Tarbiat Modares University

Winter 2023. Vol 26. Num 4

The Journal of Spatial Planning

Research Paper

Modelling of Soil Heavy Metal Contamination Using Machine Learning Techniques and Spectroscopic Data

Ali Shamsoddini[1*], Shahrbanou Esmaeili[2]

1. Associate Professor, Department of Remote Sensing and GIS, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran
2. M.Sc., Department of Remote Sensing and GIS, Islamic Azad University, Science and Research Branch of Tehran, Iran

Received: 2021/10/17
Accepted: 2022/07/12

ABSTRACT

Mines and their related-industries are able to affect their surrounding environment, not only by their activities, but also after being abandoned. Among their different harmful effects, under water and surface water contaminations, and soil contamination can be mentioned. In order to manage these environmental effects, it is necessary to use reasonable methods for modelling heavy metal concentration in soil. This study aims to present a framework for modelling heavy metal soil contamination based on spectroscopy and statistical models. For this purpose, the spectral curves of the 53 soil samples, derived from an abandoned mine and its surrounding areas in New South Wales, Australia, were collected using a spectroradiometer in visible to short wavelength infrared (SWIR) wavelengths. Calculating the second derivative of the collected spectral data, random forest feature selection method (RFFS) was used to determine the most important spectral data for modelling heavy metal concentrations including lead, silver, cadmium and mercury. Then, the modelling techniques including multiple linear regression, random forest regression, and support vector regression (SVR) were applied on the selected spectral data. The results indicated that SWIR wavelengths are the most important spectral data for modelling heavy metal concentrations. Moreover, the non-linear machine learning methods, especially random forest with RMSE of 0.8 ppm and R^2 of 0.51 for lead and RMSE of 9.4 ppm and R^2 of 0.46 for cadmium performed better than multiple linear regression.

Keywords:

Soil Contamination; Heavy Metals; Feature Selection; Machine Learning; Modelling

***Corresponding Author:** Department of Remote Sensing and GIS, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran

http://dor. 20.1001.1.16059689.1401.0.0.3.5
https://doi. 10.2022/hsmsp.26.4.6

ORCID: 0000-0003-4559-7563

ali.shamsoddini@modares.ac.ir

Extended Abstract

Introduction

Soils of abandoned mines are often contaminated with heavy metals and require a regular monitoring. One of the most important measures, in this regard, is the modeling of soil contamination with heavy metals using indirect methods, which are less expensive than the direct ones and require less time to perform. In other words, the modeling methods quantitatively predict the heavy metals in soil with a good accuracy. The most popular statistical models used in the soil pollution modeling include multiple linear regression, non-linear regression, logistic regression, artificial neural networks or random forest classification and regression. The aim of this research is to identify and extract the most important effective wavelengths in the heavy metal analysis of soil and use them to model the concentration of pollutants in the soil of a mining area near Sydney, Australia. For this purpose, random forest regression, vector regression machine, and multiple linear regressions are used.

Methodology

During the chemical analysis, the concentration of 4 heavy metals including silver (Ag), cadmium (Cd), mercury (Hg) and lead (Pb) were measured in the collected samples. For this purpose, each of the soil samples were dried well and then grounded into fine powder. After that, 1 gram of the samples was used for the chemical analysis and the rest for the spectral measurement of soil samples. The spectrum of soil samples was measured in the laboratory using FieldSpec 3 spectrometer (ASD). The spectrometer device in this research is capable of measuring the electromagnetic spectrum in wavelengths from 350 nm to 2500 nm with a spectral resolution of 1.4 nm (for wavelengths from 350 to 1000 nm) and 2 nm (for wavelengths 1000 to 2500 nm). In order to measure the relative reflectance for each sample, the radiation spectrum was normalized using a 99% white reference surface spectral. In the current research, the second derivative of the measured spectra was used as input features for modeling using different machine learning methods. Many of those methods have been proposed as feature selection, among which the random forest being the more common one. Therefore, in order to extract the most important variables, the aforementioned random forest feature selection method was used. After selecting the most important second derivative of the wavelengths, regression methods such as random forest, support vector and multiple linear were used in order to model the concentration of the heavy metals including mercury, silver, cadmium and arsenic in the soil.

Results and discussion

With regard to modeling, the second derivative of short-wavelength infrared was selected in lead, the second derivative of near and short-wavelength infrared in silver, whereas for cadmium, the second derivative of ultraviolet and short-wavelength infrared, as well as for mercury, short-wavelength infrared was selected. The main soil compounds that absorb visible, near and short-wavelength infrared are powerful variables to increase the accuracy of soil's heavy metal prediction. Although most of the heavy metals in soil with a concentration greater than 1000 $mgkg^{-1}$ are difficult to detect from a spectral point of view, several studies have shown a significant correlation between the concentration of heavy metals and spectral data acquired in near and short wavelength infrared. According to the results, mercury heavy metal modeling using the random forest regression method with a root mean square error of 0.26 ppm has less error than the modeling of other heavy metals in this study. The lowest accuracy of the modeling is related to that of cadmium using the multiple linear regression method, which has a root mean square error of 12.18 ppm. For estimating most soil pollutants, the random forest regression method has a lower root mean square error than other methods heavy metals has been confirmed in the studies of other researchers.

The random forest regression method, as such, is able to better identify relationships between predictors and pollutant concentrations with minimum number of independent variables. Also, this method has a good performance and high accuracy in identifying a complex non-linear relationship between independent and dependent variables.

Conclusion

The results of the random forest feature selection method showed that the second derivative of the spectral data derived from soil samples in short wavelength infrared are the most important in modeling the concentration of heavy metals. Also, the results of this research showed that although statistically, there is no significant difference between the methods used to model the heavy metal concentrations in soil, but overall, machine learning methods have had better performance than the multiple linear regression method.

فصلنامه برنامه‌ریزی و آمایش فضا
دانشگاه تربیت مدرس

دوره ۲۶، شماره ٤، زمستان ۱٤۰۱

مدل‌سازی آلودگی خاک به فلزات سنگین با استفاده از روش‌های یادگیری ماشین و داده‌های طیف‌سنجی

علی شمس الدینی[۱*]، شهربانو اسماعیلی[۲]

دانشیار گروه سنجش از دور و سیستم اطلاعات جغرافیایی، دانشگاه تربیت مدرس، تهران، ایران
کارشناس ارشد گروه سنجش از دور و سیستم اطلاعات جغرافیایی، دانشکدهٔ منابع طبیعی و محیط زیست، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران

تاریخ دریافت: ۱٤۰۰/۰۷/۲۵
تاریخ پذیرش: ۱٤۰۱/۰٤/۲۱

چکیده

معادن و صنایع وابسته به آن، در زمان بهره‌برداری و پس از متروکه شدن، بر محیط زیست اطراف خود تأثیرگذارند. از جملهٔ این تأثیرات می‌توان به آلودگی آب‌های زیرزمینی و سطحی، و نیز آلودگی خاک اشاره کرد. مدل‌سازی غلظت فلزات سنگین با استفاده از روش‌های مقرون‌به‌صرفه لازمهٔ مدیریت و اصلاح آسیب‌های واردشده به محیط زیست است. هدف این تحقیق ارائهٔ چارچوبی به‌منظور مدل‌سازی فلزات سنگین در خاک با استفاده از طیف‌سنجی و نیز روش‌های مدل‌سازی آماری است. بدین منظور با استفاده از طیف‌سنجی، نمودار طیفی مربوط به ۵۳ نمونه خاک مربوط به منطقه‌ای در اطراف یک معدن متروکه در ایالت نیوساوث ولز استرالیا در طول موج‌های مرئی تا مادون قرمز میانی برداشت شد و مشتق دوم این داده‌ها محاسبه شد. سپس داده‌های طیفی مناسب برای مدل‌سازی غلظت فلزات سنگین شامل سرب، نقره، کادمیوم و جیوه با استفاده از روش انتخاب ویژگی جنگل تصادفی تعیین شدند و به‌عنوان ورودی برای مدل‌سازی غلظت فلزات سنگین با استفاده از روش‌های رگرسیون خطی چندمتغیره، جنگل تصادفی رگرسیون و ماشین‌بردار رگرسیون به‌کار گرفته شدند. نتایج نشان داد که طول موج‌های مادون قرمز میانی دارای اهمیت بیشتری به‌منظور مدل‌سازی غلظت فلزات سنگین در این تحقیق هستند. همچنین روش‌های غیرخطی یادگیری ماشین به‌خصوص جنگل تصادفی رگرسیون با مقادیر مجذور میانگین مربعات خطا ۰/۸ ppm و ضریب تعیین ۰/۵۱ برای سرب و ۹/۴ ppm و ۰/۴۶ برای کادمیوم دارای عملکرد بهتری نسبت‌به روش رگرسیون خطی چندمتغیره هستند.

واژگان کلیدی: *آلودگی خاک، فلزات سنگین، انتخاب ویژگی، یادگیری ماشین، مدل‌سازی*

***نویسنده مسئول:** E-mail: ali.shamsoddini@modares.ac.ir

۱. مقدمه

فلزات سنگین از آلاینده‌های غیرآلی پایدار با پایین‌ترین سطح تجزیهٔ بیولوژیکی محسوب می‌شوند (Hegazi et al., 2013; Gupta et al., 2001; Kumar et a;., 2017; Peng et al., 2019; Tasharrofi et al., 2018; Rhouati et al., 2018; Atieh e al., 2017) و به همین در بدن موجودات زنده تجمع می‌یابند (Jin et al., 2014; lau et al., 2019; Järup et al., 2003; Khan et al., 2008). برخلاف برخی از آلاینده‌ها، فلزات سنگین حتی در حداقل‌ترین میزان غلظت می‌توانند عوارض شدیدی ایجاد کنند. آژانس حفاظت از محیط زیست ایالات متحدهٔ آمریکا فلزات سنگین شامل سرب، آرسنیک، نیکل، کروم، مس، روی، کادمیوم و جیوه را ازجمله جدی‌ترین آلاینده‌ها معرفی کرده است (Schmidt et al., 2016). آلودگی فلزات سنگین در خاک پدیده‌ای رو به گسترش در سطح جهان است که مورد توجه بسیاری از دولت‌ها و نهادهای نظارتی قرار گرفته است (Bou Kheir et al., 2014). فعالیت‌های انسانی مانند استخراج معدن، حمل و نقل، دفع فاضلاب، و کوددهی تهدیدی مداوم برای سلامت خاک در دو دههٔ گذشته است (Wang et al., 2014). خاک‌های معادن متروکه به فلزات سنگین آلوده است و نظارت دقیق و منظمی باید بر این مکان‌ها صورت گیرد (Shamssodini et al., 2014). مکانیسم‌های انتقال فلزات سنگین به خاک مدت زیادی است که از مباحث مهم در آلودگی خاک است. به‌طور کلی بسیاری از خاک‌ها به طیف وسیعی از فلزات سنگین با میزان غلظت متفاوت وابسته به محیط جغرافیایی، زمین‌شناسی، فعالیت‌های انسانی و طبیعی آلوده هستند (Kanugo, 2000). این فلزات سنگین نه‌تنها از طریق معادن فلزات سنگین روی، مس، نقره و ...، بلکه از طریق هوا در قسمت‌های بافتی گیاهان هم می‌توانند نفوذ کنند (Hong-gui et al., 2012).

ارزیابی توزیع و نیز اصلاح آلودگی فلزات سنگین در خاک‌های حاصلخیز کشاورزی جزو دغدغه‌های مهم این سال‌ها محسوب می‌شود (Hong-gui et al., 2012). آلودگی فلزات در خاک تهدیدی جدی برای سلامتی انسان و ایمنی محصولات کشاورزی است. فلزات سنگین به‌ویژه سرب، کادمیوم، و جیوه می‌توانند به سلامتی انسان آسیب جدی وارد کند. آلودگی خاک به فلزات سنگین می‌تواند موجب بیماری‌ها و کوتاه شدن طول عمر انسان شود. فلزات سنگین نه‌تنها باعث تخریب محیط زیست می‌شوند، بلکه می‌توانند در سطح خاک رسوب کنند و سپس در بافت سبزیجات و محصولات کشاورزی جذب شوند (Hong-gui et al., 2012). با توجه به عواقب شدید ناشی از آلودگی فلزات سنگین در خاک کنترل آلودگی فلزات سنگین ضروری است (Bou Kheir et al., 2014). یکی از مهم‌ترین اقدامات در این خصوص، مدل‌سازی آلودگی خاک به فلزات سنگین از طریق استفاده از روش‌های غیرمستقیم است که نسبت‌به روش‌های مستقیم دارای هزینهٔ کم‌تری است و زمان کم‌تری برای انجام نیاز دارد (Shamssodini et al., 2014). مدل‌سازی آلودگی خاک چالشی فنی و علمی است. مدل‌های آماری غالباً برای توضیح مکانی ـ فضایی به‌کار می‌روند (Nolan & Hitt, 2006).

همچنین مدل‌سازی ریاضی و کامپیوتری در درک فرایندهای رخ‌داده در خاک به ما کمک می‌کند. روش‌های مدل‌سازی از نظر کمی فلزات سنگین موجود در خاک را با صحت خوبی پیش‌بینی می‌کنند. اکثر مدل‌های آماری مورد استفاده در رابطه با مدل‌سازی آلودگی خاک شامل رگرسیون خطی چندمتغیره، رگرسیون غیرخطی، رگرسیون لجستیک، رویکردهای بیزی، شبکه‌های عصبی مصنوعی یا طبقه‌بندی و رگرسیون جنگل‌های تصادفی هستند که برای مدل‌سازی با استفاده از متغیرهای مؤثر، مورد استفاده قرار گرفته‌اند (Rawlings et al., 1998; Burow et al., 2010; Mair & El-Kadi, 2013; Gurdak & Qi, 2012; Mattern et al., 2012).

مطالعات متعددی در رابطه با مدل‌سازی آلودگی خاک صورت گرفته است. در برخی از این مطالعات به‌منظور مدل‌سازی آلاینده‌های خاک، از خصوصیات خاک نظیر بافت به‌منظور ایجاد مدلی برای تخمین میزان غلظت آلاینده‌های خاک استفاده شده است. برای مثال بازوبندی و همکاران (۱۳۹٦)، با برداشت ۲۵۰ نمونه از خاک‌های استان گیلان برای برآورد میزان غلظت کادمیوم از ویژگی‌های خاک شامل درصد سیلت، شن، کربن آلی، PH، ازت کل، فسفر به‌عنوان متغیرهای ورودی در مدل‌سازی با استفاده از روش‌های شبکهٔ عصبی مصنوعی و شبکهٔ عصبی فازی تطبیقی[1] استفاده کردند. نتایج این مطالعه نشان داد شبکهٔ عصبی مصنوعی با ضریب تبیین (R^2) ۰/۸۳ و همچنین مجذور میانگین مربعات خطا (RMSE) ۱/۰۱ ppm و میانگین خطای مطلق (MAE) برابر ۰/٥٤ ppm روش مناسب‌تری نسبت‌به شبکهٔ عصبی فازی تطبیقی است. درحالی که در برخی دیگر از مطالعات به‌منظور مدل‌سازی غلظت آلاینده‌های فلزات سنگین در خاک، از روش‌های طیفی‌سنجی استفاده شده است. محمدی منور و باقرپور (۱۳۹٦) به بررسی توزیع فلزات سنگین شامل کادمیوم[2] و سرب[3] در سطح خاک شهرستان بهار در استان همدان به‌منظور ارزیابی وضعیت آلودگی فلزات سنگین در خاک مزارع کشت سیب‌زمینی پرداختند. در نمونه‌برداری از سطح خاک در عمق ۰ تا ٤٥ سانتی‌متر از روش سیستماتیک برای برداشت ۹۵ نمونه خاک استفاده شد و بر پایهٔ طیف‌های مرئی و فروسرخ نزدیک[4] در محدودهٔ طول موج‌های ۲۰۰۰–۳۷۰ نانومتر مدل‌سازی طیفی آلودگی انجام شد. روش‌های پیش‌پردازش بر روی داده‌های طیف‌سنجی شامل MSC[5]، SNV[6] و مشتق‌گیری بود و به‌منظور مدل‌سازی مقادیر فلزات سنگین خاک از روش‌های رگرسیون حداقل مربعات جزئی[7] و شبکهٔ عصبی مصنوعی[8] استفاده شد.

---

1. Adaptive Network Based Fuzzy Inference System (ANFIS)
2. Cadmium (CA)
3. Plumbum (Pb)
4. Near Infrared
5. Mass Spectrometry Centre (MSC)
6. Standard Normal Variate (SNV)
7. Partial Least Squares Regression (PLS regression)
8. Artificial Neural Network (ANN)

نتایج این مطالعه نشان داد شبکهٔ عصبی مصنوعی با ضریب تعیین ۰/۹۵ می‌تواند برای پیش‌بینی میزان کادمیوم و سرب موجود در سطح خاک مناسب باشد. تن[1] و همکاران (2019) مطالعه‌ای بر روی منطقهٔ معدنی زغال سنگ واقع در استان جیانگ‌سو، چین انجام دادند. آن‌ها زمین‌های کشاورزی این منطقهٔ معدنی را به مناطق A، B و C تقسیم و از هر منطقه ده نمونه از خاک‌های آلوده به روی، کروم، سرب و آرسنیک را برداشت کردند. پس از تعیین میزان آلودگی نمونه‌های برداشت‌شده در آزمایشگاه، با استفاده از طیف‌سنجی میدانی ASD[2] با لامپ هالوژن ۵۰ وات و با زاویهٔ ۱۵ درجه، نمونه‌های خاک آنالیز طیفی شدند.

در این مطالعه با استفاده از روش‌های جنگل‌های تصادفی، رگرسیون حداقل مربعات، ماشین‌بردار پشتیبان، به مدل‌سازی فلزات سنگین روی، کروم، آرسنیک و سرب پرداخته شد. نتایج این تحقیق نشان داد در مدل‌سازی غلظت فلزات سنگین در خاک، روش جنگل تصادفی با مقادیر ضریب تعیین و مجذور میانگین مربعات خطا به ترتیب ۰/۹۰٦۱ و ٦/۵۰۰۸ ppm برای مس؛ ۰/۹۱۱۰ و ٤/۵٦۸۳ ppm برای کروم، ۰/۹۹۱۲ و ۰/۵۳۲۷ ppm برای آرسنیک و ۰/۹۷۵٦ و ۱/۱٦۹٤ ppm برای روی دارای عملکرد بهتری نسبت‌به سایر روش‌هاست. (Hu et al., 2020) مطالعه‌ای بروی قسمت جنوبی استان شانگهای چین انجام دادند. در این منطقه فعالیت‌های صنعتی و تجاری بسیار مشهود است. از زمین‌های کشاورزی این منطقه ۱۸۲۲ نمونه خاک در سال ۲۰۱۳ برداشت شد.

برای آنالیز طیفی فلزات سنگین شامل مس، روی، نیکل، نقره، کادمیوم، آرسنیک، سرب و کروم از روش طیف‌سنجی جرمی پلاسما[3] استفاده شد. سپس به‌منظور مدل‌سازی فلزات سنگین از روش‌های رگرسیونی جنگل تصادفی، GBM[4] و روش خطی تعمیم‌یافته[5] استفاده کردند. نتایج این تحقیق نشان داد ضریب تعیین در روش آماری جنگل تصادفی در فلزات سنگین شامل، روی، مس، کروم، نیکل، نقره، کادمیوم، آرسنیک و سرب به ترتیب ۰/۸٤، ۰/٦٦، ۰/۵۹، ۰/۵۸، ۰/۵۸، ۰/۵۱ و ۰/۳۰ است که به‌نسبت سایر روش‌های استفاده‌شده در این مطالعه از نتایج بهتری برخوردار بود (Wang et al., 2020). بر روی منطقه‌ای دونگلی در استان تیانجین چین بر روی خاک‌های زمین‌های کشاورزی مطالعاتی به‌منظور مدل‌سازی آلاینده‌های فلزات سنگین خاک انجام دادند. به‌منظور مدل‌سازی آلاینده‌های خاک سرب، کروم، کادمیوم، آرسنیک، روی و نقره از روش جنگل تصادفی و رگرسیون کاربری زمین[6] استفاده شد. نتایج این تحقیق نشان داد در روش جنگل تصادفی ضریب تعیین برابر با ۰/۹۰ است که نسبت‌به روش رگرسیون کاربری زمین از عملکرد بهتری برخوردار است.

---

1. Tan
2. Enhanced Spectral Resolution (ASD)
3. Plasma Mass Spectrometry
4. Gradient Boosted Machine (GBM)
5. Generalized Linear (GLM)
6. Land Use Regression (LUR)

با توجه به مطالعات انجام‌شدهٔ فوق، هدف تحقیق حاضر، شناسایی و استخراج بااهمیت‌ترین طول موج‌های مؤثر در آنالیز فلزات سنگین خاک و استفاده از این طول موج‌ها به‌منظور مدل‌سازی میزان غلظت آلاینده‌های موجود در خاک یک منطقهٔ معدنی در نزدیکی سیدنی، استرالیا با استفاده از روش‌های رگرسیون جنگل تصادفی[1]، ماشین‌بردار رگرسیون[2]، و رگرسیون خطی چندمتغیره[3] است. نتایج حاصل از این مدل‌سازی‌ها با استفاده از روش آماری T-Test مورد مقایسه قرار گرفتند تا بهترین روش در مدل‌سازی میزان غلظت فلزات سنگین در خاک مشخص شود.

۲. مواد و روش‌ها

۲ـ۱. منطقهٔ مورد مطالعه و داده‌های مورد استفاده در این تحقیق

منطقهٔ مورد مطالعه در این تحقیق منطقهٔ یراندری[4] واقع در جنوب والز استرالیاست که منطقهٔ استخراج معادن نقره، سرب و روی در گذشته بوده است. پس از کشف سولفید سرب (PbS) در منطقه، فعالیت‌های استخراج معادن آغاز شد و درنتیجه تعدادی معدن کوچک زیرزمینی و چهار سایت فرآوری در مجاورت شهر یراندری تأسیس شد (Harrison et al., 2003). فعالیت‌های استخراج معادن به‌دلیل ساخت سد واراگامبا[5] در سال ۱۹۲۹ که یکی از منابع مهم تأمین آب شرب شهر سیدنی است، متوقف شد (Archer & Caldwell, 2004). درنتیجه معادن و سایت‌های فرآوری محصولات معادن در شهر یراندری رها شدند و به‌صورت متروک درآمدند. قرارگیری این منطقه در مجاورت رودخانه‌ها و نیز دریاچهٔ بوراگورانگ[6] که تأمین کنندهٔ ۸۰ درصد آب شرب شهر سیدنی است، موجب شده است تا نظارت منظم این منطقه از نظر آلودگی خاک به فلزات سنگین منتشرشده از معادن متروک منطقه از اهمیت زیادی برخوردار باشد. به‌منظور انجام این تحقیق، ۵۳ نمونه خاک در تاریخ‌های ۱۳ و ١٤ نوامبر سال ۲۰۱۳ از امتداد دو رود اصلی که در مجاور معادن رهاشده در منطقهٔ یراندری قرار دارند برداشت شد. نمونه‌ها تا عمق ۱۰ سانتی‌متر خاک برداشت شدند و پس از ترکیب خاک هر نمونه، با استفاده از الک دو میلی‌متری، هر نمونه الک شد. نمونه‌های خاکی که از الک ۲ میلی‌متری به‌دست آمده بودند، قبل از تجزیه و تحلیل شیمیایی و اندازه‌گیری طیفی، به مدت ٢٤ ساعت در کوره با دمای ٤۰ درجهٔ سانتی‌گراد خشک شدند.

۲ـ۲. آنالیز شیمیایی نمونه‌ها

در طی آنالیز شیمیایی، غلظت چهار فلز در نمونه‌های برداشت‌شده شامل نقره[7] (Ag)، کادمیوم (Cd)، جیوه[8] (Hg) و سرب (Pb) اندازه‌گیری شدند. بدین منظور، هریک از نمونه‌های خاک به خوبی خشک شدند و سپس با آسیاب آزمایشگاهی به صورت پودر ریز درآمدند. پس از آن یک گرم از نمونه‌ها برای آنالیز شیمیایی استفاده شد و باقیماندهٔ آن به‌منظور اندازه‌گیری طیفی نمونه‌های خاک مورد استفاده قرار گرفتند. پیش از قرارگیری نمونه‌ها در ظرف مایکروویو کوارتز به‌منظور حرارت‌دهی، ابتدا به مدت تقریبا دو ساعت این نمونه‌ها در اسید نیتریک غلیظ حل شدند. نمونه‌ها پس از حل شدن در محلول و حرارت دیدن، برای تجزیه و تحلیل آماده شدند.

1. Random Forest Regression (RFR)
2. Support Vector Regression (SVR)
3. Multiple Linear Regression (MLR)
4. Yerranderie
5. Warragamba
6. Burragorang
7. Silver (Ag)
8. Mercury (Hg)

پس از آماده شدن نمونه‌های محلول، از طیف‌سنجی جرمی پلاسما به روش القایی جفت‌شده[1] برای مشخص کردن میزان غلظت فلزات موجود در نمونه‌ها استفاده شد. جدول ۱ خلاصه‌ای از نتایج اندازه‌گیری فلزات سنگین حاصل از تجزیه و تحلیل شیمیایی برای ۵۳ نمونه خاک را نشان می‌دهد. همانطور که در جدول ۱ مشخص است غلظت آرسنیک و سرب در نمونه‌ها بالاست. این نتایج حاصل از تجزیه و تحلیل شیمیایی با نتایج حاصل از مطالعهٔ هریسون و همکاران (۲۰۰۳) که در همین منطقه انجام شده بود، منطبق است.

جدول ۱. اطلاعات آماری مربوط به مقادیر فلزات سنگین در ۵۳ نمونهٔ برداشت‌شده از خاک منطقهٔ مورد مطالعه

فلزات سنگین	حداقل	حداکثر	میانگین	انحراف معیار
سرب (%W/W)	۰/۰٤	۵/۰	۱/۸	۱/۳
نقره (mg/kg)	۲/۳	٤۸/٤	۱٤/۱	۱۲/۷
کادمیوم (mg/kg)	٤/۰	٦٤/٤	۱۸/٤	۱۳/۳۰
جیوه (mg/kg)	۰/۰۳	۱/۷	۰/٤	۰/۳

Table 1. Statistical information related to the amounts of heavy metals in 53 samples derived from the soil of the study area

۲ـ۳. اندازه‌گیری طیفی

طیف نمونهٔ خاک‌ها با استفاده از دستگاه طیف‌سنج FieldSpec 3 (ASD) در محیط آزمایشگاه اندازه‌گیری شد. دستگاه طیف‌سنج مورد استفاده در این تحقیق، قادر به اندازه‌گیری طیف الکترومغناطیس در طول موج‌های ۳۵۰ نانومتر تا ۲۵۰۰ نانومتر با فاصلهٔ نمونه‌برداری ۱/٤ نانومتر (برای طول موج‌های ۳۵۰ تا ۱۰۰۰ نانومتر) و ۲ نانومتر (برای طول موج‌های ۱۰۰۰ تا ۲۵۰۰ نانومتر) بود. سی طیف به‌عنوان طیف‌های نهایی برای هر نمونه توسط ردیاب تماسی[2] جمع‌آوری شد. به‌منظور تولید طیف ضریب بازتاب نسبی برای هر نمونه، طیف تابش با استفاده از یک سطح مرجع سفید اسپکترالون[3] ۹۹ درصد نرمال‌سازی شد. در مطالعات گذشته مشخص شده است که استفاده از داده‌های طیفی بدون پردازش، نتایج چندان مناسبی را در مدل‌سازی آلاینده‌های خاک به همراه ندارد (Shamsoddini et al., 2014). لذا در این تحقیق از مشتق دوم طیف‌های اندازه‌گیری‌شده به‌عنوان ویژگی‌های ورودی برای مدل‌سازی با استفاده از روش‌های مختلف یادگیری ماشین استفاده شد.

۳. انتخاب مهم‌ترین طول موج‌ها در فلزات سنگین خاک

انتخاب ویژگی می‌تواند با حذف ویژگی‌های نامناسب و نویز از داده‌های ورودی موجب افزایش صحت در نتایج پیش‌بینی شود (Panthong & Srivihok, 2015). حذف نکردن این ویژگی‌ها می‌تواند بار محاسباتی را بالا ببرد و موجب کاهش صحت در مدل‌سازی شود. روش‌های انتخاب ویژگی را می‌توان به دو کلاس تقسیم کرد:

[1]. Inductively Coupled Plasma Mass Spectrometry (ICP-MS)
[2]. Contact Probe
[3]. Spectralon

دستهٔ اول شامل روش‌هایی است که اهمیت و کیفیت ویژگی‌ها را برآورد می‌کند، و دستهٔ دوم شامل روش‌هایی است که هدف آن‌ها انتخاب زیرمجموعه‌ای از ویژگی‌هایی باکیفیت و بدون نویز است (Cehovin & Bosnic, 2010; Rodriguez –Galino et al., 2018). روش‌های زیادی به‌عنوان روش‌های انتخاب ویژگی ارائه شده است که در این میان روش جنگل تصادفی[1] رایج‌تر است. دلایل برتری روش جنگل تصادفی در انتخاب ویژگی‌ها عبارت است از حداقل خطای تعمیم و زمان انجام محاسبات کم و توانایی کار با داده‌هایی با ابعاد و تعداد نمونهٔ بسیار زیاد. همچنین در این روش، قبل از انتخاب ویژگی، میزان اهمیت متغیرها برآورد می‌شود که معیاری از توانایی پیش‌بینی هر متغیر در مدل‌سازی است (شمس‌الدینی و اسماعیلی، ۱۳۹۸؛ Ahmadlo et al., 2016; Couronné et al., 2018). ازاین‌رو در این تحقیق به‌منظور استخراج مهم‌ترین متغیرهای مؤثر از روش انتخاب ویژگی جنگل تصادفی رگرسیونی استفاده شد. پارامترهای نیازمند بهینه‌سازی در این روش شامل تعداد درختان و نیز تعداد متغیرها برای جداسازی در هریک از گره‌هاست که این مقدار همیشه از بُعد فضایی داده‌های ورودی کوچک‌تر است. در رابطه با تعداد درختان، افزایش تعداد درختان تا جایی ادامه می‌یابد که خطا کاهش می‌یابد و درمورد تعداد متغیرها به‌منظور جداسازی در هریک از گره‌ها به‌طور معمول آن را برابر جذر تعداد ویژگی‌های ورودی قرار می‌دهند (Belgiu & Dragut, 2016). در این تحقیق به‌منظور تعیین تعداد درخت مناسب برای انتخاب ویژگی و مدل‌سازی، تعداد درختان ۲۰۰۰ و ۱۵۰۰ ، ۱۰۰۰، ۵۰۰ بررسی شدند و در تنظیم تعداد متغیرهایی که در هر گره باید مورد استفاده قرار گیرند، با توجه به جذر تعداد کل متغیرها اعداد ۱ تا ۱۰ مورد بررسی قرار گرفتند و مقادیر بهینه برای این دو پارامتر برای هریک از فلزات سنگین، تعیین شد.

۴. مدل‌سازی

پس از انتخاب بااهمیت‌ترین مشتق دوم طول موج‌ها با استفاده از روش انتخاب ویژگی رگرسیون جنگل تصادفی، از روش‌های رگرسیونی مانند رگرسیون جنگل تصادفی، ماشین‌بردار رگرسیون و رگرسیون خطی چندمتغیره به‌منظور مدل‌سازی فلزات سنگین جیوه، نقره، کادمیوم و آرسنیک موجود در خاک استفاده شد. سپس با روش آماری T-Test، نتایج حاصل از روش‌های مدل‌سازی با یکدیگر مقایسه شدند.

---

[1]. Random Forest

۴ـ۱. رگرسیون جنگل تصادفی

جنگل تصادفی یکی از روش‌های آماری غیرخطی و یک تکنیک یادگیری گروهی[1] است که از مجموعه‌ای از درختان تصمیم‌گیری تشکیل شده است و نتیجهٔ نهایی حاصل تجمیع نتایج حاصل از این درختان تصمیم‌گیری بوده که به کاهش واریانس در مقایسه با درختان تصمیم‌گیری مجزا منجر می‌شود (Breiman, 1996; Prasad et al., 2006; Couronné et al., 2018). جنگل‌های تصادفی نسبت‌به کمبود داده‌های آموزشی حساسیت ندارند و بیشترین استفاده از این روش در کاربردهایی است که حجم داده‌ها بالا و به انجام محاسبات در زمان کوتاه‌تری نیاز است (Breiman, 1996; Prasad et al. 2006). هر درخت تصمیم‌گیری در روش جنگل تصادفی، با توجه به انتخاب تصادفی ویژگی‌های ورودی به آن درخت، تمایل به یادگیری الگوهایی متفاوت از سایر درختان را در داده‌ها داشته و با ایجاد نتایجی متفاوتی برای هر درخت، نتیجهٔ نهایی دارای صحت بالاتری است (Hastie et., 2008). به عبارت دیگر هر شاخهٔ درخت به‌طور جداگانه با استفاده از یک زیرمجموعهٔ تصادفی از متغیرهای پیش‌بینی رشد می‌کند. یکی از مزایای روش جنگل تصادفی استفاده از روش خارج از کیسه[2] (OOB) در نمونه‌ها برای استخراج بااهمیت‌ترین متغیرها و مشخص کردن میزان خطا در مجموعهٔ درختان تولید شده است که نیاز به اعتبارسنجی را برطرف می‌کند (Breiman, 2001a; Liaw & Wiener, 2002). یکی از پارامترهای مهم در این روش، تعداد درختانی است که برای مدل‌سازی مورد استفاده قرار می‌گیرد. با افزایش تعداد درختان، خطای هم‌گرایی یا تعمیم می‌تواند اتفاق بیفتد (Rodriguez-Galiano et al. 2014).

الگوریتم یادگیری رگرسیون جنگل تصادفی با استفاده از نمونه‌برداری بوت استرپ و با تقسیم داده‌های آموزشی به زیرمجموعه‌های M شروع می‌شود. در مرحلهٔ بعد درخت رگرسیون T_i برای هر زیرمجموعه با استفاده از زیرمجموعه‌هایی از ویژگی‌ها به‌طور تصادفی تنظیم می‌شود. این فرایند تقسیم گره‌ها به جنگلی از درختان رگرسیون M منجر می‌شود. پس از برازش مدل در کل مجموعهٔ آموزشی پاسخ $(\hat{y})$ برای مجموعه دادهٔ آزمایشی $(\acute{x})$ با میانگین‌گیری پیش‌بینی می‌شود که به شرح زیر است:

$$\hat{y} = \frac{1}{M}\sum_{i=1}^{M} T_i\ (\acute{x}) \qquad (۱)$$

[1]. Ensemble
[2]. Out of Bag (OOB)

که در آن M تعداد کل درختان رگرسیون و $T_i(\acute{x})$ خروجی درخت رگرسیون است (Hafsa et al., 2020). در این مطالعه به‌منظور تنظیم پارامترهای مدل شامل تعداد از روش آزمون و خطا اعتبارسنجی تقاطعی[1] استفاده شد و در پیدا کردن بهینه‌ترین مقدار در تعیین تعداد ویژگی‌ها[2] در هر درخت، اعداد ۱، ۲، ۳، ٤ و ۵ مورد بررسی قرار گرفتند و همچنین برای تعیین تعداد درختان[3] اعداد ۱۰۰، ۲۰۰، ۳۰۰، ٤۰۰، ۵۰۰، ٦۰۰، ۷۰۰، ۸۰۰، ۹۰۰ و ۱۰۰۰ مورد بررسی قرار گرفتند و تا مقدار بهینه برای آن‌ها تعیین شود.

۴ـ۲. ماشین‌بردار رگرسیون (SVR)

ماشین‌بردار رگرسیون یکی از روش‌های یادگیری ماشین بوده که برخلاف روش ماشین‌بردار پشتیبان[4]، در مسائل رگرسیون و پیش‌بینی مورد استفاده قرار می‌گیرد (Aryafar et al., 2012). ماشین‌بردار رگرسیون براساس روش‌های یادگیری هسته عمل می‌کند و در فضای نمونه‌ها، اقدام به یافتن یک رابطهٔ غیرخطی بین ورودی‌های مدل و خروجی مدل می‌پردازد. عملکرد ماشین‌بردار رگرسیون به پارامترهای ورودی آن بستگی دارد. نقض محدودیت[5] یا C یک پارامتر منظم‌سازی است، این پارامتر میزان خطای مدل در مرحلهٔ آموزش را به حداقل می‌رساند. اگر مقدار عددی C خیلی کوچک باشد تأکید کافی بر تناسب داده‌های آموزشی وجود نخواهد داشت. اگر C خیلی بزرگ باشد خطای هم‌گرا اتفاق خواهد افتاد (Wang et al., 2003). لذا لازم است تا این پارامتر بهینه‌سازی شود. برای تنظیم کردن مقدار عددی مناسب C، پارامتر اپسیلون نیز باید در کنار C تنظیم شود. مقدار عددی مناسب اپسیلون به نوع داده‌های ورودی بستگی دارد که معمولاً نامعلوم است. اپسیلون اثر وجود نویز در داده‌های آموزشی را در مدل‌سازی کاهش می‌دهد. بهینه‌سازی مقدار مناسب اپسیلون در ماشین‌بردار رگرسیون از نظر تئوری بسیار مهم است (Aryafar et al., 2012). ماشین‌بردار رگرسیون به‌صورت رابطه‌ای ۲ محاسبه می‌شود:

$$Y_{svr}(x) = \sum_{i=1}^{n} \beta_i \; k(x; x_i) + b \qquad (۲)$$

β_i و x_i به ترتیب وزن و موقعیت هر کدام از بردارهای پشتیبان هستند. همچنین n تعداد ماشین‌های بردار و b بایاس است و $k(x; x_i)$ تابع هستهٔ مربوط به x_i است (Dewi & Chen, 2019).

---

[1]. Cross Validation
[2]. Mtry
[3]. Ntree
[4]. Support Vector Machine (SVM)
[5]. Cost of Constraints Violation (C)

در این تحقیق به‌منظور تنظیم پارامترهای C و اپسیلون، از روش اعتبارسنجی تقاطعی استفاده شد. در همین راستا جهت تعیین مناسب‌ترین مقدار نقض محدودیت یا همان C اعداد ۲، ٤، ۸، ١٦، ۳۲، ٦٤، ۱۲۸ و ۵۱۲ مورد بررسی قرارگرفتند. سپس در انتخاب بهینه‌ترین مقدار اپسیلون[1] اعداد ۰/۱، ۰/۲، ۰/۳، ۰/٤، ۰/۵، ۰/٦، ۰/۷، ۰/۸، ۰/۹ و ۱ مورد بررسی قرار گرفتند و مقادیر بهینه برای این دو پارامتر برای هریک از فلزات سنگین به‌دست آمد.

۴ـ۳. رگرسیون خطی چندمتغیره

روش‌های رگرسیونی مانند رگرسیون خطی یا رگرسیون خطی چندمتغیره به‌طور گسترده در سنجش از دور به‌منظور مدل‌سازی استفاده می‌شود (اسماعیلی و شمس‌الدینی، ۱۳۹۸). استفاده از این مدل‌ها با هدف استفاده از کم‌ترین پیش‌بینی‌کننده‌ها برای توضیح بیشترین تغییرپذیری در متغیر پاسخ است (Couronné et al., 2018). به‌منظور مدل‌سازی غلظت فلز سنگین با استفاده از متغیرهای مستقل (داده‌های طیفی)، می‌توان از رابطهٔ زیر استفاده کرد:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_n + C \qquad (۳)$$

β_n ضریب رگرسیون، C مقدار ثابت و X_n مقادیر مربوط به داده‌های طیفی برداشت شده است. n بستگی به تعداد متغیرهای طیفی دارد که در روش انتخاب ویژگی برای هریک از فلزات سنگین انتخاب می‌شوند.

۴ـ۴. مقایسهٔ روش‌های مدل‌سازی با آزمون T-Test

آزمون آماری T-Test یا همان (آزمون T دانش‌آموز[2]) برای مقایسهٔ میانگین بین دو گروه استفاده می‌شود (McDonald et al., 2014; Altman, 1990). این روش یکی از محبوب‌ترین روش‌های آزمون آماری است که برای آزمایش این‌که آیا اختلاف میانگین بین دو گروه از نظر آماری معنی‌دار است یا خیر استفاده می‌شود (شمس‌الدینی، ۱۳۹۵، ۱۳۹٦). فرضیهٔ صفر یعنی هر دو میانگین از نظر آماری برابر هستند. آزمون T-Test سه نوع است که شامل آزمون T تک‌نمونه‌ای[3]، آزمون T دو جامعهٔ مستقل[4]، آزمون T برای زوج متغیرها[5] است (Jaykaran et al., 2010). آزمون T دو نمونهٔ مستقل، میانگین دو گروه مستقل از پیش‌بینی‌ها را با یکدیگر مقایسه می‌کند. این آزمون آماری استنباطی نشان می‌دهد که بین دو گروه غیرمرتبط (برای مثال دو گروه دادهٔ پیش‌بینی‌شده توسط دو روش متفاوت) از نظر آماری تفاوت معناداری وجود دارد یا ممکن است که این دو روش عملکردی شبیه به هم داشته باشند (Gerald, 2018). در این مطالعه به‌منظور مقایسهٔ عملکرد روش‌های مختلف مدل‌سازی فلزات سنگین از این آزمون آماری استفاده شد.

---

[1]. Epsilon
[2]. the Student's T Test
[3]. One Sample T Test
[4]. Two Independent Sample T Test
[5]. Paired Sample T Test

۵. نتایج

همانگونه که در جدول ۲، که دربارهٔ بااهمیت‌ترین طول موج‌های استخراج‌شده با روش انتخاب ویژگی رگرسیون جنگل تصادفی است، مشاهده می‌شود، در مدل‌سازی سرب مشتق دوم طول موج‌های مادون قرمز میانی، و در رابطه با نقرهٔ مشتق دوم طول موج‌های مادون قرمز نزدیک و میانی، برای مدل‌سازی کادمیوم مشتق دوم طول موج‌های ماورای بنفش و مادون قرمز میانی، و همچنین برای مدل‌سازی جیوه طول موج‌های مادون قرمز میانی انتخاب شدند. پس از استخراج بااهمیت‌ترین طول موج‌ها با استفاده از روش‌های رگرسیونی شامل رگرسیون جنگل تصادفی (RFR)، ماشین‌بردار رگرسیون (SVR) و رگرسیون خطی چندمتغیره (MLR) مقادیر فلزات سنگین در نمونه‌های برداشت‌شده، مدل‌سازی شد (جدول ۳). با توجه به نتایج، مدل‌سازی فلز سنگین جیوه با استفاده از روش رگرسیون جنگل تصادفی با مجذور میانگین مربعات خطای ۰/۲٦ ppm نسبت‌به مدل‌سازی سایر فلزات سنگین در این مطالعه دارای خطای کم‌تری است. همچنین مدل‌سازی فلز سرب با استفاده از روش رگرسیون جنگل تصادفی نیز در مقایسه با فلزات کادمیوم و نقره دارای صحت بالاتری است. کم‌ترین صحت مدل‌سازی مربوط به مدل‌سازی فلز سنگین کادمیوم با استفاده از روش رگرسیون خطی چندمتغیره است که دارای مجذور میانگین مربعات خطای ۱۲/۱۸ ppm است. با توجه به جدول ۳، اگرچه صحت مدل‌سازی فلزات سنگین به صورت قابل قبولی به‌دست آمده است، ولی مدل‌های به‌دست‌آمده نتوانسته‌اند به‌خوبی میزان تغییرات فلزات سنگین در خاک را در نمونه‌های به‌دست‌آمده مشخص کنند، و بهترین ضریب تعیین به‌دست آمده برای این فلزات سنگین مقدار ۵۱ درصد است که برای فلز سرب به‌دست آمده است و این مقدار برای سایر فلزات بسیار کم‌تر است. پس از ارزیابی میزان خطای مدل‌سازی، همانگونه که در جدول ٤ مشاهده می‌شود، با استفاده از روش آزمون T برای زوج متغیرها عملکرد روش‌های مدل‌سازی مختلف با هم به‌صورت آماری مقایسه شد. با توجه به نتایج به‌دست آمده از این آزمون آماری، برای مدل‌سازی نقره، روش ماشین‌بردار رگرسیون نسبت‌به روش جنگل تصادفی رگرسیون دارای عملکردی بهتر و نیز روش جنگل تصادفی رگرسیون نسبت‌به روش رگرسیون خطی چندمتغیره دارای عملکردی بهتر از نظر آماری برای مدل‌سازی کادمیوم در سطح معنی‌داری ۵ درصد هستند. برای سایر موارد در سطح معنی‌داری ۵ درصد، تفاوتی بین عملکرد روش‌های مختلف وجود ندارد.

جدول ۲. انتخاب بااهمیت‌ترین طول موج‌ها با روش رگرسیون جنگل تصادفی

نوع آلاینده	طول موج‌های انتخاب‌شده (نانومتر)
سرب	۲۲۷۰–۲۲٤۸–۲۱۲٤–۱۹۵۰–۱۹۱۵–۱۸۵۸
نقره	۲٤٤۲–۲٤۱٦–۲۰۸۸–۱۸۱٦–۱۰۹۰
کادمیوم	۲۳۷٤–۲۱۲۰–۱۷٦۰–۱۷٤۸–۱۷۲۰–۱۳۹۵–۱۳٦٤–۳۸٤–۳٦۹
جیوه	۲۰٦۰–۱۷۲٦–۱٤٤٤

Table 2. Selection of the most important wavelengths with the random forest regression method

جدول ۳. مدل‌سازی با روش‌های رگرسیون جنگل تصادفی، ماشین‌بردار رگرسیون و رگرسیون خطی چندمتغیره

نام فلز سنگین	RFR		SVR		MLR	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
سرب	۰/۸۰	۰/۵۱	۱/۰۲	۰/۲۸	۱/۱۳	۰/٤٦
نقره	۱۰/٤٦	۰/۰٤	۹/۲۷	۰/۱۰	۱۱/۵۳	۰/۱۸
کادمیوم	۹/٤۰	۰/٤٦	۱۰/۳۰	۰/۳۸	۱۲/۱۸	۰/۳۰
جیوه	۰/۲٦	۰/۲٦	۰/۲۹	۰/۱۹	۰/۲۹	۰/۳۷

Table 3. Modeling with random forest regression, support vector regression and multiple linear regression

جدول ٤. مقایسه‌ای روش‌های مدل‌سازی با روش آماری T-Test

آلایندهٔ خاک	مدل‌سازی	P-level
نقره	RFR-SVR	۰/۰۱
نقره	RFR-MLR	۱/۹۱
نقره	SVR-MLR	۱/۷٤
کادمیوم	RFR-SVR	۱/۰٦
کادمیوم	RFR-MLR	۰/۰
کادمیوم	SVR-MLR	۰/۳۱
جیوه	SVR-RFR	۰/۳۲
جیوه	SVR-MLR	۰/۱۲
جیوه	RFR-MLR	۰/۰٦
سرب	SVR-RFR	۱/۷٦
سرب	SVR-MLR	۳/۵٦
سرب	RFR-MLR	۰/۱۰

Table 4. Comparison of modeling methods with T-Test statistical method

۶. بحث

در این تحقیق، پس از محاسبهٔ مشتق دوم مقادیر طیفی، بااهمیت‌ترین طول موج‌های آنالیز طیفی برای مدل‌سازی مقادیر فلزات سنگین سرب، جیوه، نقره و کادمیوم با استفاده از روش انتخاب ویژگی جنگل تصادفی استخراج شدند. با توجه به جدول ۲، عمدتاً طول موج‌های مادون قرمز میانی (بزرگ‌تر از ۱۳۰۰ نانومتر) به‌عنوان بهترین داده‌های طیفی به‌منظور مدل‌سازی فلزات سنگین توسط روش انتخاب ویژگی جنگل تصادفی انتخاب شدند. این درحالی است که در مطالعات دیگر نیز به داده‌های طیفی حاصل از طول موج‌های مادون قرمز نزدیک و میانی به‌عنوان طول موج‌های مناسب برای مدل‌سازی فلزات سنگین خاک اشاره شده است (Kooistra et al., 2001; Omaran et al., 2016; Bao et al., 2020). ترکیبات اصلی خاک که تابش امواج مرئی، مادون قرمز نزدیک وکوتاه را جذب می‌کنند، متغیرهای قدرتمندی برای افزایش صحت پیش‌بینی فلزات سنگین خاک هستند (Lamine et al., 2019; Omaran et al., 2016).

اگرچه بیشتر فلزات سنگین در خاک با غلظت بیشتر از $mgkg^{-1}$ ۱۰۰۰ از نظر طیفی به‌سختی قابل شناسایی هستند، مطالعات متعددی ثابت کرده‌اند هم‌بستگی قابل‌توجهی بین فلزات سنگین موجود در خاک و داده‌های طیفی با طول موج مادون قرمز نزدیک و میانی وجود دارد (Malley & Willimas, 1997; Wu et al., 2005; Bao et al., 2020).

روش مورد استفاده به‌منظور مدل‌سازی فلزات سنگین خاک می‌تواند تأثیر مهمی در صحت مدل‌سازی داشته باشد (Shamssodini et al., 2014). مدل‌سازی آماری و رایانه‌ای به درک فرایندهای رخ‌داده در خاک کمک می‌کند. نتایج حاصل از مدلی که دارای صحت مناسبی باشد، می‌تواند از نظر کمی تغییرات مکانی و زمانی فلزات سنگین را تخمین بزند (Dube et al., 2000). نتایج ارائه‌شده در جدول‌های ٤ و ۵ نشان می‌دهد که اگرچه تفاوت چندانی از نظر آماری میان عملکرد روش‌های مورد استفاده در این تحقیق وجود ندارد، روش جنگل تصادفی رگرسیون نسبت‌به سایر روش‌ها دارای مجذور میانگین مربعات خطای کم‌تری برای تخمین اکثر آلاینده‌هاست. عملکرد مناسب روش جنگل تصادفی در مدل‌سازی مقادیر فلزات سنگین در خاک در مطالعات سایر محققان نیز تأیید شده است (Wang et al., 2020; Qiu et al., 2016; Hafsa et al., 2020). روش جنگل تصادفی رگرسیون قادر به شناسایی بهتر روابط بین پیش‌بینی‌کننده‌ها و غلظت آلاینده‌ها با حداقل متغیرهای مستقل است (Brokamp et al., 2017; Qiu et al., 2016). همچنین روش جنگل تصادفی رگرسیون در شناسایی روابط پیچیدهٔ غیرخطی بین متغیرهای مستقل و وابسته از عملکرد خوب و از صحت بالایی برخوردار است (Araki et al., 2018; Qiu et al., 2016)، در حالی که مدل رگرسیون خطی چندمتغیره بر روابط خطی متمرکزند، روش جنگل تصادفی رگرسیون یک روش غیرخطی و براساس یادگیری ماشین عمل می‌کند و دارای ساختار پیچیده‌ای نیست (Wang et al., 2020).

در کنار روش جنگل تصادفی رگرسیون، روش ماشین‌بردار رگرسیون نیز دارای عملکرد مناسبی بود. عملکرد مناسب روش ماشین‌بردار رگرسیون به‌منظور مدل‌سازی فلزات سنگین در مطالعات دیگر نیز تأیید شده است (Sakizadeh et al., 2016). روش‌های یادگیری ماشین مانند ماشین‌بردار رگرسیون می‌توانند خطاهای مربوط به مدل آموزشی را با ایجاد رابطهٔ پیچیده‌ای بین متغیرهای مستقل و متغیر وابسته بهبود بخشند که این مسئله می‌تواند برای حل مسائل رگرسیون غیرخطی، مؤثر باشد (Xue et al., 2020). در مطالعهٔ حاضر نشان داده شد که روش‌های یادگیری ماشین مانند روش جنگل تصادفی رگرسیون و ماشین‌بردار رگرسیون قابلیت بالاتری نسبت‌به روش رگرسیون خطی چندمتغیره دارند.

در مطالعات متعدد دیگری نیز به برتری روش‌های یادگیری ماشین نسبت‌به روش‌های رگرسیون خطی اشاره شده است (Brokamp et al., 2017; Araki et al., 2018; Wang et al., 2020). تعیین صحیح توزیع فضایی فلزات سنگین با استفاده از روش‌های یادگیری ماشین در مناطقی که بیش از حد استاندارد آلوده به فلزات سنگین هستند یک روش قابل اعتماد برای مدل‌سازی و توضیح توزیع فضایی آلاینده‌ها برای اتخاذ راهبردهای پیشگیرانه و اقدامات اصلاحی است (Guan et al., 2019; Xue et al., 2020).

۷. نتیجه‌گیری

هدف اصلی این تحقیق، ارائهٔ چارچوبی به‌منظور مدل‌سازی فلزات سنگین موجود در خاک شامل سرب، نقره، کادمیوم و جیوه با استفاده از طیف‌سنجی و روش‌های مدل‌سازی آماری بود. بدین منظور از روش‌های مدل‌سازی رگرسیون خطی چندمتغیره، جنگل تصادفی رگرسیون و ماشین‌بردار رگرسیون به‌منظور مدل‌سازی متغیر وابسته یعنی غلظت فلزات سنگین استفاده شد. همچنین به‌عنوان متغیر مستقل از مشتق دوم داده‌های طیفی حاصل از طیف‌سنجی در طول موج‌های مرئی تا مادون قرمز میانی استفاده شد. نتایج حاصل از روش انتخاب ویژگی با استفاده از روش انتخاب ویژگی جنگل تصادفی نشان داد که مشتق دوم داده‌های طیفی برداشته‌شده از نمونه‌های خاک در طول موج‌های مادون قرمز میانی دارای بیشترین اهمیت در مدل‌سازی غلظت فلزات سنگین خاک هستند. همچنین نتایج این تحقیق نشان داد که اگرچه از نظر آماری، تفاوت معنی‌داری بین روش‌های استفاده‌شده برای مدل‌سازی غلظت برخی از فلزات سنگین وجود ندارد، ولی درمجموع روش‌های یادگیری ماشین از عملکردی بهتر نسبت‌به روش رگرسیون خطی چندمتغیره برخوردارند. به همین سبب توصیه می‌شود که در مدل‌سازی غلظت فلزات سنگین خاک، شامل فلزات سنگین جیوه، کادمیوم و نقره، از روش‌هایی شامل جنگل تصادفی رگرسیون و ماشین‌بردار رگرسیون استفاده شود.

منابع

- اسماعیلی، ش.، و شمس‌الدینی، ع. (۱۳۹۸). ادغام خصیصه‌های اجتماعی ـ اقتصادی و سنجش از دوری به‌منظور مدل‌سازی رشد فیزیکی شهر کرج. *برنامه‌ریزی و آمایش فضا*، ۲۳ (۱)، ۱۱۹-۱۵۰.
- بازوبندی ا.، امام قلی‌زاده، ص.، قربانی، ه.، و افشاری بدرلو، ت. (۱۳۹۶). برآورد میزان غلظت کادمیوم خاک با استفاده از مدل‌های ANN و ANFIS. *محیط زیست طبیعی (منابع طبیعی ایران)*، ۷۰(۳)، ۵۰۹-۵۲۳.
- حیدرپور، م.، و علیایی، م. (۱۳۹۲). انتشار آلودگی‌های نفتی در خاک تحت تأثیر شرایط مختلف خاک و آلاینده. *مهندسی عمران*، ۱۳ (۲)، ۳۹ـ ۵۱.
- شمس‌الدینی، ع. (۱۳۹۵). برآورد ساختار جنگل کاج با استفاده از تصاویر راداری، *برنامه‌ریزی و آمایش فضا*، ۲۰ (۱)، ۵۳-۷۸.
- شمس‌الدینی، ع. (۱۳۹۵). قابلیت‌سنجی کارایی داده‌های لیدار و اپتیک به‌منظور استخراج پارامترهای ساختاری جنگل. *برنامه‌ریزی و آمایش فضا*، ۲۱ (۲)، ۱۱۹-۱۴۵.
- محمدی منور، ح.، و باقرپور، ح. (۱۳۹۶). کاربرد روش طیف‌سنجی مرئی و فرو سرخ نزدیک در تشخیص آلودگی خاک به کادمیوم و سرب با مدل‌سازی رگرسیونی و شبکهٔ عصبی مصنوعی. *مهندسی بیوسیستم ایران (علوم کشاورزی ایران)*، ۴۸ (۱)، ۳۷-۴۳.

- Atieh, M. A., Ji, Y., & Kochkodan, V. (2017). Metals in the Environment: Toxic Metals Removal. Bioinorganic Chemistry and Applications, 4309198. doi:10.1155/2017/4309198
- Archer, M., & Caldwell, R. (2004). Response of six Australian plant species to heavy metal contamination at an abandoned mine site. *Water, air, and soil pollution*, *157*, 257-267.
- Aryafar, A., Gholami, R., Rooki, R., & Doulati Ardejani, F. (2012). Heavy metal pollution assessment using support vector machine in the Shur River, Sarcheshmeh copper mine, Iran. *Environmental Earth Sciences*, *67*(4), 1191-1199. doi:10.1007/s12665-012-1565-7
- Araki, S., Shima, M., & Yamamoto, K. (2018). Spatiotemporal land use random forest model for estimating metropolitan NO2 exposure in Japan. *Sci Total Environ*, *634*, 1269-1277. doi:10.1016/j.scitotenv.2018.03.324
- Ahmadlou, M., Delavar, M.R., Shafizadeh-Moghadam, H., & Tayebi, A. (2016). Modeling Urban Dynamics Using Random Forest: Implementing ROC and TOC for Model Evaluation, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B2, XXIII ISPRS Prague, Czech Republic, Congress, pp. 12–19.
- Altman, D.G. (1990). *Practical Statistics for Medical Research*. Boca Raton, Florida: CRC Press
- Bazoubandi, A., EmamGholi Zadeh, S., Ghorbani, H., & Afshari Badrlou, T. (2017). Prediction of cadmium concentration of soil using ANN and ANFIS models. *Journal of Natural Resources of Iran*, *3*, 509-523
- Bou Kheir, R., Shomar, B., Greve, M. B., & Greve, M. H. (2014). On the quantitative relationships between environmental parameters and heavy metals pollution in Mediterranean soils using GIS regression-trees: The case study of Lebanon. *Journal of Geochemical Exploration*, *147*, 250-259.

- Bao, Y., Meng, X., Ustin, S., Wang, X., Zhang, X., Liu, H., & Tang, H. (2020). Vis-SWIR spectral prediction model for soil organic matter with different grouping strategies. Catena,Vol 195, 104703 . doi:10.1016/j.catena.104703
- Breiman, L. (2001a). Random forests. *Mach Learn*, *45*, 5–32
- Belgiu,M.,&Drăguţ, L. (2016). Randomforest in remote sensing: A review of applications and future directions. *Isprs Journal of Photogrammetry & Remote Sensing*, *114*, 24–31.
- Brokamp, C., Jandarov, R., Rao, M. B., LeMasters, G., & Ryan, P. (2017). Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. *Atmospheric Environment*, *151*, 1-11. doi:https://doi.org/10.1016/j.atmosenv.2016.11.066
- Burow KR, Nolan BT, Rupert MG, Dubrovsky NM. (2010). Nitrate in groundwater of the United States. Environ SciTechnol 44(13):4988–4997, PP 1991−2003
- Couronné, R., Probst, P., Boulesteix, A.l. (2018). Random forest versus logistic regression: a large-scale benchmark experiment, BMC Bioinformatics, 19(270), pp, 1-14
- Dube, A, Z. R., Kowalkowski, T., Cukrowska, E., & Buszewski, B. (2001). Adsorption and Migration of Heavy Metals in Soil. *Polish Journal of Environmental Studies*, *10*(1), 10.
- Dewi, C., & Chen, R.C. (2019). Random forest and support vector machine on features selection for regression analysis. *Int. J. Innov. Comput. Inf. Control*, *15* (6), 2027-2037.
- Doksum, K., Tang, S., & Tsui, K.-W. (2008). Nonparametric Variable Selection: The EARTH Algorithm. *Journal of the American Statistical Association*, *103*(484), 1609-1620. doi:10.1198/016214508000000878
- Esmaeili, S., & Shamsoddini, A. (2019). Fusion of socio-economic and remote sensing-based attributes for Karaj physical growth modeling. *MJSP*, *23* (1), 119-150 URL: http://hsmsp.modares.ac.ir/article-21-25154-fa.html
- Guan, Q., Zhao, R., Wang, F., Pan, N., Yang, L., Song, N., Lin, J. (2019). Prediction of heavy metals in soils of an arid area based on multi-spectral data. *Journal of Environmental Management*, *243*, 137-143. doi: https://doi.org/10.1016/j.jenvman.2019.04.109
- Gerald, B. (2018). A Brief Review of Independent, Dependent and One Sample t-test. *International Journal of Applied Mathematics and Theoretical Physics*, *4*(2), doi:10.11648/j.ijamtp.20180402.13
- Gurdak, J.J., & Qi SL. (2012). Vulnerability of recently recharged groundwater in principle aquifers of the United States to nitrate contamination. *Environ Sci Technol*, *46*(11), PP 6004–6012
- Gupta, V. K., Gupta, M., & Sharma, S. (2001). Process development for the removal of lead and chromium from aqueous solutions using red mud--an aluminium industry waste. *Water Research*, *35*(5), 1125–1134. https://doi.org/10.1016/s0043-1354(00)00389-4
- Hastie, T, T. R., & Friedman, J. (2008). The Elements of Statistical Learning. Springer, 2nd edn.
- Harrison, J., Heijnis, H., & Caprarelli, G. (2003). Historical pollution variability from abandoned mine sites, Greater Blue Mountains World Heritage Area, New South Wales, Australia. *Environmental Geology*, *43*(6), 680-687. doi:10.1007/s00254-002-0687-8
- Hong-gui, D., Teng-feng, G., Ming-hui, L., & Xu, D. (2012). Comprehensive Assessment Model on Heavy Metal Pollution in Soil. *International Journal of Electrochemical Science*, *7*, 5286-5296.

- Hafsa, N., Rushd, S., Al-Yaari, M., & Rahman, M. (2020). A Generalized Method for Modeling the Adsorption of Heavy Metals with Machine Learning Algorithms. Water, no 12, 1-22. doi:10.3390/w12123490
- Hegazi, H. A. (2013). Removal of heavy metals from wastewater using agricultural and industrial wastes as adsorbents. *HBRC Journal*, *9*(3), 276-282. doi: https://doi.org/10.1016/j.hbrcj.2013.08.004
- Heidarpoor, M., & Oliaei, M. (2013). Oil Contamination Propagation Patterns in Soils. *MCEJ*, *2*, 39-51. URL: http://mcej.modares.ac.ir/article-16-9574-fa.html.
- Jaykaran, Ch. (2010). How to select appropriate statistical test? *Journal of Pharmaceutical Negative Results*, *1*, 61. doi:10.4103/0976-9234.75708
- Hu, B., Xue, J., Zhou, Y., Shao, S., Fu, Z., Li, Y., Chen, S., Qi, L., Shi, Z. (2020). Modelling bioaccumulation of heavy metals in soil-crop ecosystems and identifying its controlling factors using machine learning. Environ. *Pollut*, *262*, 1-11, https://doi.org/10.1016/j.envpol.2020.114308.
- Jin, L., Zhang, G., & Tian, H. (2014). Current state of sewage treatment in China. *Water Research*, *66*, 85–98. https://doi.org/10.1016/j.watres.2014.08.014
- Järup L. (2003). Hazards of heavy metal contamination. *British Medical Bulletin*, *68*, 167–182. https://doi.org/10.1093/bmb/ldg032
- Kooistra, L., Wehrens, R., Leuven, R. S. E. W., & Buydens, L. M. C. (2001). Possibilities of visible–near-infrared spectroscopy for the assessment of soil contamination in river floodplains. *Analytica Chimica Acta*, *446*(1), 97-105. doi:https://doi.org/10.1016/S0003-2670(01)01265-X
- Kellner, J., & Celisse, A. (2019). A one-sample test for normality with kernel methods. *Bernoulli*, *25*(3), 1816-1837. doi:10.3150/18-BEJ1037
- Kanungo, S. B., & Mohapatra, R. (2000). Leaching Behavior of Various Trace Metals in Aqueous Medium from Two Fly Ash Samples. *Journal of Environmental Quality*, *29*(1), 188-196. doi:https://doi.org/10.2134/jeq2000.00472425002900010024x
- Kumar, P., & Saravanan, A. (2017). Sustainable wastewater treatments in textile sector. Editor(s): Subramanian Senthilkannan Muthu, In The Textile Institute Book Series, Sustainable Fibres and Textiles, Woodhead Publishing, pp. 323-346, ISBN 9780081020418,
- Khan, S., Cao, Q., Zheng, Y.M., Huang, Y.Z., & Zhu, Y.G. (2008). Health risks of heavy metals in contaminated soils and food crops irrigated with wastewater in Beijing, China. Environ. *Pollut*, *152*, 686–692.
- Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R News*, *2*(3),18–22.
- Lamine, S., Petropoulos, G. P., Brewer, P. A., Bachari, N. E., Srivastava, P. K., Manevski, K., . . . Macklin, M. G. (2019). Heavy Metal Soil Contamination Detection Using Combined Geochemistry and Field Spectroradiometry in the United Kingdom. *Sensors (Basel)*, *19*(4). doi:10.3390/s19040762
- Lau, Y. J., Khan, F. S. A., Mubarak, N. M., Lau, S. Y., Chua, H. B., Khalid, M., & Abdullah, E. C. (2019). Chapter 10 - Functionalized carbon nanomaterials for wastewater treatment. In S. Thomas, Y. Grohens, & Y. B. Pottathara (Eds.), Industrial Applications of Nanomaterials (pp. 283-311): Elsevier.

- Mattern, S., Raouafi, W., Bogaert, P., Fasbender, D., & Vanclooster, M. (2012). Bayesian data fusion (BDF) of monitoring data with a statistical groundwater contamination model to map groundwater quality at the regional scale. *J Water Resour Prot*, *4*(11), 929–943.
- Malley, D. F., & Williams, P. C. (1997). Use of Near-Infrared Reflectance Spectroscopy in Prediction of Heavy Metals in Freshwater Sediment by Their Association with Organic Matter. *Environmental Science & Technology*, *31*(12), 3461-3467. doi:10.1021/es970214p
- Mair, A., & El-Kadi, A. (2013). Logistic regression modeling to assess groundwater vulnerability to contamination in Hawaii, USA. *Journal of Contaminant Hydrology*, *153*, 1-23
- Mcdonald, J. H. (2014). Handbook of Biolological Statistics. Third Edition. Baltimore, Maryland, U.S.A: Sparky House Publishing, University of Delaware
- Mohammadi Monavar, H., & Bagher pour, H. (2017). Application of visible and near-infrared spectroscopy for identification of cadmium (Cd) and lead (Pb) pollution in soil using regression models and ANN, *48*, 37-43. Doi: 10.22059/IJBSE.2017.61559
- Nolan, B.T., & Hitt, K.J. (2006). Vulnerability of shallow groundwater and drinking-water wells to nitrate in the United States. *Environ Sci Technol*, *40*(24), 7834–7840. https://doi.org/10.1021/es060911u
- Omran, E.-S. E. (2016). Inference model to predict heavy metals of Bahr El Baqar soils, Egypt using spectroscopy and chemometrics technique. *Modeling Earth Systems and Environment*, *2*(4), 1-17. doi:10.1007/s40808-016-0259-7
- Pacyna, J.M. (1994). Global Perspectives on Lead, Mercury and Cadmium Cycling in the Environment. Edited by T.C. Hutchingson Wiley Eastern Ltd. 315-328
- Prasad, A., Iverson, L., & Liaw, A. (2006). Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems*, *9*, 181-199. doi:10.1007/s10021-005-0054-1
- Peng, B., Fang, S., Tang, L., Ouyang, X., & Zeng, G. (2019). Chapter 8 - Nanohybrid Materials Based Biosensors for Heavy Metal Detection. In L. Tang, Y. Deng, J. Wang, J. Wang, & G. Zeng (Eds.), Nanohybrid and Nanoporous Materials for Aquatic Pollution Control (pp. 233-264): Elsevier.
- Pereira, L. A., Taylor-Rodríguez, D., & Gutiérrez, L. (2020). A Bayesian nonparametric testing procedure for paired samples. *Biometrics*, *76*(4), 1133-1146. doi: https://doi.org/10.1111/biom.13234
- Qiu, L., Wang, K., Long, W., Wang, K., Hu, W., & Amable, G. S. (2016). A Comparative Assessment of the Influences of Human Impacts on Soil Cd Concentrations Based on Stepwise Linear Regression, Classification and Regression Tree, and Random Forest Models. *PLoS One*, *11*(3), e0151131. doi:10.1371/journal.pone.0151131
- Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M., & Ribeiro, L. (2014). Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: a case study in an agricultural setting (Southern Spain). *The Science of the Total Environment*, *476-477*, 189–206. https://doi.org/10.1016/j.scitotenv.2014.01.001
- Rhouati, A., Marty, J.-L., & Vasilescu, A. (2018). Chapter 7 - Metal Nanomaterial-Assisted Aptasensors for Emerging Pollutants Detection. In D. P. Nikolelis & G.-P. Nikoleli (Eds.), Nanotechnology and Biosensors (pp. 193-231): Elsevier.

- Shamsoddini, A. (2017). LiDAR and optical data capability assessment for plantation structural parameter estimation Assessment of LiDAR and optical data capability in the estimation of structural parameters of plantations. *MJSP*, *2*, 119-145. URL: http://hsmsp.modares.ac.ir/article-21-7739-fa.html
- Shamsoddini, A., Raval, S., & Taplin, R. (2014). Spectroscopic analysis of soil metal contamination around a derelict mine site in the Blue Mountains, Australia. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *II-7*, 75-79. doi:10.5194/isprsannals-II-7-75-2014
- Sakizadeh, M., Mirzaei, R., & Ghorbani, H. (2016). Support vector machine and artificial neural network to model soil pollution: a case study in Semnan Province, Iran. Neural Computing and Applications, 28(11), 3229-3238. doi:10.1007/s00521-016-2231-x
- Szefer, P., Ikuta, K., Kushiyama, S., Frelek, K., & Geldon, J. (1997). Distribution of Trace Metals in the Pacific Oyster, Crassostrea gigas, and Crabs from the East Coast of Kyushu Island, Japan. Bulletin of Environmental Contamination and Toxicology, 58(1), 108-114. doi:10.1007/s001289900307
- Szefer, P., Ali, A. A., Ba-Haroon, A. A., Rajeh, A. A., Gełdon, J., & Nabrzyski, M. (1999). Distribution and relationships of selected trace metals in molluscs and associated sediments from the Gulf of Aden, Yemen. *Environmental Pollution*, *106*(3), 299-314. doi: https://doi.org/10.1016/S0269-7491(99)00108-6
- Schmidt, S.-A., Gukelberger, E., Hermann, M., Fiedler, F., Großmann, B., Hoinkis, J., .Bundschuh, J. (2016). Pilot study on arsenic removal from groundwater using a small-scale reverse osmosis system—towards sustainable drinking water production. *Journal of Hazardous Materials*, *318*, 671-678. doi: https://doi.org/10.1016/j.jhazmat.2016.06.005
- Tan, K., Ma, W., Wu, F., & Du, Q. (2019). Random forest-based estimation of heavy metal concentration in agricultural soils with hyperspectral sensor data. *Environ Monit Assess*, *191*(7), 446. doi:10.1007/s10661-019-7510-4
- Tasharrofi, S., Sadegh Hassani, S., Taghdisian, H., & Sobat, Z. (2018). 24 - Environmentally friendly stabilized nZVI-composite for removal of heavy metals. In C. M. Hussain & A. K. Mishra (Eds.), New Polymer Nanocomposites for Environmental Remediation (pp. 623-642): Elsevier.
- Vapnik,V. (1998). *Statistical Learning Theory*. John Wiley & Sons Inc., New York, p. 736.
- Wang, W., Xu, Z., Lu, W., & Zhang, X. (2003). Determination of the spread parameter in the Gaussian kernel for classification and regression. *Neurocomputing*, *55*(3-4), 643-663. doi:10.1016/s0925-2312(02)00632-x
- Wang, H., Yilihamu, Q., Yuan, M., Bai, H., Xu, H., & Wu, J. (2020). Prediction models of soil heavy metal (loid)s concentration for agricultural land in Dongli: A comparison of regression and random forest. *Ecological Indicators*, *119*. doi:10.1016/j.ecolind.2020.106801
- Wang, J., Cui, L., Gao, W., Shi, T., Chen, Y., & Gao, Y. (2014). Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy. *Geoderma*, *216*, 1-9. doi:10.1016/j.geoderma.2013.10.024
- Wu, Y., Chen, J., Wu, X., Tian, Q., Ji, J., & Qin, Z. (2005). Possibilities of reflectance spectroscopy for the assessment of contaminant elements in suburban soils. *Applied Geochemistry*, *20*(6), 1051-1059. doi:https://doi.org/10.1016/j.apgeochem.2005.01.009
- Wuana, R. A., & Okieimen, F. E. (2011). Heavy Metals in Contaminated Soils: A Review of Sources, Chemistry, Risks and Best Available Strategies for Remediation. *ISRN Ecology*, 2011, 1-20. doi:10.5402/2011/402647
- Xue, Y., Zou, B., Wen, Y., Tu, Y., & Xiong, L. (2020). Hyperspectral Inversion of Chromium Content in Soil Using Support Vector Machine Combined with Lab and Field Spectra. *Sustainability*, *12*(11). doi:10.3390/su12114441