A deep learning-based soil diagnostic method for simultaneous multi-parameter analysis using full-wavelength ICP spectra

Rising global food demand and accelerating climate change are intensifying the need for agricultural systems that sustain productivity while conserving limited natural resources. Maintaining healthy soils is central to this challenge, as effective fertilization and management require accurate knowledge of soil properties. However, conventional soil analyses are often costly and time-consuming, limiting their application in many agricultural regions, particularly in developing countries. Consequently, affordable and accessible soil diagnostic technologies are urgently needed. Optical sensing approaches, such as near-infrared (NIR) and mid-infrared (MIR) spectroscopy, have been explored as non-destructive alternatives, but remain limited in the number of soil properties that can be simultaneously assessed.

Here we developed a rapid and low-cost soil diagnostic approach capable of simultaneously estimating multiple soil properties by applying deep learning to the full emission spectrum obtained from an inductively coupled plasma atomic emission spectrometer (ICP-AES).

A total of 1,941 soil samples collected from seven countries, primarily in Africa and Asia, were analyzed using conventional laboratory methods and used as training data. Twelve soil parameters were included: pH (H_2O and KCl), electrical conductivity (EC), available phosphorus (Bray-1), exchangeable Ca, Mg, K, and Na, exchangeable Al, cation exchange capacity (CEC), total carbon, total nitrogen, and particle-size composition (clay and sand contents). In conventional ICP analysis, only selected wavelengths corresponding to target elements are used for quantification. In contrast, this study utilized the entire spectral information. Soil extracts obtained using 1 M ammonium acetate (pH 7) were measured, and intensity values from 2,574 pixels representing 234 wavelength targets were extracted and log-transformed for deep learning input.

The dataset was randomly divided into 80% training and 20% evaluation data, and model training was conducted using a feed-forward neural network (FFNN). Predictive performance was evaluated using the indices proposed by Malley et al. (2004) and Chang et al. (2001). The proposed method predicted all twelve soil parameters with high accuracy: most variables showed coefficients of determination (R^2) greater than 0.9, while even the lowest-performing variable, total carbon, achieved an R^2 of 0.812. Notably, soil properties not directly measurable by ICP, including pH and CEC, were also predicted with high accuracy (Fig. 1), and all parameters reached accuracy levels suitable for practical soil diagnostics (Table 1). This technology reduces costs for reagents, equipment, and time, enabling affordable soil diagnostics (Fig. 2).

Authors: Nakamura, S., Ikazaki, K. [JIRCAS], Imaya, A. [FFPRI]

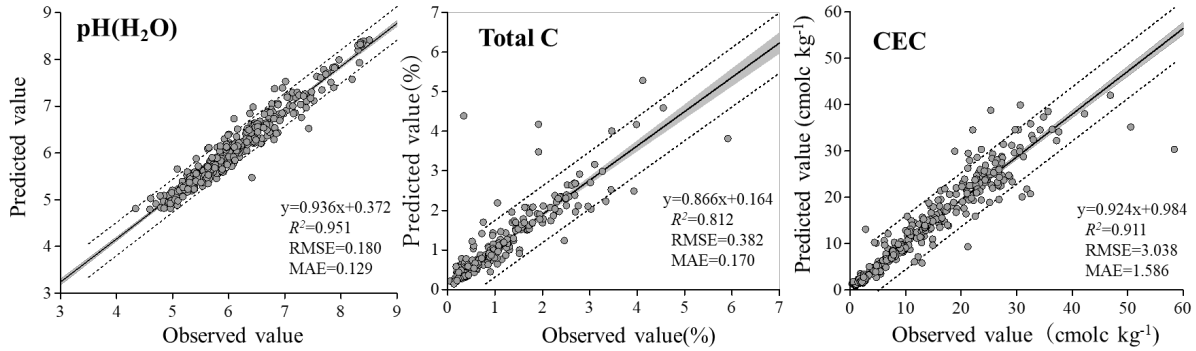

Fig. 1. The relationship between the predicted values and the observed values

RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error)

Table 1. Prediction accuracy for each parameter and its accuracy for soil diagnosis

Parameter	Unit	Determination Coefficient	Results of the Soil Diagnostic Accuracy Assessment	
		R^2	Malley's index	Chang's index
pH (H2O)		0.951	A	A
pH (KCl)		0.959	A	A
EC	mS m^{-1}	0.960	A	A
Bray1-P	mg kg^{-1}	0.963	A	A
Ex. Al	cmolc kg^{-1}	0.964	A	A
Ex. Ca		0.995	A	A
Ex. Mg		0.983	A	A
Ex. K		0.993	A	A
Ex. Na		0.946	B	A
CEC		0.911	B	A
Total N	%	0.956	A	A
Total C		0.812	C	A
Clay		0.844	C	A
Sand		0.870	D	A

Malley et al. 2004
A: Excellent, B: Successful, C: Moderately Successful, D: Moderately Useful, E: Screening

Chang et al. 2001
A: Successful, B: Possibility, C: Not Useful

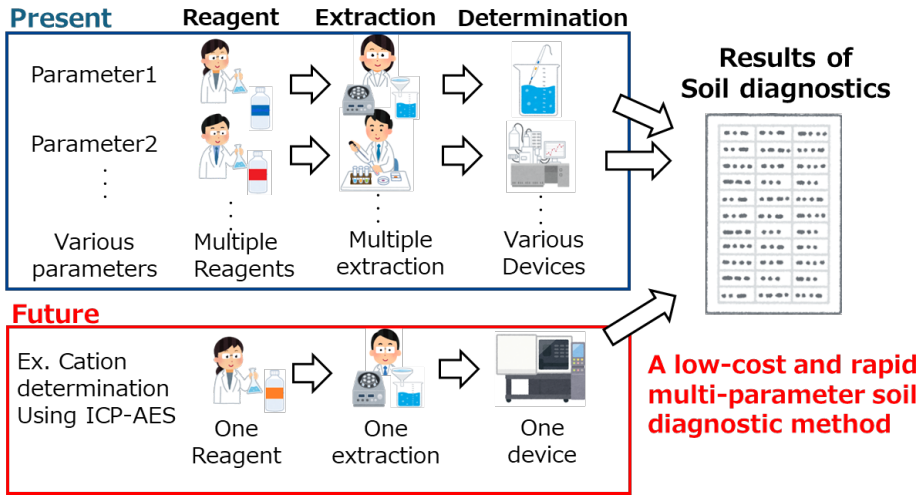

Fig. 2. Present and Future of Soil Diagnosis

Reference: Nakamura et al. (2025) *Scientific Reports* 15: 37753.