Coupling Backpropagation Neural Network and AdaBoost Algorithm for Quantitative Analysis of Nickel via Laser-Induced Breakdown Spectroscopy

Edward Harefa and Weidong Zhou*
Key Laboratory of Optical Information Detection and Display Technology of Zhejiang, Zhejiang Normal University, Jinhua 321004, China

wdzhou@zjnu.cn

Abstract. A wide area of cropland or soil might well be contaminated with heavy metals, contaminating agricultural goods and posing a risk to human health. As a result, it is required to evaluate the concentration of heavy metals in soil. The combination between laser-induced breakdown spectroscopy (LIBS) and multivariate chemometrics methods was employed to determine heavy metal Ni concentration in twelve soil samples. The comparison between univariate calibration curve, traditional backpropagation neural network (BPNN), and hybrid BPNN-AdaBoost was presented. The result revealed that BPNN-AdaBoost outperformed other models with the coefficient determination calibration ($R^2_C$), coefficient determination prediction ($R^2_P$), root mean square error calibration (RMSEC), root mean square error prediction (RMSEP) are 0.985, 0.977, 2.04, 3.18, respectively. This study indicates that BPNN-AdaBoost can be adopted as a reliable chemometric technique to enhance the quantitative analysis of heavy metals based on LIBS.

1. Introduction
Soil is an essential aspect of the natural ecosystem and a critical commodity for human growth and survival [1]. Along with industrialization, manufacturing, mineral reserves expansion and mining activities, wastewater irrigation and fertilization, as well as related human activities, have made a significant contribution to a number of heavy metal contaminants to the soil via a variety of pathways. This has resulted in an increase in the level of heavy metal pollution in soils. Any amount of heavy metal contamination introduced into the human body via the food chain will have a detrimental effect on human health [2]. As a result, reliable identification of heavy metals in soils and management of inadequate source emissions of heavy metals are critical components of ensuring agricultural goods and food safety.

Among the conventional analytical techniques for heavy metals, namely atomic absorption spectrometry (AAS), inductively coupled plasma-atomic emission spectrometry (ICP-AES), and X-ray fluorescence (XRF). While AAS and ICP-AES are highly sensitive and accurate methods for detecting heavy metal elements in soil, they involve complicated sample preparation. Additionally, XRF enables the simultaneous detection of many heavy metals. The energy and intensity of the elements' characteristic-fluoresced radiation are utilized to identify and quantify their quantities in a particular
sample; nonetheless, X-rays pose substantial radiation dangers to human health [3–5]. As a novel and promising approach for element identification, laser-induced breakdown spectroscopy (LIBS) may indeed be offered as a substitute for extensive sample preparation in rapid analysis. LIBS has a number of advantages over other analytical techniques, including a lower sample requirement, no complicated pretreatment, multi-element joint measurement, and rapid installation [6–8].

Recently, many studies have concentrated on LIBS deployment to detect heavy metal contamination combined with chemometrics methods. Liu et al. [9] proposed a least-squares support vector machine (LS-SVM) and partial least-squares regression (PLSR) to determine Cd content in the soil in air and Ar condition. The experiment showed that LS-SVM demonstrated flawless capability for quantitative detection of Cd in soil, with $R^2$ values greater than 0.98 in both the calibration and prediction sets. Partial least square (PLS) and artificial neural network (ANN) was employed by Sirven et al. [10] and discovered that ANN had a better prediction performance to detect Cr in soil. Determination of Fe and Cu contents on the surface water had been successfully explored by Zhang et al. [11] using backpropagation neural network (BPNN) and obtained the correlation coefficient of 0.8.

In this study, the combination of BPNN and adaptive boosting (AdaBoost) is proposed to undertake a quantitative analysis of Ni in twelve soil samples. We explore the performance of BPNN-AdaBoost to enhance the prediction concentration of Ni and calculate the coefficient correlation ($R^2$), root mean square error calibration (RMSEC), and root mean square error prediction (RMSEP).

2. Methodology

2.1. Sample Preparation and LIBS Setup

Figure 1 illustrates the experimental setup used to generate plasma and collect the emission spectra of strontium and nickel elements. The Q-switched Nd:YAG laser (Vlite-200, Beamtech, China) with wavelength of 1064 nm, operating frequency of 1 Hz, and laser energy of 100 mJ was focused by a lens (f=50 mm) onto the surface of twelve standard soils (National Institute of Metrology of China). The standard soils were formed into pellet samples with 33 mm in diameter and 2.5 mm in thickness. The concentration of Ni in each of the soil samples is shown in Table 1. In this study, soil samples #10, #11, and #12 were selected as prediction sets and the remaining samples as calibration sets.

| Sample label | Concentration (µg/g) | Sample label | Concentration (µg/g) |
|--------------|----------------------|--------------|----------------------|
| #1           | 53                   | #7           | 38                   |
| #2           | 23                   | #8           | 47                   |
| #3           | 23                   | #9           | 32                   |
| #4           | 41                   | #10          | 33                   |
| #5           | 38                   | #11          | 29                   |
| #6           | 39                   | #12          | 32                   |

The pulsed laser was focused on the surface of the sample placed on a two-dimensional moving stage to generate laser plasma and a multi-channel fiber optic spectrometer (Avantes, AvaSpec-ULS2048-2-USB2, Netherlands) was used to detect the plasma radiation spectrum. The fiber probe was placed about 25 mm away from the target (the angle between the probe axis and the laser beam is about 45°), and it conducted the received plasma light radiation to the entrance of the spectrometer. The detection wavelength range was 195–510 nm, the spectral sampling interval was less than 0.08 nm, and the integration time was 2 ms. The spectrometer was equipped with a 2048-element charged-couple device (CCD) camera and started to collect the spectrum after a delay time of 1 µs. The spectrum was transmitted to the computer for storage and processing of the spectrum data. In order to ensure that each laser pulse can hit a different position of the soil test sample, after each laser pulse bombarded the sample...
surface, the electric displacement platform automatically moved to a new position. In the experiment, in order to improve the stability of the plasma signal, the laser focus was focused at about 2 mm below the surface of the experimental sample. Each sample was examined 25 times in total and yielding 300 spectral data.

Figure 1. LIBS experimental setup.

2.2. **BPNN-AdaBoost in LIBS and Performance Evaluation**

The BPNN approach is essentially a local search optimization method, which is prone to local extremes. Coupling BPNN and AdaBoost strong classification is presented to increase prediction performance and resolve the element interference issue. To achieve an optimal output, the BPNN-AdaBoost approach optimizes output by combining the classification effects of various weak classifiers to form a strong classifier. Each BPNN is built as a weak classifier. The prediction error $e_m$ is used to compute the weight $w_m$ of each weak classifier in the assessment phase as explained in the equation (1) and (2):

$$e_m = \frac{1}{2} \sum_m (p_m - o_m)^2,$$

$$w_m = \frac{1}{2} \ln \left( \frac{1 - e_m}{e_m} \right),$$

where $p_m$ is the forecasted output result following $m$ iterations, while $o_m$ is the target output result following $m$ iterations. The more effective a weak classifier is in predicting, the higher its fraction in the strong classifier will be. Thus, the prediction effect of numerous BPNN weak classifiers $f_1, f_2, \ldots, f_m$ is merged to generate a strong classifier $I_m$:

$$I_m = \text{sign} \left[ \sum_{m=1}^{M} w_m g f \left( p_m, y_m \right) \right].$$

The correlation coefficient $R^2$ is employed for obtaining a fair approximation of the univariate and multivariate (BPNN-AdaBoost) method’s prediction effect. Assessment of the fit of the regression line to the measured concentration indicated by the $R^2$ value. The root mean square error (RMSE) is performed to determine the accuracy of the assessment by comparing the expected value to the true value. The following equations are the evaluation parameters.
\[ R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}, \]  
\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \]

where \( y_i \) denotes the sample’s actual concentration, \( \hat{y}_i \) defines the predicted concentration, and \( \bar{y}_i \) is the average value of predicted concentration. RMSE was explored in the calibration prediction sets, which are denoted by RMSEC and RMSEP, accordingly.

3. Results and Discussion

Figure 2 exhibits a representative region of Ni element from the full spectrum to show the effectiveness and capability of LIBS for detecting Ni elements. The spectrum generated by LIBS is composed of a strong continuum and various neutral and ionic lines associated with the individual elements. It is noted that the spectrum is dominated by continuum emission at the initial plasma produced. This condition resulted in electron collisions with ions and atoms, as well as electron recombination with ions. As time passes, the continuum backdrop becomes less prominent, while atomic and ionic emission lines grow more prevalent [12]. Using the atomic NIST Atomic Spectra Database [13], we can determine singly neutral line Ni I (234.52 nm, 279.82 nm, 309.72 nm, 324.86 nm, 337.42 nm, 357.18 nm, 358.74 nm) and ionic line Ni II (226.44 nm, 227.08 nm, 233.44 nm, 237.5 nm, 243.79 nm).

![Figure 2. Typical spectra of soil sample #3 in 320-360 nm.](image)

This study first performs a univariate calibration curve using three selected lines, namely 234.52 nm, 237.5 nm, and 324.86 nm, that are free from self-absorption and overlapping peaks. The univariate calibration model is developed by plotting the Ni line intensity against its concentration. The calibration curves for three distinctive Ni lines are shown in Figure 3(a) – (c). Figure 3(a) depicts the linear relationship between line intensity and concentration, which has a correlation coefficient \( R^2 \) of 0.785. The achieved correlation coefficients signify poor linearity (\( R^2 < 0.9 \)) and may affect the prediction’s
accuracy and precision as well as the model's robustness. The low correlation coefficient in the univariate calibration curve is due to the variated matrix effect on the experimented element [4], which should be enhanced to conduct an extensive quantitative analysis.

![Graphs showing univariate calibration curves for different Ni emission lines.](image)

**Figure 3.** The univariate calibration curve and value of correlation coefficient $R^2$ in the selected Ni emission lines at (a) 234.52 nm, (b) 237.5 nm, and (c) 324.86 nm.

The acquired spectra are normalized using standard normal variate transformation and implemented wavelet transform denoising to reduce background signal before implementing BPNN-AdaBoost. BPNN-AdaBoost is employed with the number of hidden layer neurons was set between 10 and 63. During training, the number of trees is 75, the maximum number of iterations is 300, the learning rate is 0.01, and the momentum factor of 0.8, which optimize by grid search and ten-fold cross-validation. Performing classical BPNN, we get $R^2_C$, $R^2_P$, RMSEC, RMSEP by 0.943, 0.928, 7.35, 8.11, respectively. On the other hand, the value of $R^2_C$, $R^2_P$, RMSEC, RMSEP in BPNN-AdaBoost is increased to 0.985, 0.977, 2.04, 3.18, respectively. As illustrated in Figure 4, the majority of the calibration and prediction data points in BPNN-AdaBoost are concentrated around the fitting curve. It is noticeable that the achieved result BPNN and BPNN-AdaBoost has significantly improved compared to the univariate calibration analysis. As typical nonlinear models, BPNN and BPNN-AdaBoost have self-adjusting and greater adaptive capabilities and effectively predict the nonlinear relationship between Ni concentration and spectral emission lines intensity [14,15]. However, the combination between BPNN and AdaBoost is increasing prediction performance by iteratively training and forecasting the output of samples using
an aggregate of BPNN decisions and proper assignment number of trees. Thus, the BPNN-AdaBoost algorithm outperforms the BPNN algorithm in terms of the quantification ability of Ni content in the soil sample.

![Figure 4](image)

**Figure 4.** Comparison between Ni concentration predicted by BPNN-AdaBoost based on LIBS in and actual concentration in (a) calibration set and (b) prediction set.

4. **Conclusion**
The investigation of LIBS calibration alternatives in order to suggest a method for determining pollutants demonstrates that when compared to standard linear fit calibration, the usage of BPNN-AdaBoost is an effective option. The LIBS technology and the BPNN-AdaBoost technique produce accurate and precise elemental predictions for each of the twelve soil types studied. The results indicated that BPNN-AdaBoost is capable of overcoming matrix effects, which are extremely prevalent in LIBS spectra and significantly improved prediction of Ni concentration. Thus, applying BPNN-AdaBoost to detect heavy metals in soil based on LIBS spectra proved to be beneficial.

5. **Acknowledgment**
This study was supported by the National Natural Science Foundation of China (Grant No. 61975186).

**References**
[1] Nicolodelli G, Cabral J, Menegatti C R, Marangoni B and Senesi G S 2019 Recent advances and future trends in LIBS applications to agricultural materials and their food derivatives: An overview of developments in the last decade (2010–2019). Part I. Soils and fertilizers *TrAC - Trends in Analytical Chemistry* 115 70–82
[2] Sharafi K, Nodahi R N, Yunesian M, Hossein Mahvi A, Pirsheb M and Nazmara S 2019 Human health risk assessment for some toxic metals in widely consumed rice brands (domestic and imported) in Tehran, Iran: Uncertainty and sensitivity analysis *Food Chemistry* 277 145–155
[3] Han P, Dong D, Du X, Jiao L and Zhao X 2016 Fast determination of calcium concentration in the internal tissues of a single seed using laser-induced breakdown spectroscopy *Analytical Methods Methods* 8(37) 6705–6710
[4] Zaytsev S M, Krylov I N, Popov A M, Zorov N B and Labutin T A 2018 Accuracy enhancement
of a multivariate calibration for lead determination in soils by laser-induced breakdown spectroscopy *Spectrochimica Acta* - Part B Atomic Spectroscopy. **140** 65–72

[5] Peng J, Liu F, Zhou F, Song K, Zhang C, Ye L and He Y 2016 Challenging applications for multi-element analysis by laser-induced breakdown spectroscopy in agriculture: A review *TrAC - Trends in Analytical Chemistry* **85** 260–272

[6] Cremers D A and Radziemski L J 2013 *Handbook of Laser-induced Breakdown Spectroscopy* John Wiley & Sons

[7] Markiewicz-Keszycka M, Cama-Moncunill X, Casado-Gavalda M P, Dixit Y, Cama-Moncunill R, Cullen P J and Sullivan C 2017 Laser-induced breakdown spectroscopy (LIBS) for food analysis: A review *Trends in Food Science & Technology* **65** 80–93

[8] Yu K, Ren J and Zhao Y 2020 Principles, developments and applications of laser-induced breakdown spectroscopy in agriculture: A review *Artificial Intelligence in Agriculture*. **4** 127–139

[9] Liu X, Liu F, Huang W, Peng J, Shen T and He Y 2018 Quantitative determination of Cd in soil using laser-induced breakdown spectroscopy in air and ar conditions *Molecules* **23(10)** 1–15

[10] Sirven J B, Bousquet B, Canioni L and Sarger L 2006 Laser-Induced Breakdown Spectroscopy of Composite Samples: Comparison of Advanced Chemometrics Methods *Analytical Chemistry* **78(5)** 1462–1469

[11] Zhang X, Zhang F, Kung H Te, Shi P, Yushanjiang A and Zhu S 2018 Estimation of the Fe and Cu contents of the surface water in the Ebinur lake basin based on LIBS and a machine learning algorithm *International Journal of Environmental Research and Public Health* **15(11)** 1–20

[12] Anderson D E, Ehliamann B L, Forni O, Clegg S M, Cousin A, Thomas N H, Lasue J, Delapp D M, McInroy R E, Gasnault O, Dyar M D, Schröder S, Maurice S and Wiens R C 2017 Characterization of LIBS emission lines for the identification of chlorides, carbonates, and sulfates in salt/basalt mixtures for the application to MSL ChemCam data *Journal of Geophysical Research: Planets* **122(4)** 744–770

[13] Kramida A, Ralchenko Y and Reader J 2020 NIST Atomic Spectra Database [Online]. Available: https://physics.nist.gov/asd National Institute of Standards and Technology, Gaithersburg (accessed: 16 July 2021)

[14] He J, Pan C, Liu Y and Du X 2019 Quantitative Analysis of Carbon with Laser-Induced Breakdown Spectroscopy (LIBS) Using Genetic Algorithm and Back Propagation Neural Network Models *Applied Spectroscopy* **73(6)** 678–86

[15] Suliman A and Zhang Y 2015 A Review on Back-Propagation Neural Networks in the Application of Remote Sensing Image Classification *Journal of Earth Science and Engineering* **5** 52–65