



Dimension reduction of the remote sensing data to estimate soil organic carbon

N. Mozafarinahr, H. Sadat Hasani, M. Jafari *

Department of Geodesy and Surveying Engineering, Tafresh University, Tafresh, Iran

ABSTRACT: Soil is a very complex phenomenon that includes organic materials, minerals, water, and air. The distribution of organic matter in the soil has a profound effect on biological activity, nutrient availability, soil and soil seed structure, and water holding capacity, and soil management in general. In this research, the relation between soil spectral reflectance using the Landsat 8 satellite data as well as the SRTM Elevation data and soil organic carbon has been investigated. In the proposed method, spectral reflection of data in the main bands of the Landsat 8 satellite is investigated and processed. In addition to the main bands, vegetation and lighting indices, and topographic features have been studied. In this study, a method for selecting effective indexes in increasing the accuracy of soil organic carbon modeling is presented. For this purpose, in the first step of modeling, Linear regression, Support Vector Machine regression, and Neural Network methods have been used for the connection between remote sensing data and soil organic carbon. To implement the proposed method, 100 soil samples in East Azerbaijan province have been used. According to RMSE and R2 statistical indices, which are the basis for evaluating the models, the neural network model was selected as the final model, and with the values of $RMSE = 0.404$, $R^2 = 0.254$, and $RRMSE = 46.597$ is more accurate than the regression method. Due to the importance of dimensionality to increase accuracy and reduce the complexity of calculations, a genetic algorithm was proposed in this study. This efficient algorithm increases the accuracy of soil organic carbon modeling and eliminates additional indicators. After applying the genetic algorithm (GA) to the neural network model, we were able to achieve better accuracy, and the values of the baseline statistical indices were changed to $RMSE = 0.279$, $R^2 = 0.718$, and $RRMSE = 27.116$. Also, to check the efficiency of the genetic algorithm, the PCA algorithm was also implemented on the data and the comparison results showed that the genetic algorithm was successful in reducing dimensions along with increasing accuracy.

Review History:

Received: Dec. 25, 2022

Revised: Apr. 29, 2024

Accepted: May, 05, 2024

Available Online: May, 31, 2024

Keywords:

Remote Sensing

Soil Organic Carbon

Neural Network

Dimension Reduction

Genetic Algorithm

1- Introduction

To determine soil organic carbon, field observing and remote sensing methods are used. Considering the time-consuming field operations and laboratory investigations and the unavailability of some points, we need indirect, fast, and low-cost methods. The use of remote sensing data, digital terrain models, and topographic features in the estimation of soil indices is a reliable, cheap, and accurate method, which is based on the surface reflection of light by phenomena [1, 2]. Recent studies show that the measurement data of two indices extracted from it is a suitable source of information in determining soil organic carbon [3]. Considering the importance of soil organic carbon, its modeling is very important for unavailable areas. Due to the progress of remote sensing systems, a huge range of different data is available to everyone. Due to the high dimensions of the data as a result

of the complexity of the calculations, it is very important to choose the correct data [4].

Height, slope, direction of slope, shape and position of slope, and natural conditions of soil drainage are factors that are attributed to topography [5]. Since topography affects the movement of water and the distribution of materials carried by water flow and thus controls the intensity and type of different processes in the soil, topography is one of the most important indices affecting soil organic carbon changes and its spatial distribution [6, 7]. Studies show that a large number of soil characteristics depend on the slope and the direction of the slope [8]. The relationship between soil characteristics and several slope factors such as degree, curvature, distance from the flat part, longitudinal direction, and height compared to the flat part has been investigated and the results of recent research show a significant relationship between different

*Corresponding author's email: jafari@tafreshu.ac.ir



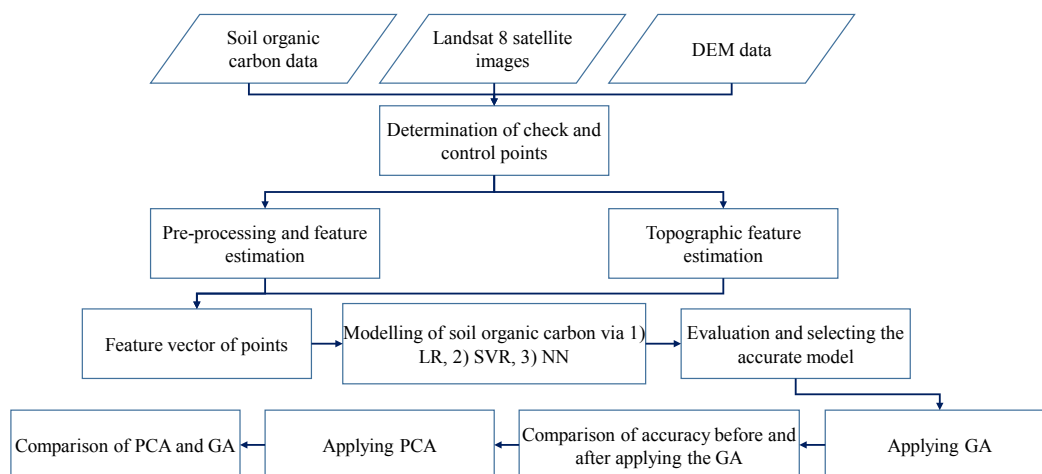


Fig. 1. The flowchart of the proposed method.

slope values and soil characteristics [9].

Besides the topographic data, soil spectral reflectance with different wavelengths can also be used to estimate soil organic carbon. Measurement of soil spectral reflectance is done by three methods laboratory, field, and remote sensing. Since the spectral reflectance of the soil is affected by its constituent materials, various factors such as moisture, percentage of organic matter, percentage of iron oxide, the relative percentage of clay, silt, and roughness coefficient of the soil surface affect the spectral reflectance of the soil [10, 11].

Due to the impact of spectral and topographical information in the estimation of soil organic carbon and the availability of different remote sensing data, data processing faces challenges. Despite the valuable results of previous research, there are limitations due to the large amount of data in conducting the research. Finding the features of effective remote sensing in soil organic carbon modeling can avoid wasting time and money. Feature selection techniques can simplify modeling by reducing the number of input variables and improving prediction accuracy. To remove additional and unimportant features, several feature selection and extraction methods have been developed, including the most important of them are Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Artificial Bee Colony (ABC) algorithm, and Principal Component Analysis (PCA).

In this research, in the first step, soil organic carbon modeling has been done using an artificial neural network based on spectral and topographical data. In the next step, the selection of optimal and effective indices is done using the genetic algorithm, and with the help of the optimal indices, we will try to increase the accuracy of modeling. In other words, the innovation of this research is in applying the genetic algorithm to the modeling process using remote sensing data to avoid wasting time and money in addition to increasing the accuracy of the model.

2- Methodology

The process of implementing the proposed method takes place in four main steps. In the first step, after preparing the desired satellite images, the necessary pre-processing will be done. In the second step, the desired indices are extracted from the images, and the feature vector is determined for specific points. In the third step, soil organic carbon modeling is done using the feature vector of points, and in the last step, using the dimension reduction technique, we try to reduce and eliminate some features to achieve higher accuracy for modeling. The flowchart of the proposed method is shown in "Figure 1".

3- Results and Discussion

The studied area is located in the northwest of Iran and East Azarbaijan province, between longitude 45.49 to 48.02 and latitude 36.93 to 39.30. The region area is about 47830 m². According to the size of the area and the existing conditions, 100 points with latitude and longitude coordinates along with the known soil organic carbon were provided by the Soil and Water Research Institute of Iran. To implement the method, the Landsat 8 satellite images were selected for this research due to their availability, and spectral and spatial resolution [12]. In addition, the topographical data were extracted from the SRTM¹ model.

For all data coordinates, soil organic carbon, spectral indices, values in the main band, and topographical indices are calculated. Then, we modeled soil organic carbon based on points with known coordinates and organic carbon, as well as based on 21 features collected, using three methods: Linear Regression (LR), Support Vector linear Regression (SVR), and Neural Network (NN). According to the numerical results, the evaluation criteria for NN modeling are superior to the other two models. So NN modelling is more accurate

¹ Shuttle Radar Topography Mission

Table 1. Comparison of accuracy evaluation criteria resulting from dimension reduction based on the GA and PCA.

Method	RMSE	R ²	RRMSE
GA	0.279	0.718	27.116
PCA	0.417	0.372	40.469

and predicts soil organic carbon better.

After implementing the proposed method based on combining the GA with the soil organic carbon modeling method, we will evaluate the modeling results. To the effectiveness of the GA, the PCA method was implemented on the data so that the results can be compared with the results obtained from the GA. According to the eigenvalues, the first 10 principal components were used in this study. The comparison of the desired evaluation indices can be seen in "Table 1".

4- Conclusions

The obtained results indicate that the GA is more capable of dimension reduction of the input data compared to the PCA method. According to the presented results the efficiency of the GA to increase the accuracy is clear. Therefore, by applying this algorithm, you can reach results closer to reality.

According to the results of the GA application, among the 31 input features that include the spectral-topographic indices obtained along with the main bands of the Landsat satellite, the application of the 15 spectral-topographic features (Coastal, Green, NIR, SWIR1, SWIR2, MSR, OSAVI, NDSI, SI1, SI2, SI32, SI4, SI5, Slope, Aspect) has increased the accuracy of the modeling; about the rest of the indexes, it is better to exclude them from the modeling process.

References

- [1] F. Chen, D.E. Kissel, L.T. West, W. Adkins, Field-scale mapping of surface soil organic carbon using remotely sensed imagery, *Soil Science Society of America Journal*, 64(2) (2000) 746-753.
- [2] M. Ladoni, H.A. Bahrami, S.K. Alavipanah, A.A. Norouzi, Estimating soil organic carbon from soil reflectance: a review, *Precision Agriculture*, 11 (2010) 82-99.
- [3] X. He, L. Yang, A. Li, L. Zhang, F. Shen, Y. Cai, C. Zhou, Soil organic carbon prediction using phenological parameters and remote sensing variables generated from Sentinel-2 images, *Catena*, 205 (2021) 105442.
- [4] G.S. Bhunia, P. Kumar Shit, H.R. Pourghasemi, Soil organic carbon mapping using remote sensing techniques and multivariate regression model, *Geocarto International*, 34(2) (2019) 215-226.
- [5] H. Emami, Investigation the effects of aspect and degree of slope on soil quality in the South East of Mashhad, *Journal of Water and Soil Conservation*, 23(2) (2016) 301-310, (in Persian).
- [6] J. Kumhálová, F. Kumhála, M. Kroulík, Š. Matějková, The impact of topography on soil properties and yield and the effects of weather conditions, *Precision Agriculture*, 12 (2011) 813-830.
- [7] S. Singh, Understanding the role of slope aspect in shaping the vegetation attributes and soil properties in Montane ecosystems, *Tropical Ecology*, 59(3) (2018) 417-430.
- [8] S. Jakšić, J. Ninkov, S. Milić, J. Vasin, M. Živanov, D. Jakšić, V. Komlen, Influence of slope gradient and aspect on soil organic carbon content in the region of Niš, Serbia, *Sustainability*, 13(15) (2021) 8332.
- [9] P.H. Walker, Contributions to the understanding of soil and landscape relationships, *Soil Research*, 27(4) (1989) 589-605.
- [10] H. Aïchi, Y. Fouad, C. Walter, R.V. Rossel, Z.L. Chabaane, M. Sanaa, Regional predictions of soil organic carbon content from spectral reflectance measurements, *Biosystems engineering*, 104(3) (2009) 442-446.
- [11] H. Bartholomeus, M.E. Schaepman, L. Kooistra, A. Stevens, W. Hoogmoed, O. Spaargaren, Spectral reflectance based indices for soil organic carbon quantification, *Geoderma*, 145(1-2) (2008) 28-36.
- [12] T.R. Loveland, J.R. Irons, Landsat 8: The plans, the reality, and the legacy, *Remote Sensing of Environment*, 185 (2016) 1-6.



کاهش ابعاد داده‌های سنجش از دوری به منظور برآورد کربن آلی خاک

نیوشا مظفری نهر، حدیثه سادات حسنی، مرضیه جعفری^{1b}

گروه ژئودزی مهندسی نقشه برداری، دانشگاه تفرش، تفرش، ایران.

تاریخچه داوری:

دریافت: ۱۴۰۱/۱۰/۰۴
بازنگری: ۱۴۰۳/۰۲/۱۰
پذیرش: ۱۴۰۳/۰۲/۱۶
ارائه آنلاین: ۱۴۰۳/۰۳/۱۱

کلمات کلیدی:

سنجش از دور
کربن آلی خاک
شبکه عصبی
کاهش ابعاد
الگوریتم ژنتیک

خلاصه: توزیع مواد آلی در خاک، بر فعالیت بیولوژیکی، در دسترس بودن مواد مغذی، ساختمان خاک و خاکدانه‌ها و ظرفیت نگهداری آب و بر مدیریت خاک بسیار موثر است. در این پژوهش به بررسی ارتباط بازتاب طیفی خاک با استفاده از داده‌های ماهواره لندست ۸ و همچنین داده‌های ارتفاعی SRTM و کربن آلی خاک پرداخته شده است. در روش پیشنهادی، انعکاس طیفی باندهای ماهواره لندست ۸ در کنار شاخص‌های گیاهی و ویژگی‌های توپوگرافی با هدف تعیین شاخص‌های موثر بر کربن آلی خاک، مورد بررسی قرار گرفته است. برای این منظور، از شبکه عصبی جهت ارتباط بین داده‌های سنجش از دوری و کربن آلی خاک استفاده شده است. جهت پیاده‌سازی روش پیشنهادی از ۱۰۰ نمونه خاک در استان آذربایجان شرقی استفاده شده است. مقادیر شاخص‌های آماری، RMSE و R2 به ترتیب ۰/۴۰۴، ۰/۲۵۴ و ۴۶/۵۹۷ بدست آمده و روش شبکه عصبی در مقایسه با روش‌های رگرسیون خطی و رگرسیون بردار پشتیبان خطی به دقت بالاتری دست یافت. در ادامه الگوریتم ژنتیک جهت کاهش ابعاد و تعیین شاخص‌های بهینه، افزایش دقت و کاهش پیچیدگی محاسبات پیشنهاد شد. این الگوریتم با حذف شاخص‌های اضافی، منجر به افزایش دقت مدل سازی کربن آلی خاک می‌شود. در این مرحله مقادیر شاخص‌های آماری R2، RMSE و RRMSE با مقادیر ۰/۲۷۹، ۰/۷۱۸ و ۲۷/۱۱۶ بهبود یافت. همچنین به منظور بررسی کارایی الگوریتم ژنتیک، الگوریتم آنالیز مولفه‌های اصلی نیز بر روی داده‌ها پیاده سازی شد و نتایج مقایسه نشان داد که الگوریتم ژنتیک در کاهش ابعاد همراه با افزایش دقت موفق بوده است.

۱- مقدمه

حفاظت و ارتقاء خصوصیات خاک نیازمند اقدامات همه جانبه است که تنها با مطالعه و درک عمیق عملکرد خاک در دو بعد مکان و زمان میسر می‌باشد. به منظور تعیین کربن آلی خاک بازدید میدانی (نمونه‌برداری شبکه‌ای یا منطقه‌ای) و روش‌های سنجش از دوری مورد استفاده قرار می‌گیرند. با توجه به هزینه بالای جمع‌آوری نمونه و بررسی‌های آزمایشگاهی، در دسترس نبودن برخی نقاط و زمانبر بودن عملیات زمینی، احتیاج به روش‌های غیر مستقیم، سریع و کم هزینه داریم. استفاده از داده‌های سنجش از دور، مدل رقومی زمین و ویژگی‌های توپوگرافی در برآورد شاخص‌های خاک روشی مطمئن، ارزان و دقیق می‌باشد که اساس آن بر پایه بازتاب سطحی نور توسط پدیده‌ها است [۳، ۴]. مطالعات اخیر نشان می‌دهد که داده‌های سنجش از دور و شاخص‌های مستخرج از آن، یک منبع اطلاعاتی مناسب‌تر تعیین کربن آلی خاک است [۵]. با توجه به اهمیت کربن آلی خاک، مدل سازی آن برای مناطق دور از دسترس بسیار حائز اهمیت است. به واسطه پیشرفت سامانه‌های سنجش از دوری، طیف عظیمی از داده‌های متنوع در دسترس

خاک یک منبع طبیعی بسیار مهم برای تولید غذا، فیبر و انرژی می‌باشد. علاوه بر این خاک یک فیلتر زیست محیطی برای فلزات، مواد معدنی و آلاینده‌های دیگر بوده و به عنوان تنظیم کننده حرکت آب در طبیعت در نظر گرفته می‌شود. توانایی یک خاک بستگی به ترکیب شیمیایی، بیولوژیکی و خواص فیزیکی آن دارد که بسته به مکان و زمان متغیر می‌باشد. این پدیده به عنوان یک مخزن بالقوه برای ذخیره کربن و گرمایش جهانی به حساب می‌آید. اساساً خاک یک پدیده بسیار پیچیده است که شامل مواد آلی و معدنی، آب و هوا می‌باشد [۱]. توزیع مواد آلی در خاک، بر روی فعالیت بیولوژیکی، در دسترس بودن مواد مغذی، ساختمان خاک و خاکدانه‌ها و ظرفیت نگهداری آب و به طور کلی بر روی مدیریت و حاصلخیزی خاک بسیار موثر است. مقدار کربن آلی خاک اهمیت بسیار زیادی در مسائلی مانند کشاورزی، بیابان زایی، گرم شدن کره زمین و مقدار CO₂ هوا دارد [۲].

* نویسنده عهده‌دار مکاتبات: jafari@tafreshu.ac.ir



تحدب در میزان کربن آلی خاک پرداختند. در تمام موارد در اراضی با پستی و بلندی کم، خاک‌ها در موقعیت‌های مقعر، مقدار کربن آلی خاک بیشتری نسبت به موقعیت‌های محدب دارند و همچنین در تمام موارد با افزایش درجه شیب مقدار کربن آلی خاک کمتر شده است. این موضوع احتمالاً به خروج سریع تر آب از این خاک‌ها مرتبط است [۱۷]. یکی دیگر عوامل تاثیرگذار بر خصوصیات خاک، کاربری اراضی است. Forsberg و همکاران در سال ۲۰۱۲ در تحقیق خود به بررسی تاثیر کاربری اراضی، آب و هوا، توپوگرافی و ویژگی های خاک بر میزان کربن آلی و نیتروژن کل در بالا دست رودخانه مایون در حوزه آبخیز شمال چین پرداختند. نتایج نشان می دهد که حجم کربن آلی خاک و نیتروژن کل در جنگل‌های طبیعی و چراگاه‌ها بیشتر از مزارع و زمین‌های کشت و کار شده می‌باشد. ارتفاع، شیب، رس خاک و میزان رطوبت تاثیرگذارترین ویژگی های کنترل میزان کربن آلی خاک و نیتروژن کل در این منطقه است [۱۸].

در سه دهه اخیر، در داخل کشور نیز مطالعاتی در زمینه تاثیر توپوگرافی بر روی ویژگی های خاک انجام شده است. خادمی و خیر در سال ۱۳۸۰ در بررسی تغییرپذیری برخی از خصوصیات خاک در اراضی شیب دار سمیرم به این نتیجه رسیدند که در مورد خصوصیات کیفی، قسمت‌های پایین و انتهای شیب حداکثر مقدار مواد آلی و فعالیت آنزیم فسفاتاز را دارا می‌باشد و بخش بالای آن در درجه بعدی قرار دارد. در مقابل، خاک‌های واقع بر روی شانه شیب و شیب پستی حداقل خصوصیات فوق را دارا هستند [۱۹]. محمدی و همکاران در سال ۱۳۸۲ نیز بررسی‌هایی در زمینه تاثیر شیب روی شاخص های خاک در حوزه سرآب حوزه آبی زاینده رود انجام دادند. نتایج بدست آمده نشان داد که با افزایش شیب، میزان مواد آلی خاک کمتر می شود [۲۰]. ابراهیمی و همکاران در سال ۱۳۹۶ تغییرات پوشش گیاهی و ویژگی های فیزیکی و شیمیایی خاک را در مراتع منطقه مورچه خورت مورد بررسی قرار دادند. آن‌ها نشان دادند که میزان چرای دام باعث افزایش معنی دار سدیم، آهن و هدایت الکتریکی و کاهش میزان پتاسیم، رطوبت اشباع و کربن آلی خاک شده است [۲۱].

علاوه بر اطلاعات توپوگرافی، از بازتاب طیفی خاک در طول موج‌های مختلف نیز به منظور برآورد کربن آلی خاک می‌توان استفاده کرد. اندازه‌گیری بازتاب‌های طیفی خاک از سه روش مشاهدات آزمایشگاهی، میدانی و فضایی انجام می‌شود. از آنجائیکه بازتاب طیفی خاک تحت تاثیر مواد تشکیل دهنده آن می‌باشد، فاکتورهای گوناگونی نظیر رطوبت، درصد ماده آلی، درصد اکسید آهن، درصد نسبی رس، سیلت و ضریب زبری سطح خاک بر بازتاب طیفی

همگان قرار دارد. به دلیل بالا بودن ابعاد داده‌ها در نتیجه پیچیده شدن محاسبات، انتخاب درست و صحیح این داده‌ها بسیار مهم است [۶]. تشکیل خاک نتیجه تاثیر متقابل بسیاری از فرآیندها بوده که هر یک از این فرآیندها در قالب پنج عامل تشکیل دهنده خاک عمل کرده و خاک‌های گوناگون را روی سطح زمین ایجاد می‌کنند. عوامل تشکیل دهنده خاک عبارتند از مواد مادری (p)، اقلیم (cl)، پستی و بلندی (r)، موجودات زنده (o) و زمان (t) که طبق رابطه $S = f (cl, p, r, o, t)$ بیان می‌شود [۷]. ارتفاع نسبت به سطح آب‌های آزاد، شیب، جهت شیب، شکل و موقعیت شیب و شرایط طبیعی زهکشی خاک از عواملی هستند که به توپوگرافی (پستی و بلندی) نسبت داده می‌شوند [۸]. از آنجائیکه توپوگرافی، حرکت آب و توزیع مواد حمل شده توسط جریان آب را تحت تاثیر قرار داده و اینگونه شدت و نوع فرآیندهای مختلف در خاک را کنترل می‌کند، توپوگرافی یکی از مهمترین شاخص‌های موثر بر تغییرات کربن آلی خاک و توزیع مکانی آن است [۹، ۱۰]. مطالعات نشان می‌دهند که تعداد زیادی از ویژگی‌های خاک به شیب و جهت شیب بستگی دارند [۱۱]. رابطه بین ویژگی های خاک و تعدادی از عوامل شیب همچون درجه، انحناء، فاصله از قسمت مسطح، جهت طولی و ارتفاع نسبت به قسمت مسطح مورد بررسی قرار گرفته و نتایج تحقیقات اخیر نشان دهنده ارتباط معناداری بین مقادیر مختلف شیب و ویژگی های خاک است [۱۲].

در یکی از نخستین مطالعات، Dalal و همکاران در سال ۱۹۸۶ رابطه بین ویژگی های خاک (شامل کربن آلی، رطوبت و میزان نیتروژن) و شیب منطقه در جنوب تایلند را مورد مطالعه قرار دادند. آن‌ها به این نتیجه رسیدند که با افزایش ارتفاع، مقدار کربن آلی کاهش پیدا خواهد کرد [۱۳]. در دهه ۱۹۹۰ مطالعات گسترده‌ای به بررسی اثر ویژگی های توپوگرافی و فرسایش بر روی خصوصیات خاک از جمله کربن آلی خاک پرداختند. Pierson و همکارانش در ۱۹۹۰ با بررسی و مقایسه ویژگی های خاک در موقعیت‌های متفاوت شیب متوجه شدند که خاک‌های واقع بر پایه شیب و انتهای شیب با مواد آلی بیشتر و فرسایش کمتر، دانه بندی بالاتری دارند [۱۴]. از نظر Cheng و همکارانش در سال ۱۹۹۸، فرسایش در قسمت‌های بالای شیب شدید می‌باشد و مواد آلی در قسمت‌هایی با شیب کمتر، کمترین مقدار و در پایه و انتهای شیب بیشترین مقدار را داراست [۱۵]. Garten و همکارانش در سال ۱۹۹۹ رابطه‌ای میان افزایش مقدار کربن آلی خاک و کاهش ارتفاع پیدا کردند. آنها این موضوع را به دمای سردتر خاک در ارتفاعات بالاتر نسبت دادند [۱۶]. در سال ۲۰۰۳، COX و همکاران به بررسی اثر تقعر و

خاک تاثیرگذار هستند [۲۲، ۲۳].

مواد آلی نقش مهمی در فرآیندهای شیمیایی و فیزیکی خاک و در نتیجه اثر زیادی بر ویژگی های طیفی خاک دارد. ماده آلی اثر قابل توجهی بر ناحیه طیفی مادون قرمز نزدیک^۱ و مادون قرمز با طول موج های کوتاه^۲ دارد. با افزایش میزان ماده آلی خاک، بازتابش در باندهای محدوده مرئی کمتر گردد. هنگامی که میزان ماده آلی به حدود ۹٪ برسد، معمولاً بازتاب پارامترهای دیگر خاک کم اثر می شوند [۲۴، ۲۵]. Bajwa و همکاران در سال ۲۰۰۱ ارتباط میان ویژگی های خاک و داده های تصویر ابرطیفی را مورد بررسی قرار داده و بیان کردند که مقدار ماده آلی رابطه قوی با بازتاب طیفی دارد [۲۶]. Saxena و همکاران در سال ۲۰۰۳ در مطالعه ای به وجود ارتباط بین شاخص های فیزیکی و شیمیایی خاک های هندوستان با انعکاس های طیفی پی بردند. آن ها با استفاده از همبستگی بین این ویژگی ها و انعکاس طیفی به این نتیجه دست یافتند که انعکاس طیفی خاک با افزایش مقدار ماده آلی و رطوبت خاک در باند مادون قرمز نزدیک کمتر می شود [۲۷]. Sullivan و همکاران در سال ۲۰۰۵ با استفاده از تصاویر آیکنوس بررسی کردند که باند مادون قرمز حرارتی به تنهایی ۳۸٪ و به همراه باندهای مرئی و مادون قرمز ۹۳٪ تغییرات در ماده آلی در عمق یک سانتی متری را آشکار می کند [۲۸]. Debaene و همکاران در سال ۲۰۱۰ برای تخمین مقدار کربن آلی خاک، اسیدیته و مقادیر منیزیم، پتاسیم، فسفر و درصد رس خاک سطحی روش طیف سنجی مرئی و مادون قرمز نزدیک را مورد استفاده قرار دادند. نتایج آن ها مبنی بر این است که بهترین ضریب تبیین^۳ (R²) برای درصد کربن آلی خاک (۰/۶۵) و منیزیم خاک (۰/۶۹) به دست آمده، در حالی که ضرایب تبیین برای سایر پارامترها پایین تر می باشد [۲۹]. در سال ۲۰۱۹، Zizala و همکاران به بررسی توانایی سنجنده های چندطیفی ماهواره ای (شامل سنتینل-۲، لندست-۸ و PlanetScope) به منظور تعیین کربن آلی خاک در موراوای جنوبی پرداختند. آن ها از تصاویر فراطیفی CASI و SASI به عنوان مرجع استفاده کردند. نتایج بدست آمده توانایی یکسان سنجنده های مختلف فضا برد را نشان می دهد [۳۰]. در همین سال، An-gelopoulou و همکاران مروری جامع بر توانایی سنجنش از دور در برآورد کربن آلی خاک داشتند. برای این منظور سکوه های مختلف پهپاد، هوایی و ماهواره ای و همچنین الگوریتم های یادگیری ماشین مختلف را بررسی کردند. این مطالعه نشان داد به منظور جبران کاهش دقت در سکوه های ماهواره ای

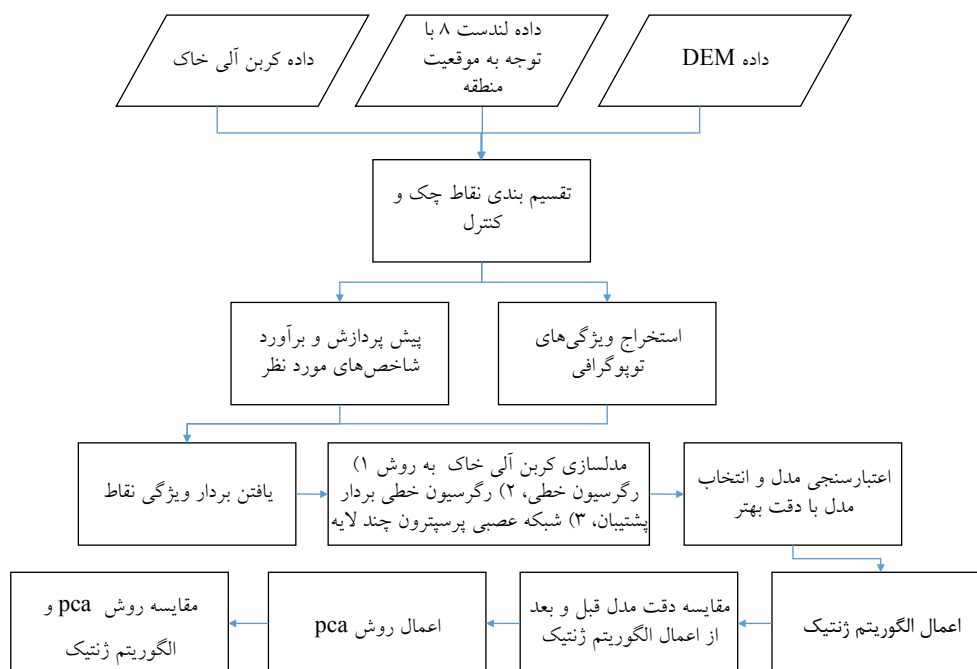
لازم است از تصحیحات اتمسفری و الگوریتم های یادگیری ماشین کارآتر استفاده شود. همچنین آن ها نشان دادند که استفاده از اطلاعات طیفی به تنهایی کافی نبوده و اطلاعات کمکی نظیر ناهمواری می تواند به بهبود تخمین کربن آلی کمک کند [۳۱]. در سال ۲۰۲۱، He و همکاران با استفاده از ویژگی های زمانی و مکانی مستخرج از تصاویر سنتینل-۲ و پارامترهای فنولوژی به پیش بینی کربن آلی پرداختند. نتایج بدست آمده حاکی از افزایش دقت پیش بینی کربن آلی با افزودن اطلاعات فنولوژی و ویژگی های بدست آمده می باشد [۵]. با توجه به اهمیت موضوع، Pouladi و همکاران در سال ۲۰۲۳ نیز مروری سیستماتیک بر تهیه نقشه کربن آلی خاک بر اساس داده های سنجنش از دوری کردند. مرور مطالعات انجام شده در این حوزه نشان می دهد که اطلاعات توپوگرافی به همراه تصاویر سنجنش از دور دومین مهم در تهیه نقشه کربن آلی خاک هستند [۳۲].

با توجه به تاثیر اطلاعات طیفی و توپوگرافی در برآورد کربن آلی خاک و در دسترس بودن داده های سنجنش از دوری مختلف، پردازش داده ها با چالش هایی مواجه می شود. تکنیک های انتخاب ویژگی می تواند مدل سازی را با کاهش تعداد متغیرهای ورودی مدل سازی را ساده تر کند و دقت پیش بینی را بهبود بخشد. برای حذف ویژگی های اضافی و غیر مهم روش های انتخاب و استخراج ویژگی متعددی از جمله مهمترین آنها الگوریتم بهینه سازی ازدحام ذرات^۴ (PSO)، الگوریتم ژنتیک^۵ (GA)، الگوریتم کلونی زنبور عسل مصنوعی^۶ (ABC) و تحلیل مولفه اصلی^۷ (PCA) گسترش داده شده اند.

مطالعات گسترده ای در حوزه کاهش ابعاد داده های سنجنش از دوری در کاربردهای مختلف انجام شده است که الگوریتم ژنتیک یک ابزار بسیار کارآمد در این حوزه می باشد [۳۳]. صمدزادگان و حسنی در سال ۱۳۹۱ از توانایی الگوریتم ژنتیک به عنوان یک تکنیک بهینه سازی فرا ابتکاری در تعیین مقادیر بهینه پارامترهای ماشین های بردار پشتیبان و همچنین انتخاب زیرمجموعه ویژگی های بهینه با هدف طبقه بندی تصاویر فراطیفی استفاده کردند. نتایج نشان داد که ویژگی های ورودی و پارامترها هر کدام جداگانه تاثیر زیادی بر عملکرد ماشین های بردار پشتیبان دارند ولی بهترین عملکرد طبقه بندی کننده تصاویر فراطیفی با حل همزمان آن دو بدست آمد [۳۴]. صادق و همکاران در سال ۱۳۹۸ از رگرسیون بردار پشتیبان و شبکه عصبی

4 Particle swarm optimization
5 Genetic algorithm
6 Artificial bee colony
7 Principal component analysis

1 Near Infrared (NIR)
2 Short-wave Infrared (SWIR)
3 Coefficient of determination



شکل ۱. فلوجارت روش پیشنهادی

Fig. 1. Flowchart of the proposed method

الگوریتم ژنتیک بر روند مدل سازی انجام شده با استفاده از داده های سنجش از دوری می باشد تا علاوه بر افزایش دقت مدل از هدر رفت زمان و هزینه جلوگیری کند.

۲- روش پیشنهادی

این مطالعه به منظور ارائه راهکاری بهینه برای کاهش بعد داده های سنجش از دوری جهت انتخاب شاخص های بهینه در مدل سازی کربن آلی خاک صورت گرفته است. روند اجرای روش پیشنهادی در چهار گام اصلی صورت می پذیرد. در گام نخست پس از تهیه تصاویر ماهواره ای مورد نظر پیش پردازش های لازم انجام خواهد شد. در گام دوم شاخص های مورد نظر از تصاویر استخراج شده و برای نقاط مشخص بردار ویژگی تعیین می گردد. در گام سوم مدل سازی کربن آلی خاک با استفاده از بردار ویژگی نقاط صورت می پذیرد و در گام آخر با استفاده از تکنیک کاهش ابعاد سعی بر کاهش و حذف برخی از ویژگی ها داریم تا به دقت بالاتر برای مدل سازی برسیم. فلوجارت روش پیشنهادی در شکل ۱ نمایش داده شده است. در ادامه هر یک از بخش های مربوط به فلوجارت به تفصیل شرح داده شده است.

به منظور برآورد رطوبت خاک در تصاویر سنتینل-۱ و ۲ استفاده کردند. الگوریتم ژنتیک با انتخاب باندهای بهینه تصاویر سنتینل-۲ به بهبود عملکرد مدل سازی رگرسیون منجر شد [۳۵]. Mohamed و همکاران در ۲۰۲۳ به بررسی الگوریتم های یادگیری ماشین و انتخاب ویژگی در مساله برآورد شوری خاک از تصاویر فعال و غیرفعال سنجش از دوری پرداختند. الگوریتم ژنتیک با انتخاب زیرمجموعه بهینه از ویژگی های طیفی در طول موج مرئی و همچنین ویژگی های بافت تصاویر راداری منجر به افزایش دقت و کاهش هزینه محاسباتی گردید [۳۶].

علیرغم نتایج ارزشمند تحقیقات پیشین، محدودیت هایی به دلیل حجم بالای داده ها در انجام پژوهش وجود دارد. یافتن ویژگی های سنجش از دوری موثر در مدل سازی کربن آلی خاک، از اتلاف هزینه و زمان بسیار جلوگیری کند. در این پژوهش در گام اول به مدل سازی کربن آلی خاک با استفاده از شبکه عصبی مصنوعی بر اساس داده های طیفی و توپوگرافی پرداخته شده است. در گام بعدی با استفاده از الگوریتم ژنتیک انتخاب شاخص های بهینه و موثر انجام می شود که با کمک شاخص های بهینه سعی در افزایش دقت مدل سازی خواهیم داشت. به عبارت بهتر، نوآوری این تحقیق در اعمال

۲-۱- پیش پردازش

این تحقیق از دو ویژگی شیب و جهت شیب استفاده می شود، باید محاسبه گردد. برای به دست آوردن این ویژگی ها از DEM منطقه استفاده می شود. اگر بخواهیم شیب بین دو نقطه را محاسبه کنیم، به راحتی اختلاف ارتفاع دو نقطه ابتدا و انتها را بر فاصله آن دو تقسیم می کنیم. اما زمانی که قصد داریم برای یک پیکسل از DEM منطقه شیب را تخمین بزنیم، با مشکل رو به رو هستیم، چراکه در هریک از جهات هشت گانه یک شیب محاسبه می شود. با استفاده از روابط (۱) و (۲) می توان شیب (slope) و جهت شیب (aspect) را با استفاده از میانگین تغییرات ارتفاع در راستای محور X و Y محاسبه کرد.

$$slop = \sqrt{\left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2} \quad (1)$$

$$Aspect = \arctan\left(\frac{\frac{dZ}{dY}}{\frac{dZ}{dX}}\right) \quad (2)$$

در روابط ذکر شده dx معادل تغییرات هر پیکسل DEM در راستای محور x، dy معادل تغییرات هر پیکسل DEM در راستای محور y، dz معادل تغییرات هر پیکسل DEM در راستای محور Z می باشد. به این ترتیب ما برای نقاط مورد نظر که دارای کربن آلی خاک مشخصی هستند یک بردار ویژگی شامل ویژگی ها، بازتاب در باندهای مختلف و ویژگی توپوگرافی داریم که بر اساس آن ها باید مدل سازی صورت گیرد.

۲-۳- مدل سازی

گام سوم این تحقیق به مدل سازی کربن آلی خاک اختصاص دارد. به این منظور باید به دنبال رابطه ای بین بردار ویژگی نقاط و کربن آلی خاک در آن نقطه بود. برای دستیابی به نتیجه بهتر و دقیق تر در مرحله آموزش، آزمون و اعتبارسنجی، داده ها چندین بار دسته بندی می شوند تا تمامی داده ها در دسته آموزشی، آزمون و اعتبارسنجی قرار گیرند و مقدار معیارهای ارزیابی نهایی از میانگین گیری مقادیر این معیارها در هر کدام از دسته بندی ها بدست می آید. به منظور مدل سازی از سه روش (۱) رگرسیون خطی، (۲) رگرسیون خطی بردار پشتیبان، و (۳) روش شبکه عصبی پرسپترون چندلایه^۱ (MLP) استفاده می شود. اعمال مدل سازی توسط داده های آموزشی از طریق روش

داده های ورودی به این الگوریتم پیشنهادی سه دسته است. (۱) داده های واقعی کربن آلی خاک که از نمونه های خاک منطقه مورد مطالعه بدست آمده، (۲) داده های سنجش از دوری که در اینجا از تصاویر ماهواره لندست ۸ بدست آمده، (۳) مدل ارتفاعی رقومی (DEM) منطقه. یکی از مهم ترین مراحل در پردازش تصاویر ماهواره ای، انجام مرحله پیش پردازش داده ها می باشد. تصاویر ماهواره لندست ۸ به صورت زمین مرجع شده در اختیار کاربران قرار می گیرد، بنابراین نیاز به تصحیح هندسی ندارد. گام اول برای پیش پردازش داده ها، انجام تصحیحات رادیومتریک می باشد. گام بعدی در انجام پیش پردازش داده ها، انجام تصحیحات اتمسفریک که ناشی از تاثیر جذب و پخش اتمسفری است. تصحیح رادیومتریک و پس از آن الگوریتم FLAASH بر روی تصاویر برای جبران اثرات اتمسفریک اعمال می شود که یکی از انواع روشهای حذف اثر اتمسفریک است [۳۷]. در شبکه عصبی مصنوعی جهت طراحی نیازمند سه دسته داده آموزشی، اعتبارسنجی و آزمون می باشیم. از داده های آموزشی برای یافتن رابطه بین داده های ورودی و خروجی های مشاهده شده استفاده می شود. از داده های اعتبارسنجی برای نظارت و ارزیابی بر یادگیری درست شبکه استفاده می گردد. از داده های آزمون برای ارزیابی عملکرد شبکه استفاده می شود. در این مرحله باید داده ها، به دسته های آموزشی، آزمون و اعتبارسنجی تقسیم شوند، بنابراین از بین تمام داده ها درصد مشخصی از نقاط به عنوان نقاط آموزشی، و درصدی به عنوان نقاط اعتبارسنجی و آزمون تعیین می شود.

۲-۲- استخراج ویژگی ها

پس از اعمال پیش پردازش ها، اولین اقدام محاسبه شاخص های مورد نظر است. شاخص های پوشش گیاهی، خاک و روشنایی از تصاویر تصحیح شده بدست می آید. مقدار این شاخص ها برای هر نقطه با طول و عرض مشخص قابل محاسبه است و طبق روابط ریاضی ساده بین باندهای تصاویر محاسبه می شود. تعداد ۱۹ شاخص های مورد مطالعه در این پژوهش در جدول ۱ ارائه شده است.

در روابط ذکر شده B، G، R، معادل باندهای آبی، سبز و قرمز به ترتیب در باندهای ۲، ۳ و ۴، NIR بازتاب مادون قرمز نزدیک در باند ۵، SWIR بازتاب مادون قرمز با طول موج کوتاه در باند ۶ تصاویر ماهواره لندست ۸ می باشد.

به غیر از موارد ذکر شده در جدول، ویژگی های توپوگرافی نیز که در

1 Multi Layer perceptron

جدول ۱. شاخص‌های پوشش گیاهی، خاک و روشنایی

Table 1. Vegetation, soil and brightness indices

ردیف	نام اختصاری شاخص	نام شاخص به تفصیل	رابطه
۱	NDVI	Normalized Difference Vegetation Index	$NDVI = \frac{NIR - R}{NIR + R}$
۲	DVI	Difference Vegetation Index	$DVI = NIR - R$
۳	EVI	Enhanced vegetation index	$EVI = \frac{NIR - R}{(NIR + 6R - 7.5B) + 1}$
۴	IPVI	Infrared percentage vegetation index	$IPVI = (NDVI + 1) \frac{NIR}{NIR + R}$ 2
۵	GEMI	Global Environment Monitoring Index	$GEMI = NIR(1 - 0.25NIR) - \frac{R - 0.125}{1 - R}$
۶	SAVI	Soil Adjusted Vegetation Index	$SAVI = (1 + L) \frac{NIR - R}{NIR + R + L}; L = 0.5$
۷	OSAVI	Optimized Soil Adjusted Vegetation Index	$OSAVI = (1 + 0.16) \frac{NIR - R}{NIR + R + 0.16}$
۸	MSAVI	Modified Soil Adjusted Vegetation Index	$MSAVI = \frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - R)}}{2}$
۹	NDBI	Normalized difference built-up index	$NDBI = \frac{SWIR - NIR}{SWIR + NIR}$
۱۰	NDSI	Normalized Difference Salinity Index	$NDSI = \frac{G - SWIR}{G + SWIR}$
۱۱	GOSAVI	Green Optimized Soil Adjusted Vegetation Index	$GOSAVI = \frac{NIR - G}{NIR + G + Y}; Y = 0.16$
۱۲	BI	Brightness index	$BI = \sqrt{NIR^2 + R^2}$
۱۳	SI1	Salinity index 1	$SI1 = \frac{R \times NIR}{G}$
۱۴	SI2	Salinity index 2	$SI2 = \frac{B - R}{B + R}$
۱۵	SI3	Salinity index 3	$SI3 = \sqrt{(G^2 + R^2 + NIR^2)}$
۱۶	SI32	Salinity index 32	$SI32 = \sqrt{(G^2 + R^2)}$
۱۷	SI4	Salinity index 4	$SI4 = \sqrt{B \times R}$
۱۸	SI5	Salinity index 5	$SI5 = \frac{B}{R}$
۱۹	CRSI	Canopy Response Salinity Index	$CRSI = \frac{(R \times NIR) - (B \times G)}{(R \times NIR) + (B \times G)}$

بهترین افراد هر نسل دارای مقدار RMSE کمتر و مقدار نزدیک‌تر کربن آلی خاک تخمین زده شده به واقعیت می‌باشند. سپس اعضای نسل بعد با ایجاد یک حوضچه تکثیر و قرار دادن اعضای نسل فعلی با احتمال انتخاب بیشتر در آن آغاز می‌شود. والدین از این حوضچه انتخاب شده و با عملگر تلفیق تک نقطه‌ای فرزندان جدید تولید می‌شود. با استفاده از عملگر جهش، راه‌حل‌های تولید شده تغییر خواهند کرد و اعضای نسل بعد ایجاد می‌شوند. انتظار می‌رود بعد از چندین تکرار این الگوریتم ما به نسلی با جواب بهینه و دقت بیشتر برسیم. بررسی کروموزم بهینه نشانگر این مسئله خواهد بود که چه ویژگی‌هایی باید از روند مدل سازی حذف شوند، تا با استفاده از کاهش بعد فضای ویژگی به برآورد مطلوب‌تری از کربن آلی خاک برسیم.

به منظور بررسی کارایی الگوریتم پیشنهادی ژنتیک، روش تحلیل مولفه اصلی (PCA) به عنوان یک روش کاهش بعد مرسوم پیاده سازی می‌شود. معیارهای ارزیابی محاسبه شده و با مقادیر این معیارها حاصل از الگوریتم ژنتیک مقایسه خواهد شد.

۳- مطالعه موردی

به منظور ارزیابی عملکرد روش پیشنهادی پژوهش، مدل سازی کربن آلی خاک بر روی یکسری نقاط انجام شده و دقت مدل سازی قبل و بعد از الگوریتم کاهش ابعاد ژنتیک مورد ارزیابی قرار خواهد گرفت.

۳-۱- منطقه مورد مطالعه و مجموعه داده مورد استفاده

منطقه مورد مطالعه در شمال غربی ایران و در استان آذربایجان شرقی، بین طول جغرافیایی ۴۵/۴۹ الی ۴۸/۰۲ و عرض جغرافیایی ۳۶/۹۳ الی ۳۹/۳۰ قرار دارد. مساحت منطقه حدود ۴۷/۸۳۰ مترمربع است. با توجه به وسعت منطقه و شرایط موجود تعداد ۱۰۰ نقطه مختصات دار با طول و عرض جغرافیایی همراه با مقادیر معلوم کربن آلی خاک توسط سازمان آب و خاک در اختیار قرار گرفت. نقشه پراکندگی نقاط در شکل ۲ آمده است.

جهت پیاده سازی روش نیاز به تصاویر ماهواره‌ای به عنوان یکی از داده های ورودی است. تصاویر ماهواره لندست ۸ با توجه به در دسترس بودن، قدرت تفکیک طیفی و قدرت تفکیک مکانی مناسب برای این پژوهش انتخاب شدند [۳۸]. تصاویر لندست ۸ دارای ۱۱ باند می باشد که به شرح جدول ۲ می‌باشد.

این اطلاعات از طریق سایت^۱ USGS اخذ گردید. تصاویر مربوط به ماه مه سال ۲۰۱۹ است. نمونه‌ای از تصاویر ماهواره‌ای منطقه مورد نظر

رگرسیون خطی و رگرسیون خطی بردار پشتیبان انجام شده و سپس با استفاده از داده‌های اعتبارسنجی و تست ارزیابی می‌شوند. پیاده سازی شبکه عصبی MLP نیز جهت مدل سازی انجام شده و با داده‌های تست ارزیابی صورت می‌گیرد.

به منظور ارزیابی عملکرد روشهای مدل سازی رگرسیون خطی و بردار پشتیبان و همچنین شبکه عصبی MLP، از معیارهای میانگین مجذور خطای مربعی (RMSE)، ضریب تبیین (R^2) و میانگین مجذور خطای مربعی نسبی (RRMSE) که روابط آن بصورت زیر ارائه شده است استفاده می‌شود.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - Y_i)^2} \quad (3)$$

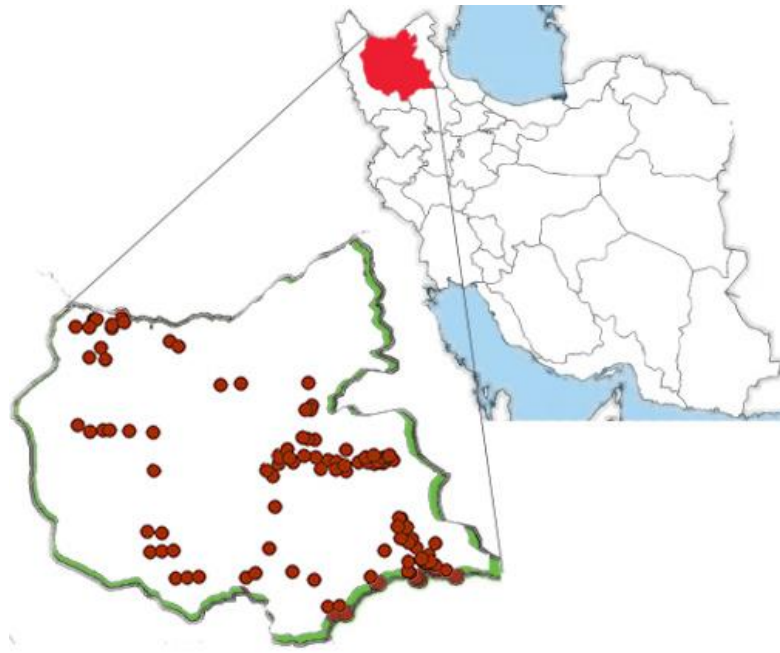
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - Y_i)^2}{\sum_{i=1}^n (y_i - y_{mean_i})^2} \quad (4)$$

$$RRMSE = \left(\frac{RMSE}{y_{mean}} \right) \times 100 \quad (5)$$

در روابط فوق y_i مقدار کربن آلی اندازه گیری شده، Y_i مقدار کربن آلی پیش بینی شده، y_{mean} میانگین کربن آلی اندازه گیری شده و n تعداد نمونه‌ها است.

۳-۲- کاهش تعداد شاخص ها به منظور کاهش ابعاد ویژگی ها بر مبنای الگوریتم ژنتیک

در این پژوهش به منظور کاهش تعداد ویژگی‌ها و کاهش ابعاد ویژگی به هدف دستیابی به دقت بالاتر در مدل سازی، از الگوریتم ژنتیک استفاده می‌شود. جهت پیاده سازی الگوریتم ژنتیک، نسل اولیه‌ای به صورت تصادفی تشکیل شده که در آن هر عضو به صورت دودویی کدگذاری می‌شود. ابعاد هر عضو به تعداد ویژگی‌های تولید شده بوده و یک به معنای انتخاب آن ویژگی و صفر به معنای عدم انتخاب آن است. کیفیت تمام اعضاء بر اساس مقدار RMSE آن‌ها در برآورد کربن آلی خاک محاسبه می‌شود.



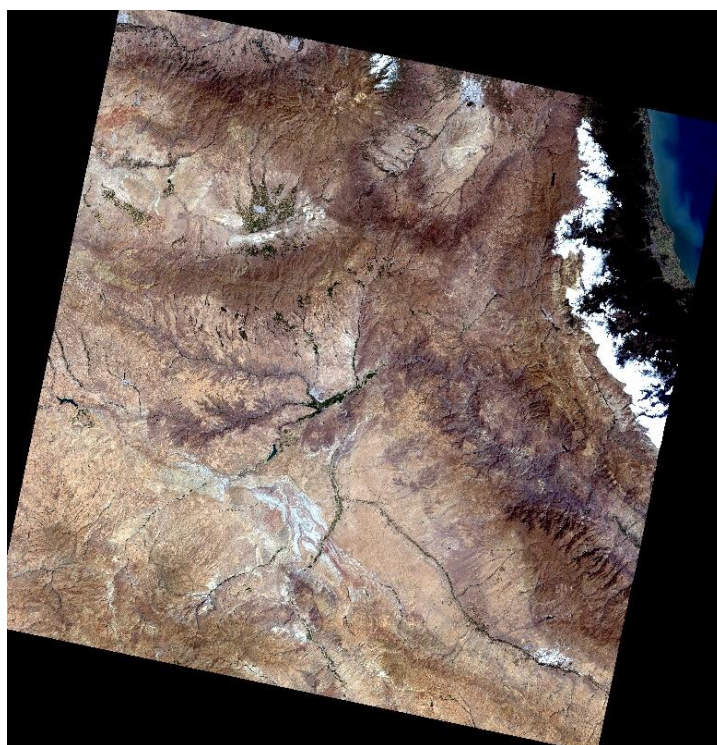
شکل ۲. منطقه مورد مطالعه و پراکندگی نقاط مشاهداتی

Fig. 2. Case study area and observed points distribution

جدول ۲. اطلاعات باندهای لندست ۸

Table 2. Information of Landsat 8 bands

شماره باند	طول موج (میکرومتر)	رزولوشن (متر)
۱	۰,۴۵۳-۰,۴۳۳	۳۰
۲	۰,۵۱۵-۰,۴۵۰	۳۰
۳	۰,۶۰۰-۰,۵۲۵	۳۰
۴	۰,۶۸۰-۰,۶۳۰	۳۰
۵	۰,۸۸۵-۰,۸۴۵	۳۰
۶	۱,۶۶۰-۱,۵۶۰	۳۰
۷	۲,۳۰۰-۲,۱۰۰	۳۰
۸	۰,۶۸۰-۰,۵۰۰	۱۵
۹	۱,۳۹۰-۱,۳۶۰	۳۰
۱۰	۱۰,۶-۱۱,۲	۱۰۰
۱۱	۱۱,۵-۱۲,۵	۱۰۰



شکل ۳. تصویر ماهواره‌های اخذ شده توسط ماهواره لندست ۸ از سایت USGS مربوط به ماه مه ۲۰۱۹

Fig. 3. Landsat 8 satellite image acquired from USGS site at May 2019

اصلی و ویژگی‌های توپوگرافی محاسبه شدند. سپس کربن آلی خاک را براساس نقاط با مختصات و کربن آلی معلوم، و همچنین بر اساس ۲۱ ویژگی جمع آوری شده، به سه روش رگرسیون خطی، رگرسیون خطی بردار پشتیبان و شبکه عصبی مدل سازی کردیم.

برای مدل سازی به روش رگرسیون خطی داده‌ها در نرم افزار SPSS وارد شده و رگرسیون خطی اعمال شده است و با کمک این نرم افزار مدل خطی بین کربن آلی خاک و بردار ویژگی برآزش داده شده است. مدل سازی به روش رگرسیون بردار پشتیبان در محیط برنامه نویسی MATLAB پیاده سازی شده و مدل برآزش داده شده در تعدادی از نقاط مورد آزمون قرار گرفتند. برای مدل سازی به روش شبکه عصبی از ابزار نرم افزار MATLAB استفاده شده و شبکه با پارامترهای بهینه طراحی شده است. در این نرم افزار نقاط به طور تصادفی انتخاب شده و به سه دسته آموزشی، آزمون و اعتبارسنجی تقسیم شدند. در این پژوهش، ۷۰ درصد داده‌ها به عنوان داده آموزش به منظور طراحی شبکه، ۱۵ درصد داده‌ها برای آزمون و ۱۵ درصد داده‌ها برای اعتبارسنجی در نظر گرفته شد. این دسته بندی چندین بار تکرار شده به

در شکل ۳ نمایش داده شده است. تصاویر استفاده شده در این پژوهش زمین مرجع شده بوده و احتیاجی به تصحیح هندسی نداشته است. تصحیح اتمسفریک و رادیومتریکی توسط نرم افزار ENVI و الگوریتم FLAASH انجام گرفته و تمامی پردازش‌ها بر روی تصاویر تصحیح شده انجام می شود. برای تصاویر ماهواره‌ای ۱۹ ویژگی طیفی (جدول ۱) با توجه به مقدار هر پیکسل در باندهای مختلف در نرم افزار ENVI محاسبه گردید. با توجه به مختصات نقاط مورد نظر مقدار بازتابندگی در هر باند و مقدار ویژگی‌های محاسبه شده گردآوری شد.

داده‌های توپوگرافی از مدل SRTM^۱ منطقه مورد مطالعه استخراج شد و در نرم افزار ARCGIS مورد بررسی قرار گرفته است. با توجه به تحقیقات انجام شده و اهمیت شیب و جهت شیب در تعیین کربن آلی خاک، این دو ویژگی مهم برای تمامی نقاط مورد نظر در نرم افزار ARCGIS محاسبه شده و در کنار ویژگی‌های طیفی قرار می گیرد.

تمامی داده‌ها با مختصات، کربن آلی خاک، ویژگی‌ها، مقادیر در باند

1 Shuttle Radar Topography Mission

جدول ۳. پارامترهای تنظیم شده الگوریتم ژنتیک

Table 3. Parameters setting for Genetic Algorithm

تعداد اعضاء نسل	تعداد تکرار الگوریتم	احتمال انتخاب	سایز حوضچه تکثیر	احتمال جهش
۱۰	۳۰	٪۷۰	۱۰	٪۰/۱

جدول ۴. معیارهای ارزیابی دقت حاصل از شبکه عصبی در مقایسه با روش خطی و بردار پشتیبان

Table 4. Evaluation metrics of MLP in comparison with linear and support vector regression

داده تست			داده اعتبارسنجی			روش
RMSE	R ²	RRMSE	RMSE	R ²	RRMSE	
۰/۵۹۶	-۰/۶۲۶	۶۸/۷۷۵	۰/۶۵۶	-۰/۸۱۳	۷۰/۵۶۳	رگرسیون خطی
۰/۴۴۵	۰/۰۹۶	۵۱/۲۹۴	۰/۴۶۵	۰/۰۸۹	۴۹/۹۹۸	رگرسیون بردار پشتیبان
۰/۴۰۴	۰/۲۵۴	۴۶/۵۹۷	۰/۲۷۸	۰/۶۷۴	۲۹/۹۳	شبکه عصبی

۳-۲- نتایج

پس از مدل سازی کربن آلی خاک می بایست مقدار دقت این مدل ها مورد ارزیابی قرار گیرد و مدل با دقت بیشتر به عنوان مدل نهایی انتخاب شود. ارزیابی مدل های موجود بر اساس معیار ارزیابی RMSE, RRMSE و R2 با استفاده از داده های در نظر گرفته شده برای آزمون، انجام شده و نتایج آماری در جدول ۴ ارائه شده است.

با توجه به نتایج جدول معیارهای ارزیابی برای مدل سازی به روش شبکه عصبی نسبت به دو مدل دیگر (رگرسیون خطی و بردار پشتیبان) برتری دارد بنابراین مدل سازی به روش شبکه عصبی دارای دقت بیشتر بوده و کربن آلی خاک را بهتر پیش بینی کرده است. به منظور ارزیابی توانایی ویژگی های طیفی و ارتفاعی در برآورد کربن آلی خاک، تاثیر هر یک از این ویژگی ها به تنهایی و همچنین در فضای ویژگی ترکیبی در جدول ۵ آورده شده است.

آنالیز جدول فوق نشان می دهد که هر دو فضای ویژگی طیفی و ارتفاعی بر برآورد کربن آلی خاک تاثیرگذار هستند؛ هرچند اطلاعات طیفی عملکرد مطلوب تری بر روی داده اعتبارسنجی دارند. با اینحال

صورتی که تمامی نقاط در هر سه دسته بندی قرار بگیرند و در نهایت با میانگین گیری از معیارهای ارزیابی به دست آمده از هر دسته بندی نتیجه نهایی مشخص گردید. بردار ویژگی نقاط شامل ویژگی های محاسبه شده، ویژگی توپوگرافی و مقدار نقاط در باندهای مختلف به عنوان ورودی شبکه و مقدار کربن آلی نقاط به عنوان تابع هدف در نظر گرفته شده است. در این پژوهش از دو لایه پنهان استفاده شد. توابع فعال ساز برای لایه های میانی، تابع سیگموئید و در لایه خروجی خطی است.

پس از ارزیابی نتایج اعتبارسنجی، روش مدل سازی با دقت بالاتر جهت مدل سازی کربن آلی خاک بر اساس ویژگی های تعیین شده انتخاب می گردد. در گام بعدی طبق روش پیشنهادی، روش انتخابی مدل سازی کربن آلی خاک با الگوریتم ژنتیک تلفیق می شود.

به منظور بررسی کارایی الگوریتم ژنتیک، الگوریتم PCA نیز بر روی داده ها پیاده سازی می شود و نتایج آن با نتایج حاصل از الگوریتم ژنتیک مقایسه می شود. پارامترهای تنظیم

جدول ۵. ارزیابی فضای ویژگی در برآورد کربن آلی خاک

Table 5. Feature space evaluation in soil organic carbon estimation

داده تست			داده اعتبارسنجی			روش
RMSE	R2	RRMSE	RMSE	R2	RRMSE	
۰/۵۸۲	۰/۰۳۷	۶۳/۸۴۷	۰/۳۴۱	۰/۶۲۸	۳۶/۳۸	فضای ویژگی طیفی
۰/۵۵۹	-۰/۱۳۱	۵۴/۳۰۱	۰/۵۳۱	۰/۱۷۵	۵۲/۷۴۷	فضای ویژگی ارتفاعی
۰/۴۰۴	۰/۲۵۴	۴۶/۵۹۷	۰/۲۷۸	۰/۶۷۴	۲۹/۹۳	فضای ویژگی ترکیبی

جدول ۶. مقایسه معیارهای ارزیابی دقت ناشی از کاهش ابعاد بر اساس روش PCA و GA.

Table 6. Comparison of accuracy evaluation metric resulting from dimension reduction based on the GA and PCA.

RMSE	R ²	RRMSE	
۰/۲۷۹	۰/۷۱۸	۲۷/۱۱۶	روش پیشنهادی
۰/۴۱۷	۰/۳۷۲	۴۰/۴۶۹	کاهش ابعاد با PCA

مشخص است. بنابراین با اعمال این الگوریتم می‌توان به نتایج نزدیک‌تری نسبت به واقعیت رسید.

با توجه به نتایج اعمال الگوریتم ژنتیک از بین ۳۱ ویژگی ورودی که شامل شاخص‌های طیفی-توپوگرافی بدست آمده در کنار باندهای اصلی ماهواره لندست است، کاربرد ۱۵ ویژگی طیفی-توپوگرافی زیر به افزایش دقت مدل سازی کمک بسیار زیادی کرده است و بقیه شاخص‌ها بهتر است از روند مدل سازی حذف شوند.

Coastal, Green, NIR, SWIR1, SWIR2, MSR, OSA-VI, NDSI, SI1, SI2, SI32, SI4, SI5, Slope, Aspect

به این ترتیب با اعمال الگوریتم ژنتیک دقت بهتری در مدل سازی حاصل شده و با شناسایی ویژگی‌هایی بهینه با تاثیر بیشتر و حذف بقیه شاخص‌ها، به کاهش بعد ویژگی همراه با افزایش دقت دست پیدا کرده‌ایم.

استفاده همزمان از اطلاعات طیفی و توپوگرافی دارای بالاترین دقت در مقایسه با هر یک از دو فضا به تنهایی می‌باشد.

پس از پیاده سازی روش پیشنهادی براساس تلفیق الگوریتم ژنتیک با روش مدل سازی کربن آلی خاک، به ارزیابی نتایج مدل سازی می‌پردازیم. به منظور بررسی کارایی الگوریتم ژنتیک روش PCA بر روی داده‌ها پیاده سازی شد تا بتوان نتایج حاصل را با نتایج حاصل شده از الگوریتم ژنتیک مقایسه نمود. با توجه به مقادیر ویژه، در این مطالعه از ۱۰ مولفه اصلی نخست استفاده شد. مقایسه شاخص‌های ارزیابی مورد نظر در جدول ۶ قابل مشاهده می‌باشد.

نتایج بدست آمده حاکی از توانایی بیشتر الگوریتم ژنتیک در مقایسه با روش PCA در کاهش ابعاد داده ورودی می‌باشد. با توجه به نتایج و اشکال ارائه شده کارایی الگوریتم ژنتیک به منظور افزایش دقت به طور واضح

۴- نتیجه گیری

در این پژوهش روشی برای انتخاب ویژگی‌های موثر در افزایش دقت مدل سازی کربن آلی خاک ارائه شده است. به همین منظور در گام اول مدل سازی، از روشهای رگرسیون خطی و رگرسیون خطی بردار پشتیبان و شبکه عصبی جهت ارتباط بین داده‌های سنجش از دوری و کربن آلی خاک استفاده شده است. در گام دوم، جهت کاهش ابعاد و تعیین ویژگی‌های بهینه که از اهمیت بالایی برخوردار است اعمال الگوریتم ژنتیک بر روی مدل شبکه عصبی به منظور افزایش دقت و کاهش پیچیدگی محاسبات پیشنهاد شد.

پیاده سازی روش پیشنهادی این پژوهش بر روی داده های منطقه‌ای در شمال غربی ایران و در استان آذربایجان شرقی صورت گرفت. داده های منطقه مورد مطالعه شامل تصویر ماهواره لندست ۸ مربوط به ماه مه ۲۰۱۹، ۱۰۰ نقطه مختصات دار با کربن آلی خاک معلوم توسط سازمان آب و خاک و همچنین مدل ارتفاعی SRTM منطقه است.

ارزیابی نتایج مدل سازی نشان داد که مدل شبکه عصبی بسیار دقت بالاتری نسبت به دو روش مدل سازی دیگر جهت مدل سازی کربن آلی خاک داشته که می‌توان از آن به عنوان یک روش کارآمد از آن استفاده کرد. در ادامه به تلفیق روش کاهش بعد با استفاده از الگوریتم ژنتیک با الگوریتم منتخب برای مدل سازی که همان روش شبکه عصبی است، پرداخته شد. نتایج ارزیابی ارائه شده نشان می دهد که با حذف شاخص های غیر ضروری، مدل سازی کربن آلی خاک با افزایش دقت، کاهش ابعاد ویژگی و زمان پردازش در مدل سازی روبرو شده است.

به عنوان پیشنهاد می توان در آینده به بررسی دقیق تری در مورد اینکه از کدامیک از داده‌های سنجش از دوری برای مدل سازی کربن آلی استفاده شود تا بتوان از پیچیدگی بی دلیل محاسبات و حجم بالای داده‌ها جلوگیری کرد، پرداخت.

تشکر و قدردانی

از موسسه تحقیقات خاک و آب کشور که با در اختیار قرار دادن داده های کربن آلی خاک، امکان انجام این تحقیق را مسیر ساختند، کمال تشکر و قدردانی را داریم.

منابع

- [2] R. Lal, Challenges and opportunities in soil organic matter research, *European Journal of Soil Science*, 60(2) (2009) 158-169.
- [3] F. Chen, D.E. Kissel, L.T. West, W. Adkins, Field-scale mapping of surface soil organic carbon using remotely sensed imagery, *Soil Science Society of America Journal*, 64(2) (2000) 746-753.
- [4] M. Ladoni, H.A. Bahrami, S.K. Alavipanah, A.A. Norouzi, Estimating soil organic carbon from soil reflectance: a review, *Precision Agriculture*, 11 (2010) 82-99.
- [5] X. He, L. Yang, A. Li, L. Zhang, F. Shen, Y. Cai, C. Zhou, Soil organic carbon prediction using phenological parameters and remote sensing variables generated from Sentinel-2 images, *Catena*, 205 (2021) 105442.
- [6] G.S. Bhunia, P. Kumar Shit, H.R. Pourghasemi, Soil organic carbon mapping using remote sensing techniques and multivariate regression model, *Geocarto International*, 34(2) (2019) 215-226.
- [7] H. Jenny, *Factors of soil formation: a system of quantitative pedology*, Courier Corporation, 1994.
- [8] H. Emami, Investigation the effects of aspect and degree of slope on soil quality in the South East of Mashhad, *Journal of Water and Soil Conservation*, 23(2) (2016) 301-310, (in Persian).
- [9] J. Kumhálová, F. Kumhála, M. Kroulík, Š. Matějková, The impact of topography on soil properties and yield and the effects of weather conditions, *Precision Agriculture*, 12 (2011) 813-830.
- [10] S. Singh, Understanding the role of slope aspect in shaping the vegetation attributes and soil properties in Montane ecosystems, *Tropical Ecology*, 59(3) (2018) 417-430.
- [11] S. Jakšić, J. Ninkov, S. Milić, J. Vasin, M. Živanov, D. Jakšić, V. Komlen, Influence of slope gradient and aspect on soil organic carbon content in the region of Niš, Serbia, *Sustainability*, 13(15) (2021) 8332.
- [12] P.H. Walker, Contributions to the understanding of soil and landscape relationships, *Soil Research*, 27(4) (1989) 589-605.

- [1] S. Ellis, T. Mellor, *Soils and environment*, Routledge, 2002.

- [22] H. Aïchi, Y. Fouad, C. Walter, R.V. Rossel, Z.L. Chabaane, M. Sanaa, Regional predictions of soil organic carbon content from spectral reflectance measurements, *Biosystems engineering*, 104(3) (2009) 442-446.
- [23] H. Bartholomeus, M.E. Schaepman, L. Kooistra, A. Stevens, W. Hoogmoed, O. Spaargaren, Spectral reflectance based indices for soil organic carbon quantification, *Geoderma*, 145(1-2) (2008) 28-36.
- [24] O.L. Montgomery, AN INVESTIGATION OF THE RELATIONSHIP BETWEEN SPECTRAL REFLECTANCE AND THE CHEMICAL, PHYSICAL AND GENETIC CHARACTERISTICS OF SOILS, Purdue University, 1976.
- [25] M.F. Baumgardner, L.F. Silva, L.L. Biehl, E.R. Stoner, Reflectance properties of soils, *Advances in agronomy*, 38 (1986) 1-44.
- [26] S.G. Bajwa, L. Tian, D. Bullock, K. Sudduth, N. Kitchen, H. Palm, Soil characterization in agricultural fields using hyperspectral image data, in: 2001 ASAE Annual Meeting, American Society of Agricultural and Biological Engineers, 1998, pp. 1.
- [27] R. Saxena, K. Vermal, R. Srivastava, A. Av, A. Shiwalkar, S. Londhel, Spectral reflectance properties of some dominant soils occurring on different altitudinal zones in Uttaranchal Himalayas, *Agropedology*, 13(2) (2003) 35-43.
- [28] D.G. Sullivan, J. Shaw, D. Rickman, IKONOS imagery to estimate surface soil property variability in two Alabama physiographies, *Soil Science Society of America Journal*, 69(6) (2005) 1789-1798.
- [29] G. Debaene, J. Niedzwiecki, A. Pecio, Visible and near-infrared spectrophotometer for soil analysis: preliminary results, *Polish Journal of Agronomy*, (03) (2010).
- [30] D. Žížala, R. Minařík, T. Zádorová, Soil organic carbon mapping using multispectral remote sensing data: Prediction ability of data with different spatial and spectral resolutions, *Remote Sensing*, 11(24) (2019) 2947.
- [31] T. Angelopoulou, N. Tziolas, A. Balafoutis, G. Zalidis, D. Bochtis, Remote sensing techniques for soil organic
- [13] R.C. Dalal, R.J. Henry, Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry, *Soil Science Society of America Journal*, 50(1) (1986) 120-123.
- [14] F. Pierson, D. Mulla, Aggregate stability in the Palouse region of Washington: effect of landscape position, *Soil Science Society of America Journal*, 54(5) (1990) 1407-1412.
- [15] W. Cheng, R. Virginia, S.F. Oberbauer, C.T. Gillespie, J.F. Reynolds, J. Tenhunen, Soil nitrogen, microbial biomass, and respiration along an arctic toposequence, *Soil Science Society of America Journal*, 62(3) (1998) 654-662.
- [16] C.f. Garten, W.M. Post, P.J. Hanson, L.W. Cooper, Forest soil carbon inventories and dynamics along an elevation gradient in the southern Appalachian Mountains, *Biogeochemistry*, 45 (1999) 115-145.
- [17] M. Cox, P.D. Gerard, M. Wardlaw, M. Abshire, Variability of selected soil properties and their relationships with soybean yield, *Soil Science Society of America Journal*, 67(4) (2003) 1296-1302.
- [18] K.J. Forsberg, A. Reyes, B. Wang, E.M. Selleck, M.O. Sommer, G. Dantas, The shared antibiotic resistome of soil bacteria and human pathogens, *science*, 337(6098) (2012) 1107-1111.
- [19] H. Khademi, H. Khayyer, Landscape-scale Variability of Selected Surface Soil Quality Attributes in a Rangeland in Semirom Area, *Isfahan University of Technology- Journal of Crop Production and Processing*, 8(2) (2004) 59-74, (in Persian).
- [20] H. Ghayoumi Mohammadi, M. Ramesht, N. Tomanian, M. Moayeri, Space and spatial view in soil and geomorphology studies, *Geography and Environmental Planning*, 35(3) (2009) 1-20 (in Persian).
- [21] M.-J. Ebrahimi, H. Bashari, M. Bassiri, M. Borhani, A. Mohajeri, Evaluating vegetation and soil physico-chemical characteristics changes along a grazing gradient using non-metric multi-dimensional scaling analysis (Case study: Morchehkhort rangelands-Isfahan), *Rangeland*, 11(1) (2017) 106-115, (in Persian).

- estimation using Sentinel 1 and 2 satellite images, *Journal of Geospatial Information Technology*, 7(4) (2020) 215-232, (in Persian).
- [36] S.A. Mohamed, M.M. Metwaly, M.R. Metwalli, M.A. AbdelRahman, N. Badreldin, Integrating Active and Passive Remote Sensing Data for Mapping Soil Salinity Using Machine Learning and Feature Selection Approaches in Arid Regions, *Remote Sensing*, 15(7) (2023) 1751.
- [37] J. Yuan, Z. Niu, Evaluation of atmospheric correction using FLAASH, in: 2008 International Workshop on Earth Observation and Remote Sensing Applications, IEEE, 2008, pp. 1-6.
- [38] T.R. Loveland, J.R. Irons, Landsat 8: The plans, the reality, and the legacy, *Remote Sensing of Environment*, 185 (2016) 1-6.
- carbon estimation: A review, *Remote Sensing*, 11(6) (2019) 676.
- [32] N. Pouladi, A. Gholizadeh, V. Khosravi, L. Borůvka, Digital mapping of soil organic carbon using remote sensing data: A systematic review, *Catena*, 232 (2023) 107409.
- [33] X. Zhang, Q. Sun, J. Li, Optimal band selection for high dimensional remote sensing data using genetic algorithm, in: Second International Conference on Earth Observation for Global Changes, SPIE, 2009, pp. 522-528.
- [34] F. Samadzadegan, H.S. Hasani, Determination of Optimum SVMs Based on Genetic Algorithm in Classification of Hyperspectral Imagery, *Journal of Information and Communication Technology*, 4(13) (2019) 9-23, (in Persian).
- [35] S. Ranjbar, M. Akhoondzadeh, Volumetric soil moisture

چگونه به این مقاله ارجاع دهیم

N. Mozafarinahr, H. Sadat Hasani, M. Jafari, *Dimension reduction of the remote sensing data to estimate soil organic carbon*, *Amirkabir J. Civil Eng.*, 56(5) (2024) 589-606.

DOI: [10.22060/ceej.2024.22053.7890](https://doi.org/10.22060/ceej.2024.22053.7890)



