

برآورد عیار پتاسیم شورابه زیرزمینی با استفاده از تصاویر ماهواره‌ای سنتینل و الگوریتم جنگل تصادفی (مطالعه موردی پلایای خور و بیابانک، استان اصفهان)

چکیده

یکی از عناصر پرمصرف که نقش مهمی در تولید پایدار کشاورزی دارد، پتاسیم است. پتاسیم خاک سطحی در پلایا از پتاسیم موجود در آب زیرزمینی نشأت می‌گیرد و در نتیجه، بین پتاسیم خاک سطحی و عیار پتاسیم شورابه زیرزمینی همبستگی وجود دارد. هدف این پژوهش، استفاده ترکیبی از الگوریتم جنگل تصادفی (RF) و تصویر ماهواره‌ای برای یافتن ارتباط بین پتاسیم سطحی خاک و شاخص‌های سنجش‌ازدور تعریفی مختص این مطالعه به‌منظور پیش‌بینی عیار پتاسیم شورابه زیرزمینی در پلایای خور و بیابانک استان اصفهان است. بدین منظور تعداد ۶۰ نمونه خاک از لایه ۵-۰ سانتی‌متری جهت اندازه‌گیری پتاسیم لایه سطحی (متغیر وابسته) نمونه‌برداری شد. به‌منظور تعیین مختصات نمونه‌گیری‌ها از روش ابر مکعب لاتین استفاده شد. همچنین ۱۲ گمانه جهت استخراج و اندازه‌گیری عیار پتاسیم شورابه زیرزمینی حفر شد. از ۱۲ باند ماهواره سنتینل ۲ و چهار عمل اصلی ریاضی برای تعریف شاخص (متغیرهای مستقل) به‌منظور مدل‌سازی پتاسیم لایه سطحی و درنهایت برآورد عیار پتاسیم شورابه زیرزمینی استفاده شد. داده‌ها به دو دسته ۷۰ درصد برای واسنجی (آموزش) و ۳۰ درصد برای اعتبار سنجی (آزمون) دسته‌بندی شده و با الگوریتم RF در محیط Google Colab و با استفاده از زبان برنامه‌نویسی پایتون مدل‌سازی شدند. نتایج این الگوریتم با شاخص‌های آماری R^2 ، MSE، RMSE و MAE به ترتیب ۰/۵۱، ۰/۱۷۹، ۰/۱۳۳۸ و ۰/۱۱۳۰ به دست آمد. نتایج این پژوهش تأیید کننده کارایی داده‌های سنجش‌ازدور و الگوریتم یادگیری ماشین در پیش‌بینی عیار پتاسیم شورابه زیرزمینی است.

واژه‌های کلیدی: پایتون، سنجش از دور، کفه‌های نمکی، مدل‌سازی، یادگیری ماشین.

ABSTRACT

One of the widely used elements that plays an important role in sustainable agricultural production is potassium. The potassium in the surface soil of the playa originates from the potassium present in the underground water. As a result, there is a correlation between the surface soil potassium and the potassium grade of the groundwater. The aim of this research is to utilize a combination of the random forest (RF) algorithm and satellite imagery to establish the relationship between soil surface potassium and remote sensing indicators. This will enable the prediction of the potassium grade of the underground in Koor and Biabank playa in Isfahan province. For this purpose, 60 soil samples were taken from the 0-5 cm layer to measure potassium in the surface layer (dependent variable). In order to determine the sampling coordinates, the Latin supercube method was used. Twelve boreholes were drilled to extract and measure the potassium grade of underground saline water. The 12 bands of the Sentinel-2 satellite and four main mathematical operations were used to define the index (independent variables) to model the potassium content of the surface soil layer and ultimately estimate the rate of potassium grade in the underground saline water. The data were categorized into two groups: 70% for calibration (training) and 30% for validation (testing). The data were modeled using the RF algorithm in the Google Colab environment and implemented with the Python programming language. The results of this algorithm were obtained with R^2 , MSE, RMSE and MAE statistical indices of 0.51, 0.0179, 0.1338 and 0.1130 respectively. The results of this research confirm the effectiveness of remote sensing data and machine learning algorithms in predicting the potassium grade of saline groundwater.

Keywords: Python, Remote sensing, Salt pans, Modeling, Machine learning.

مقدمه

پتاسیم یکی از عناصر پرمصرف مورد نیاز گیاه است؛ زیرا حفظ آب، ارزش غذایی، مقاومت در برابر بیماری محصولات غذایی و عملکرد را بهبود می‌بخشد (Al Rawashdeh, 2016). در چند دهه‌ی اخیر با افزایش رشد جمعیت و نیاز روزافزون به تولید مواد غذایی بیشتر امروز بیش از ۹۰ درصد تولید پتاسیم به‌عنوان کود استفاده می‌شود. عنصر پتاسیم در شورابه‌های زیرزمینی کفه‌های نمکی به‌وفور یافت می‌شود. کفه‌های نمکی به نام پلایا از ویژگی‌های بارز چشم‌انداز حوزه‌های هیدرولوژیکی در مناطق خشک و نیمه‌خشک است. سطح کلیه پلایاها از یک لایه نسبتاً ضخیم نمکی به نام پوسته نمک تشکیل شده که فضای خالی بین بلورهای آن با شورابه زیرزمینی حاوی پتاسیم پر شده است. یکی از عناصر معدنی اصلی موجود در شورابه‌های زیرزمینی پتاسیم است؛ که می‌تواند برای مصارف کشاورزی و صنعتی مورد بهره‌برداری قرار گیرد (Shaw et al., 2011). به دلیل تعادل آب منفی برای بیش از نیمی از سال و داشتن خاصیت مویبندی

نزدیک به سطح، تبخیر آب بیشتر بوده و همین امر منجر به رسوب املاح محلول در سطح می‌گردد (Gulibert et al., 1997)؛ بنابراین با بررسی این رسوبات امکان جانمایی مناطق دارای عیار بالای پتاسیم جهت استخراج شورابه زیرزمینی در پلایا فراهم می‌آید. پراکنش مکانی رسوبات نمکی در پلایاها و وسعت بسیار بالای کویر نمک و شرایط پیچیده اقلیمی حاکم بر آن، اندازه‌گیری‌های میدانی جهت تخمین عیار یک ماده معدنی را مشکل می‌سازد. داده‌های سنجش‌ازدور برای پایش محیط‌های پلایا با سطوح تبخیری در مقیاس‌ها و حسگرهای مختلف استفاده می‌شوند. تصاویر ماهواره‌ای می‌توانند تخمین عیار یک ماده معدنی را در سطح منطقه تسهیل و دسترسی به نقاط دور یا غیرقابل دسترس را فراهم سازند (Hewson et al., 2017). یکی از روش‌های نوین جهت تخمین منابع معدنی استفاده ترکیبی از الگوریتم یادگیری ماشین و سنجش‌ازدور است. در سال‌های اخیر پیشرفت‌های زیادی در تکنیک‌های مربوط به نقشه‌برداری زمین‌شناسی، نقشه‌برداری ژئوفیزیکی و نقشه‌برداری ژئوشیمیایی صورت گرفته است. ظهور کامپیوترهای قدرتمند و نرم‌افزارهای تخصصی به تکنیک‌های مختلف ترکیب این داده‌ها رونق بخشیده که منجر به تخمین ذخایر معدنی شده است. امروزه در بسیاری از پژوهش‌ها جهت برآورد عیار عناصر معدنی از سنجش‌ازدور و الگوریتم‌های یادگیری ماشین استفاده می‌گردد. متداول‌ترین مدل‌های یادگیری ماشینی که همراه نقشه‌برداری دیجیتال برای تخمین عیار مقادیر مواد معدنی و ویژگی‌های خاک سطحی استفاده می‌شوند عبارت‌اند از جنگل تصادفی، ماشین بردار پشتیبان، خطوط رگرسیون تطبیقی چند متغیره و مدل شبکه عصبی مصنوعی (Moghaddam et al., 2020). الگوریتم‌های یادگیری ماشینی می‌توانند با دقت بالا پیش‌بینی انجام دهند و این مدل‌ها خطاها را به حداقل می‌رسانند و نتایج دقیق‌تری نسبت به روش‌های سنتی ارائه می‌دهند. الگوریتم جنگل تصادفی یک مدل یادگیری گروهی نظارت‌شده است که به‌طور گسترده در مسائل طبقه‌بندی و رگرسیون استفاده می‌شود و به دلیل سرعت پردازش بالا و نتایجی با دقت عالی، توجه روزافزونی را به خود جلب کرده است (Breiman, 2001).

از آنجاکه بخش وسیعی از کشور ایران را مناطق خشک و نیمه‌خشک تشکیل می‌دهد؛ بیش از ۶۰ پلایای بزرگ و کوچک در کشور وجود دارد که این پلایاها سرشار از شورابه‌هایی است که می‌توانند مورد استحصال و بهره‌برداری قرار گیرند (موسوی و همکاران، ۱۳۹۵). پلایای منطقه خوروبایانک در استان اصفهان یکی از مناطقی است که دارای ذخایر خوب و مناسب پتاسیم است و با توجه به اینکه عنصر پتاسیم یکی از عناصر مهم و ضروری در بخش کشاورزی و صنعتی است. یکی از راهکارهای نوین و کم‌هزینه تخمین منابع معدنی استفاده ترکیبی از تکنیک سنجش‌ازدور و الگوریتم‌های یادگیری ماشین است. هدف اصلی این پژوهش استفاده از سنجش‌ازدور و الگوریتم جنگل تصادفی به منظور تخمین پتاسیم سطحی خاک پلایا و ارزیابی رابطه بین پتاسیم و شاخص‌های تصاویر ماهواره‌ای جهت تخمین عیار پتاسیم شورابه زیرزمینی است؛ که خود نوآوری این تحقیق نسبت به سایر پژوهش‌های قبلی است.

پیشینه پژوهش

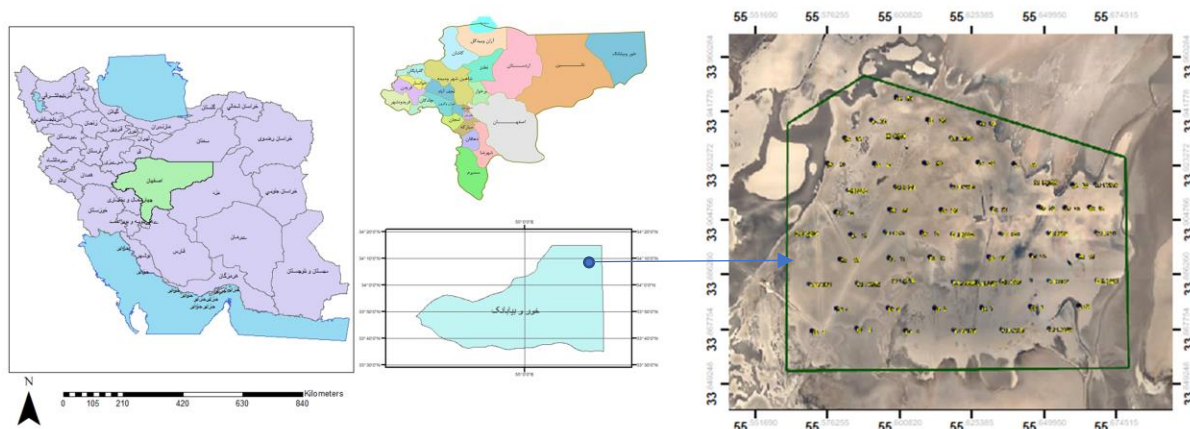
مطالعاتی در زمینه تخمین مواد معدنی به‌وسیله سنجش‌ازدور و الگوریتم‌های یادگیری ماشین انجام گرفته است و این روش کاملاً کاربردهای خود را در بخش معدنی و تحقیقات زیرزمینی پیدا کرده است. به‌عنوان نمونه کاربرد شبکه‌های عصبی مصنوعی برای نقشه‌برداری پتانسیل معدنی، غرب هند (Porwal et al., 2003)، استفاده شبکه‌های عصبی مصنوعی به‌عنوان ابزاری برای نقشه‌برداری پتانسیل معدنی با GIS اسپانیا (Rigol-Sanchez et al., 2003)، کاربرد شبکه‌های عصبی مصنوعی برای تهیه نقشه پتانسیل معدنی برای کانی‌سازی مس - طلا برزیل (Leite et al., 2009)، الگوریتم‌های یادگیری ماشین و کاربرد آن‌ها در تخمین ذخیره کانسار طلا (Dutta et al., 2010)، نقشه‌برداری پتانسیل ذخایر طلا - نقره، مطالعه موردی کره جنوبی (Oh et al., 2010)، استفاده از الگوریتم ژنتیک برای نقشه‌برداری فلز پایه در هند (Lewkowski et al., 2010)، برآورد غلظت آهن با استفاده از یک بردار پشتیبان شبکه عصبی مصنوعی کانسار آهن در شرق ایران (Maleki et al., 2014)، مقایسه مدل جنگل‌های تصادفی مبتنی بر داده و یک روش دانش‌محور برای نقشه‌برداری آینده‌نگری معدنی: مطالعه موردی برای ذخایر طلا کانادا (McKay & Harris, 2016)، مقایسه روش‌های یادگیری ماشینی برای تخمین عیار کانسار مس (Jafarsteh et al., 2018)، الگوریتم جنگل تصادفی برای ارزیابی غلظت سدیم در یک منطقه معدنی در

کانادا (Schnitzler et al., 2019). Nwaila et al. (2020) یک الگوریتم یادگیری ماشین به نام GS-Pred برای تعیین عیار ذخایر طلا در آفریقای جنوبی ارائه کردند. این الگوریتم داده‌های رسوب‌شناسی و عیار طلا را برای پیش‌بینی دقیق عیار طلا و شناسایی خودکار رخساره‌های کنگلومرا ترکیب می‌کند. نتایج اعتبارسنجی این الگوریتم نشان می‌دهد که GS-Pred دقیق‌تر از تکنیک‌های یادگیری ماشین معمول و کریجینگ است. (Safae et al. (2020 با کاربرد تصاویر لندست ۸، نقشه کفه‌های نمکی و پلایاها را در غرب ایالات متحده آمریکا تهیه کرده و از یک شاخص طیفی و الگوریتم ماشین بردار پشتیبان برای مشخص کردن مناطق پلایاها استفاده کردند. نتایج ارزیابی این الگوریتم نشان داد که کیفیت نقشه‌برداری برای اکثر مناطق کفه نمکی، دارای ضریب کاپا بیش از ۹۰ درصد بود. (Radwin et al. (2021 با استفاده از داده‌های چند طیفی لندست، نقشه‌برداری کانی‌شناسی در حوضچه‌های تبخیری بونویل و یوتا در آمریکا انجام دادند و با استفاده از سنسورهای Landsat-5 Thematic Mapper (TM) و Landsat-8 Operational Land Imager (OLI)، مجموعه‌ای از شاخص‌های ریاضی برای نقشه‌برداری مناطق کانی‌شناسی غالب هالیت، گچ و کربنات‌ها در این حوضه ایجاد کردند. نتایج این شاخص‌ها منبعی برای نقشه‌برداری کانی‌شناسی در حوضچه‌های تبخیری در کل جهان با کاربردهای متنوع برای موضوع‌هایی درباره کاربری زمین و تغییرات محیطی فراهم می‌کنند. (Naimi et al. (2021 با استفاده از سنجش پروگزیمال برخی خصوصیات ذاتی خاک را در اراضی خشک با کاربرد طیف‌سنجی Vis-NIR، MIR و pXRF برای توصیف و تخمین میزان رس، شن و سیلت خاک، کربنات کلسیم، گچ، کربن آلی و هدایت الکتریکی خاک را بررسی کرده و از روش رگرسیون حداقل مربعات جزئی برای پیش‌بینی خواص خاک استفاده کردند. نتایج ایشان نشان داد که طیف Vis-NIR-SWIR برای پیش‌بینی بافت خاک، شاخص RPI برای پیش‌بینی کربنات کلسیم و داده‌های طیف Vis-NIR-SWIR و mid-IR برای پیش‌بینی گچ و pXRF برای پیش‌بینی شوری خاک می‌تواند مورد استفاده قرار گیرند.

مواد و روش‌ها

منطقه مطالعاتی

این پلایا در ۴۰ کیلومتری شرق شهرستان خور و بیابانک و در استان اصفهان واقع شده و یکی از شاخص‌ترین پلایاهای ایران از منظر وجود معادن پتاسیم است. پلایا در مختصات جغرافیایی ۳۳ درجه و ۳۷ دقیقه تا ۳۴ درجه و ۱۱ دقیقه عرض شمالی و ۵۴ درجه و ۱۶ دقیقه تا ۵۵ درجه و ۱۴ دقیقه طول شرقی می‌باشد. خور و بیابانک به دلیل آن که در فلات مرکزی ایران قرار گرفته است دارای آب‌وهوای گرم و خشک است. این پلایا از پلیگون نمک است که برای استخراج شورابه زیرزمینی مناسب هستند. وسعت منطقه مورد مطالعه ۴۹/۰۵ کیلومتر مربع می‌باشد که در قسمت شرقی پلایای خور و بیابانک واقع شده است (شکل ۱). ارتفاع متوسط منطقه مورد مطالعه از سطح دریا ۷۰۵ متر، متوسط بارندگی ۸۲ میلی‌متر در سال و تعداد ساعات آفتابی ۳۳۰۰ ساعت در سال می‌باشد.



شکل ۱. موقعیت جغرافیایی منطقه مورد مطالعه و نمونه برداری شده در نقشه ایران و استان اصفهان، منطقه خور و بیابانک

نمونه برداری، آماده سازی و اندازه گیری پتاسیم سطحی و عیار پتاسیم شورابه زیرزمینی

در این مطالعه، جهت تعیین محل نمونه ها از روش ابرمکعب لاتین^۱ بهره گرفته شده است. این روش، یک طرح تصادفی طبقه بندی شده است که باعث نمونه برداری مؤثر به کمک توزیع چند متغیر می شود. یک شبکه مربع حاوی موقعیت های نمونه، یک شبکه لاتین است؛ اگر فقط یک واحد نمونه برداری در هر ردیف و هر ستون وجود داشته باشد. یک هایپرکیوب لاتین، تعمیم این مفهوم به تعداد اختیاری از ابعاد است. به طوری که هر واحد نمونه برداری فقط در یک صفحه آن قرار گرفته باشد. هایپرکیوب لاتین شامل نمونه برداری n مقدار از توزیع تشریح شده هر متغیر است. توزیع تجمعی هر متغیر به n فاصله با احتمال مساوی تقسیم می شود و یک مقدار از هر فاصله به طور تصادفی انتخاب می شود. سپس، n مقدار به دست آمده برای هر متغیر با متغیرهای دیگر جفت می شود. یک روش یک پوشش کامل از هر متغیر را ایجاد می کند (Minasny & McBratney, 2006). با توجه به روش ابرمکعب لاتین و مساحت منطقه مورد مطالعه، تعداد ۶۰ نمونه از مرکز پلیگون (دارای مختصات و غلظت پتاسیم سطحی) از کل محدوده مورد مطالعه برای اندازه گیری پتاسیم لایه سطحی و برآزش آن ها به شاخص های سنجش از دور حاصل از تصاویر ماهواره ای، برداشت شد. همچنین ۱۲ عدد از این نمونه ها از مرکز دوازده پلیگون انتخابی جهت مغزه گیری برای اندازه گیری عیار پتاسیم شورابه زیرزمینی انتخاب شد. مغزه گیری، اصلی ترین روش نمونه گیری برای مناطقی مانند پلایاها و کفه های نمکی است که دارای رسوبات سخت می باشند. مغزه، نمونه ای استوانه ای شکل که از طبقات درون زمین در عمق پائین و اندازه های مختلف گرفته می شود. قطر مغزه ها، به قطر مته حفاری بستگی دارد. برای نمونه برداری فیزیکی در سطح پلایا به دلیل شرایط سخت محیطی و آب و هوایی و در برخی مناطق باتلاقی؛ همچنین از آنجایی که مغزه گیری و حفاری بسیار هزینه بر و واریانس فیزیکی نمونه ها در پلایا کم است، به همین تعداد نمونه برداری سطحی و مغزه گیری، بسنده شد. مغزه گیری برای استخراج شورابه زیرزمینی توسط دستگاه حفاری تا عمق ۱۰ متری تهیه شده و نمونه های سطحی از عمق ۵ سانتیمتری برداشت شدند. نمونه ای از پلیگون، دستگاه مغزه گیری و مغزه های برداشت شده در شکل (۲) ارائه شده است. برای انجام آزمایش اندازه گیری پتاسیم لایه سطحی (درصد) و عیار پتاسیم شورابه زیرزمینی (گرم بر لیتر) مناسب ترین روش استفاده از دستگاه فیلم فتومتر^۲ است (Estefan & Ryan 2013).

^۱ Latin Hypercube Sampling

^۲ Flame photometer



شکل ۲. پلایا، دستگاه مغزه‌گیری، نمونه‌های مغزه، پلیگون‌های منطقه مورد مطالعه و نمونه‌برداری از آن‌ها به منظور اندازه‌گیری پارامترهای مورد نیاز

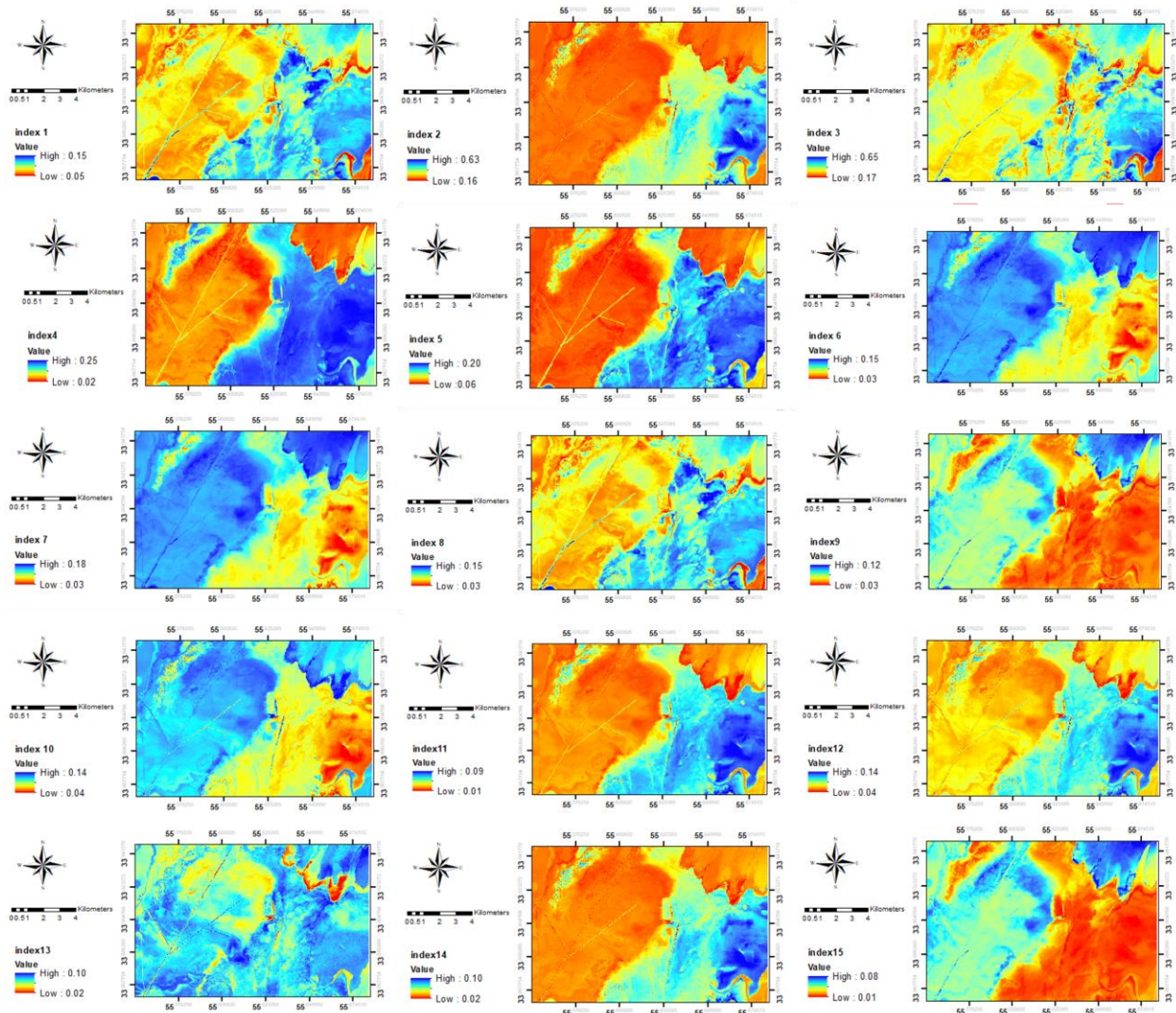
محاسبه شاخص‌های ماهواره‌ای و برآورد پتاسیم سطحی خاک

در این پژوهش ابتدا شاخص‌های ماهواره‌ای موجود مانند RSCI، NDSI، SI و ... به منظور مدل‌سازی پتاسیم سطحی خاک بررسی شدند که همبستگی و نتایج قابل قبولی نداشته و نتایج آن در این مطالعه ارائه نشده است. سپس با استفاده از ۴ عمل پایه‌ای ریاضی (جمع، تفریق، ضرب و تقسیم) و ۱۲ باند ماهواره SENTINEL-2، حدود ۶۱ میلیون عبارت ریاضی (شاخص) تولید و ضریب همبستگی پیرسون بین پتاسیم سطحی خاک و شاخص‌های تولیدی بررسی شدند. ۱۵ شاخص تولیدشده‌ای که حداقل باند در آن استفاده شده و همبستگی بالاتری با پتاسیم سطحی خاک اندازه‌گیری شده داشتند؛ انتخاب و به عنوان ورودی الگوریتم RF استفاده شدند. معادله ریاضی ۱۵ شاخص انتخاب شده در جدول (۱) ارائه شده است. باندهای ۱ تا ۱۲ با عبارات b_1 تا b_{12} نشان داده شده است.

جدول ۱. شاخص‌های سنجش‌ازدور تعریف شده منتج از باندهای ماهواره SENTINEL-2 برای منطقه مورد مطالعه

نام شاخص	معادله ریاضی
Index1	$b_1/b_3/b_4 * b_5 * b_7/b_9 + b_8a * b_{12}$
Index2	$b_1 - b_4 + b_5/b_6 * b_7/b_8 - b_9 * b_{11} + b_{12}$
Index3	$b_1 + b_3/b_4/b_6 * b_7/b_8 * b_8a$
Index4	$b_1 + b_2/b_4 * b_5/b_6/b_8 * b_8a - b_{11} + b_{12}$
Index5	$b_1/b_2/b_4 * b_5 * b_7/b_9 + b_{11}$
Index6	$b_2 * b_3 - b_4 + b_5/b_6/b_8 * b_8a$
Index7	$b_1 * b_3/b_4 * b_5/b_6/b_8 * b_8a$
Index8	$b_1 * b_2/b_4/b_5 * b_7/b_8 * b_8a$
Index9	$b_1 * b_2/b_4 + b_5/b_6/b_8 * b_8a$
Index10	$b_3 - b_4 + b_5/b_6/b_8 * b_8a$
Index11	$b_1/b_2/b_4 * b_5 * b_7/b_8/b_9 * b_8a$
Index12	$b_1 * b_2/b_4/b_8 * b_8a$
Index13	$b_1 * b_2/b_4 + b_5/b_6 * b_7/b_8 + b_9 - b_8a$
Index14	$b_5/b_6/b_8 * b_8a - b_{11} + b_{12}$
Index15	$b_1 * b_3 - b_4 + b_5/b_6/b_8 * b_8a$

تصویر ماهواره‌ای SENTINEL-2 در آذرماه سال ۱۴۰۰ هم‌زمان با برداشت نمونه‌های میدانی تهیه شد. در این مرحله به محاسبه و تبدیل‌های رادیومتری و انعکاس و تصحیح‌های اتمسفری نبود؛ زیرا تصویر در Google Earth Engine پردازش شده است. تصاویر ماهواره‌ای شاخص‌های انتخاب شده در شکل (۳) ارائه شده است.



شکل ۳. نقشه ۱۵ شاخص ماهواره‌ای مورد استفاده در مطالعه بر اساس جدول ۱



مدل‌سازی

الگوریتم جنگل تصادفی

روش‌های یادگیری ماشینی ابزارهای قدرتمند پیش‌بینی هستند. استفاده از آن‌ها در بخش معدن، جهت تخمین عیار منابع معدنی، بیشتر می‌شود. الگوریتم مورد استفاده در این مطالعه الگوریتم جنگل تصادفی (RF^۱) است. مدل جنگل تصادفی جزئی از روش‌های طبقه‌بندی نظارت شده گروهی در حوزه داده‌کاوی است. الگوریتم RF، به‌عنوان یک الگوریتم محبوب در طبقه‌بندی، دارای پتانسیل بسیار زیادی است. این الگوریتم با تولید چندین درخت تصمیم مختلف به‌عنوان طبقه‌بندی‌های پایه و استفاده از رأی اکثریت برای ترکیب نتایج،

^۱ Random Forest

کارایی بیشتری نسبت به سایر روش‌های گروهی از خود نشان می‌دهد. در این الگوریتم، داده‌های نمونه به تعداد n مجموعه تقسیم می‌شوند و از هر مجموعه، یک درخت برای آموزش استفاده می‌شود. درخت‌ها دارای یک عمق مشخص هستند و در هر گره، یک ویژگی به‌طور تصادفی انتخاب شده و بر اساس آن، تقسیم و شاخه‌بندی انجام می‌شود. با توجه به استفاده از چندین مجموعه نمونه، این الگوریتم مشکل داده‌های پرت را ندارد (سعادت و زمانی بروجنی، ۱۳۹۵).

در این پژوهش برای بررسی عملکرد الگوریتم RF در تخمین عیار پتاسیم شورابه زیرزمینی از نتایج آزمایشگاهی اندازه‌گیری ۶۰ نمونه پتاسیم لایه سطحی و داده‌های حاصل از تصاویر ماهواره‌ای (شاخص‌های ارائه شده در بخش قبلی) استفاده گردید. برای تهیه کردن مدل ابتدا تمام شاخص‌ها (داده‌های آموزشی و تست) در قالب فایل اکسل آماده شده و سپس برای انجام مدل‌سازی از زبان برنامه‌نویسی پایتون در بستر Google Colab استفاده شد. ۷۰ درصد تعداد نمونه‌های برداشت شده به‌منظور آموزش مدل و تعداد ۳۰ درصد باقیمانده برای آزمایش و تست مدل در نظر گرفته شده است (Kisi et al., 2006). برای تقسیم داده‌ها به مجموعه‌های آموزش و تست از تابع `train_test_split` موجود در کتابخانه Scikit-learn استفاده شد. هابیر پارامترهای الگوریتم جنگل تصادفی، تعداد درخت^۱، حداکثر تعداد ویژگی‌ها^۲، حداقل تعداد برگ‌های مورد نیاز برای تقسیم یک گره داخلی^۳، حداکثر گره‌های برگ در هر درخت^۴ و حداکثر عمق مجاز برای هر درخت تصمیم^۵ می‌باشند که لازم است برای آن‌ها مقادیر مناسبی تعیین شوند (Harris et al., 2015). در این مطالعه به‌منظور برآورد درست و مناسب هابیر پارامترهای الگوریتم RF از تابع `GridSearchCV` در کتابخانه Scikit-learn استفاده شده است. کد مورد استفاده برای بهینه‌سازی هابیر پارامترهای الگوریتم جنگل تصادفی در کد (۱) ارائه شده است.

کد ۱. کد پایتون مورد استفاده برای بهینه‌سازی هابیر پارامترهای الگوریتم جنگل تصادفی

```
# Import necessary libraries
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import joblib

# Define the parameter grid for tuning
param_grid = {
    'n_estimators': [50, 100, 150, 200, 250],
    'max_depth': [None, 10, 20, 30],
    'min_samples_split': [2, 5, 10],
    'max_leaf_nodes': [None, 10, 50, 100],
    'max_features': ['auto', 'sqrt', 'log2']
}

# Initialize Random Forest Regressor
rf = RandomForestRegressor(random_state=42)

# Initialize GridSearchCV
grid_search = GridSearchCV(estimator=rf, param_grid=param_grid, cv=5, n_jobs=-1,
                           verbose=2, scoring='r2')

# Fit the GridSearchCV to the data
grid_search.fit(x_train, y_train)

# Get the best model from GridSearchCV
best_model = grid_search.best_estimator_

# Make predictions on the test data
y_pred = best_model.predict(x_test)

# Calculate evaluation metrics
```

^۱ n_estimators

^۲ max_features

^۳ min_sample_split

^۴ max_leaf_nodes

^۵ max_depth

```

best_mse = mean_squared_error(y_test, y_pred)
best_mae = mean_absolute_error(y_test, y_pred)
best_rmse = np.sqrt(best_mse)
best_r2 = r2_score(y_test, y_pred)

# Print the best model's performance metrics
print("Best MSE:", best_mse)
print("Best MAE:", best_mae)
print("Best RMSE:", best_rmse)
print("Best R-squared:", best_r2)
print("Best parameters:", grid_search.best_params_)

# Save the best model to a file
filename = 'best_RF_model_after_tuning.pkl'
joblib.dump(best_model, filename)
print("Best model saved as", filename)

```

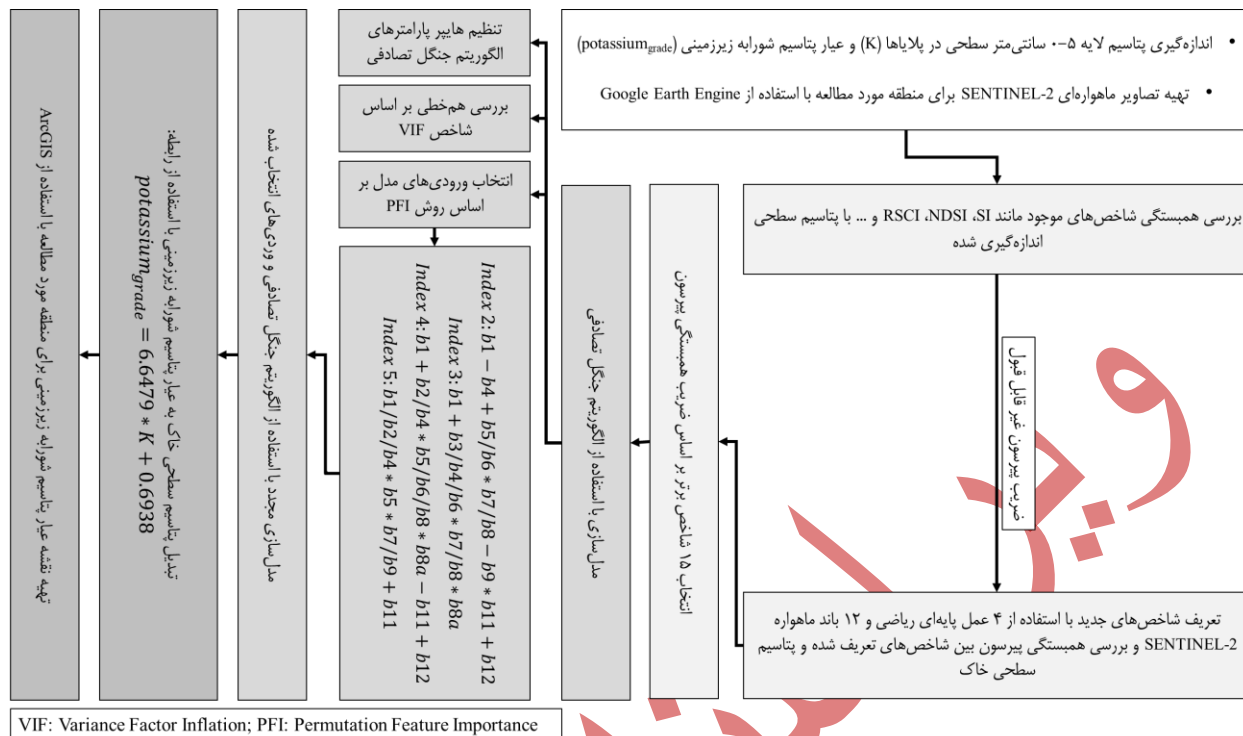
به منظور بررسی چندخطی بین متغیرهای مستقل که در استفاده از مدل‌های رگرسیونی مشکل ساز بوده و باعث برآزش بیش از حد می‌شود؛ از عامل تورم واریانس (VIF¹) استفاده می‌شود. اما در الگوریتم RF که یک روش یادگیری گروهی مبتنی بر درخت‌های تصمیم است و هر درخت با استفاده از زیرمجموعه‌ای تصادفی از ویژگی‌ها ساخته می‌شود و پیش‌بینی نهایی، مجموعه‌ای از پیش‌بینی‌های انجام‌شده توسط درختان است که بر مفروضات مشابه رگرسیون خطی تکیه نمی‌کند. برای جلوگیری از همبستگی درختان مختلف RF تنوع درختان را با رشد آن‌ها از زیرمجموعه داده‌های آموزشی مختلف که از طریق روشی به نام بسته‌بندی ایجاد شده‌اند؛ افزایش می‌دهد. بنابراین، پایداری بیشتری حاصل شده و در مواجهه با تغییرات جزئی در داده‌های ورودی، قوی‌تر شده و دقت پیش‌بینی را افزایش می‌دهد. به این ترتیب RF می‌تواند یک تخمین بی‌طرفانه از خطای تعمیم بدون استفاده از زیرمجموعه داده‌های متنی خارجی محاسبه کند و به همین دلیل مشکل چندخطی در آن وجود ندارد و VIF در یک مدل جنگل تصادفی معنادار نیست (Breiman, 2001). بنابراین می‌توان نتیجه گرفت که هیچ‌گونه هم‌خطی و چندخطی بین متغیرهای مستقل انتخاب‌شده برای مدل‌سازی وجود ندارد.

برای تخمین عیار پتاسیم شورابه زیرزمینی با استفاده از پتاسیم سطحی خاک از یک مدل رگرسیونی خطی طبق رابطه (۱) استفاده شد (ایرجی و همکاران، ۱۴۰۲).

$$potassium_{grade} = 6.6479 * K + 0.6938 \quad (1)$$

در این معادله $potassium_{grade}$ عیار پتاسیم شورابه زیرزمینی و K پتاسیم لایه سطحی می‌باشد. در نهایت مقدار عیار پتاسیم شورابه زیرزمینی پیش‌بینی شده در نرم‌افزار ArcGis فراخوانده شده و بر این اساس نقشه عیار پتاسیم برای کل منطقه مورد مطالعه ترسیم شد. شکل (۴) خلاصه، مراحل و توابع استفاده شده در مدل‌سازی پتاسیم سطحی در مطالعه حاضر را نشان می‌دهد.

¹ Variance Factor Inflation



شکل ۴. خلاصه، مراحل و توابع استفاده شده در مدل‌سازی عیار پتاسیم در مطالعه حاضر

ارزیابی مدل

مدل‌های پیش‌بینی کننده باید برای تعیین عملکرد آن‌ها ارزیابی شوند. یک روش رایج در یادگیری ماشین این است که پایگاه داده را به سه گروه جدا می‌شود یک قسمت از آن برای آموزش، دیگری برای آزمایش و آخرین برای اعتبارسنجی مدل استفاده شده و پس از آن، یک خطای تخمین با اعتبارسنجی متقاطع با توجه به پارامترهای خاص محاسبه می‌شود (Vabalas et al., 2019). از آماره‌های مجذور میانگین مربعات خطا (RMSE)، قدر مطلق خطای نسبی (MAE) خطای میانگین مربعات (MSE) و ضریب تعیین (R^2) بر اساس روابط ۲ تا ۵ استفاده شد (Moriassi et al., 2007).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$MAE = \sqrt{\frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}} \quad (3)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

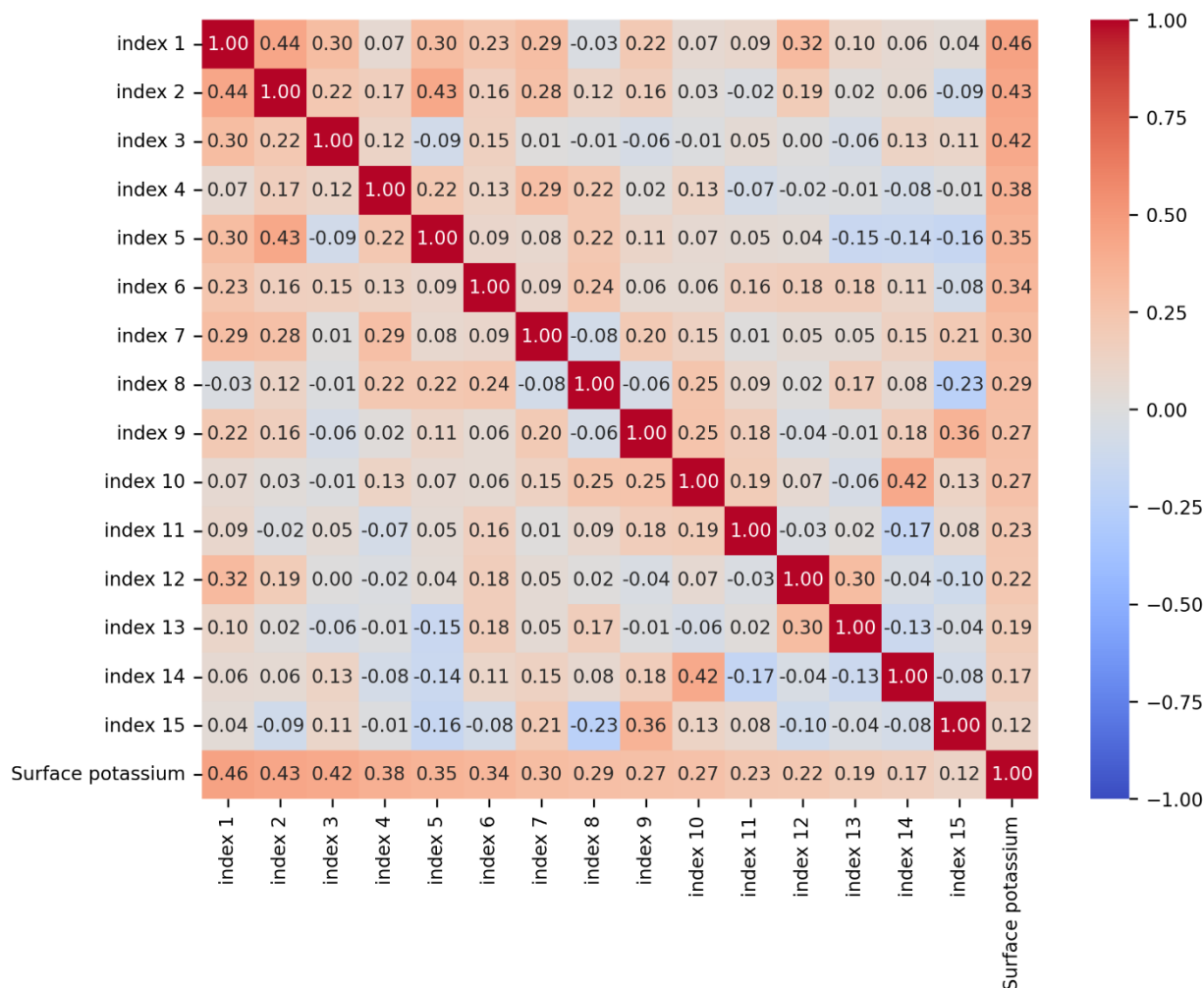
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

که در رابطه‌های بالا y_i و \hat{y}_i به ترتیب مقادیر اندازه‌گیری واقعی و پیش‌بینی شده، n تعداد نمونه‌ها و \bar{y} میانگین مقادیر اندازه‌گیری شده واقعی است؛ در مجموع، مدلی که بیش‌ترین مقدار ضریب تعیین R^2 و کمترین مقادیر خطاها RMSE، MAE و MSE را داشته باشد، از عملکرد بهتری برخوردار است.

نتایج و بحث

ارزیابی کارایی الگوریتم جنگل تصادفی و تهیه نقشه پیش‌بینی پتاسیم لایه سطحی

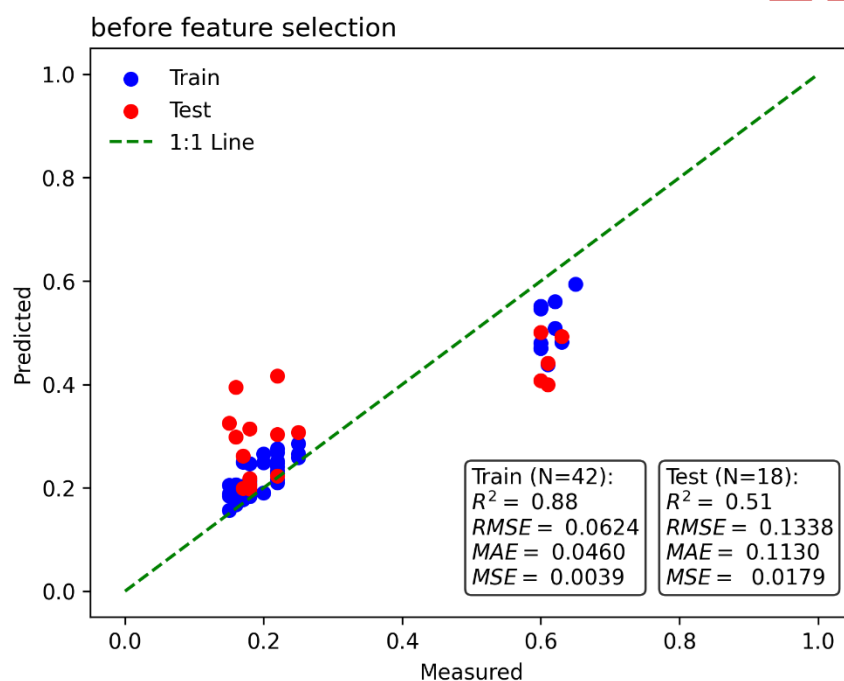
در شکل (۵) ضریب همبستگی پیرسون موجود بین متغیرهای مستقل (Index1-15) و وابسته (Surface potassium) موجود ارائه شده است. با توجه به نتایج ارائه شده در شکل (۵) مشاهده می‌شود که شاخص‌های که بیشترین همبستگی پیرسون (انتخاب شده از بین ۶۱ میلیون شاخص تعریف و بررسی شده) با پتاسیم سطحی خاک را داشته‌اند انتخاب شده است. به طوری که بالاترین ضریب همبستگی مربوط به Index 1 با مقدار ۰/۴۶ می‌باشد.



شکل ۵. همبستگی موجود بین متغیرهای مستقل و وابسته بر اساس ضریب همبستگی پیرسون

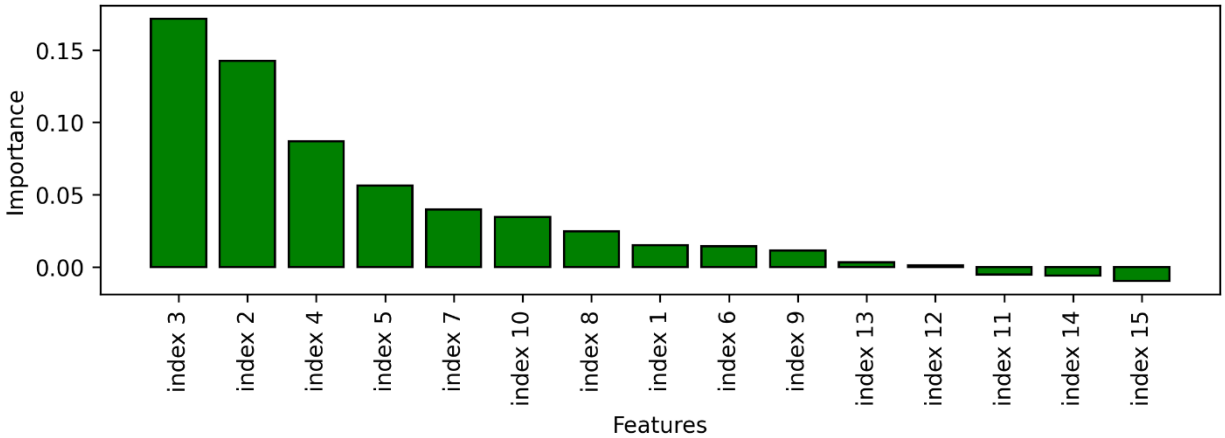
همانطور که در بخش مواد و روش توضیح داده شده، ابتدا با استفاده از کل ۱۵ شاخص انتخابی مدل‌سازی انجام شده و نتایج آن در شکل (۶) ارائه شده است. با توجه به شکل (۶) که مقادیر واقعی اندازه‌گیری شده و مقادیر پیش‌بینی شده پتاسیم سطحی خاک با مدل RF برای داده‌های واسنجی و صحت‌سنجی ارائه شده است؛ مشخص است که مدل برای داده‌های واسنجی با دارا بودن R^2 معادل ۰/۸۸ و RMSE، MSE و MAE به ترتیب معادل ۰/۰۰۳۹، ۰/۰۶۲۴ و ۰/۰۴۶۰ و همچنین شاخص‌های آماری R^2 ، MSE، RMSE و MAE برای داده‌های صحت‌سنجی به ترتیب ۰/۰۵۱، ۰/۰۱۷۹، ۰/۱۳۳۸ و ۰/۱۱۳۰ درصد است که با توجه به مرور منابع مشابه در این

زمینه (مدل‌سازی عناصر معدنی) از دقت خوبی برخوردار است (Boroh et al., 2022؛ Zhang et al., 2019؛ Yang et al., 2016؛ بندک و همکاران، ۱۴۰۲؛ موسوی و همکاران، ۱۴۰۰؛ صادقی و احمدی ندوشن، ۱۴۰۰؛ موسوی و همکاران، ۱۳۹۹). به عنوان مثال Boroh et al. (2022) شاخص‌های آماری R^2 ، MAE و RMSE را برای الگوریتم جنگل تصادفی در مدل‌سازی آهن به ترتیب ۰/۸۸ و ۲/۷۷ و گزارش کردند. همچنین صادقی و ندوشن در مطالعه خود برای مدل‌سازی نیتروژن خاک در شهرستان لنجان استان اصفهان، شاخص R^2 را ۰/۴۴ گزارش کردند. همان‌طور که در شکل (۶) نیز به خوبی دیده می‌شود، نتایج مدل RF به خط نیم‌ساز نسبتاً نزدیک هستند و می‌توان گفت مقادیر را با دقت قابل قبولی پیش‌بینی کرده است. از لحاظ تئوری اگر R^2 برابر ۱ درصد شود تمامی مقادیر مشاهده‌شده با مقادیر برآزش شده یکسان خواهند بود و همه نقاط داده‌ها بر روی خط برآزش شده قرار دارند (Devore, 2015). شکل (۶) نشان می‌دهد که در مقادیر کم پتاسیم سطحی، مدل بیش‌برآورد داشته و برعکس با افزایش پتاسیم سطحی اندازه‌گیری شده، مدل کم‌برآورد داشته است.



شکل ۶. نتایج واسنجی و صحت‌سنجی الگوریتم RF با استفاده از کل ۱۵ شاخص‌های تعریف‌شده در پیش‌بینی پتاسیم سطحی

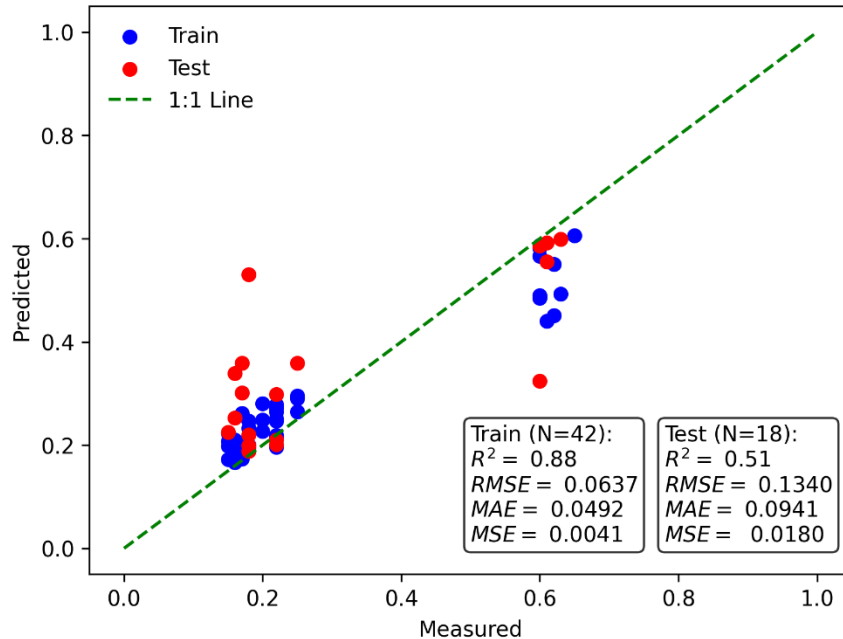
مطابق نتایج به‌دست‌آمده و ارائه‌شده در شکل (۶) برای انتخاب و اولویت‌بندی مهم‌ترین شاخص‌های ماهواره‌ای که بیشترین اثر بر میزان پتاسیم سطحی خاک را داشته‌اند، از روش‌های PFI (Kaneko, 2022) استفاده شد؛ که نتایج آن در شکل (۷) ارائه شده است. سپس برای مدل‌سازی مجدد از ۴ شاخص برتر مشخص‌شده در شکل (۷) که استفاده شده است. با توجه به نتایج ارائه شده در شکل (۷) مشخص است که به ترتیب شاخص‌های Index 3، Index 2، Index 4 و Index 5 با اهمیت نسبی ۰/۱۷۲، ۰/۱۴۳، ۰/۰۸۷ و ۰/۰۵۶ مهم‌ترین شاخص‌های موثر بر پتاسیم سطحی خاک می‌باشند و شاخص‌های Index 15، Index 14، Index 11 و Index 12 دارای کم‌ترین اثر می‌باشند.



شکل ۷. نتایج بررسی اهمیت ویژگی‌ها و انتخاب آن‌ها بر اساس روش PFI برای مدل‌سازی مجدد با استفاده از الگوریتم RF

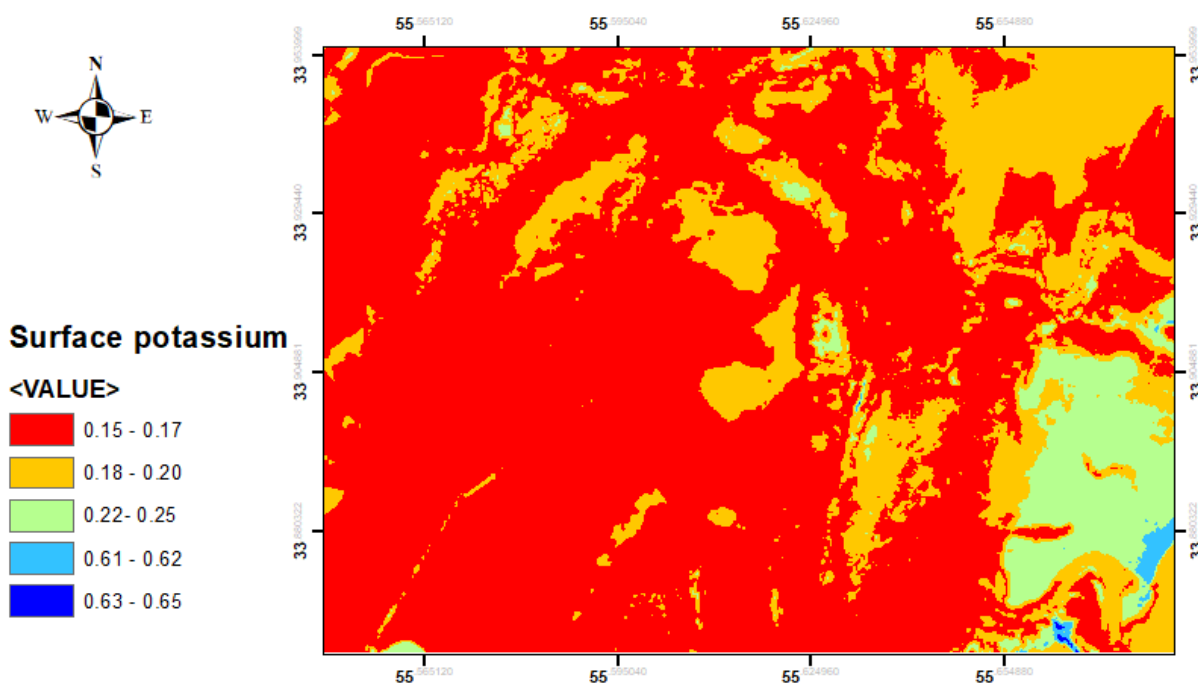
بعد از مدل‌سازی مجدد با استفاده از ۴ شاخص پراهمیت Index 3، Index 2، Index 4 و Index 5 که در شکل (۷) نیز مشخص است، مدل توسعه داده شده، مورد ارزیابی قرار گرفت که نتایج آن در شکل (۸) ارائه شده است. با توجه به نتایج، با کاهش شاخص‌های ورودی مدل و انتخاب ویژگی‌های پراهمیت باز هم می‌توان به مدلی با دقت خوب و قابل قبول دست یافت. به طوری که نتایج نشان داد، مدل دوم نیز در داده‌های واسنجی با دارا بودن R^2 معادل ۰/۸۸ و MSE ، $RMSE$ و MAE به ترتیب معادل ۰/۰۴۱، ۰/۰۶۳۷ درصد، ۰/۰۴۹۲ و همچنین شاخص‌های آماری R^2 ، MSE ، $RMSE$ و MAE برای داده‌های صحت‌سنجی مدل به ترتیب ۰/۵۱، ۰/۰۱۸۰، ۰/۱۳۴۰ درصد و ۰/۰۹۴۱ است که از دقت قابل قبولی برخوردار است.

After feature selection



شکل ۸. نتایج واسنجی و صحت‌سنجی الگوریتم RF با استفاده از ۴ شاخص انتخاب شده بر اساس روش PFI

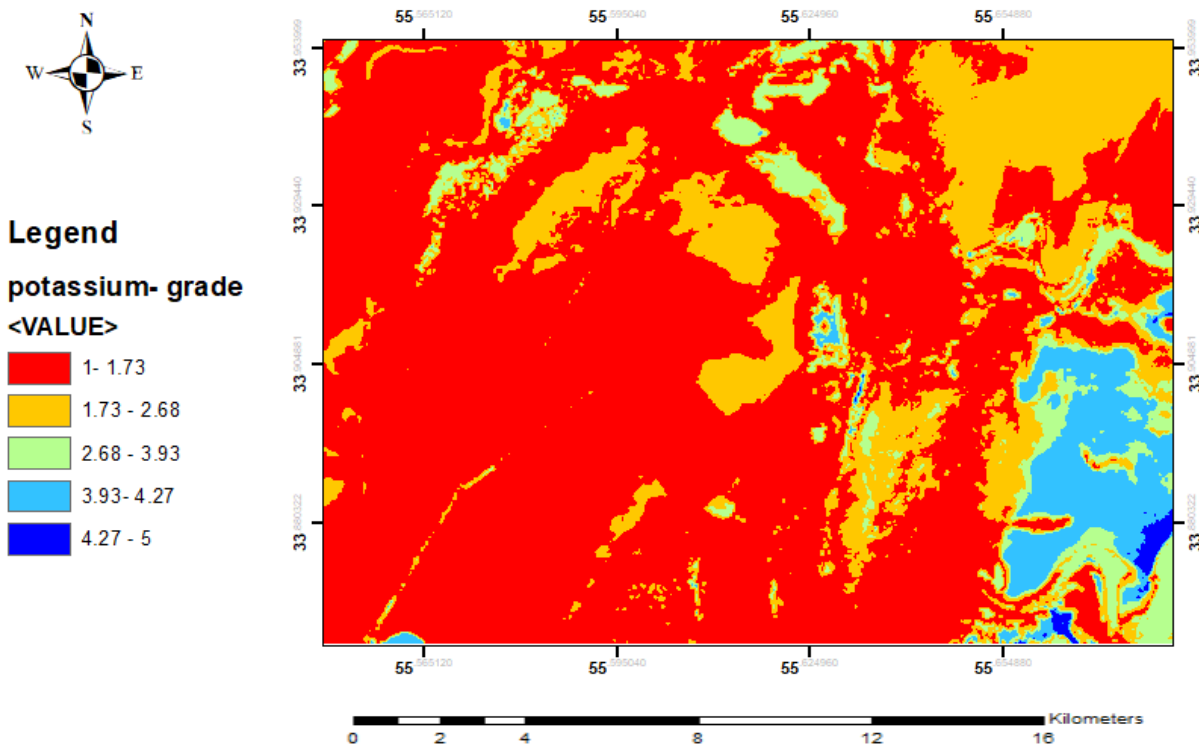
در کل، بررسی نتایج حاصل نشان‌دهنده کارآمدی الگوریتم RF استفاده‌شده برای پیش‌بینی پتاسیم سطحی است. همچنین محققان دیگری نیز به عملکرد قابل قبول الگوریتم‌ها در برآورد منابع معدنی اشاره کرده‌اند. تخمین منبع پلاتین (Chatterjee & Bandopadhyay, 2011)، تلفیق شبکه‌های عصبی مصنوعی و زمین‌آمار در تخمین ذخیره معدنی (Jalloh et al., 2016). مطابق نتایج به‌دست‌آمده الگوریتم RF، دارای این توانایی است که پتاسیم سطحی خاک را به‌خوبی در نقاط دیگر پلایا تخمین بزند. در واقع با استفاده از یک الگوریتم مناسب، امکان اظهارنظر در مورد برنامه‌های آبی استخراج مواد معدنی، به صورتی که امکان برنامه‌ریزی درازمدت وجود داشته باشد، ممکن خواهد بود. در آخر با استفاده از نتایج مدل توسعه داده شده برای ۶۰ نقطه، نقشه پیش‌بینی پتاسیم لایه سطحی برای کل منطقه در شکل (۹) تهیه و ارائه شده است. نتایج نشان می‌دهد که قسمت شرق و جنوب شرقی نسبت به دیگر مناطق دارای پتاسیم بیشتری می‌باشند.



شکل ۹. نقشه پیش‌بینی پتاسیم لایه سطحی (درصد) تهیه‌شده با الگوریتم RF برای کل منطقه مورد مطالعه

تهیه نقشه پیش‌بینی عیار پتاسیم شورابه زیرزمینی

بعد از تهیه نقشه پیش‌بینی پتاسیم لایه سطحی برای کل منطقه، با استفاده از رابطه (۱) حاصل از مطالعه ایرجی و همکاران (۱۴۰۲) برای منطقه مورد مطالعه، پتاسیم سطحی برآورد شده به عیار پتاسیم شورابه زیرزمینی تبدیل شده و نقشه آن برای کل منطقه مورد مطالعه در شکل (۱۰) ارائه شده است. همان‌گونه که در شکل (۱۰) مشاهده می‌شود در قسمت‌های شرق و جنوب شرقی عیار پتاسیم شورابه زیرزمینی بالاتر است و ارتباط معناداری بین این شاخص‌ها و پتاسیم خاک سطحی مشاهده می‌شود. به طوری که بیش‌ترین مقادیر عیار پتاسیم شورابه زیرزمینی در نقشه با رنگ آبی نشان داده شده است. این موضوع را می‌توان به این صورت بیان کرد که در این مناطق نیز مقدار پتاسیم سطحی نسبت به سایر مناطق بالاتر بوده و با توجه به روابط رگرسیونی بین پتاسیم لایه سطحی و عیار پتاسیم شورابه زیرزمینی، عیار پتاسیم شورابه زیرزمینی هم در این مناطق بالاتر است. قسمت‌های مرکز و جنوب غربی منطقه مطالعاتی کم‌ترین میزان عیار پتاسیم شورابه زیرزمینی را دارا بوده و این مناطق با رنگ قرمز و نارنجی مشاهده می‌شوند.



شکل ۱۰. نقشه پیش‌بینی عیار پتاسیم شورابه زیرزمینی (گرم بر لیتر) تهیه‌شده با الگوریتم RF برای کل منطقه مورد مطالعه

در پلایا ممکن است مواردی همچون درجه حرارت و رطوبت، جریان باد، بارندگی و غرقاب شدن بخشی از پلایا (وجود دریاچه‌های غیردائمی قسمت‌هایی از پلایا در برخی سال‌ها و فصل‌هایی از سال) بر میزان املاح محلول سطح اثرگذار باشد که بر میزان پتاسیم اثر می‌گذارند (Stout, 2022). این موارد از دلایل تغییر میزان پتاسیم در مناطق مختلف پلایا می‌باشد. ولی بر مبنای این تحقیق، بین پتاسیم لایه سطحی و عیار پتاسیم در شورابه زیرزمینی، همبستگی بالایی وجود دارد، به نحوی که می‌توان از پتاسیم لایه سطحی در تخمین عیار آن در شورابه زیرزمینی استفاده کرد. علت این همبستگی این است که پتاسیم لایه سطحی، با صعود مویبندی از شورابه زیرزمینی موجود در آب زیرزمینی نشات می‌گیرد و آب‌های زیرزمینی حداقل در بخشی از سال از محدوده ارتفاع مویبندی بحرانی^۱ برای صعود پتاسیم، بالاتر قرار می‌گیرند (Sass & Viles, 2022). به‌طور کلی می‌توان بیان کرد که الگوریتم جنگل تصادفی در ترکیب با داده‌های سنجش از دور با عملکرد خوب مناطق دارای عیار پتاسیم شورابه زیرزمینی بالاتر را پیش‌بینی کرده است. همچنین عمق آب زیرزمینی در منطقه مورد مطالعه در عمق ۳ تا ۳/۵ متر قرار دارد که برای بافت‌های متوسط و ریز خاک، بالاتر از ارتفاع مویبندی بحرانی بوده و جریان مویبندی از سطح ایستابی به سطح خاک، بیشتر از ۱ میلی‌متر بر روز بوده که موجب انتقال پتاسیم و سایر املاح موجود در شورابه زیرزمینی به سطح خاک می‌شود. با توجه به اینکه پتانسیل تبخیر روزانه در کویر در غالب ایام سال از یک میلی‌متر بر روز بیشتر است، این جریان املاح از آب زیرزمینی به سطح خاک غالباً یک جریان پیوسته^۲ محسوب می‌شود که در شرایط غیر اشباع در منافذ ریز انجام می‌شود و با تبخیر سطحی، غلظت املاح در سطح خاک از شورابه زیرزمینی بیشتر است (Wang, 2021). بارندگی‌های نادر شدید، ممکن است باعث حرکت نزولی املاح از منافذ درشت در خاک سطحی شوند که با قطع بارندگی و ادامه صعود مویبندی، تعادل قبلی برقرار می‌شود. این حرکت نزولی املاح، جزئی است؛ زیرا در مناطق مسطح کویر، آب باران معمولاً منافذ خاک سطحی را که در اثر تبخیر تخلیه شده‌اند، پر می‌کند و به آب زیرزمینی نمی‌رسد. آب زیرزمینی در کویر، غالباً از تشکیل آب زیرزمینی

^۱ Critical water depth for upward capillary flow

^۲ Step Through

در کوهستان‌های اطراف آن و حرکت زیرزمینی به سمت کویر تامین می‌شود که در این صورت تبخیر و تعرق حقیقی از بارندگی بیشتر است و حتی می‌توان گفت که جریان تدریجی نمک‌ها به سمت سطح خاک با مرور زمان طولانی، تجمعی است و بتدریج نمک‌های جدید موجود در آب زیرزمینی و حاصل از هوازدگی کانی‌ها در سطح خاک به صورت پوسته‌های نمک تجمع می‌کنند (Goldberg- et al., 2022). با جریان پیوسته املاح از آب زیرزمینی به سطح خاک و تعادل نسبی سالانه املاح بین آب زیرزمینی و سطح خاک انتظار می‌رود که اثر تبادل ترجیهِی^۱ در انتقال املاح مختلف پر رنگ نباشد و املاح مختلف با سرعت تقریباً یکسانی در فاصله آب زیرزمینی تا سطح خاک حرکت کنند و در نتیجه بین غلظت املاح مختلف در سطح خاک و غلظت آن‌ها در شورابه زیرزمینی همبستگی بالایی باشد (Bolt & Bruggenwert, 1978). چون املاح مختلف موجود در آب زیرزمینی ایزوترم‌های تبدالی و جذب متفاوتی دارند؛ سرعت سالانه حرکت آنها در شرایط غیر تعادلی (مانند شروع آبیاری در تبدیل زمین‌های دیم به آبی) متفاوت است؛ ولی در شرایط تعادلی وقتی غلظت املاح مختلف (از جمله پتاسیم) در محلول خاک با کمپلکس تبدالی به تعادل می‌رسند، با سرعت‌های مختلف جریان محلول خاک و املاح، غلظت آنها تغییر نمی‌کند و همه املاح با سرعت سالانه تقریباً یکسانی حرکت می‌کنند. در این تحقیق، همبستگی بین پتاسیم در آب زیرزمینی و سطح خاک بالا بود که نتیجه همین تعادل نسبی سالانه و عدم جریان ترجیهِی املاح است که در نتیجه می‌توان عیار پتاسیم شورابه زیرزمینی را از غلظت پتاسیم سطحی برآورد کرد و نقشه عیار پتاسیم شورابه زیرزمینی در پلایا را تهیه نمود.

در تعداد معدودی از مطالعات به بررسی کارایی الگوریتم‌های یادگیری ماشین در مدل‌سازی و برآورد یک عنصر با استفاده از شاخص‌های سنجش‌ازدوری پرداخته شده است. ارزیابی کارایی قابلیت‌های چهار الگوریتم رگرسیون یادگیری ماشین شبکه عصبی مصنوعی، درخت تصمیم، جنگل تصادفی و ماشین بردار پشتیبان برای مدل‌سازی پیش‌بینی پتانسیل طلا در اسپانیا (Rodriguez- et al., 2016)، استفاده از رگرسیون ماشین بردار پشتیبان و رگرسیون فرآیند گاوسی برای ارزیابی عیار معدنی یکی از معادن سنگ آهن در منطقه شرقی کشور هند (Goswami et al., 2016)، کاربرد شبکه‌های عصبی مصنوعی در تخمین ذخایر سنگ معدن در کانسار سنگ آهن چگارت یزد (Nezamolhosseini et al., 2017)، کاربرد الگوریتم ماشین بردار پشتیبان برای جداسازی پهنه‌های معدنی در کانسار پورفیری تخت گنبد در جنوب شرق ایران (Mohammadi & Hezarkhani, 2018)، ارزیابی عملکرد الگوریتم‌های یادگیری ماشین مبتنی بر کرنل RBF الگوریتم SVM برای مدل‌سازی آینده‌نگری معدنی ذخایر مس پورفیری در بخش ورزقان در شمال غرب ایران (Ghezelbash et al., 2019)، نقشه‌برداری مواد معدنی از طریق جنگل تصادفی نیمه نظارت شده (Wang et al., 2020). به‌طور کلی بر اساس یافته‌ها و نتایج این پژوهش و نتایج تحقیقات دیگر درمی‌یابیم که استفاده از داده‌های تصاویر ماهواره‌ای و الگوریتم‌های یادگیری ماشین برای تخمین عیار پتاسیم شورابه زیرزمینی روشی مناسب و مقرون‌به‌صرفه است.

نتیجه‌گیری و پیشنهادات

در این پژوهش از تکنیک ترکیبی سنجش‌ازدور و الگوریتم یادگیری ماشین جهت تخمین عیار شورابه زیرزمینی استفاده شد، به طوری که با پارامتر قابل‌اندازه‌گیری در سطح خاک (پتاسیم لایه سطحی) و ارتباط آن با شاخص‌های تصاویر ماهواره‌ای، غلظت پتاسیم سطحی برآورد شد و با یک رابطه رگرسیونی، پتاسیم لایه سطحی به عیار پتاسیم شورابه زیرزمینی تبدیل شده است. همچنین نتایج نشان داد که از بین شاخص‌های مستخرج از تصویر، شاخص‌های Index 3، Index 2، Index 4 و Index 5 به‌عنوان مهم‌ترین شاخص‌ها شناخته شده‌اند و می‌توانند به تخمینی خوب و قابل قبولی از پراکندگی مقدار پتاسیم لایه سطحی نائل شود. مدل توسعه داده شده جنگل تصادفی با دارا بودن R^2 برابر با ۰/۸۸ و ۰/۵۱ به ترتیب برای داده‌های واسنجی و صحت‌سنجی با دقت مناسبی توانسته است پتاسیم

^۱ Favorable Exchange

سطحی خاک را برآورد و پیش‌بینی کند. الگوریتم جنگل تصادفی با تلفیق داده‌های سنجش از دور با اولویت‌بندی شاخص‌های مؤثر و یافتن ارتباط معنی‌دار بین متغیرها و مشخص کردن پارامترهای پراهمیت به‌عنوان ابزاری کارآمد در جهت نقشه‌برداری وسیع از مناطق بزرگ برای مواردی که پیش‌بینی یک متغیر مهم به روش سنتی به دلیل تغییرپذیری مکانی و زمانی پیش‌بینی آن‌ها سخت و پرهزینه می‌باشد؛ بسیار کارآمد باشد. مدل توسعه داده شده با مدت‌زمان کم و صرف هزینه بسیار پایین، تهیه نقشه پتاسیم سطحی و عیار آن در شورابه زیرزمینی را بسیار آسان می‌سازد. با توجه به اینکه پلایاهای زیادی در کشور وجود دارند که دارای منابع پتاسیم هستند؛ تعیین پارامترهای پراهمیت با روش یادگیری ماشین و فناوری سنجش‌از‌دور ابزار مفیدی در تصمیم‌گیری مدیران در جهت سرمایه‌گذاری برای حفاری در مناطق امیدبخش جهت استحصال پتاسیم شورابه زیرزمینی، نقش مؤثری را ایفا می‌کند. از آنجایی که شرایط پلایا در کویر خیلی متفاوت نیست، احتمال می‌رود که نتایج این تحقیق برای تعیین عیار پتاسیم شورابه زیرزمینی به سایر پلایا در کویر قابل تعمیم باشد که در این صورت تخمین عیار پتاسیم شورابه زیرزمینی در کل پلایا با استفاده از تصاویر ماهواره‌ای مقدور می‌شود. به‌منظور بررسی این احتمال پیشنهاد می‌شود در تحقیقات آتی در تعداد دیگری از پلایاها، عیار پتاسیم شورابه زیرزمینی و پتاسیم خاک سطحی در چند نمونه اندازه‌گیری شوند و با مقادیر آن‌ها که با پارامترهای پیشنهادی در این تحقیق از تصاویر ماهواره‌ای استخراج شدند؛ مقایسه شود. احتمالاً بتوان عیار سایر عناصر موجود در شورابه زیرزمینی مثل منیزیم را نیز با تهیه نقشه‌های سطحی بدست آورد؛ زیرا احتمالاً همین تعادل نسبی بین غلظت پتاسیم سطح خاک و عیار پتاسیم شورابه زیرزمینی بین سایر عناصر محلول در شورابه زیرزمینی و سطح خاک هم وجود دارد و در خاک با سرعت مشابهی با پتاسیم جایجا می‌شوند.

منابع

ایرجی، مریم، موحدی نائینی، سید علیرضا، کمکی، چوقی بایرام، ابراهیمی، سهیلا، و یغمایی، بامشاد. (۱۴۰۲). ارزیابی پارامترهای مؤثر جهت پیش‌بینی عیار پتاسیم شورابه با استفاده از الگوریتم‌های ماشین بردار پشتیبان و جنگل تصادفی (مطالعه موردی: پلایای شهرستان خور و بیابانک، استان اصفهان). تحقیقات آب و خاک ایران، پذیرفته شده انتشار آنلاین از ۵ دی ۱۴۰۲.

بندک، ثریا، موحدی نائینی، سید علیرضا، کمکی، چوقی بایرام، کاکویی، محمد. و وولست، جوشم. (۱۴۰۲). پیش‌بینی و نقشه‌برداری کربن آلی خاک با استفاده از سنجش از دور و الگوریتم‌های یادگیری ماشین، علوم آب و خاک، ۳۷(۳): ۱۷-۳۴.

سعادت، محمد. و زمانی بروجنی، فرساد. (۱۳۹۵). مروری بر روش‌های بهبود کارایی تکنیک جنگل تصادفی، اولین همایش ملی فناوری اطلاعات، ارتباطات و محاسبات، اصفهان، ایران

صادقی، محبوبه. و احمدی ندوشن، مژگان. (۱۴۰۰). مدل‌سازی نیتروژن خاک با استفاده از سنجش از دور و مدل‌های رگرسیون و جنگل تصادفی، حفاظت منابع آب و خاک، ۱۱(۲): ۹۷-۱۱۱.

موسوی، سید روح اله، سرمیدیان، فریدون، امید، محمود. و بوگارت، پاتریک. (۱۴۰۰). کاربرد مدل‌های یادگیری ماشین در برآورد مکانی فسفر و پتاسیم خاک در بخشی از اراضی دشت آبیک، پژوهش‌های خاک، ۳۵(۴): ۳۹۸-۴۱۲.

موسوی، سیدروح‌اله، پارسائی، فرزانه، رحمانی، اصغر، سدري، محمد حسین. و کوهسار بستانی، محمد. (۱۳۹۹). پیش‌بینی مکانی برخی از ویژگی‌های خاک سطحی با استفاده از مدل‌های درون‌یابی و یادگیری ماشین، مدیریت خاک و تولید پایدار، ۱۰(۳): ۲۷-۴۹.

موسوی، سیدعلی، نقوی، رعنا. و شجاعی، سعید. (۱۳۹۵). استفاده صنعتی از شورابه و پلایا جهت استخراج پتاسیم و پی‌جویی آن در بین چندین پلایا در ایران. ایده‌های نو در علوم و فناوری، ۱(۲).

REFERENCES

- Al Rawashdeh, R., Xavier-Oliveira, E., & Maxwell, P. (2016). The potash market and its future prospects. *Resources Policy*, 47, 154-163.
- Alavipanah, S. K., Matinfar, H. R., Sarmasti, N., Jafarbeglou, M., & Goodarzimehr, S. (2011). Evaluation of ASTER and LISS III data in identification of saline soils, case study: regions of Iran. *Geocomputation, London, UK*, 20-22.
- Allbed, A., & Kumar, L. (2013). Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: a review. *Advances in remote sensing*, 2013.

- Bandak S, Movhedei naeani A, Komaki C, kakooei M, Verrlest J. (2023). Predicting and Mapping Soil Organic Carbon Using Remote Sensing and Machine Learning Algorithms. *iwss*; 27(3):17-34. (inPersian)
- Bolt, G. H., & Bruggenwert, M. G. M. (1978). *Soil chemistry. A. Basic elements*. Elsevier Scientific Publishing Company.
- Boroh, A. W., Lawou, S. K., Mfenjou, M. L., & Ngounouno, I. (2022). Comparison of geostatistical and machine learning models for predicting geochemical concentration of iron: case of the Nkout iron deposit (south Cameroon). *Journal of African Earth Sciences*, 195, 104662.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Chatterjee, S., & Bandopadhyay, S. (2011). Goodnews Bay Platinum resource estimation using least squares support vector regression with selection of input space dimension and hyperparameters. *Natural Resources Research*, 20, 117-129.
- Devore, J. L. (2015). *Probability and Statistics for Engineering and the Sciences*. Cengage Learning.
- Dutta, S., Bandopadhyay, S., Ganguli, R., & Misra, D. (2010). Machine learning algorithms and their application to ore reserve estimation of sparse and imprecise data. *Journal of Intelligent Learning Systems and Applications*, 2(02), 86-96.
- Estefan, G., Sommer, R., & Ryan, J. (2013). Methods of soil, plant, and water analysis. *A manual for the West Asia and North Africa region*, 3, 65-119.
- Geranian, H., Tabatabaei, S. H., Asadi, H. H., & Carranza, E. J. M. (2016). Application of discriminant analysis and support vector machine in mapping gold potential areas for further drilling in the Sari-Gunay gold deposit, NW Iran. *Natural Resources Research*, 25, 145-159.
- Ghezelbash, R., Maghsoudi, A., & Carranza, E. J. M. (2019). Performance evaluation of RBF-and SVM-based machine learning algorithms for predictive mineral prospectivity modeling: integration of SA multifractal model and mineralization controls. *Earth Science Informatics*, 12, 277-293.
- Goldberg-Yehuda, N., Assouline, S., Mau, Y., & Nachshon, U. (2022). Compaction effects on evaporation and salt precipitation in drying porous media. *Hydrology and Earth System Sciences*, 26(9), 2499-2517.
- Goswami, A. D., Mishra, M. K., & Patra, D. (2016, October). Adapting pattern recognition approach for uncertainty assessment in the geologic resource estimation for Indian iron ore mines. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPE5)* (pp. 1816-1821). IEEE.
- Gulibert, j. park, ch. (1997). The geology of or deposite , *New york : fereeman*, p. 70-729
- Harris, J. R., & Grunsky, E. C. (2015). Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. *Computers & geosciences*, 80, 9-25.
- Hewson, R., Robson, D., Carlton, A., & Gilmore, P. (2017). Geological application of ASTER remote sensing within sparsely outcropping terrain, Central New South Wales, Australia. *Cogent Geoscience*, 3(1), 1319259.
- Iraji, M., Movahedi naeini, S. A., Komaki, C. B., Ebrahimi, S., & Yaghmaei, B. (2023). Evaluation of effective parameters for predicting the potassium grade of saline water by using support vector machine and random forest algorithms (case study: playa of Koor and Biabank area city, Isfahan province). *Iranian Journal of Soil and Water Research*, 1. (inPersian)
- Jafrasteh, B., Fathianpour, N., & Suárez, A. (2018). Comparison of machine learning methods for copper ore grade estimation. *Computational Geosciences*, 22, 1371-1388.
- Jalloh, A. B., Kyuro, S., Jalloh, Y., & Barrie, A. K. (2016). Integrating artificial neural networks and geostatistics for optimum 3D geological block modeling in mineral reserve estimation: A case study. *International Journal of Mining Science and Technology*, 26(4), 581-585.
- Kaneko, H. (2022). Cross-validated permutation feature importance considering correlation between features. *Analytical Science Advances*, 3(9-10), 278-287.
- Kisi, O., Karahan, M. E., & Şen, Z. (2006). River suspended sediment modelling using a fuzzy logic approach. *Hydrological Processes: An International Journal*, 20(20), 4351-4362.
- Leite, E. P., & de Souza Filho, C. R. (2009). Artificial neural networks applied to mineral potential mapping for copper- gold mineralizations in the Carajás Mineral Province, Brazil. *Geophysical Prospecting*, 57(6), 1049-1065.
- Lewkowski, C., Porwal, A., & González-Álvarez, I. (2010, May). Genetic programming applied to base-metal prospectivity mapping in the Aravalli Province, India. In *EGU general assembly conference abstracts* (p. 523).
- Maleki, S., Ramazia, H. R., & Moradi, S. (2014). Estimation of Iron concentration by using a support vector machine and an artificial neural network-the case study of the Choghart deposit southeast of Yazd, Yazd, Iran. *Geopersia*, 4(2), 201-212.

- McKay, G., & Harris, J. R. (2016). Comparison of the data-driven random forests model and a knowledge-driven method for mineral prospectivity mapping: A case study for gold deposits around the Huritz Group and Nueltin Suite, Nunavut, Canada. *Natural Resources Research*, 25(2), 125-143.
- Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & geosciences*, 32(9), 1378-1388.
- Moghaddam, D. D., Rahmati, O., Panahi, M., Tiefenbacher, J., Darabi, H., Haghizadeh, A., ... & Bui, D. T. (2020). The effect of sample size on different machine learning models for groundwater potential mapping in mountain bedrock aquifers. *Catena*, 187, 104421.
- Mohammadi, N. M., & Hezarkhani, A. (2018). Application of support vector machine for the separation of mineralised zones in the Takht-e-Gonbad porphyry deposit, SE Iran. *Journal of African Earth Sciences*, 143, 301-308.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885-900.
- Mousavi, S. R., Parsayi, F., Rahmani, A., Sedri, M. H., & Kohsar Bostani, M. (2020). Spatial Prediction Some of the Surface Soil Properties Using Interpolation and Machine Learning Models. *Journal of Soil Management and Sustainable Production*, 10(3), 27-49. (inPersian)
- Mousavi, S. R., Sarmadian, F., Omid, M., & Bogaert, P. (2022). Application of Machine Learning Models in Spatial Estimation of Soil Phosphorus and Potassium in Some Parts of Abyek Plain. *Iranian Journal of Soil Research*, 35(4), 397-411. (inPersian)
- Mousavi, Seyyed Ali, Naqvi, Rana, and Shojaei, Saeed (2015). Industrial use of shourabeh and playa for extracting potassium and searching for it among several playas in Iran. *New Ideas in Science and Technology*, 1(2) (inPersian)
- Nezamolhosseini, S. A., Mojtahedzadeh, S. H., & Gholamnejad, J. (2017). The application of artificial neural networks to ore reserve estimation at choghart iron ore deposit.
- Nwaila, G. T., Zhang, S. E., Frimmel, H. E., Manzi, M. S., Dohm, C., Durrheim, R. J., ... & Tolmay, L. (2020). Local and target exploration of conglomerate-hosted gold deposits using machine learning algorithms: a case study of the Witwatersrand gold ores, South Africa. *Natural Resources Research*, 29, 135-159.
- Oh, H. J., & Lee, S. (2010). Application of artificial neural network for gold-silver deposits potential mapping: A case study of Korea. *Natural Resources Research*, 19, 103-124.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Porwal, A., Carranza, E. J. M., & Hale, M. (2003). Artificial neural networks for mineral-potential mapping: a case study from Aravalli Province, Western India. *Natural resources research*, 12, 155-171.
- Rigol-Sanchez, J. P., Chica-Olmo, M., & Abarca-Hernandez, F. (2003). Artificial neural networks as a tool for mineral potential mapping with GIS. *International Journal of Remote Sensing*, 24(5), 1151-1156.
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71, 804-818.
- Saadat, Mohammad., and Zamani Borojni, Farsad. (2015). A Review of Methods to Improve the Efficiency of Random Forest Technique, First National Conference on Information, Communication and Computing Technology. (inPersian)
- Sadeghi, M. & Ahmadi Nadoushan, M. (2022). Modeling soil nitrogen using remote sensing, regression and random forest models. *water and soil resources conservation*, 11(2): 97-111. (inPersian)
- Sass, O., & Viles, H. (2022). Heritage hydrology: a conceptual framework for understanding water fluxes and storage in built and rock-hewn heritage. *Heritage Science*, 10(1), 66.
- Schnitzler, N., Ross, P. S., & Gloaguen, E. (2019). Using machine learning to estimate a key missing geochemical variable in mining exploration: Application of the Random Forest algorithm to multi-sensor core logging data. *Journal of Geochemical Exploration*, 205, 106344.
- Shaw, P. A., & Bryant, R. G. (2011). Pans, playas and salt lakes. *Arid zone geomorphology: process, form and change in drylands*, 373-401.
- Stout, J. E. (2022). Playa dynamics and salinity: a study of yellow lake on the high plains of Texas. *The Texas Journal of Science*, 74(1), Article-6.
- Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PloS one*, 14(11), e0224365.

- Wang, C., Chen, J., Wu, J., Tang, Y., Shi, P., Black, T. A., & Zhu, K. (2017). A snow-free vegetation index for improved monitoring of vegetation spring green-up date in deciduous ecosystems. *Remote sensing of environment*, 196, 1-12
- Wang, J., Zuo, R., & Xiong, Y. (2020). Mapping mineral prospectivity via semi-supervised random forest. *Natural Resources Research*, 29, 189-202.
- Wang, Z. (2021). Research on desert water management and desert control. *European Journal of Remote Sensing*, 54(sup2), 42-54.
- Yang, R. M., Zhang, G. L., Liu, F., Lu, Y. Y., Yang, F., Yang, F., ... & Li, D. C. (2016). Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological indicators*, 60, 870-878.
- Zhang, Y., Sui, B., Shen, H., & Ouyang, L. (2019). Mapping stocks of soil total nitrogen using remote sensing data: A comparison of random forest models with different predictors. *Computers and Electronics in Agriculture*, 160, 23-30.

Estimating the potassium grade of saline underground water using Sentinel satellite images and random forest algorithm (case study of Koor and Biabank playa, Isfahan province)

EXTENDED ABSTRACT

Introduction:

In recent decades, with the increase in population growth and the growing need to produce more food, today more than 90% of potassium production is used as fertilizer. One of the main sources of potassium fertilizers is saline water underground. One of the main mineral elements in saline water underground is potassium, which is found in the playa. Due to the environmental conditions of the playa, there is a lot of evaporation and it leads to the precipitation of soluble solutes on the surface. By examining these sediments, it is possible to determine areas with grade high potassium levels for extracting salt water underground in the playa but the complex climatic conditions that govern it make field measurements to estimate the grade of difficult. One of the new methods to estimate mineral resources is the combined use of machine learning algorithm and remote sensing.

Objective:

The main purpose of this research is to use remote sensing and random forest algorithm to estimate the surface potassium of playa soil and to evaluate the relationship between potassium, and index of satellite images to estimate the grade potassium saline water underground, which is the innovation of this research compared to other previous researches.

Materials and method:

In this research, the use of remote sensing and random forest algorithm was used to estimate the surface potassium of playa soil and to evaluate the relationship between potassium, and the index of satellite images to estimate the potassium grade of underground saline water. For this purpose, 60 samples of surface layer potassium (dependent variable) were sampled from the 0 - 5 cm layer using Latin hypercube method. Also, in 12 drilling boreholes, the potassium grade of saline water was measured in December 1400. because there was no related satellite index that has a high correlation with soil surface potassium. By using 4 basic arithmetic operations (addition - subtraction - multiplication and division) between SENTINEL 2 satellite image bands and by writing a new code (specific to the study) 61 million times, the code was executed with different combinations to produce new index. A regression model was used to estimate potassium grade of underground saline water, which was converted to the potassium grade of underground saline with a potassium equation of the surface layer. The Sentinel 2 satellite image and the resulting indicators from this satellite (independent variables) were used to predict the potassium of the surface layer and finally estimate the potassium grade of underground saline water. Also, Permutation Feature Importance (PFI) method was used in the RF model to prioritize and select parameters for modeling. The data were divided into two categories: 70% for calibration (training) and 30% for validation (testing) and were implemented in the random forest model in the Python programming environment.

Results and discussion:

results of the actual measured values and the predicted values of surface potassium with the RF model is based on the statistical indicators of the evaluation of the ML models including R^2 , MSE, RMSE and MAE. The results of the model showed that the calibration data with R^2 equal to 0.88 and MSE, RMSE and MAE equal to 0.0039, 0.0624 and 0.0460, respectively, as well as statistical indicators of R^2 , MSE, RMSE and MAE for the validation data of the model. It is 0.51, 0.0179, 0.1338 and 0.1130 respectively. The results show that Index 3, Index 2, Index 4, Index 5 have the greatest effect on the estimation of soil surface potassium and potassium grade of saline water and Index 15, Index 14, Index 11 and Index 12 have the least effect.

Conclusion:

Random forest algorithm by combining remote sensing technology with prioritizing effective indicators and finding meaningful relationships between variables and specifying important parameters as an efficient tool for extensive mapping of large areas for cases where predicting an important variable in the traditional way due to Spatial diversity. And when it is difficult and expensive to predict them, it will be very efficient and it will make it very easy to determine the parameters and prepare the map with a short period of time and spending much less money. Considering that there are many playas in the country that have potassium resources, determining the most important parameters with machine learning technology and remote sensing is a useful tool in managers' decision making in order to invest in drilling in promising areas for saline water underground extraction. It has an effective role. Since the conditions of the playa in the are not very different, it is possible that the results of this research can be generalized to other playas in the desert to determine the potassium grade of saline water in the desert, in which case it is possible to estimate the potassium grade of saline water in the will be playa using satellite images.

Keywords: *Python, Remote sensing, Salt pans, Modeling, Machine learning.*

فیلد استادی