



Print ISSN: 2251-7480
Online ISSN: 2251-7400

Journal of
Water and Soil
Resources Conservation
(WSRCJ)

Web site:

<https://wsrcj.srbiau.ac.ir>

Email:

iauwsrcj@srbiau.ac.ir
iauwsrcj@gmail.com

**Vol. 14
No. 4 (56)**

**Received:
2024-10-11**

**Accepted:
2024-12-08**

Pages: 45-56

Predicting Regional Spatial Distribution of Soil Texture Fractions in Sistan Flood Plain using Random Forest Method

Mohammad Shahriari^{*1}, Masumeh Delbari², Peyman Afrasiab³ and Mohammadreza Pahlavan-Rad⁴

- 1) Researcher, Agricultural Engineering Research Department, Agricultural and Natural Resources Research and Education Center, AREEO, Zabol, Iran.
- 2) Associate Professor of Irrigation and Drainage Engineering Department of Water Engineering Faculty of Water and Soil University of Zabol, Iran.
- 3) Associate Professor of Irrigation and Drainage Engineering Department of Water Engineering Faculty of Water and Soil University of Zabol, Iran.
- 4) Soil and Water Research Department, Golestan Agricultural and Natural Resources Research and Education Center, AREEO, Gorgan, Iran.

*Corresponding author emails: m.shahriari@areeo.ac.ir

Abstract:

Introduction: Soil properties are highly spatially variable in flood plains. Soil texture is an important physical soil property that affects many agricultural and environmental activities, so it has strongly influenced water retention curve, fertility, drainage and porosity. So, knowledge on its spatial distribution is essential especially in alluvial plains and large scale. Field studies of spatial distribution of soil properties, especially on a large scale and in floodplains, a large number of soil samples may need to be collected, which is cost- and time-consuming. Digital soil mapping (DSM) method using remote sensing data are coupled as significant impact on predicting spatial distribution soil properties.

Methods: The aim of this study was to predict the spatial distribution of soil texture fractions in Sistan flood plain at a regional scale (area 1300 km²). In this study, 160 soil samples collected under different of various soil series of the surface layer (0-30 cm) in the agriculture land of Sistan plain and soil texture fractions including percentage of sand, silt and clay content were measured. So, remotely sensed data including Landsat 8's Band 1, Band 2, Band 3, Band 4, Band 5, Band 6, Band 7, Band 8 and Band 4/ Band 8, Band 4/ Band 3, NDVI index, brightness index, clay index, grain size index were used as auxiliary variables for interpolation of soil texture fractions. Random forest technique was used to examine the relation between auxiliary variables and the soil texture components. Random forest is a developed model of classification and regression tree (CART). In the RF model, hundreds or thousands of classification trees are produced. 80 % of data was used for prediction and 20 % of data was used for validation, and RMSE, nRMSE, Willmott index (d_r), Effectiveness index (EF), MBE and MAE were used for evaluation.

Results: Pearson's correlation analysis showed that among soil texture fractions, sand content has the highest significant relationship with the most environmental variables. Band 8 had the highest correlation with sand, silt and clay content. The findings of the research show that the use of remote sensing data has increased the accuracy of predictions. The results show that the values of RMSE and MAE are lower for prediction set than validation set whereas the values of ME are similar for both sets. The values of RMSE of estimating percentage of sand, silt, clay at validation sites using RF method were 15.42, 12.56 and 8.97 %, respectively. Also, the values of RMSE of estimation by ordinary kriging were 18.2, 9.53 and 15.1% for sand, silt and clay, respectively that were 18, 5.9 and 11.2 % higher than those obtained by RF model. Also, the values of nRMSE were 0.19, 0.13 and 0.2 for prediction dataset and it was 0.39, 0.21 and 0.34 for validation dataset for sand, silt and clay fractions, respectively. The results of d_r coefficient value shows that the modeling has been done with acceptable accuracy. Also value of EF shows that spatial maps of soil texture fraction produced by using RF model has good accurate.

So, RF method when combined by remotely sensed data is a suitable method for mapping soil texture fractions in a regional scale. Also, between auxiliary variables, results showed that the clay index and grain size index were the most important environmental variables for predicting soil texture by the random forest method in the study area. The results of Willmott's coefficient of agreement (d_r) show that the modeling has been carried out with acceptable accuracy. Also, the evaluation of the efficiency factor (EF) values of the model shows that the random forest method has correctly produced the maps of soil texture components in the studied area. Other environmental variables such as Band 4 - Band 8 ratio, Band 1, Band 8 and Band 7 also influenced soil texture fractions prediction.

Conclusion: Remote sensing data combined with the random forest model can be applied for an appropriate prediction of spatial distribution pattern of soil texture fractions in large scale floodplains with a hot and dry climate condition. Highly of RMSE value for sand and silt than clay, which could be due to the wider range of silt and sand over the study region. Another reason for this could be related to the number of samples used. Therefore, it is recommended that for better accuracy in soil property maps, especially physical properties, the number of soil sampling points be increased, and optimal sampling points in these areas be determined. For future works, the use of other co-variables such as land use map, distance from the river, soil series, and salinity map or remote sensing data of smaller resolution, as well as hyperspectral visible and near-infrared reflectance spectroscopy should be evaluated for a regional spatial prediction of soil fractions in floodplains.

Keywords: Environmental variable; Random forest; Remote sensing; Soil texture; Spatial variation



شاپا چاپی: ۷۴۸۰-۲۲۵۱
شاپا الکترونیکی: ۷۴۰۰-۲۲۵۰
* ایمیل نویسنده مسئول: m.shahriari@areo.ac.ir

نشریه حفاظت منابع آب و خاک

آدرس تارنما:

<https://wsrj.srbiau.ac.ir>

پست الکترونیک:

iawwsrj@srbiau.ac.ir
iawwsrj@gmail.com

سال چهاردهم

شماره ۴ (۵۶)

تاریخ دریافت:

۱۴۰۳/۰۷/۲۰

تاریخ پذیرش:

۱۴۰۳/۰۹/۱۸

صفحات: ۵۶-۴۵

پیش‌بینی توزیع منطقه‌ای اجزاء بافت خاک دشت سیلابی سیستان با استفاده از روش جنگل تصادفی

محمد شهریاری^{۱*}، معصومه دلبری^۲، پیمان افراسیاب^۳ و محمدرضا پهلوان راد^۴

- (۱) محقق بخش تحقیقات فنی و مهندسی، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی سیستان، سازمان تحقیقات، آموزش و ترویج کشاورزی، زابل.
(۲) دانشیار گروه مهندسی آب، دانشکده آب و خاک، دانشگاه زابل.
(۳) دانشیار گروه مهندسی آب، دانشکده آب و خاک، دانشگاه زابل.
(۴) دانشیار بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی گلستان، سازمان تحقیقات، آموزش و ترویج کشاورزی، گرگان.
* ایمیل نویسنده مسئول: m.shahriari@areo.ac.ir

چکیده:

زمینه و هدف: خصوصیات خاک در دشت‌های سیلابی دارای تغییرات مکانی زیادی هستند بافت خاک یکی از خصوصیات مهم خاک است که تأثیر زیادی بر بسیاری از فعالیت‌های کشاورزی و مدیریت خاک دارد به‌طوری‌که مقدار نگهداری آب و عناصر غذایی، نفوذپذیری، زهکشی، تخلخل و بسیاری از خواص مکانیکی و هیدرولیکی خاک تحت‌تأثیر آن است؛ لذا داشتن آگاهی از توزیع مکانی آن به‌ویژه در دشت‌های سیلابی که تغییرات آن زیاد است، ضرورتی مضاعف دارد. مطالعات صحرایی تغییرات مکانی خصوصیات خاک به علت برداشت تعداد زیاد نمونه به‌ویژه در مقیاس بزرگ پرهزینه و زمان‌بر است. روش نقشه‌برداری رقومی خاک در تلفیق با داده‌های سنجش‌ازدور به‌عنوان راهکاری سودمند برای تولید نقشه رقومی خصوصیات خاک با دقت بالا و صرف هزینه و زمان کمتر است.

روش پژوهش: در این پژوهش پیش‌بینی تغییرات مکانی اجزاء بافت خاک در دشت سیلابی سیستان در مقیاس بزرگ (مساحت ۱۳۰۰ کیلومتر مربع) انجام شده است. برای این منظور بر اساس روش تصادفی طبقه‌بندی شده در سری‌های مختلف خاک اراضی کشاورزی دشت سیستان تعداد ۱۶۰ نقطه انتخاب گردید و سپس نمونه‌برداری‌ها از لایه سطحی خاک (۳-۰ cm) انجام گردید. همچنین از داده‌های سنجش از دور تصویر ماهواره لندست ۸ شامل تصاویر باندهای یک تا هشت، نسبت تصاویر باند چهارم به هفتم، باند چهارم به نهم و ششم و شاخص پوشش گیاهی، شاخص روشنایی، شاخص رس و شاخص اندازه ذرات خاک به عنوان متغیرهای کمکی برای میان‌یابی اجزاء بافت خاک کمک گرفته شد. برای یافتن ارتباط بین متغیرهای کمکی و اجزاء بافت خاک از روش جنگل تصادفی استفاده شد. این تکنیک، مدل توسعه یافته‌ای از روش طبقه‌بندی و رگرسیون درختی می‌باشد که در آن به جای رشد یک درخت صدها یا هزاران درخت طبقه‌بندی تولید می‌شود. ۸۰ درصد داده‌ها به داده‌های آموزش و ۲۰ درصد به داده‌های آزمون اختصاص یافتند. مدل‌سازی بر روی داده‌های آموزش و ارزیابی‌ها بر اساس داده‌های آزمون انجام گردید. از معیارهای RMSE, MBA, MAE و nRMSE و همچنین شاخص توافق ویلموت (df) و ضریب کارایی (EF) برای ارزیابی و کارایی مدل استفاده شد.

یافته‌ها: تجزیه و تحلیل همبستگی پیرسون نشان داد که در بین کسرهای بافت خاک، محتوای شن و ماسه با بیشترین متغیرهای محیطی دارای بیشترین رابطه معنی‌دار است. باند ۸ بیشترین همبستگی را با میزان شن، سیلت و رس داشت. یافته‌های پژوهش بیانگر این است که استفاده از داده‌های سنجش‌ازدور موجب افزایش دقت پیش‌بینی‌ها شده‌اند. همچنین باین‌حال که روش جنگل تصادفی مقادیر کمتری برای RMSE در مقایسه با یک مدل ساده کریجینگ برای پیش‌بینی مکانی ذرات شن، سیلت و رس خاک ارائه داد؛ اما به علت تغییرات زیاد خصوصیات خاک در دشت‌های سیلابی، مقادیر RMSE نسبتاً بزرگ‌تر مقادیر شن و سیلت نسبت به رس به دلیل فراگیرتر بودن آنها در منطقه مورد مطالعه است. مقدار RMSE در روش جنگل تصادفی برای پیش‌بینی شن، سیلت و رس در داده‌های اعتبارسنجی به ترتیب برابر ۱۵/۴۲، ۱۲/۵۶ و ۸/۹۷ درصد به دست آمد. این در حالی است که RMSE برای مدل کریجینگ معمولی به ترتیب ۱۸/۲، ۱۲/۳ و ۹/۵۳ برآورد شد که نسبت به روش جنگل تصادفی به ترتیب ۱، ۵/۹ و ۱۱/۲ درصد بیشتر است. مقادیر نسبتاً زیاد RMSE در این مطالعه ناشی از تغییرات زیاد خصوصیات خاک و شرایط تشکیل رسوبات در دشت‌های سیلابی است. مقادیر nRMSE برای اجزاء شن، سیلت و رس خاک به ترتیب برابر ۰/۱۹، ۰/۱۳ و ۰/۲۱ برای داده‌های پیش‌بینی و برابر ۰/۳۹، ۰/۲۹ و ۰/۳۴ برای داده‌های اعتبارسنجی است. همچنین بین متغیرهای کمکی، نتایج نشان داد که شاخص رس و شاخص اندازه دانه مهم‌ترین متغیرهای محیطی برای پیش‌بینی بافت خاک به روش جنگل تصادفی در منطقه مورد مطالعه بودند. نتایج مقادیر ضریب توافق ویلموت (df) نشان می‌دهد که مدل‌سازی با دقت قابل‌قبولی انجام شده است. همچنین بررسی مقادیر ضریب کارایی مدل (EF) نیز نشان می‌دهد که روش جنگل تصادفی به‌دراستی نقشه‌های اجزاء بافت خاک را در محدوده مورد مطالعه تولید کرده است. سایر متغیرهای محیطی مانند نسبت باند ۴-باند ۸، باند ۸، باند ۱، باند ۸ و باند ۷ نیز بر پیش‌بینی کسر بافت خاک تأثیر گذاشتند.

نتایج: نتایج نشان داد در بین اجزاء بافت خاک شن دارای بالاترین ضریب همبستگی پیرسون با متغیرهای محیطی بود و در بین متغیرهای کمکی باند ۸ بیشترین ضریب همبستگی را با اجزاء بافت خاک دارد. شاخص رس و شاخص اندازه ذرات خاک مهم‌ترین متغیرهای محیطی در جریان مدل‌سازی با روش جنگل تصادفی بودند. همچنین متغیرهای محیطی نسبت باند ۴ به باند ۸، باند ۸، باند ۷ و باند ۲ ماهواره لندست ۸ از دیگر متغیرهایی هستند که بر پیش‌بینی توزیع مکانی اجزاء بافت خاک تأثیر داشتند. در مجموع به علت تغییرات زیاد خصوصیات خاک در دشت‌های سیلابی مقادیر معیارهای ارزیابی نسبتاً زیاد برآورد گردید که این به دلیل تغییرپذیری زیاد خصوصیات فیزیکی خاک در دشت سیستان است. دلیل دیگر آن می‌تواند به تعداد نمونه‌های مورد استفاده ارتباط داشته باشد؛ بنابراین پیشنهاد می‌شود که برای بدست‌آوردن نقشه‌های با دقت بهتر برای خصوصیات خاک به‌ویژه خصوصیات فیزیکی آن در دشت‌های آبرفتی، تعداد نمونه‌برداری‌های خاک افزایش یابد و همچنین تعداد نقاط بهینه در این مناطق تعیین گردد. همچنین به دلیل اینکه خصوصیات خاک در دشت‌های سیلابی تابع نحوه رسوب‌گذاری است که خود تابع نحوه پراکنش مسیرهای جریان آب منطقه است، پیشنهاد می‌شود در مطالعات آتی از شاخص‌های که به نحوی به این موضوع مرتبط است نظیر فاصله تا رودخانه‌ها، شبکه انهار منطقه، نقشه کاربری اراضی به‌عنوان متغیرهای کمکی استفاده شود.

کلید واژه‌ها: بافت خاک؛ تغییرات مکانی؛ سنجش‌ازدور؛ جنگل تصادفی؛ متغیر محیطی

مقدمه

دانستن تغییرات مکانی خصوصیات فیزیکی و شیمیایی خاک مهم است؛ زیرا که موجب تغییرپذیری تولید می‌گردد. تغییرات مکانی خصوصیات خاک در دشت‌های سیلابی به دلیل تنوع بالای شرایط محیطی به طور قابل توجهی متغیر است و درک این تغییرات به منظور برنامه‌ریزی بهتر کشاورزی و حفاظت از محیط‌زیست ضروری است؛ لذا اهمیت دانستن تغییرات مکانی خصوصیات خاک در دشت‌های سیلابی بسیار زیاد است (Welter et al. 2008; Bui et al., 2020). بافت خاک یکی از خصوصیات مهم خاک است که تأثیر زیادی بر بسیاری از فعالیت‌های کشاورزی، مدیریت خاک و حفاظت از زیست محیط دارد. مقدار نگهداری آب و عناصر غذایی، نفوذپذیری، زهکشی، تخلخل و بسیاری از خواص مکانیکی خاک و پیش‌بینی پارامترهای هیدرولیکی خاک نظیر نقطه پژمردگی، ظرفیت زراعی و هدایت هیدرولیکی اشباع خاک تحت تأثیر بافت خاک است (Chagas et al. 2016). کمبود اطلاعات بافت خاک منجر به مدیریت نامناسب خاک و مدیریت نادرست ناشی از کمبود اطلاعات در مورد بافت خاک می‌تواند منجر به کاهش بهره‌وری کشاورزی شود (Zhao et al., 2018) لذا تولید نقشه‌های پیوسته خصوصیات خاک برای برنامه‌ریزی مدیریت کشاورزی و حفاظت از محیط‌زیست ضروری است. مطالعات صحرایی تغییرات مکانی خصوصیات خاک به علت برداشت تعداد زیاد نمونه به‌ویژه در مقیاس بزرگ پرهزینه و زمان‌بر است. استفاده از روش نقشه‌برداری رقومی خصوصیات خاک با استفاده از داده‌های سنجش‌از‌دور به‌ویژه در مناطق خشک و نیمه‌خشک به‌عنوان راهکاری سودمند برای تولید نقشه پیوسته خصوصیات خاک بادقت بالا و هزینه و زمان کمتر توصیه شده است (Chagas et al. 2016). در روش نقشه‌برداری رقومی خصوصیات خاک از تکنیک‌های مختلفی نظیر شبکه عصبی^۱، توابع انتقالی^۲، زمین‌آمار^۳ و درخت تصمیم^۴ برای یافتن ارتباط بین پارامترهایی که به‌آسانی از داده‌های سنجش‌از‌دور به دست می‌آیند، با پارامترهای اندازه‌گیری شده در نقاط نمونه‌برداری و تخمین مقدار آن در مناطق فاقد اطلاعات استفاده می‌شود (Heung et al., Mirzaee et al 2016; Pringle et al. 2010; Poggio et al., 2021; تکنیک درخت تصمیم از روش‌های جدیدی است که در سال‌های اخیر برای ارزیابی پراکنش مکانی خصوصیات خاک‌ها استفاده از داده‌های سنجش‌از‌دور به‌عنوان داده‌های ارزان‌قیمت و قابل‌دسترس در روش‌های نقشه‌برداری رقومی خاک توسعه‌یافته است درخت تصمیم نیز یک روش ناپارامتری است که قادر به پیش‌بینی متغیرهای کمی یا متغیرهای طبقه‌بندی شده بر اساس مجموعه‌ای از متغیرهای پیش‌بینی‌کننده کمی و کیفی است. در این روش، مجموعه‌ای از شرط‌های منطقی به‌صورت یک الگوریتم با ساختار درختی برای طبقه‌بندی یا پیش‌بینی کمی یک متغیر به کار می‌رود (kheir et al. 2010; Taghizadeh-Mehrjardi et al.,

2020; Jafari et al. 2012) در این زمینه مطالعاتی در مورد خصوصیات خاک وجود دارد از جمله جعفری^۵ و همکاران (2012) در منطقه خشک اقدام به پیش‌بینی کلاس‌های خاک با استفاده از روش رگرسیون درختی کردند و شاخص خیسی و همواری دره با درجه تفکیک بالا را به‌عنوان پارامترهای مهم برای مدل معرفی کردند. کاروالو جونو^۶ و همکاران (۲۰۱۴) ارزیابی منطقه‌ای نقشه‌برداری رقومی خصوصیات خاک را در یک اقلیم گرمسیری با استفاده از مدل‌های کریجینگ، رگرسیون درختی و خطی با در نظر گرفتن ۱۸ متغیر کمکی استخراج شده از DEM و تصاویر ماهواره لندست ۵ انجام دادند. یکی از روش‌های درخت تصمیم روش درخت کلاسه‌بندی و رگرسیون (CART) است. تکنیک جنگل تصادفی (RF) مدل توسعه‌یافته‌ای از روش CART است (Breiman et al., 1984).

کاروالو جونو و همکاران (۲۰۲۰) در اقلیم‌های گرمسیری از داده‌های سنجش‌از‌دور و تکنیک RF برای تهیه نقشه‌های خاک استفاده کردند و دقت بالای این روش را تأیید نمودند. آکپا^۷ و همکاران (۲۰۱۴) نقشه‌های رقومی اجزاء بافت خاک را در نیجریه با استفاده از روش جنگل تصادفی تولید کردند و نتیجه گرفتند که نقشه‌های تولید شده توسط این مدل از دقت خوبی برخوردار است. ویسه و لجاچری^۸ (۲۰۱۵) پیش‌بینی اجزاء بافت خاک را با استفاده از روش جنگل تصادفی انجام دادند و بیان کردند این مدل توانایی خوبی برای تولید نقشه‌های مکانی بافت خاک دارد. لایب^۹ و همکاران (۲۰۱۲) مدل‌های RF و رگرسیون درختی (RT) را برای پیش‌بینی توزیع بافت خاک استفاده کردند و مشاهده نمودند که روش RF دقت بالاتری نسبت به مدل RT است. چاکاز^{۱۰} و همکاران (۲۰۱۵) تکنیک‌های RF و رگرسیون خطی چندگانه (MLR) را برای پیش‌بینی مکانی بافت خاک در یک منطقه خشک بکار بردند. در این پژوهش نتایج نشان داد که روش RF نقشه‌های بافت خاک را واقعی‌تر نسبت به روش MLR تولید می‌کند. گامبیل^{۱۱} و همکاران (۲۰۱۶) پیش‌بینی طبقه‌بندی بافت خاک به روش یونیفاید (USCS) را با استفاده از روش جنگل تصادفی انجام دادند که در این تحقیق متغیرهای کمکی لحاظ شده شامل بافت خاک، درصد مواد آلی خاک، آب قابل‌استفاده خاک بودند و در نتیجه خطای پیش‌بینی با این روش به کمتر از دو درصد رسید.

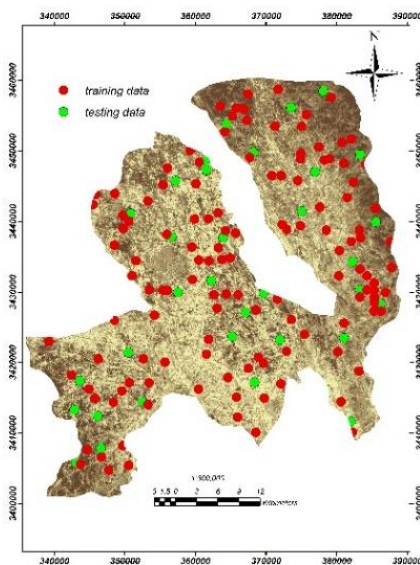
مروری بر مطالعات پیشین نبود یا کمبود مطالعات کاربرد مدل جنگل تصادفی را در برآورد خصوصیات فیزیکی خاک در دشت‌های سیلابی و مقیاس منطقه‌ای نشان می‌دهد. بنابراین هدف از این تحقیق پیش‌بینی تغییرات مکانی منطقه‌ای اجزاء بافت خاک در منطقه نسبتاً مسطح و آبرفتی اراضی کشاورزی دشت سیستان با استفاده از روش جنگل تصادفی می‌باشد.

مواد و روش‌ها

منطقه مورد مطالعه

محدوده مورد مطالعه در شمال استان سیستان و بلوچستان در اراضی کشاورزی منطقه سیستان در محدوده جغرافیایی ۶۰/۲۵ درجه تا ۶۱/۸۳ درجه طول شرقی و ۳۰/۰۸ درجه تا ۳۱/۴۶ درجه عرض شمالی و مساحت حدود ۱۳۰۰ کیلومتر مربع قرار دارد. بر اساس طبقه بندی کوپن منطقه دارای آب و هوای گرم و خشک با میانگین دراز مدت سالیانه بارش، درجه حرارت و تبخیر و تعرق پتانسیل به ترتیب ۵۵ میلی‌متر، ۲۱/۷ سانتی‌گراد و ۴۵۰۰ میلی‌متر می‌باشد. کاربری اراضی منطقه زمین‌های بایر، مرتع، جنگل کاری، زمین‌های کشاورزی، مجموعه- های آبی، شن‌زار است. دشت مسطح بوده و از لحاظ ژئومورفولوژی دشتی آبرفتی متأثر از رسوبات انشعابات رودخانه هیرمند می‌باشد. در این تحقیق برای پیش‌بینی الگوی منطقه‌ای تغییرات مکانی ذرات شن، سیلت و رس منطقه مورد مطالعه، براساس روش تصادفی طبقه‌بندی شده در سری‌های مختلف خاک تعداد ۱۶۰ نقطه انتخاب گردید و سپس نمونه‌برداری‌ها از لایه سطحی خاک (۰-۳۰ cm) انجام شد. موقعیت نقاط نمونه‌برداری در شکل (۱) نشان داده شده است. نمونه‌های خاک برای تعیین درصد توزیع اندازه ذرات خاک به آزمایشگاه منتقل و اندازه‌گیری درصد توزیع ذرات خاک با استفاده از روش هیدرومتری و مطابق با استاندارد بین‌المللی خاکشناسی انجام گردید. همچنین در این تحقیق برای پیش‌بینی الگوی منطقه‌ای تغییرات مکانی اجزاء بافت خاک از داده‌های سنجش‌ازدور به عنوان داده‌های کمکی (متغیرهای محیطی) استفاده شد در این مطالعه از ارزش رقومی باندهای مختلف تصویر ماهواره لندست ۸ مورخه ۱۵ دسامبر ۲۰۱۵ شامل باندهای ۱ تا ۸ و نسبت باند ۴ به باند ۸، نسبت باند ۴ به باند ۳ و شاخص پوشش گیاهی^{۱۲}

(NDVI)، شاخص روشنایی^{۱۳}، شاخص رس^{۱۴} و شاخص اندازه ذرات خاک^{۱۵} (GSI) استفاده شد. (جدول ۱) به این طریق که پس از مشخص کردن موقعیت نقاط نمونه‌برداری بر روی تصاویر، ارزش رقومی^{۱۶} (DN) هر پیکسل استخراج و به‌عنوان متغیر مستقل با استفاده از نرم افزارهای Arc Gis 10.3 و ENVI 5.3 استخراج و از داده‌های اندازه‌گیری شده اجزاء بافت خاک به عنوان متغیر وابسته استفاده شد. تصاویر مورد استفاده بدون اشکالات اتمسفریک و تابش‌سنجی از قبیل اثر سایه، تیرگی تصویر، نوارشدگی و غیره بود و منطقه نیز چون دشتی و مسطح است بنابراین تصحیحات اتمسفریک و رادیومتریکی و توپوگرافی نیاز نداشت.



شکل ۱. موقعیت نقاط نمونه‌برداری شده

جدول ۱. متغیرهای محیطی مورد استفاده برای پیش‌بینی اجزاء بافت خاک

| منبع | تعریف | متغیرهای محیطی |
|-------------------------|----------------------------------|---|
| | coastal | Band 1 |
| | Blue Band | Band 2 |
| | Green Band | Band 3 |
| | Red Band | Band 4 |
| | NIR Band | Band 5 |
| | SWIR Band | Band 6 |
| | SWIR Band | Band 7 |
| | Panchromatic Band | Band 8 |
| | Ratio Red and Green Bands | Band 4 / Band 3 |
| | Ratio Red and Panchromatic Bands | Band 4 / Band 8 |
| | $((Red)2+(NIR)^2)^{0.5}$ | Brightness index |
| Xiao et al (2006) | $(Red-Blue) / (Red+Blue+Green)$ | Grain Size Index (GSI) |
| Boettinger et al (2008) | $(NIR-Red) / (NIR+Red)$ | Normalized Difference Vegetation Index (NDVI) |
| Boettinger et al (2008) | Ratio Band 6 and Band 8 | Clay Index |

مدل‌سازی جنگل تصادفی

جنگل تصادفی (RF) یک روش جدید و قدرتمند است که پیشرفت‌های قابل‌ملاحظه‌ای را در تکنولوژی داده‌کاوی ارائه داده است. تکنیک جنگل تصادفی (RF) مدل توسعه‌یافته‌ای از روش طبقه‌بندی و رگرسیون درختی CART است. در RF به‌جای رشد یک درخت صدها یا هزار درخت طبقه‌بندی تولید می‌شود (Breiman and Cutler, 2004). در این تحقیق، روش جنگل تصادفی با استفاده از داده‌های ذرات نسبی بافت خاک سطحی در نرم‌افزار (۲۰۱۳) R 3.0.1 با به‌کارگیری بسته نرم‌افزاری randomforest آموزش داده شد.

اساس کار روش جنگل تصادفی ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه خود راه‌انداز از داده‌ها شرکت دارد و در ساخت هر درخت به‌طور تصادفی تعدادی از متغیرهای ورودی شرکت می‌کنند؛ لذا با تکرار عملیات نمونه‌گیری، تعدادی مجموعه داده OOB از مجموعه آموزشی به وجود می‌آید که می‌توان برای هر مجموعه داده آموزشی یک درخت تصمیم محاسبه نمود. (Liaw and Wiener, 2002) در روش جنگل تصادفی اگر فرض شود تعداد داده‌هایی که به‌عنوان آموزش در ساخت مدل حضور دارند، N باشد و تعداد کل متغیرهای پیش‌بینی‌کننده موجود M و تعداد کل درخت‌های حاضر در مدل n_{tree} فرض شود، برای ساخت درخت i ام (i=1, ..., n_{tree}) کافی است الگوریتم زیر دنبال گردد:

۱- ابتدا نمونه اصلی به دو قسمت نمونه آموزش و نمونه امتحان تقسیم می‌شود. سپس از نمونه آموزش نمونه‌گیری مجددی به حجم N گرفته می‌شود و یک سوم از نمونه آموزشی جدید، به عنوان نمونه OOB از مجموعه داده‌های آموزشی جدا می‌گردد. در واقع OOB در هر درخت، به نوعی نقش نمونه امتحانی را برای آن درخت ایفا می‌کند.

۲- در مرحله بعد تعداد m_{try} به تصادف از بین M متغیر توضیحی انتخاب می‌گردد (m_{try} < M) معمولاً تعداد متغیرها بایستی توسط کاربر انتخاب گردد. وقتی که درخت ساخته شد، داده OOB در درخت ساخته شده قرار داده می‌شود و میزان خطا محاسبه می‌گردد. از آنجا که تعداد درخت‌های ساخته شده بر روی نمونه‌های خودراه‌انداز زیاد است این میزان خطا ناریب می‌شود. از داده‌های OOB برای برآورد ناریب خطا و برآورد متغیر با اهمیت استفاده می‌گردد. بنابراین در الگوریتم جنگل تصادفی در نرم افزار R بایستی دو متغیر تعریف شود یکی تعداد درختان موجود (n_{tree}) در جنگل تصادفی و دیگری تعداد متغیرهای مورد استفاده (m_{try}) در هر

درخت که بر اساس کمترین مقدار OOB مشاهده شده بصورت سعی و خطا بدست می‌آید. مقدار خطای نمونه‌های OOB و همچنین درصد واریانس کل مدل برای هر متغیر با استفاده از رابطه (۱) و (۲) برآورد می‌شود (Breiman and Cutler, 2004).

$$MSE_{OOB} = \frac{1}{n} \sum_{i=1}^n (Z_i - Z_i^{oob})^2 \quad (1)$$

$$Var_{exp} = 1 - (MSE_{OOB} / Var_z) \quad (2)$$

Z_i مقادیر اندازه‌گیری شده متغیر و Z_i^{oob} متوسط خطای مشاهده شده برای داده‌های OOB و Var_{exp} درصد واریانس کل مدل و Var_z واریانس متغیر است.

معیارهای ارزیابی

در این تحقیق برای بررسی دقت و ارزیابی مدل، ابتدا داده‌ها به دودسته داده‌های آموزش و داده‌های تست تقسیم شدند، سپس ۸۰ درصد داده‌ها به داده‌های آموزش و ۲۰ درصد به داده‌های آزمون اختصاص یافتند. مدل‌سازی بر روی داده‌های آموزش انجام شد و ارزیابی‌ها داده‌های آزمون بر اساس ریشه دوم میانگین مربعات خطا (RMSE)، ریشه دوم میانگین مربعات خطای نرمال شده (nRMSE)، شاخص توافق ویلموت (dr) (Willmott et al., 2011) و ضریب کارایی (EF) (Chen et al., 2003) و همچنین شاخص‌های MBE و MAE انجام شد:

$$d_r = 1 - \frac{\sum_{i=1}^n |P_i - O_i|}{c \sum_{i=1}^n |O_i - \bar{O}|} \quad (1)$$

$$EF = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (2)$$

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

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - O_i| \quad (4)$$

$$nRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}}{\bar{O}} \quad (5)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (6)$$

که O_i مقدار اندازه‌گیری شده و P_i مقدار پیش‌بینی شده اجزاء بافت خاک توسط مدل است. همچنین \bar{O} میانگین مقدار اندازه‌گیری شده هستند.

نتایج و بحث
تحلیل‌های آماری

نرمال بودن توزیع فراوانی داده‌ها با استفاده از آزمون معنی‌داری چولگی انجام گردید و سپس متغیرهایی که از چولگی معنی‌دار و توزیع فراوانی غیرنرمال برخوردار بودند با استفاده از تابع لگاریتمی یا توابع مناسب دیگر نرمال شدند. خلاصه اطلاعات آماری درصد ذرات بافت خاک شامل شن، سیلت و رس در جدول (۲) ارائه شده است. همچنین خلاصه آماری داده‌های متغیرهای محیطی مورد استفاده در جدول (۳) نشان داده شده است. نتایج بیانگر این است که تفاوت غیرمعنی‌داری در داده‌های درصد اجزاء بافت خاک وجود ندارد. شن و رس از تغییرات زیادی در منطقه برخوردار بودند به طوری که ضریب تغییرات آنها به ترتیب ۵۳ و ۵۱ درصد برآورد شده است. همچنین ضریب تغییرات سیلت برابر ۳۵ درصد است که بیانگر ناهمگنی بافت خاک منطقه است. میانگین مقدار شن و سیلت خاک ۳۷ و ۴۳ درصد و میانگین مقدار رس خاک ۲۰ درصد است که بیانگر بافتی متوسط در منطقه است. مشابه درصد اجزاء بافت

خاک متغیرهای محیطی مورد استفاده نیز مورد تحلیل‌های آماری در سطح ۹۵ درصد قرار گرفت که نتایج بیانگر این است تفاوت معنی‌داری در داده‌ها وجود ندارد. همه متغیرهای محیطی به جز شاخص پوشش گیاهی (NDVI) با ۳۲ درصد و شاخص اندازه ذرات خاک (GSI) با ۱۱۳ درصد، دارای ضریب تغییرات اندکی هستند که بیانگر همگن بودن متغیرهای محیطی مورد استفاده در منطقه است.

رابطه بین متغیرهای محیطی و اجزاء بافت خاک

ضریب همبستگی پیرسون بین درصد اجزاء بافت خاک شامل شن، سیلت و رس و متغیرهای محیطی برآورد شده و در جدول (۴) نشان داده شده است. شن خاک با همه متغیرهای محیطی به جز نسبت باند ۴ به باند ۸ رابطه معنی‌داری دارد. بیشترین و کمترین همبستگی داده‌های شن خاک مربوط به باند ۸ با ۰/۳۳۸- و NDVI با ۰/۲۱ است. چاگاز و همکاران در سال ۲۰۱۶ رابطه همبستگی قوی بین شن خاک و نسبت باند ۳ به باند ۷ و همبستگی متوسط

جدول ۲. خلاصه آماری داده‌های پیش‌بینی و اعتبارسنجی اجزاء بافت خاک

| خصوصیات خاک | داده‌های پیش‌بینی | | | | | داده‌های اعتبارسنجی | | | | |
|-------------|-------------------|-----|------|----|----|---------------------|-----|------|----|----|
| | Min | Max | Mean | SD | CV | Min | Max | Mean | SD | CV |
| | gr/kg | | | | | % | | | | |
| clay | ۴ | ۴۸ | ۲۰ | ۱۰ | ۵۱ | ۶ | ۴۰ | ۲۱ | ۹ | ۴۵ |
| silt | ۲ | ۷۹ | ۴۳ | ۱۵ | ۳۵ | ۷ | ۷۶ | ۴۳ | ۱۴ | ۳۳ |
| sand | ۶ | ۹۱ | ۳۸ | ۲۰ | ۵۳ | ۱۰ | ۸۲ | ۳۶ | ۱۹ | ۵۲ |

جدول ۳. خلاصه آماری داده‌های پیش‌بینی و اعتبارسنجی متغیرهای محیطی

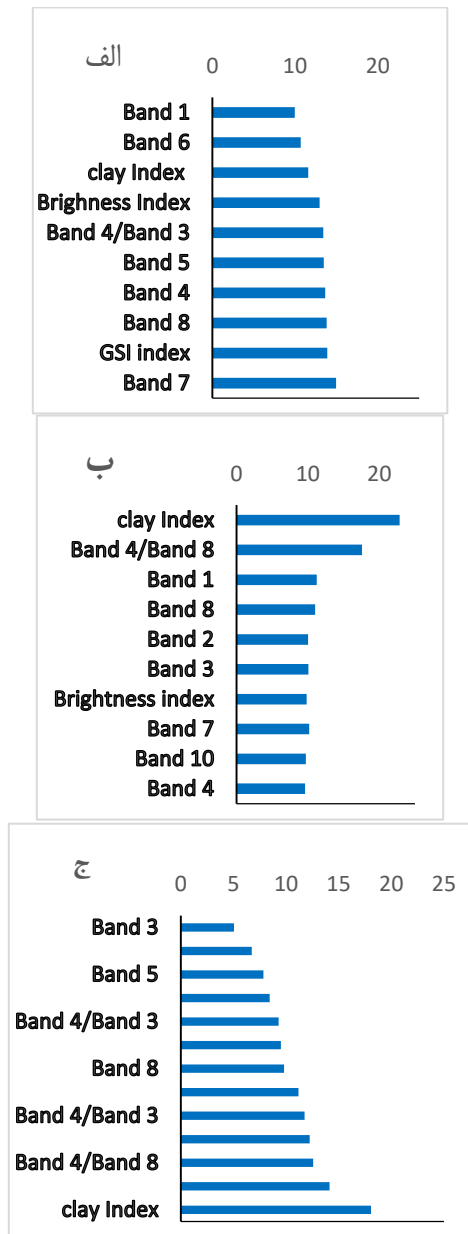
| متغیرهای محیطی | داده‌های پیش‌بینی | | | داده‌های اعتبارسنجی | | |
|----------------|-------------------|------|-----|---------------------|------|-----|
| | Mean | SD | CV | Mean | SD | CV |
| Band 1 | ۷۹ | ۸ | ۱۰ | ۸۰ | ۸ | ۱۰ |
| Band 2 | ۸۰ | ۱۰ | ۱۲ | ۸۱ | ۱۰ | ۱۳ |
| Band 3 | ۷۹ | ۱۲ | ۱۵ | ۸۰ | ۱۳ | ۱۶ |
| Band 4 | ۷۷ | ۱۳ | ۱۷ | ۷۸ | ۱۴ | ۱۸ |
| Band 5 | ۵۹ | ۸ | ۱۴ | ۵۹ | ۹ | ۱۵ |
| Band 6 | ۱۶ | ۲ | ۱۴ | ۱۶ | ۲ | ۱۵ |
| Band 7 | ۵ | ۱ | ۱۶ | ۵ | ۱ | ۱۷ |
| Band 8 | ۷۹ | ۱۳ | ۱۷ | ۸۰ | ۱۳ | ۱۷ |
| Band 4 /Band 3 | ۰/۹۶ | ۰/۰۳ | ۳ | ۰/۹۶ | ۰/۰۳ | ۳ |
| Band 4 /Band 8 | ۰/۹۷ | ۰/۰۳ | ۳ | ۰/۹۷ | ۰/۰۵ | ۵ |
| Brightness | ۹۷ | ۱۵ | ۱۶ | ۹۷ | ۱۶ | ۱۶ |
| GSI | -۰/۰۲ | ۰/۰۲ | ۱۱۳ | -۰/۰۲ | ۰/۰۲ | ۱۲۵ |
| NDVI | -۰/۱۳ | ۰/۰۴ | ۳۲ | -۰/۱۴ | ۰/۰۴ | ۳۳ |
| clay index | ۰/۲ | ۰/۰۱ | ۶ | ۰/۲ | ۰/۰۱ | ۶ |

است. در این تحقیق بیشترین تعداد ضرایب همبستگی غیرمعنی‌دار مربوط به رس است به‌طوری‌که برای باند ۷ مقدار ۰/۱۵ و برای نسبت باند ۴ به باند ۸ برابر ۰/۱، برای باند ۶ و برابر ۰/۱۹۶، برای NDVI برابر ۰/۱۶۵- برآورد شده است. نتایج نشان می‌دهد که سایر متغیرهای محیطی نیز از همبستگی ضعیفی با رس برخوردار هستند. بالاترین ضریب همبستگی به مقدار ۰/۲۶ مربوط به باند ۸ و کمترین آن مربوط به شاخص NDVI برابر ۰/۱۶ برآورد شد. در مجموع رس کمترین همبستگی را با متغیرهای محیطی در مقایسه با سیلت و شن داشت. (جدول ۴) چاگاز و همکاران (۲۰۱۶) رابطه همبستگی متغیرهای محیطی با رس را مشابه شن گزارش دادند. کاروالو جونیور و همکاران (۲۰۱۴) رابطه معنی‌داری بین رس و متغیرهای نسبت باند ۵ به باند ۷، نسبت باند ۳ به باند ۲ و شاخص NDVI گزارش کردند و رابطه غیرمعنی‌داری با نسبت باند ۳ به باند ۷ تصاویر ماهواره لندست ۵ گزارش دادند. همچنین لیائو و همکاران (۲۰۱۳) گزارش دادند که رس خاک با داده‌های باند ۱ تا ۵ و باند ۵ و باند ۸ و با بقیه شاخص روشنایی رابطه همبستگی متوسطی دارد و با بقیه پارامترهای محیطی رابطه همبستگی ضعیفی دارد. (جدول ۴) همچنین سیلت با نسبت باند ۴ به باند ۸ و باند ۷ رابطه معنی‌داری ندارد. چاگاز و همکاران (۲۰۱۶) رابطه همبستگی غیرمعنی‌دار بین داده‌های سیلت و نسبت باند ۵ به باند ۷ و باندهای ۱ و ۲ حاصل شده از تصاویر ماهواره لندست ۷ گزارش کردند. سوزا جونیور^{۱۸} و همکاران (۲۰۱۱) همبستگی بالایی بین سیلت و باندهای (۱ تا ۸) تصاویر ماهواره ASTER گزارش دادند. همچنین همبستگی پایین سیلت با متغیرهای مطالعه شده توسط ایسلام^{۱۹} و همکاران (۲۰۰۳) و وترلیند و استنبرگ^{۲۰} (۲۰۱۰) گزارش شده است. لیائو و همکاران (۲۰۱۳) گزارش دادند که سیلت خاک با باندهای ۲ تا ۴ و باند ۷ تصاویر ماهواره لندست ۷ رابطه همبستگی معنی‌دار منفی دارد به‌طوری‌که بیشترین مقدار آن با باند ۷ و به مقدار ۰/۴۵۳-

با GSI و NDVI تصاویر لندست ۷ گزارش کردند. همچنین لیائو^{۱۷} و همکاران در سال ۲۰۱۳ گزارش دادند که شن خاک با باندهای ۲ تا ۵ و باند ۷ تصاویر ماهواره لندست ۷ رابطه همبستگی معنی‌دار مثبت دارد به‌طوری‌که بیشترین مقدار آن با باند ۷ و به مقدار ۰/۵۶۸ است. از طرفی کاروالو جونیو و همکاران در سال ۲۰۱۴ در مطالعه خود یافتند که شن خاک با نسبت باند ۳ به باند ۷، نسبت باند ۵ به باند ۷ و شاخص NDVI به‌دست‌آمده از تصاویر ماهواره لندست ۵ رابطه معنی‌داری ندارد. به‌طورکلی نتایج بیانگر این است که ضریب همبستگی شن با متغیرهای محیطی مورد استفاده از همبستگی بالاتری نسبت به سیلت و رس برخوردار است. نتایج نشان داد که سیلت خاک با متغیرهای محیطی باندهای ۱ تا ۵ و باند ۸ و شاخص روشنایی رابطه همبستگی متوسطی دارد و با بقیه پارامترهای محیطی رابطه همبستگی ضعیفی دارد. (جدول ۴) همچنین سیلت با نسبت باند ۴ به باند ۸ و باند ۷ رابطه معنی‌داری ندارد. چاگاز و همکاران (۲۰۱۶) رابطه همبستگی غیرمعنی‌دار بین داده‌های سیلت و نسبت باند ۵ به باند ۷ و باندهای ۱ و ۲ حاصل شده از تصاویر ماهواره لندست ۷ گزارش کردند. سوزا جونیور^{۱۸} و همکاران (۲۰۱۱) همبستگی بالایی بین سیلت و باندهای (۱ تا ۸) تصاویر ماهواره ASTER گزارش دادند. همچنین همبستگی پایین سیلت با متغیرهای مطالعه شده توسط ایسلام^{۱۹} و همکاران (۲۰۰۳) و وترلیند و استنبرگ^{۲۰} (۲۰۱۰) گزارش شده است. لیائو و همکاران (۲۰۱۳) گزارش دادند که سیلت خاک با باندهای ۲ تا ۴ و باند ۷ تصاویر ماهواره لندست ۷ رابطه همبستگی معنی‌دار منفی دارد به‌طوری‌که بیشترین مقدار آن با باند ۷ و به مقدار ۰/۴۵۳-

جدول ۴. ضریب همبستگی بین اجزاء بافت خاک و متغیرهای محیطی

| متغیرهای محیطی | اجزاء بافت خاک | | | | | |
|------------------|----------------|--------|-----------|-------|-----------|--------|
| | clay | | silt | | sand | |
| | p-values: | r | p-values: | r | p-values: | r |
| Band 1 | ۰/۰۰۳ | ۰/۲۳۲ | . | ۰/۲۷۵ | < ۰/۰۰۰۱ | -۰/۳۲ |
| Band 2 | ۰/۰۰۳ | ۰/۲۳۷ | . | ۰/۲۷۷ | < ۰/۰۰۰۱ | -۰/۳۲۳ |
| Band 3 | ۰/۰۰۲ | ۰/۲۴۳ | . | ۰/۲۷۲ | < ۰/۰۰۰۱ | -۰/۳۲۳ |
| Band 4 | ۰/۰۰۱ | ۰/۲۴۹ | . | ۰/۲۷۳ | < ۰/۰۰۰۱ | -۰/۳۲۶ |
| Band 5 | ۰/۰۰۷ | ۰/۲۱۰ | ۰/۰۰۳ | ۰/۲۲۹ | . | -۰/۲۷۷ |
| Band 6 | ۰/۰۱۳ | ۰/۱۹۶ | /۰۰۸ | ۰/۲۰۸ | ۰/۰۰۱ | -۰/۲۵۴ |
| Band 7 | ۰/۰۵۱ | ۰/۱۵۴ | ۰/۰۲۶ | ۰/۱۷۶ | ۰/۰۰۷ | -۰/۲۱۱ |
| Band 8 | ۰/۰۰۱ | ۰/۲۵۸ | . | ۰/۲۸۳ | < ۰/۰۰۰۱ | -۰/۳۳۸ |
| Band 4 / Band 3 | ۰/۰۱۵ | ۰/۱۹۲ | ۰/۰۲۵ | ۰/۱۷۶ | ۰/۰۰۴ | -۰/۲۲۴ |
| Band 4 / Band 8 | ۰/۲۰۴ | ۰/۱۰۱ | ۰/۲۵ | ۰/۰۹۱ | ۰/۱۵۷ | -۰/۱۱۲ |
| Brightness index | ۰/۰۰۲ | ۰/۲۴ | ۰/۰۰۱ | ۰/۲۶۲ | < ۰/۰۰۰۱ | -۰/۳۱۵ |
| GSI | ۰/۰۰۵ | ۰/۲۱۹ | ۰/۰۰۹ | ۰/۲۰۷ | ۰/۰۰۱ | -۰/۲۶۳ |
| NDVI | ۰/۰۳۷ | -۰/۱۶۵ | ۰/۰۲۵ | ۰/۱۷۶ | ۰/۰۰۸ | ۰/۲۰۹ |
| clay index | ۰/۰۱ | -۰/۲۰۳ | ۰/۰۰۹ | ۰/۲۰۶ | ۰/۰۰۱ | ۰/۲۵۲ |



شکل ۲. اهمیت متغیرها در روش جنگل تصادفی برای پیش‌بینی رس (الف)، سیلت (ب) و شن (ج)

نتایج ارزیابی روش جنگل تصادفی

نتایج ارزیابی روش جنگل تصادفی (جدول ۶) نشان می‌دهد که مقادیر RMSE و MAE داده‌های پیش‌بینی کمتر از داده‌های اعتبار-سنجی است. مقادیر RMSE برای شن و سیلت زیاد به دست آمد به دلیل دامنه بزرگتر ضریب تغییرات این دو خصوصیت در منطقه مورد مطالعه است. مقادیر نسبتاً زیاد RMSE در این مطالعه ناشی از تغییرات زیاد خصوصیات خاک و شرایط تشکیل رسوبات در دشت‌های سیلابی است (Walder et al., 2008; Shahriari et al, 2019). پهلوان راد و همکاران (۲۰۲۰) نیز مقادیر بالای RMSE را برای پیش‌بینی خصوصیات خاک در دشت سیستان گزارش کردند. تفاوت RMSE بین داده‌های پیش‌بینی و اعتبارسنجی نشان می‌دهد

تعیین پارامترهای بهینه مدل جنگل تصادفی

نتایج اجرای مدل جنگل تصادفی برای تعیین پارامترهای بهینه مدل در جدول (۵) ارائه شده است نتایج نشان می‌دهد که درصدهای واریانس کل مدل برای شن، سیلت و رس خاک به ترتیب برابر ۲۶/۳، ۱۶/۸۳ و ۹/۳۲ است. لایپ و همکاران (۲۰۱۳) توزیع ذرات نسبی خاک را با استفاده از روش جنگل تصادفی بررسی کردند و واریانس مدل برای شن خاک ۳۰ درصد، سیلت خاک ۴۳ درصد و رس خاک برابر ۲۶ درصد برآورد کردند. چاگاز و همکاران (۲۰۱۶) پیش‌بینی بافت خاک سطحی را با روش جنگل تصادفی انجام دادند و نتیجه گرفتند که درصد واریانس کل مدل برای شن خاک ۴۷/۶۵ و برای سیلت خاک برابر ۸/۶۱ و برای رس خاک برابر ۴۸/۹۴ درصد است. آکپا و همکاران (۲۰۱۴) گزارش دادند که واریانس مدل برای شن، ۴۸-۴۹ درصد، سیلت، ۲۶-۲۷ درصد و برای رس ۵۳-۵۶ درصد در لایه سطحی خاک (0-15 cm) است. ویسه و لجاچری (۲۰۱۵) در مطالعه خود گزارش دادند که درصد واریانس مدل برای شن بین ۳۳ تا ۳۵ درصد، سیلت بین ۲۳ تا ۲۹ درصد و برای رس بین ۳۱ تا ۳۵ درصد می‌باشد. نتایج پژوهش حاضر نشان داد که mtry برای هر خصوصیات مورد مطالعه ۵ تا ۱۱ و ntree 500 تا ۸۰۰ می‌باشد. در مطالعات مشابه مقادیر mtry و ntree را به ترتیب ۵ و ۳۵۰ برای پیش‌بینی سری‌ها و خصوصیات خاک گزارش شده است (Camera et al. 2017).

جدول ۵. خلاصه پارامترهای بهینه مدل

| properties | m _{try} | n _{tree} | Var _{exp} |
|------------|------------------|-------------------|--------------------|
| sand | ۱۱ | ۸۰۰ | ۲۶/۳ |
| silt | ۱۰ | ۷۰۰ | ۱۶/۸۳ |
| clay | ۵ | ۵۰۰ | ۹/۳۲ |

نتایج اهمیت متغیرهای محیطی

اهمیت متغیرها در روش جنگل تصادفی برای شن، سیلت و رس خاک در شکل ۲ نشان داده شده است اهمیت متغیرهای کمکی که برای پیش‌بینی درصد ذرات نسبی بافت خاک در مدل منظور شده است با لحاظ حد آستانه ۱۰ درصد به شرح زیر می‌باشد:

برای شن خاک > Clay Index > Band 4 / Band 8 > Band 1 > Band 8 / Band 7 > Band 2 > band 3
 برای سیلت خاک > Clay Index > GSI > Band 4/ Band 8 > Band 1 > Band 4 / Band 3 > Band 7
 برای رس خاک > band 7 > GSI > Band 8 > Band 4 > Band 5 / Band 4 / Band 3 > Brightness index > band 6
 چاگاز و همکاران (۲۰۱۶) و آکپا و همکاران (۲۰۱) در مطالعات خودشان نیز نتایج مشابهی گزارش کرده‌اند.

نقشه‌های پیش‌بینی بافت خاک

نقشه پراکنش مکانی شن، سیلت و رس خاک در شکل (۳) نشان داده شده است. نقشه رقومی شن نشان می‌دهد که بیشتر اراضی منطقه مورد مطالعه بین ۲۴ تا ۵۶ درصد شن دارند. ۳۷ درصد اراضی دارای شن ۲۴-۳۵ درصد، ۳۱ درصد اراضی دارای شن ۳۵-۴۵ درصد و ۸/۵ درصد اراضی هم دارای شن ۴۶-۵۶ درصد هستند. توزیع مکانی شن خاک در شکل ۳ (الف) نشان می‌دهد که بیشترین مقدار شن در نیمه مرکزی و جنوب‌شرقی منطقه مورد مطالعه دیده می‌شود. این اراضی به دلیل قرارگرفتن در حاشیه رودخانه سیستان و رسوب ذرات درشت‌تر شن در آنها سبب شده مقدار شن در آنها بیشتر باشد. شکل ۳ (ب) نقشه توزیع مکانی رس خاک را نشان می‌دهد. مقدار رس خاک‌های اراضی مورد مطالعه پایین است به‌طوری‌که ۵۶/۴ درصد اراضی بین ۱۶ تا ۲۴ درصد و ۳۴/۳ درصد اراضی ۸ تا ۱۶ درصد رس دارند. بیشترین مقدار رس خاک در اراضی نیمه شمال شرقی منطقه مورد مطالعه مشاهده می‌شود. قسمتی از اراضی این بخش دارای رس ۳۰ تا ۴۰ درصد هستند. پراکنش مکانی سیلت خاک در شکل ۳ (ج) نشان داده شده است. حدود ۵۱/۱ درصد از سطح مورد مطالعه دارای سیلت ۴۰-۵۰ درصد و ۳۰/۴ درصد منطقه دارای سیلت ۴۰-۳۰ است. بر اساس نتایج بدست آمده در منطقه مورد مطالعه اکثر اراضی (شکل ۳ د) دارای بافت‌های لوم شنی (نسبتاً درشت بافت)، لوم (بافت متوسط) و لوم شنی (بافت متوسط) هستند همچنین در قسمت‌هایی که مقدار رس زیاد است بافت‌های لوم رسی شنی و لوم رسی نیز مشاهده می‌شود. به دلیل سیلابی بودن، و وجود فرسایش بادی خصوصیات خاک در محدوده مورد مطالعه، نقشه‌های اجزاء بافت خاک دارای تغییرپذیری مکانی زیادی هستند.

نتیجه‌گیری

در این مطالعه پیش‌بینی تغییرات مکانی اجزاء بافت خاک شامل شن، سیلت و رس در مقیاس منطقه‌ای در یک اقلیم گرم و خشک و دشت سیلابی با استفاده از روش جنگل تصادفی و در نظر گرفتن داده‌های سنجش‌ازدور به‌عنوان متغیر کمکی انجام شد. نتایج بیانگر این است که استفاده از داده‌های سنجش‌ازدور باوجود داشتن رابطه همبستگی متوسط و ضعیف با ذرات نسبی خاک (درصد شن، سیلت و رس) موجب افزایش دقت پیش‌بینی‌ها می‌شود. تحلیل همبستگی نشان داد که شن خاک باارزش رقومی باندهای ۲ تا ۸، شاخص روشنایی، شاخص اندازه ذرات خاک و نسبت باند ۴ به باند ۳ تصویر ماهواره لندست ۸ رابطه همبستگی منفی و با شاخص پوشش گیاهی و شاخص رس رابطه همبستگی مثبت دارد. همچنین سیلت خاک با ارزش رقومی باندهای ۱ تا ۶ باند ۸، شاخص پوشش گیاهی، شاخص

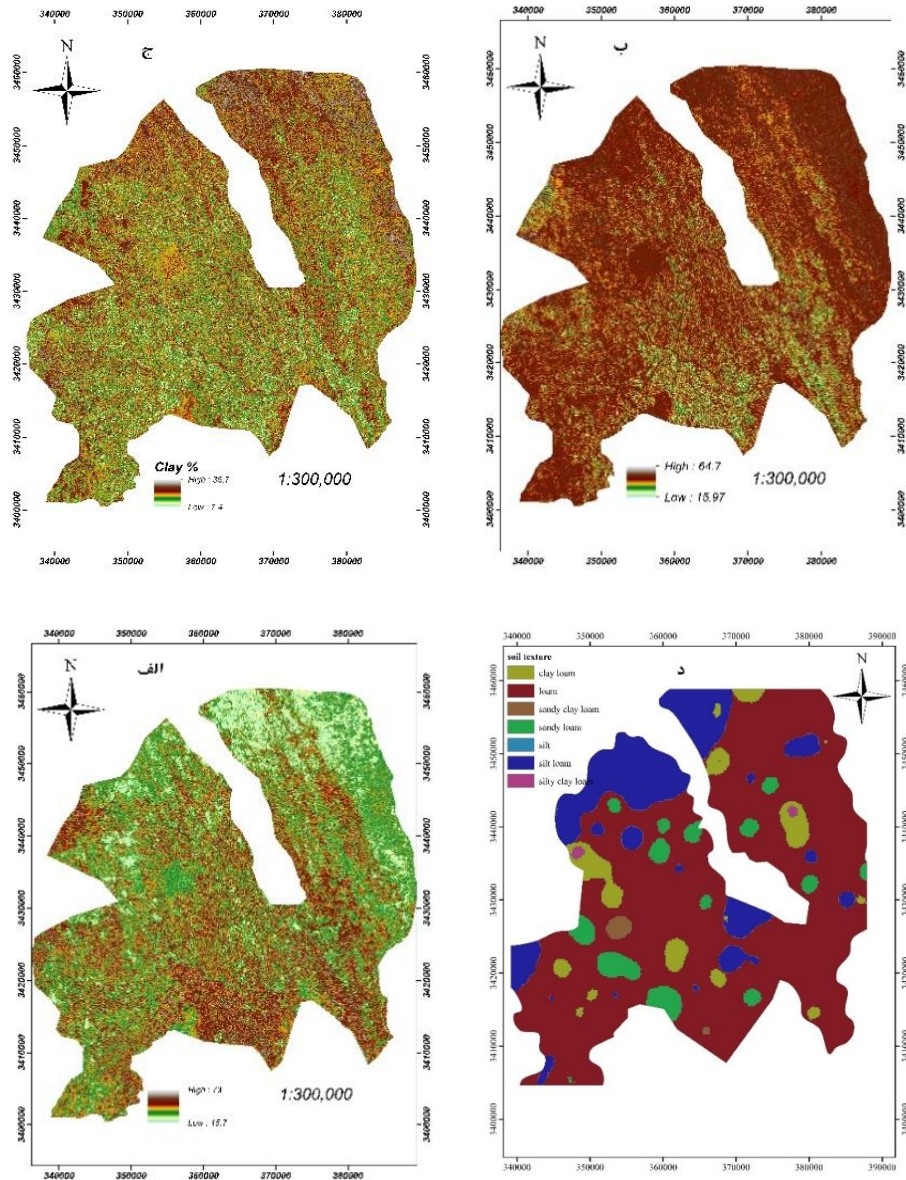
که مدل جنگل تصادفی خصوصیات خاک را بیشتر پیش‌بینی می‌کند. نتایج نشان می‌دهد که مقادیر RMSE داده‌های اعتبارسنجی برای شن، سیلت و رس خاک در روش جنگل تصادفی به ترتیب برابر ۱۵/۴۲، ۱۲/۵۶ و ۸/۹۷ درصد است. همچنین RMSE برای مدل ساده کریجینگ به ترتیب ۱۸/۲، ۱۳/۳ و ۹/۵۳ برآورد شده است که این مقادیر به ترتیب ۵/۹ و ۱۱/۲ درصد بیشتر نسبت به روش جنگل تصادفی می‌باشد. چاگاز و همکاران (۲۰۱۶) مقدار RMSE را در مرحله اعتبار سنجی برای شن، سیلت و رس خاک به ترتیب برابر ۹۹/۶۵، ۴۸/۹۳ و ۷۶/۴۴ گزارش دادند. همچنین آکپا و همکاران (۲۰۱۴) مقدار RMSE را برای شن بین ۱۹/۶۷-۱۹/۲۶، برای سیلت بین ۱۲/۲۲-۱۱/۷۲ و برای رس خاک بین ۱۳/۵۹-۱۳/۱۱ در مطالعه خود برآورد کردند. مقادیر nRMSE برای اجزاء شن، سیلت و رس خاک به ترتیب برابر ۰/۱۹، ۰/۱۳ و ۰/۲۱ برای داده‌های پیش‌بینی و برابر ۰/۳۹، ۰/۲۹ و ۰/۳۴ برای داده‌های اعتبارسنجی می‌باشد که این بیانگر برآورد متوسط پیش‌بینی‌های مدل می‌باشد. وجود تفاوت در مقادیر این شاخص بین داده‌های پیش‌بینی و اعتبارسنجی به دلیل سیلابی و آبرفتی بودن منطقه مورد مطالعه می‌باشد به‌طوری‌که این تفاوت در مطالعات سایر محققین نیز گزارش شده است.

بررسی مقادیر ضریب توافقی و بلموت (dr) داده‌های پیش‌بینی نشان می‌دهد که مدل‌سازی با دقت قابل‌قبولی انجام شده است به‌طوری‌که مقادیر این شاخص برای اجزاء شن، سیلت و رس خاک به ترتیب برابر ۰/۹۶، ۰/۹۵ و ۰/۹۴ برای داده‌های آموزش و همچنین به ترتیب برابر ۰/۷۴، ۰/۷۸ و ۰/۷۷ برای داده‌های اعتبارسنجی برآورد شده است. همچنین بررسی مقادیر ضریب کارایی مدل (EF) نیز نشان می‌دهد که روش جنگل تصادفی به‌درستی نقشه‌های اجزاء بافت خاک را در محدوده مورد مطالعه تولید کرده است.

جدول ۶: نتایج ارزیابی روش جنگل تصادفی برای پیش‌بینی

اجزاء بافت خاک

| داده‌ها | معیار ارزیابی | sand | silt | clay |
|------------|----------------|-------|-------|-------|
| پیش‌بینی | RMSE | ۷/۷۴ | ۵/۶۴۵ | ۴/۰۴۲ |
| | nRMSE | ۰/۱۹ | ۰/۱۳ | ۰/۲۱ |
| | d _r | ۰/۹۶ | ۰/۹۵ | ۰/۹۴ |
| | EF | ۰/۸۸ | ۰/۸۶ | ۰/۸۴ |
| | MBE | -۰/۱۹ | ۰/۱۵۷ | ۰/۰۸۶ |
| | MAE | ۵/۵۴ | ۴/۲۶ | ۳/۰۷۷ |
| اعتبارسنجی | RMSE | ۱۵/۴۲ | ۱۲/۵۶ | ۸/۹۷ |
| | nRMSE | ۰/۳۹ | ۰/۲۹ | ۰/۳۴ |
| | d _r | ۰/۷۴ | ۰/۷۸ | ۰/۷۷ |
| | EF | ۰/۳۲ | ۰/۳۴ | ۰/۲۹ |
| | MBE | ۲/۷۵ | -۰/۶۵ | -۰/۱۶ |
| | MAE | ۱۲/۱۳ | ۹/۴۸ | ۷/۱۶ |



شکل ۳. نقشه رقمی شن (الف)، سیلت (ب)، رس (ج) و (د) بافت خاک.

دشت سیستان به عنوان یک دشت آبرفتی است. دلیل دیگر آن می‌تواند به تعداد نمونه‌های ارتباط داشته باشد. بنابراین پیشنهاد می‌شود که برای به دست آوردن نقشه‌های با دقت بالاتر خصوصیات خاک به ویژه خصوصیات فیزیکی در دشت‌های آبرفتی، تعداد نمونه برداری‌های خاک افزایش یابد و همچنین تعداد نقاط بهینه در این مناطق تعیین گردد. همچنین به دلیل اینکه خصوصیات خاک در دشت‌های سیلابی تابع نحوه رسوب-گذاری است که خود متأثر از نحوه پراکنش مسیره‌های جریان آب می‌باشد، پیشنهاد می‌شود در مطالعات آتی از شاخص‌های که به نحوی به این موضوع مرتبط است نظیر فاصله تا رودخانه ها، شبکه انهار منطقه، نقشه کاربری اراضی به عنوان متغیرهای کمکی استفاده شود.

اندازه ذرات خاک و شاخص روشنایی تصویر ماهواره لندست ۸ رابطه همبستگی مثبت و با شاخص رس رابطه همبستگی منفی دارد. رس خاک نیز با ارزش رقمی باندهای ۱ تا ۵، باند ۸، شاخص روشنایی و شاخص اندازه ذرات خاک تصویر ماهواره لندست ۸ رابطه همبستگی مثبت و با شاخص رس رابطه همبستگی منفی دارد. متغیرهای محیطی شاخص رس، شاخص اندازه ذرات خاک و نسبت باند ۴ به باند ۸، باند ۱، باند ۸ و باند ۷ متغیرهای با اهمیت در اجرای مدل بودند. روش جنگل تصادفی مقادیر کمتری برای RMSE در مقایسه با یک مدل ساده برای پیش بینی مکانی ذرات شن، سیلت و رس خاک ارائه داد. در مجموع به علت تغییرات زیاد خصوصیات خاک در دشت‌های سیلابی مقادیر RMSE نسبتاً زیاد برآورد گردید که این به دلیل تغییرپذیری زیاد خصوصیات فیزیکی خاک در

Reference:

- Akpa, S.I.C., Odeh, I.O.A., Bishop, T.F.A., Hartemink, A.E., 2014. Digital mapping of soil particle size fractions for Nigeria. *Soil Sci. Soc. Am. J.* 78, 1953–1966.
- Boettinger, J.L., Ramsey, R.D., Bodily, J.M., Cole, N.J., Kienast-Brown, S., Nield, S.J., Saundes, A.M., Stum, A.K., 2008. Landsat spectral data for digital soil mapping. In: Hartemink, A.E., McBratney, A.B., Mendonca Santos, M.L. (Eds.), *Digital Soil Mapping With Limited Data*. Springer science, Australia, pp. 193–203
- Bui, E. N., Ballabio, C., & Panagos, P., 2020. Soil texture mapping over Europe: Machine learning and its potential for sustainable soil management. *European Journal of Soil Science*, 71(1), 13–28.
- Carvalho Junior, W., Chagas, C. D. S., Pereira, G. T. 2020. Regional-scale digital soil mapping in tropical areas using machine learning: A case study in Brazil. *Geoderma*, 363, 114151.
- Carvalho Junior, W., Lagacherie, P., Chagas, C.S., Calderano Filho, B., Bhering, S.B., 2014. A regional-scale assessment of digital mapping of soil attributes in a tropical hillslope environment. *Geoderma* 232, 479–486.
- Chagas, C.S., Junior, W.C., Bhering, S.B and Filho, B.C. 2016. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. *Catena*, 139: 232–240
- Chen, J.C., Chang, N.B., and Shieh, W.K. (2003). Assessing wastewater reclamation potential by neural network model. *Journal of Engineering Application of Artificial Intelligence*, 16:149-157.
- Demattê, J.A.M., Fiorio, P.R., Ben-Dor, E., 2009. Estimation of soil properties by orbital and laboratory reflectance means and its relation with soil classification. *Open Remote Sens. J.* 2, 12–23.
- Gambill, D.R., Wall, W.A., Fulton, A.J. and Howard, H.R. 2016. Predicting USCS soil classification from soil property variables using Random Forest. *Journal of Terramechanics*, 65:85–92
- Heung, B., Bulmer, C. E., Schmidt, M. G., 2016. Predictive soil parent material mapping at a regional-scale: A random forest approach. *Geoderma*, 279, 68-80.
- Jafari A., Finke P.A., de Wauw J.V., Ayoubi S., and Khademi H. 2012. Spatial prediction of USDA- great soil groups in the arid Zarand region, Iran: comparing logistic regression approaches to predict diagnostic horizons and soil types. *E. J. Soil Sci*, 63: 284–298.
- Islam, K., Singh, B., McBratney, A., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near infrared reflectance spectroscopy. *Aust. J. Soil Res.* 41, 1101–1114.
- Liao, K., Xu, S., Wu, J., Zhu, Q., 2013. Spatial estimation of surface soil texture using remote sensing data. *Soil Sci. Plant Nutr.* 59, 488–500.
- Liaw, A. and Wiener, M. 2002. Classification and regression by random Forest. *R news The newsletter of the R project*, 2(3): 18-22.
- Ließ, M., Glaser, B., and Huwe, B. 2012. Uncertainty in the spatial prediction of soil texture comparison of regression tree and random forest models. *Geoderma*, 170: 70-79.
- Mirzaee, S.; Ghorbani-Dashtaki, S.; Mohammadi, J.; Asadi, H.; Asadzadeh, F. Spatial variability of soil organic matter using remote sensing data. *Catena* 2016, 145, 118–127.
- Pahlavan-Rad, M.R., K.H. Dahmardeh, M. Hadizadeh, G. Keykha, N. Mohammadnia, M. Gangali, M. Keikha, N. Davatgar, C. Brungard. 2020. Prediction of soil water infiltration using multiple linear regression and random forest in a dry flood plain, eastern Iran. *Catena*, 194(2020) 104715.
- Poggio, L., de Sousa, L. M., Kempen, B., Heuvelink, G. B. M., Brus, D. J. 2021. Soil property mapping using machine learning models: A review and challenges. *Earth-Science Reviews*, 223, 103786.
- Pringle, M.J., Romano, N., Minasny, B., Chirico, G.B., and Lark, R.M. 2007. Spatial evaluation of pedotransfer functions using wavelet analysis. *Hydrology*, 333: 182-198.
- R Development Core Team, 2007. *R: A Language and Environment for Statistical Computing*.
- Zhao, W., Zhang, G. L., Zhao, Y. G., & Pan, W., 2018. Predicting soil properties using remote sensing data and machine learning models. *Geoderma*, 322, 1-10.
- Shahriari, M., Delbari, M., Afrasiab, P., & Pahlavan-Rad, M. R. 2019. Predicting regional spatial distribution of soil texture in floodplains using remote sensing data: A case of southeastern Iran. *Catena*, 182, 104149.
- Taghizadeh-Mehrjardi, R., Schmidt, K., Mahmoodi, S., Omid, M. 2020. Spatial prediction of soil properties using machine learning techniques: A case study from Iran. *Soil Systems*, 4(4), 63.
- Vaysse, K., Lagacherie, P., 2015. Evaluating Digital Soil Mapping approaches for mapping Global Soil Map soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Reg.* 4, 20–30.
- Willmott, C.J., Robeson, S.M., Matsuura, K. (2011). A refined index of model performance. *International Journal of Climatology*, 32(13): 2088-2094.
- Xiao, J., Shen, Y., Tateishi, R., Bayaer, W., 2006. Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. *Int. J. Remote Sens.* 12, 2411–2422.

- ¹ *Artificial Neural Network*
- ² *Pedotransfer*
- ³ *Geostatistic*
- ⁴ *Dicision Tree*
- ⁵ *Jafari*
- ⁶ *Carvalho junior*
- ⁷ *Akpa*
- ⁸ *Vaysse and Lagacherie*
- ⁹ *Ließ*
- ¹⁰ *Chagas*
- ¹¹ *Gambill*
- ¹² *Normalized Vegetation Index*
- ¹³ *Brightness Index*
- ¹⁴ *Clay Index*
- ¹⁵ *Grain Size Index*
- ¹⁶ *Digital Number*
- ¹⁷ *liao*
- ¹⁸ *Sauza juniore*
- ¹⁹ *Islam*
- ²⁰ *Weterlined and Stenberg*
- ²¹ *Dematte*