (2023). Machine learning-based approaches to enhance the soil fertility—A review. *Expert Systems with Applications*, 240, 122557. https://doi.org/10.1016/j.eswa.2023.122557

Tucker, C. J. (1980). A spectral method for determining the percentage of green herbage material in clipped samples. *Remote Sensing of Environment*, 9(2), 175–181. https://doi.org/10.1016/0034-4257(80)90007-3

Vescovo, L., & Gianelle, D. (2008). Using the MIR bands in vegetation indices for the estimation of grassland biophysical parameters from satellite remote sensing in the Alps region of Trentino (Italy). *Advances in Space Research*, 41(11), 1764–1772. https://doi.org/10.1016/j.asr.2007.07.043

Vieira, R. M. da S. P., Tomasella, J., Barbosa, A. A., Polizel, S. P., Ometto, J. P. H. B., Santos, F. C., Ferreira, Y. da C., & Toledo, P. M. de. (2021). Land degradation mapping in the MATOPIBA region (Brazil) using remote sensing data and decision-tree analysis. *Science of The Total Environment*, 782, 146900. https://doi.org/10.1016/J.SCITOTENV.2021.146900

Wadoux, A. M. J. C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, 210, 103359. https://doi.org/10.1016/J.EARSCIREV.2020.103359

Wang, F., Huang, J., Tang, Y., & Wang, X. (2007). New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice. *Rice Science*, 14(3), 195–203. https://doi.org/10.1016/s1672-6308(07)60027-4

Watt, M. S., Pearse, G. D., Dash, J. P., Melia, N., & Leonardo, E. M. C. (2019). Application of remote sensing technologies to identify impacts of nutritional deficiencies on forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149, 226–241. https://doi.org/10.1016/j.isprsjprs.2019.01.009

Wu, W. (2014). The Generalized Difference Vegetation Index (GDVI) for dryland characterization. *Remote Sensing*, 6(2), 1211–1233. https://doi.org/10.3390/rs6021211

Yang, L., Li, X., Shi, J., Shen, F., Qi, F., Gao, B., & Chen, Z. (2020). Geoderma Evaluation of conditioned Latin hypercube sampling for soil mapping based on a machine learning method. *Geoderma*, 369, 114337. https://doi.org/10.1016/j.geoderma.2020.114337