



Digital mapping of soil texture components in part of Khuzestan plain lands using machine learning models

Nasim Sahraee¹, Ahmad Landi², Saeid Hojati³✉

1. Department of Soil Science and Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran, Email: sahraee.n@gmail.com
2. Department of Soil Science and Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran, Email: landi@scu.ac.ir
3. Corresponding Author, Department of Soil Science and Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, Iran, Email: s.hojati@scu.ac.ir

Article Info

Article type: Research Article

Article history:

Received: Sep. 10, 2022

Revised: Nov. 1, 2022

Accepted: Nov. 27, 2022

Published online: Dec. 22, 2022

Keywords:

Spatial Modeling,
Remote Sensing,
Soil Texture,
Support Vector Machine,
Random Forest.

ABSTRACT

This study aims to evaluate and to compare the efficiency of support vector machine (SVM) and random forest (RF) models using digital soil mapping approach to predict soil texture in part of Khuzestan province. In February 2021, before determining soil texture, 200 soil samples were taken using stratified random sampling from the surface layer (0-10 cm). Auxiliary variables included primary and secondary derivatives of digital elevation model (DEM), remote sensing spectral indices (RS), from which the appropriate category was selected using principal component analysis (PCA). Based on PCA method, nine topographic variables from DEM and eight vegetation indices and spectra from RS were selected to predict soils texture components (sand, silt, and clay). The efficiency of the models was evaluated using the coefficient of determination (R^2) and the root mean squared of the error (RMSE). The results indicated that the random forest model had higher accuracy and less error than the support vector machine model (SVM), so that values of R^2 in this model were 0.80 for sand, 0.81 for silt, and 0.78 for clay, and the RMSE in the prediction of these particles were 6.02, 5.89 and 6.02, respectively. While the R^2 and RMSE in the support vector machine model for prediction of sand, silt and clay were (0.39, 13.70), (0.45, 10.70), and (0.46, 9.32), respectively. Also, the results of this evaluation showed that salinity index, brightness index, and channel network in addition of the 6-band Landsat 8 satellite or the far infrared band were the most important environmental variables predicting clay, silt, and sand particles. In conclusion, we suggest using Random Forest model as a useful and reliable method in preparing digital maps of soil texture in the study area.

Cite this article: Sahraee, N., Landi, A., & Hojati, S. (2022) Digital mapping of soil texture components in part of Khuzestan plain lands using machine learning models. *Iranian Journal of Soil and Water Research*, 53 (10), 2261-2276. <https://doi.org/10.22059/ijswr.2022.348442.669360>

© The Author(s).

Publisher: University of Tehran Press.

DOI: <https://doi.org/10.22059/ijswr.2022.348442.669360>



نقشه برداری رقومی اجزا بافت خاک در بخشی از اراضی دشت خوزستان با استفاده از برخی مدل‌های یادگیری ماشین

نسیم صحرایی^۱؛ احمد لندی^۲؛ سعید حجتی^۳

۱. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران، ایمیل: sahraee.n@gmail.com

۲. گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران، ایمیل: landi@scu.ac.ir

۳. نویسنده مسئول، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه شهید چمران اهواز، اهواز، ایران، ایمیل: s.hojati@scu.ac.ir

اطلاعات مقاله

چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۰/۰۶/۱۹

تاریخ بازنگری: ۱۴۰۱/۰۸/۱۰

تاریخ پذیرش: ۱۴۰۱/۰۹/۰۶

تاریخ انتشار: ۱۴۰۱/۱۰/۰۱

واژه‌های کلیدی:

مدل‌سازی مکانی،
سنجش از دور،
ماشین بردار پشتیبان،
جنگل تصادفی.

مطالعه حاضر با هدف ارزیابی و مقایسه کارایی مدل‌های ماشین بردار پشتیبان (SVM) و جنگل تصادفی (RF) با استفاده از رویکرد نقشه‌برداری رقومی خاک (DSM) برای پیش‌بینی اجزا بافت خاک در بخشی از اراضی استان خوزستان انجام شد. در بهمن سال ۱۳۹۹، به منظور تعیین بافت خاک، ۲۰۰ نمونه از خاک سطحی (عمق ۱۰-۰ سانتی متری) به صورت تصادفی طبقه‌بندی شده جمع‌آوری شدند. متغیرهای کمکی شامل مشتقات اولیه و ثانویه مدل رقومی ارتفاع (DEM) شامل (شیب، جهت شیب، شاخص شبکه آبراهه‌ای و ...) و شاخص‌های طیفی و گیاهی سنجش از دور (RS) بودند که انتخاب دسته مناسب از آنها با استفاده از روش تجزیه مولفه‌های اصلی (PCA) انجام گرفت. بر اساس روش PCA، نه متغیر توپوگرافی از DEM و هشت شاخص پوشش گیاهی از RS برای پیش‌بینی اجزا بافت خاک (شن، سیلت و رس) انتخاب گردیدند. کارایی مدل‌ها با استفاده از آماره‌های ضریب تبیین (R^2) و ریشه میانگین مربعات خطا (RMSE) مورد ارزیابی قرار گرفت. نتایج نشان داد مدل جنگل تصادفی از دقت بالاتر و خطا کمتری نسبت به مدل ماشین بردار پشتیبان (SVM) برخوردار است، به طوری که میزان R^2 در این مدل برای شن ۰/۸۰، سیلت ۰/۸۱ و رس ۰/۷۸ و ریشه‌ی میانگین مربعات خطا (RMSE) در پیش‌بینی این ذرات به ترتیب ۶/۰۲، ۵/۸۹ و ۶/۰۲ بود. این در حالی است که R^2 و RMSE در مدل ماشین بردار پشتیبان به ترتیب برای شن ۰/۳۹ و ۱۳/۷۰، سیلت ۰/۴۵ و ۱۰/۷۰ و رس ۰/۴۶ و ۹/۳۲ بود. همچنین اهمیت نسبی متغیرهای استفاده شده در پیش‌بینی اجزای بافت خاک نشان داد شاخص شوری، شاخص روشنایی و شبکه آبراهه‌ای به همراه باند ۶ ماهواره لندست ۸ مهم‌ترین متغیرهای محیطی پیش‌بینی‌کننده ذرات رس، سیلت و ماسه بودند. بنابراین پیشنهاد می‌شود از مدل جنگل تصادفی به عنوان روشی مفید و قابل اعتماد در تهیه نقشه‌های رقومی بافت خاک در منطقه مورد مطالعه استفاده شود.

استناد: صحرایی؛ نسیم، لندی؛ احمد، حجتی؛ سعید، (۱۴۰۱) نقشه‌برداری رقومی اجزا بافت خاک در بخشی از اراضی دشت خوزستان با استفاده از برخی مدل‌های یادگیری

ماشین. مجله تحقیقات آب و خاک ایران، ۵۳ (۱۰)، ۲۲۷۶-۲۲۶۱. <https://doi.org/10.22059/ijswr.2022.348442.669360>



© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2022.348442.669360>

مقدمه

بافت خاک یک ویژگی مهم خاک است که بیشتر فرآیندهای فیزیکی، شیمیایی و بیولوژیکی در خاک را از جمله ظرفیت حرارتی، نفوذپذیری، ظرفیت نگهداری آب، حرکت املاح و عناصر سنگین در خاک را تحت تاثیر قرار می دهد. در علوم خاک تهیه نقشه توزیع مکانی بافت خاک با وضوح مکانی بالا برای برنامه ریزی مناسب کاربری اراضی، مدیریت خاکها و سایر مداخلات مرتبط با حفاظت از محیط زیست مورد نیاز است (Dharumarajan and Hegde., 2020).

در روش های معمول نقشه برداری بافت خاک در مقیاس ها و کاربری های مختلف، نیازمند جمع آوری و تجزیه و تحلیل تعداد نمونه زیاد است تا تغییرپذیری مکانی بافت خاک برآورد شود. از این رو در سطح جهانی، شیوه های نقشه برداری خاک تا حد زیادی به سمت استفاده از روش های نقشه برداری رقومی خاک (DSM) و استفاده از فناوری های نوین مرتبط با داده های سنجش از دور (RS) به عنوان ابزاری برای کاهش هزینه و زمان مورد نیاز برای توسعه نقشه های خاک تغییر یافته است (Yang et al., 2016; Khaledian and Miller, 2020).

داده های سنجش از دور به دلیل دسترسی ارزان، قدرت تفکیک مکانی بالا، قابلیت دسترسی در مقیاس های مختلف و ارائه اطلاعات دقیق از زمین نمای منطقه در نقشه برداری رقومی خاک مورد توجه قرار گرفته اند (Wang et al., 2018). در حقیقت داده های سنجش از دور با استفاده از فناوری نقشه برداری رقومی و مدل های پیش بینی کننده فرصت های جدیدی برای پایش مستمر ویژگی های خاک و بررسی توزیع مکانی آنها فراهم آورده است (Taghizadeh-Mehrjardi et al., 2014). مطالعات انجام شده در سالیان اخیر نشان می دهد داده های ماهواره ای می تواند متغیر کمکی مفیدی برای پیش بینی ویژگی های سطحی خاک باشند (Castaldi et al., 2016; Gomez et al., 2018; Vagen et al., 2016). به طور معمول، این متغیرهای کمکی علاوه بر داده های سنجش از دور از داده های ژئومورفولوژیکی و داده های اقلیمی و مدل های رقومی ارتفاع (DEM) به دست می آیند، به طور کلی، می توان گفت هدف انواع روش های نقشه برداری رقومی خاک پیش بینی ویژگی های مختلف خاک از طریق مرتبط نمودن مشاهدات زمینی به پیش بینی کننده های محیطی می باشد. (McBratney et al., 2003).

در نقشه برداری رقومی خاک برای توسعه مدل سازی خاک-زمین نما از روش های مختلف یادگیری ماشین، از جمله روش های درخت های تصمیم گیری، منطق فازی، شبکه های عصبی و غیره استفاده می شود (Heung et al., 2016). در این رویکرد، الگوریتم های یادگیری ماشین، سیستم های اطلاعات جغرافیایی و روش های مختلف آماری برای یافتن روابط بین پاسخ متغیرهای خاک و مجموعه ای از متغیرهای محیطی برای پیش بینی ویژگی های خاک در مکان های نمونه برداری نشده استفاده می شوند. مدل های مختلف یادگیری ماشین توسط محققان برای پیش بینی بافت خاک و کلاس های آن به کار گرفته شده است. رایج ترین این مدل ها عبارتند از ماشین بردار پشتیبان (Zhang et al., 2019; Wu et al., 2018; Gomez et al., 2019; Mousavi et al., 2022) ۲، رگرسیون لجستیک چند جمله ای ۴ (Camera et al., 2017)، تقویت گرادیان ۵ (Zhang and Shi., 2019)، شبکه های عصبی مصنوعی ۶ (Bagheri Bodaghabadi et al., 2015; Taalab, et al., 2015) و مدل جنگل تصادفی ۷ (Camera et al., 2017). نتایج به دست آمده توسط این محققان نشان می دهد که مدل های غیر خطی مانند جنگل تصادفی نسبت به مدل های خطی مانند رگرسیون های چند جمله ای در پیش بینی و برآورد کلاس های بافتی خاک از عملکرد بهتری برخوردار بودند. مطالعه Hengl, et al (2017) نشان داد که الگوریتم های جنگل تصادفی (RF) و تقویت گرادیان (GBM) در پیش بینی ذرات بافت خاک در مطالعات با تعداد مشاهده زیاد بهتر از سایر مدل های خطی عمل می کند. Bousbih et al. (2019) در مطالعه ای در منطقه نیمه خشک مرکزی کشور تونس با هدف تخمین بافت خاک با استفاده از داده های ماهواره های سنتینل ۱ و ۲ گزارش نمودند که الگوریتم جنگل تصادفی (RF) عملکرد بهتری از ماشین بردار پشتیبان (SVM) برای پیش بینی میزان رس خاک دارد.

در طی چند سال گذشته تحقیقات مختلفی در نقاط مختلف ایران با هدف نقشه برداری بافت خاک با استفاده از رویکرد نقشه برداری

1 Digital Soil Mapping

2- Support vector machines

3- k-nearest neighbors algorithm

4- Multinomial logistic regression

5- Gradient boosting

6- Artificial Neural Networks

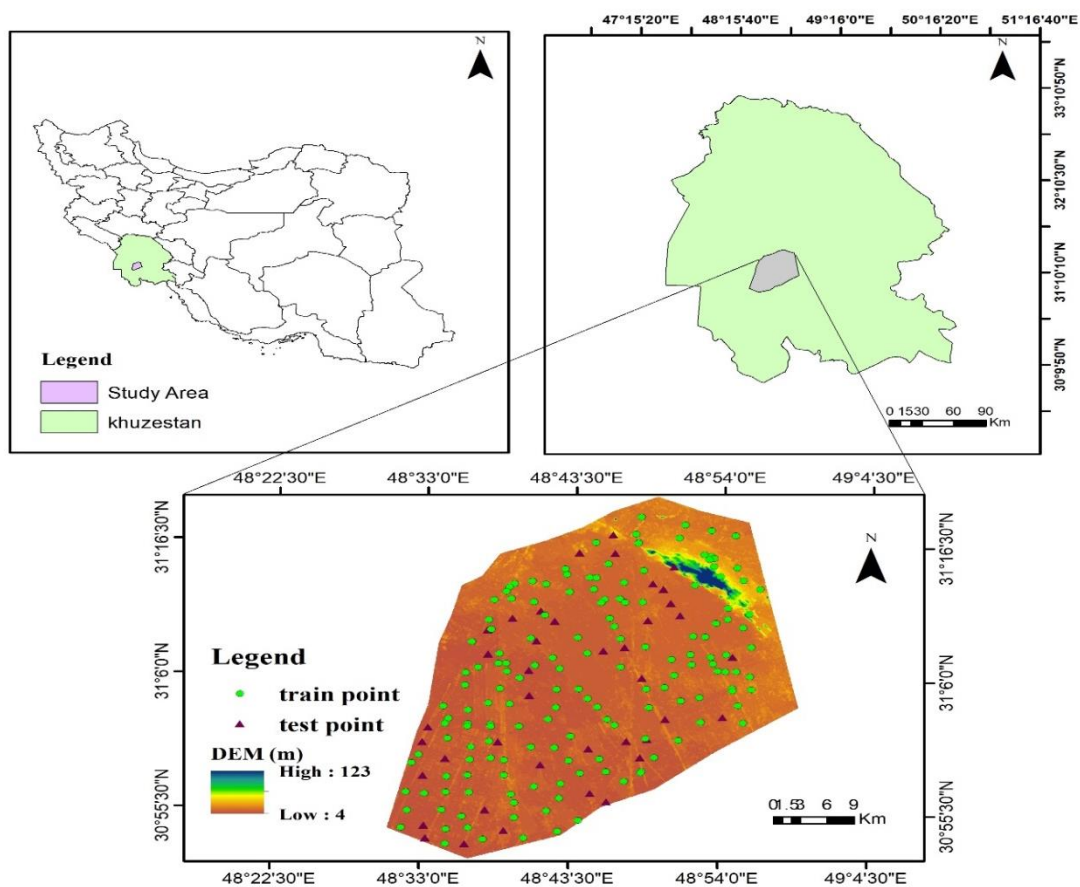
7- Random forest

رقومی خاک صورت گرفته است که حاکی از عملکرد قابل قبول الگوریتم‌های یادگیرنده ماشین در پیش‌بینی اجزای بافت خاک می‌باشد (Mousavi et al., 2020; Pahlavan-Rad and Akbari-moghaddam., 2018; Mehrabi-Gohari et al., 2019; Taghizadeh-Mehrjardi, et al., 2016, 2020). با توجه به اینکه استان خوزستان یکی از قطب‌های صنعتی و کشاورزی ایران می‌باشد، تعیین تغییرات مکانی اجزاء بافت خاک به منظور مدل‌سازی اکولوژیکی و مدیریت کشاورزی و محیط زیست از اهمیت ویژه‌ای برخوردار است. از سوی دیگر، تهیه نقشه تغییرات بافت خاک می‌تواند نقش مهمی در تصمیم‌گیری دقیق، سریع و با هزینه‌ی کمتر در ارزیابی‌های محیطی و مدیریت صحیح در سطح منطقه داشته باشد. از این‌رو، در این پژوهش مدنظر است، تا پتانسیل استفاده از داده‌های سنجش از دور در کنار سایر متغیرهای محیطی برای پیش‌بینی اجزای بافت خاک، در بخشی از اراضی استان خوزستان مورد ارزیابی قرار گیرد. از دیگر اهداف این پژوهش تهیه نقشه توزیع مکانی اجزای بافت خاک در منطقه مورد مطالعه به همراه مقایسه کارایی مدل‌های جنگل تصادفی (RF) و ماشین بردار پشتیبان (SVM) برای پیش‌بینی اجزای بافت خاک می‌باشد.

مواد و روش‌ها

منطقه مورد مطالعه

منطقه مورد مطالعه در استان خوزستان و در ۱۵ کیلومتری جنوب شهر اهواز به مختصات جغرافیایی "۳۰°۵۲'۳۰" تا "۲۰°۲۰'۳۱" عرض شمالی و "۱۰'۵۸'۴۸" تا "۰۱'۳۰'۴۸" طول شرقی با مساحتی حدود ۱۵۵ هزار هکتار واقع شده است (شکل ۱). تغییرات دامنه ارتفاع در منطقه مورد مطالعه از ۴ متر تا ۱۲۳ متر از سطح دریای آزاد می‌باشد. بر اساس آمار و اطلاعات موجود، متوسط بارش سالانه ۲۱۳ میلی‌متر و درجه حرارت متوسط ۳۱/۲ درجه سانتی‌گراد است. رژیم رطوبتی و حرارتی خاک‌های منطقه به ترتیب یوستیک و هایپرترمیک می‌باشد (Banni, M.H., 1998). کاربری غالب اراضی منطقه شامل اراضی کشاورزی با زراعت آبی و مراتع چند ساله می‌باشد. دشت رسوبی سیلابی زمین‌نمای غالب منطقه می‌باشد که از نظر مواد مادری دارای منشا رسوبات آبرفتی می‌باشد (Abyaat et al., 2017).



شکل ۱: موقعیت نقاط نمونه‌برداری در منطقه مورد مطالعه

مطالعات میدانی و تجزیه و تحلیل آزمایشگاهی

در این مطالعه با استفاده از روش نمونه‌برداری تصادفی طبقه‌بندی شده، تعداد ۲۰۰ نمونه خاک سطحی (۰ تا ۱۰ سانتی متر) با سیستم موقعیت یاب جهانی (GPS) برداشت شد. قبل از تعیین بافت در آزمایشگاه، نمونه‌ها هوا خشک و از الک ۲ میلی‌متری عبور داده شدند. توزیع اندازه ذرات خاک در آزمایشگاه به روش هیدرومتر (Gee and Boudier 1986) اندازه‌گیری گردید.

متغیرهای محیطی و انتخاب آنها

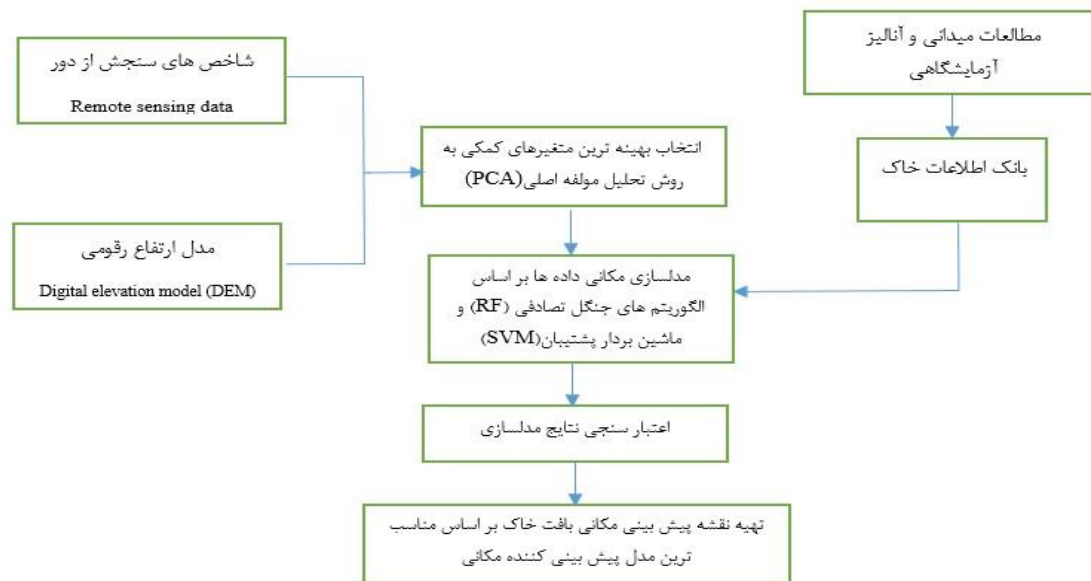
در این مطالعه از شاخص‌های سنجش از دور و مدل رقومی ارتفاع (DEM) به عنوان داده‌های کمکی پیش‌بینی کننده اجزای بافت خاک استفاده شد. هم‌زمان با برداشتن نمونه‌های زمینی در بهمن سال ۱۳۹۹، تصاویر سنجنده‌ی OLI ماهواره لندست ۸ از پایگاه داده مکانی USGS Earth Explorer بارگیری شد. پیش‌پردازش تصاویر شامل تصحیحات رادیومتری و اتمسفریک با استفاده از الگوریتم Flaash در محیط نرم افزار ENVI 5.3 انجام شد. ۴۰ متغیر محیطی شامل ۲۲ شاخص از مشتقات مدل رقومی ارتفاع در محیط نرم‌افزار SAGA GIS 7.3 و ۱۸ شاخص سنجش از دور در محیط نرم افزار ENVI 5.3 تهیه شدند. تمامی متغیرهای محیطی مورد استفاده به وضوح مکانی ۳۰ متر باز نمونه‌گیری گردیدند. برای انتخاب مناسب‌ترین متغیرهای محیطی از روش تجزیه مؤلفه‌های اصلی (PCA) استفاده شد. روش PCA برای متراکم‌سازی داده‌ها و بهبود تفسیر داده‌های ماهواره‌ای استفاده می‌شود. با استفاده از این رویکرد ۱۷ متغیر محیطی به منظور ورودی مدل‌های پیش‌بینی انتخاب شدند که شامل ۹ متغیر توپوگرافی از DEM و ۸ شاخص پوشش گیاهی و طیفی بودند (جدول ۱). این مراحل در محیط نرم افزار SPSS نسخه ۲۶ انجام گردید.

جدول ۱: متغیرهای محیطی مورد استفاده برای مدلسازی مکانی

منبع	قدرت تفکیک	نماد متغیر	متغیرهای کمکی محیطی
Landsat satellite	m ^{۳۰}	Reflectance value of Landsat (NIR) satellite band	بانده ۵- مادون قرمز نزدیک
Landsat satellite	m ^{۳۰}	(SWIR1): Reflectance value of Landsat satellite band	بانده ۶- مادون قرمز کوتاه
Landsat satellite	m ^{۳۰}	Reflectance value of (SWIR2) Landsat satellite band	بانده ۷- مادون قرمز کوتاه
Foody et al., 2001	m ^{۳۰}	(NIR- Green) / (NIR+ =(NDVI) Green)	شاخص تفاضلی پوشش گیاهی نرمال شده
Douaoui et al., 2006	m ^{۳۰}	(Red -NIR) / (Red + NIR)=(NDSI)	شاخص شوری تفاضلی نرمال شده
Khan et al., 2005	m ^{۳۰}	(Green× Red) ^ 0.5=(SI)	شاخص شوری
Khan et al., 2005	m ^{۳۰}	((Red × Red) + (NIR × NIR))^ = (BI) 0.5	شاخص روشنایی
Rondeaux et al., 1996	m ^{۳۰}	(NIR- Red) / (NIR+ =(OSVI) Red+0.16)	شاخص گیاهی تعدیل کننده اثر خاک
Conrad et al., 2015	m ^{۳۰}	Elevation=DEM (m)	مدل ارتفاع رقومی
Conrad et al., 2015	m ^{۳۰}	Multi Resolution of Ridge To Flatness Index (MRRTF)	شاخص خط الراس با وضوح مکانی بالا
Conrad et al., 2015	m ^{۳۰}	Aspect	جهت شیب
Conrad et al., 2015	m ^{۳۰}	Wind effect	شاخص اثر باد
Conrad et al., 2015	m ^{۳۰}	Flow accumulation	شاخص تجمع جریان
Conrad et al., 2015	m ^{۳۰}	Mass Balance	شاخص توازن جرم
Conrad et al., 2015	m ^{۳۰}	Terrain Surface Texture	شاخص بافت سطح زمین
Conrad et al., 2015	m ^{۳۰}	Channel Network (m)	شبکه آبراهه‌ای (زهکشی)
Conrad et al., 2015	m ^{۳۰}	Channel Network Base Level	سطح مبنای شبکه زهکشی
Conrad et al., 2015	m ^{۳۰}	Terrain Ruggedness Index (TRI)	شاخص ناهمواری زمین

روندنمای پژوهش

روندنمای کلی این پژوهش در شکل ۲ ارائه شده است. مراحل انجام این پژوهش به ترتیب شامل: ۱- تعیین موقعیت و نمونه برداری از خاک به روش نظام‌دار تصادفی ۲- تهیه شاخص‌های سنجش از دور بر اساس تصاویر ماهواره لندست ۸ سنجنده OLI و مدل رقومی ارتفاع (DEM) با قدرت تفکیک مکانی ۳۰ متر ۳- انتخاب بهینه‌ترین متغیرهای محیطی به روش تحلیل مولفه‌های اصلی (PCA) ۴- مدل‌سازی مکانی داده‌ها با استفاده از الگوریتم‌های RF و SVM ۵- اعتبار سنجی نتایج مدل‌سازی و تهیه نقشه پیش‌بینی مکانی اجزای بافت خاک بر اساس مناسب‌ترین مدل پیش‌بینی کننده مکانی.



شکل ۲: روندنمای کلی تحقیق

مدل‌سازی مکانی

به منظور نمایش تغییرپذیری و مدلسازی توزیع مکانی ذرات نسبی خاک در محدوده‌ی مطالعاتی و ارائه نقشه پهنه‌بندی آن از دو الگوریتم یادگیری ماشین شامل جنگل تصادفی (RF) و مدل ماشین بردار پشتیبان (SVM) استفاده شد. در انتها مدلی که دارای بیشترین دقت در برآورد ذرات اولیه خاک بود، برای تهیه نقشه توزیع مکانی ذرات خاک مورد استفاده قرار گرفت.

جنگل تصادفی (RF)

مدل جنگل تصادفی به عنوان یکی از کارآمدترین طبقه‌بندی‌کننده‌ها شناخته شده است این الگوریتم در تحقیقات مختلف و برای کاربردهای مختلفی از جمله تجزیه و تحلیل و طبقه‌بندی بافت خاک و کاربری اراضی در بخش کشاورزی استفاده شده است. در این مدل هر درخت با اجرای یک الگوریتم یادگیری منحصر به فرد ساخته می‌شود که مجموعه متغیرهای ورودی را بر اساس آزمون مقدار ویژگی به زیر مجموعه‌هایی تقسیم‌بندی می‌کند. هر درخت یک رای به بهترین کلاس در هر نمونه می‌دهد و رای اکثریت همه درختان، خروجی طبقه‌بندی را تعیین می‌کند (Bousbih et al, 2019).

پارامترهای اصلی مدل RF شامل تعداد درختان (ntree)، تعداد متغیرهای پیش‌بینی کننده برای هر درخت (mtry)، و اندازه گره است (Friedman and Meulman, 2003). این الگوریتم با توجه به اینکه مهم‌ترین متغیر را در هر تقسیم‌گره انتخاب می‌کند، در مقایسه با بسیاری از مدل‌های پیش‌بینی نسبت به نویز یا پیش‌بینی ضعیف حساس نیست و هر درخت تصمیم را جداگانه با تعداد مجموعه داده‌های از پیش تعریف شده متناسب می‌کند (Okun and Priisalu, 2007). از دیگر ویژگی‌های مهم این مدل، توانایی مدلسازی روابط غیرخطی با ابعاد بالا، مدیریت همزمان پیش‌بینی‌کننده‌های طبقه‌بندی شده و پیوسته، اندازه‌گیری میزان خطا و مقدار اهمیت متغیرها می‌باشد. این ویژگی‌ها ابزار مفیدی برای روشن کردن روابط خاک - محیط است و به محققین اجازه می‌دهد تا اثرات تغییرات محیطی احتمالی آینده را بر ویژگی‌های خاک پیش‌بینی کنند (Barthold et al., 2013). در مطالعه حاضر، با استفاده از بسته تخصصی "random Forest" در محیط نرم افزار Rstudio پارامترهای اصلی مدل بهینه‌سازی شدند. برای همگرایی و افزایش دقت مدل، تعداد درختان جنگل (ntree)

۸۰۰ درخت و تعداد گره ۱۴، در نظر گرفته شد. همچنین تعداد متغیرهای تصادفی هر گره (mtry) بر اساس کمترین میزان ریشه‌ی میانگین مربعات خطا (RMSE) و متناسب با فرآیند رگرسیون و تعداد ورودی‌ها، برای کلیه اجزا بافت خاک مقدار ۲ انتخاب گردید.

ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان (SVM) یک رویکرد داده محور یادگیری ماشین بر پایه‌ی نظریه یادگیری آماری است که از اصل استقرای کمینه سازی خطای ساختاری استفاده می‌کند (Vapnik et al., 1995). این مدل به‌ویژه در حوزه سنجش از دور به دلیل قابلیت تعمیم، حتی با تعداد محدودی از نمونه‌های آموزشی می‌تواند برای طبقه‌بندی، پیش‌بینی و تشخیص عدم قطعیت استفاده شود (Mountrakis et al., 2011). الگوریتم SVM برخلاف سایر الگوریتم‌های خوشه‌بندی در گروه یادگیری با نظارت قرار می‌گیرد و دارای دو فاز آموزش و آزمون می‌باشد. مبنای این مدل، فرضیه جداسازی خطی داده‌ها است، به این طریق که در فاز آموزش، مرز تصمیم‌گیری به نحوی انتخاب می‌شود که کمترین فاصله آن با هر یک از گروه‌های دیگر بیشترین مقدار یا به عبارت دیگر بیشترین حاشیه اطمینان را داشته باشد. در این روش انتخاب مرز تصمیم‌گیری بر اساس نقاطی به نام بردار پشتیبان صورت می‌گیرد. بر این اساس، الگوریتم SVM یک منحنی با پهنای مشخص به داده‌ها برازش می‌دهد، به صورتی که کمترین میزان خطا در داده‌های آموزشی دیده شود. به‌طور کلی، هدف الگوریتم SVM ایجاد یک فرا صفحه (مرز تصمیم) است که فاصله هر کلاس خاص را تا این فرا صفحه بیشینه نماید و طبقه‌بندی اشتباه را به کمترین مقدار ممکن برساند. از قابلیت‌های دیگر ماشین بردار پشتیبان، توانایی غلبه بر مشکل توزیع غیرخطی داده‌های آموزشی است؛ به نحوی که با استفاده از توابع کرنل، داده‌ها را به فضای با بعد بزرگتری انتقال می‌دهند که در آن تفکیک‌پذیری بهتر انجام می‌شود و ابر صفحه جداکننده در آن فضا تعیین می‌شود (Li et al., 2009).

در این الگوریتم فرض می‌شود که رابطه بین متغیرهای مستقل و وابسته با تابع جبری مانند $f(x)$ به همراه مقداری نویز (خطای مجاز) مشخص شود.

$$f(x) = W^T \cdot \phi(x) + b \quad \text{رابطه (۱)}$$

$$y = f(x) + \text{noise} \quad \text{رابطه (۲)}$$

در معادلات فوق W : بردار ضرایب، b : ثابت بایاس و ϕ تابع کرنل می‌باشد. تابع $f(x)$ نیز تابعی است که دارای کمترین انحراف از اهداف واقعی برای همه داده‌های آموزشی است.

در این مدل، کرنل‌های مختلفی از جمله پایه‌های خطی، شعاعی، چند جمله‌ای و سیگموئید مورد استفاده قرار گرفته است که رایج‌ترین آن، کرنل تابع پایه شعاعی گوسی ۱ (RBF) می‌باشد و در این مطالعه نیز از این کرنل استفاده شد.

$$K(x, x_j) = \exp\left(\frac{-|x_i - x_j|^2}{2\sigma^2}\right) \quad \text{رابطه (۳)}$$

برای مدل‌سازی و برآورد مشخصه‌های کمی توسط مدل SVM و کرنل تابع پایه شعاعی انتخاب مقدار پارامترهای ثابت خطای مجاز (σ) پارامتر تنظیم (C) لازم است. این پارامترها وابسته به توزیع داده‌ها هستند و از یک طبقه‌بندی به طبقه‌بندی دیگر، متفاوت می‌باشند. در این مطالعه فرآیند محاسبات ماشین بردار پشتیبان بر اساس کدنویسی در محیط نرم‌افزار Rstudio انجام شد و مقدار بهینه پارامترها از طریق این نرم‌افزار تعیین گردید (جدول ۲).

جدول ۲: مقدار بهینه پارامترهای مدل SVM برای اجزا بافت خاک

ویژگی	پارامتر تنظیم (C)	ثابت خطای مجاز (σ)
Sand	۰/۲۵	-/۰۵
Silt	۰/۵	-/۰۵
Clay	۱	-/۰۵

ارزیابی کارایی مدل‌سازی

به‌منظور ارزیابی عملکرد مدل‌های مختلف در برآورد متغیرهای وابسته (رس، شن و سیلت) از شاخص‌های آماری ضریب تبیین (R^2) و ریشه میانگین مربعات خطا (RMSE) استفاده گردید (روابط ۴ و ۵). برای بررسی دقت پیش‌بینی مدل‌ها، مشاهدات به‌طور تصادفی به دو دسته آموزش (۸۰ درصد از داده‌های نمونه‌برداری شده) و آزمون (۲۰ درصد از داده‌های نمونه‌برداری شده) تقسیم‌بندی شدند.

$$R^2 = \frac{\sum_{i=1}^n (Y_i - P_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad \text{رابطه (۴)}$$

$$RMSE = \sqrt{1/n \sum_{i=1}^n (Y_i - P_i)^2} \quad \text{رابطه (۵)}$$

در معادلات فوق Y_i : مقادیر اندازه‌گیری شده، P_i : مقادیر پیش‌بینی شده‌ی رس، سیلت و شن خاک، \bar{Y} : میانگین مقادیر اندازه‌گیری شده و n : تعداد کل مشاهدات است.

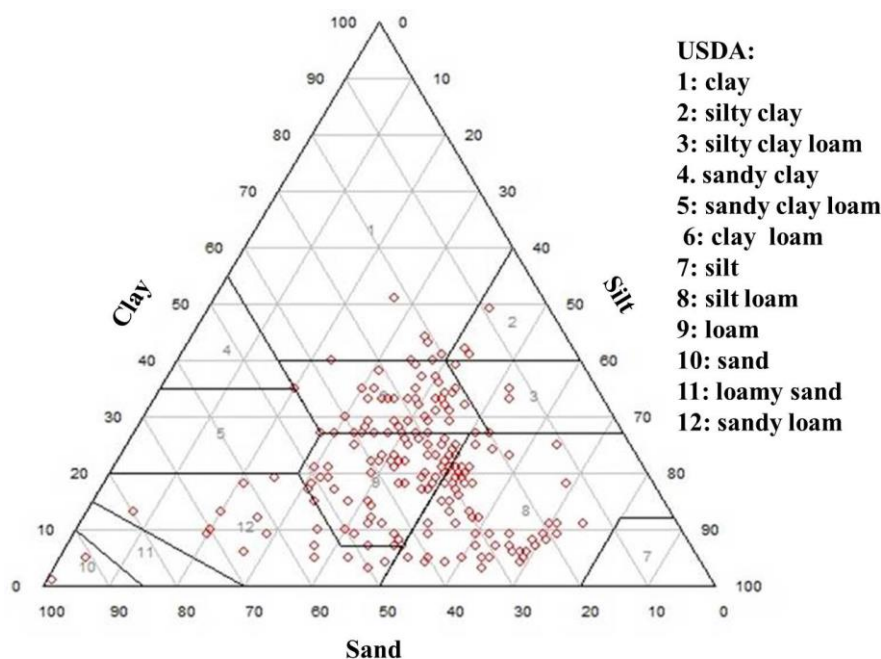
نتایج و بحث

توصیف آماری بافت خاک

در این پژوهش مدل‌سازی مکانی اجزا بافت خاک سطحی (شن، سیلت و رس) مورد بررسی قرار گرفت، که توصیف آماری این ویژگی‌ها در جدول ۳ ارائه شده است. نتایج نشان می‌دهد میانگین اجزا شن، سیلت و رس در منطقه‌ی مطالعاتی به ترتیب ۳۴/۲، ۴۵ و ۲۰/۸ درصد بود. فراوانی شن از ۹٪ تا ۹۸٪ با میانگین و انحراف معیار به ترتیب ۳۴/۲٪ و ۱۳/۳٪ متغیر بود. همچنین میانگین ذرات سیلت و رس به ترتیب ۱۲/۸ ± ۴۵ درصد و ۱۰/۸ ± ۲۰/۸ درصد را نشان می‌دهند. در مثلث بافت خاک (شکل ۳) نتایج حاصل از توزیع کلاس بافتی خاک‌های منطقه مورد مطالعه نشان داده شده است، همانطور که مشاهده می‌شود غالب‌ترین بافت خاک مربوط به کلاس بافتی لوم و در ادامه کلاس‌های بافتی لوم رسی، لوم سیلتی و لوم شنی فراوانترین کلاس‌های بافت خاک منطقه می‌باشند. بر اساس طبقه‌بندی ویلدینگ (1985) متغیرهای با ضریب تغییرات (CV) کمتر از ۱۵ درصد دارای تغییرات کم، متغیرهای با ضریب تغییرات ۱۵-۳۵ درصد دارای تغییرات متوسط و متغیرهای با ضریب تغییرات بیش از ۳۵ درصد دارای تغییرات زیاد هستند. با توجه به این طبقه‌بندی ذرات شن و رس در کلاس تغییرات زیاد و سیلت در کلاس تغییرات متوسط قرار دارند. ضریب تغییرپذیری زیاد در ویژگی‌های خاک می‌تواند حاصل از تغییر در محیط‌های رسوبگذاری و یا اختلاف در شکل اراضی باشد (Momtaz et al., 2009).

جدول ۳: خلاصه آماری ذرات بافت خاک

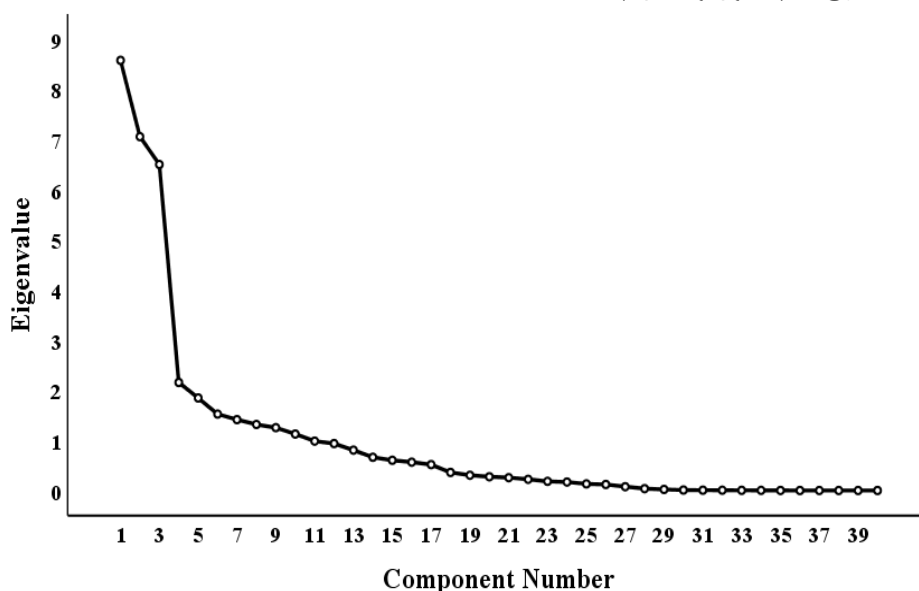
پارامتر	واحد	کمینه	میانگین	بیشینه	انحراف معیار	ضریب تغییرات (%)
شن	درصد	۹/۰۰	۳۴/۲	۹۸/۰	۱۳/۳	۳۸/۹
سیلت	درصد	۱/۰۰	۴۵/۰	۷۵/۰	۱۲/۸	۲۸/۴
رس	درصد	۱/۰۰	۲۰/۸	۵۱/۰	۱۰/۸	۵۲/۰



شکل ۳. توزیع کلاس‌های بافتی خاک در منطقه مورد مطالعه

متغیرهای محیطی

به منظور انتخاب مناسب‌ترین متغیرهای ورودی به مدل‌های ماشین بردار پشتیبان و جنگل تصادفی از روش تجزیه‌ی مولفه‌های اصلی (PCA) استفاده شد. بر همین اساس، ۱۰ مولفه اصلی با ارزش ویژه بزرگتر از یک و با توجه بیش از ۸۱ درصد واریانس تجمعی را تحت پوشش قرار می‌دادند، استخراج شد (جدول و شکل ۴).



شکل ۴: نمودار اسکری (بازویی) برای تعیین تعداد مولفه‌های مناسب

جدول ۴: واریانس تجزیه‌ی مولفه‌های اصلی متغیرهای محیطی

مولفه	مقادیر ویژه اولیه			مجموع مربعات مولفه‌های استخراج شده			مجموع مربعات مولفه‌ها پس از چرخش		
	کل	واریانس (%)	درصد تجمعی (%)	کل	واریانس (%)	درصد تجمعی (%)	کل	واریانس (%)	درصد تجمعی (%)
PC۱	۸/۵۷۶	۲۱/۴۴۰	۲۱/۴۴۰	۸/۵۷۶	۲۱/۴۴۰	۲۱/۴۴۰	۷/۷۹۰	۱۹/۴۷۴	۱۹/۴۷۴
PC۲	۷/۰۵۴	۱۷/۶۳۶	۳۹/۰۷۶	۷/۰۵۴	۱۷/۶۳۶	۳۹/۰۷۶	۷/۰۳۴	۱۷/۵۸۶	۳۷/۰۶۰
PC۳	۶/۴۹۹	۱۶/۲۴۷	۵۵/۳۲۳	۶/۴۹۹	۱۶/۲۴۷	۵۵/۳۲۳	۴/۳۵۶	۱۰/۸۹۱	۴۷/۹۵۱
PC۴	۲/۱۵۴	۵/۳۸۴	۶۰/۷۰۸	۲/۱۵۴	۵/۳۸۴	۶۰/۷۰۸	۴/۱۷۱	۱۰/۴۲۸	۵۸/۳۷۹
PC۵	۱/۸۴۴	۴/۶۱۱	۶۵/۳۱۸	۱/۸۴۴	۴/۶۱۱	۶۵/۳۱۸	۲/۰۶۷	۵/۱۶۷	۶۳/۵۴۶
PC۶	۱/۵۲۲	۳/۸۰۶	۶۹/۱۲۴	۱/۵۲۲	۳/۸۰۶	۶۹/۱۲۴	۱/۶۷۳	۴/۱۸۱	۶۷/۷۲۸
PC۷	۱/۴۱۳	۳/۵۳۲	۷۲/۶۵۶	۱/۴۱۳	۳/۵۳۲	۷۲/۶۵۶	۱/۵۲۳	۳/۸۰۷	۷۱/۵۳۴
PC۸	۱/۳۱۶	۳/۲۹۰	۷۵/۹۴۶	۱/۳۱۶	۳/۲۹۰	۷۵/۹۴۶	۱/۵۰۲	۳/۷۵۴	۷۵/۲۸۸
PC۹	۱/۲۵۳	۳/۱۳۲	۷۹/۰۷۸	۱/۲۵۳	۳/۱۳۲	۷۹/۰۷۸	۱/۴۵۶	۳/۶۶۱	۷۸/۹۵۰
PC۱۰	۱/۱۲۶	۲/۸۱۵	۸۱/۸۹۳	۱/۱۲۶	۲/۸۱۵	۸۱/۸۹۳	۱/۱۷۷	۲/۹۴۳	۸۱/۸۹۳

درون هر مولفه اصلی متغیرهای کمکی که دارای بیشترین مقادیر بودند، به همراه سایر شاخص‌هایی که ۱۰ درصد آنها بودند، شناسایی شدند و پس از انطباق با ضریب همبستگی بین آنها، در صورت وجود همبستگی بالا (بیشتر از ۰/۶)، متغیر با بالاترین مقدار نگه داشته شد و بقیه حذف گردیدند (جدول ضریب همبستگی به دلیل عدم وجود فضای کافی ارائه نشد). بر این اساس از ۴۰ متغیر ورودی ۱۷ متغیر کمکی انتخاب شدند که از این میان شاخص خط‌الراس با وضوح مکانی بالا، جهت شیب، اثر باد، تجمع جریان، توازن جرم، بافت سطح زمین، شبکه‌ی آبراهه‌ای، ناهمواری زمین، شاخص مبنای شبکه زهکشی، از DEM و باندهای ۵، ۶ و ۷ ماهواره لندست ۸ و شاخص‌های شوری تفاضلی نرمال شده، روشنایی، شوری، شاخص گیاهی تعدیل کننده اثر خاک و شاخص تفاضلی پوشش گیاهی نرمال شده به عنوان متغیرهای محیطی منتخب جهت پیش‌بینی اجزا بافت خاک مورد استفاده قرار گرفتند. نقشه‌ی تغییرات چهار مورد از مهم‌ترین آنها در شکل ۴ ارائه شده است.

کارایی مدل‌های رگرسیون ماشینی

نتایج اعتبار سنجی الگوریتم‌های یادگیری ماشینی RF و SVM در پیش‌بینی تغییرات مکانی بافت خاک در جدول (۵) ارائه شده است. میزان ضریب تبیین (R^2) در مدل جنگل تصادفی برای ذرات شن ۰/۷۸، سیلت ۰/۸۱ و رس ۰/۷۸ و میانگین ریشه‌ی مربعات خطا (RMSE) در پیش‌بینی این ذرات به ترتیب ۶/۰۲، ۵/۸۹، ۶/۰۲ بود، این درحالی است که ضریب تبیین (R^2) و میانگین خطا (RMSE) در مدل ماشینی بردار پشتیبان به ترتیب برای ذرات شن ۰/۳۹ و ۱۳/۷۰، سیلت ۰/۴۵ و ۱۰/۷۰ و برای ذرات رس ۰/۴۶ و ۹/۳۲ بود. به‌طور کلی در هر دو مدل کمترین و بیشترین میزان خطا به ترتیب مربوط به ذرات رس و شن می‌باشد. این نتایج نشان داد در منطقه مورد مطالعه، مدل RF با توجه به میزان صحت و کمینه خطا عملکرد خوبی در پیش‌بینی مکانی تمامی ذرات بافت خاک دارد، این در حالی است که مدل SVM از دقت متوسط جهت برآورد و پیش‌بینی مکانی ذرات بافت خاک سطحی برخوردار است. عملکرد بهتر مدل RF نسبت به مدل SVM را می‌توان به این دلیل دانست که مدل‌های ساده‌تر مانند SVM از مقادیر گسسته در گره‌های ترمینال برای تقسیم درخت استفاده می‌کنند و به لحاظ تغییرپذیری، عدم قطعیت بیشتری را نسبت به مدل‌های پیچیده‌تر مانند جنگل تصادفی نشان می‌دهند، بنابراین، در مدل SVM تعمیم روابط پایه‌ای بین ذرات خاک و پیش‌بینی‌های آنها دقت کمتری را نسبت به مدل RF دارا می‌باشد (Zolfaghari et al., 2019). براین اساس و با توجه به این که مدل جنگل تصادفی با کاهش بیش برآزش می‌تواند عملکرد مدل را در پیش‌بینی بهبود بخشد (Zhang et al., 2019) از این مدل برای پهنه‌بندی و تهیه نقشه اجزا بافت خاک سطحی منطقه مورد مطالعه استفاده گردید (شکل ۷).

جدول ۵. مقایسه اعتبار سنجی مدل‌های پیش‌بینی کننده ذرات بافت خاک بر اساس داده‌های آزمون

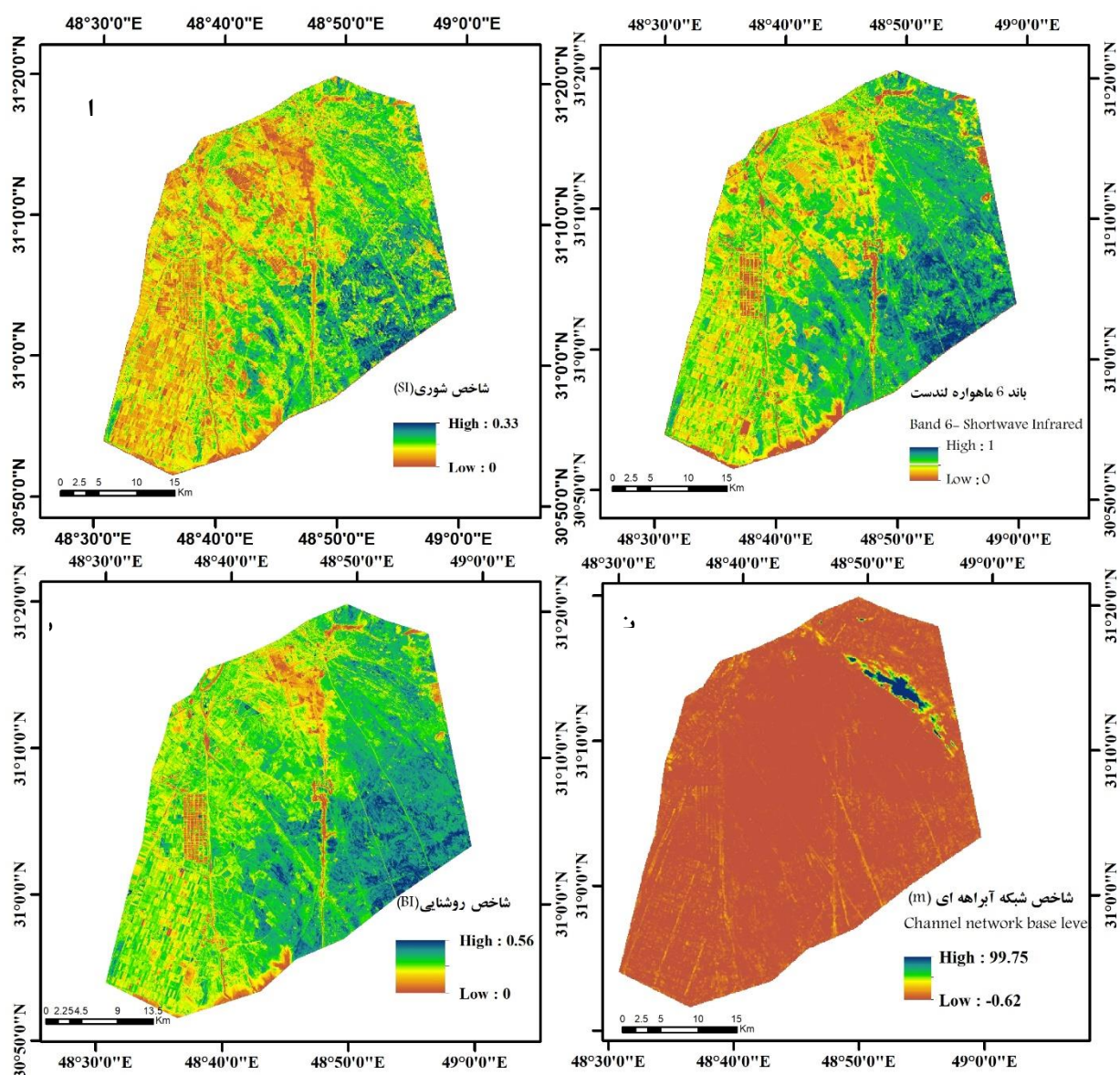
مدل	شاخص‌های اعتبار سنجی	شن	رس	سیلت
جنگل تصادفی (RF)	ریشه میانگین مربعات خطا (RMSE)	۶/۰۲	۵/۲۵	۵/۸۹
	ضریب تبیین (R^2)	۰/۸۰	۰/۷۸	۰/۸۱
ماشین بردار پشتیبان (SVM)	ریشه میانگین مربعات خطا (RMSE)	۱۳/۷۰	۹/۳۲	۱۰/۷۰
	ضریب تبیین (R^2)	۰/۳۹	۰/۴۶	۰/۴۵

اهمیت متغیرهای محیطی

نتیجه اهمیت نسبی متغیرهای محیطی پیش‌بینی کننده با استفاده از مدل جنگل تصادفی در شکل (۶) ارائه شده است. بر این اساس شاخص شوری (SI)، شاخص روشنایی (BI) و شاخص شبکه آبراهه‌ای ۱ به همراه باند ۶ تصاویر ماهواره‌ای مهم‌ترین متغیرهای محیطی پیش‌بینی کننده‌ی فراوانی ذرات خاک هستند. بر این اساس، به‌نظر می‌رسد متغیرهای محیطی استخراج شده از داده‌های سنجش از دور تأثیر بیشتری را در تخمین دقیق رس و سیلت نسبت به متغیرهای محیطی مشتق شده از مدل رقومی ارتفاع داشته‌اند. این درحالی است که در پیش‌بینی میزان شن، ویژگی‌های فیزیوگرافی منطقه و شاخص‌های به دست آمده از سنجش از دور از سهم تقریباً مساوی برخوردارند. احتمالاً با توجه به رسوبی بودن و شیب کم منطقه می‌توان بیان داشت ویژگی‌های توپوگرافیکی و شاخص‌های مربوط به آن از درجه اهمیت کمتری در تخمین میزان رس و سیلت نسبت به شاخص‌های سنجش از دور برخوردار باشند.

شاخص شوری مهم‌ترین شاخص در تخمین میزان رس بود و علت آن می‌تواند به این دلیل باشد که بخش رس، شوری بیشتری نسبت به شن و سیلت دارد (Pahlavan Rad et al., 2018). در مقابل، شاخص شبکه آبراهه‌ای مهم‌ترین متغیر محیطی به‌منظور پیش‌بینی شن بود. این موضوع نشان می‌دهد که رسوبات سیلابی و شبکه کانال‌های موجود در منطقه نقش مهمی در ایجاد تغییرات خاک در منطقه مورد مطالعه دارند. (Pahlavan Rad et al., 2018) نیز با استفاده از شاخص‌های سنجش از دور و توپوگرافی و مدل RF به نتایج مشابهی در پیش‌بینی بافت خاک و pH در دشت‌های سیلابی سیستان و بلوچستان دست یافتند در این مطالعه مهم‌ترین متغیرهای کمکی فاصله تا رودخانه، شبکه آبراهه‌ای، شاخص شوری نرمال شده، عمق دره، شاخص روشنایی و فاصله عمودی تا شبکه آبراهه‌ای بودند. باند ۶ تصاویر ماهواره لندست ۸، دومین متغیر محیطی مهم در برآورد میزان تمامی اجزای بافت خاک بود. (Brown et al., 2006)

از مقادير رقومي باندهای ۶ و ۷ به عنوان متغير محيطی برای تخمين بافت خاک استفاده کردند، نتايج اين مطالعه نشان داد که بين اين دو باند و اجزای بافت خاک همبستگي خوبی وجود دارد. ايشان دليل اصلی آن را حساس بودن طول موج قرمز کوتاه به مقدار رطوبت قابل دسترس خاک و وابسته بودن اين مقدار آب به بافت خاک دانستند. همچنين بيان داشتند در مناطق خشک و نيمه خشک به دليل رطوبت پايين خاک و تاج پوشش کم گیاهی، طول موج قرمز کوتاه به خوبی ویژگی های خاک را منعکس می کند. Parviz (2017) نیز در مطالعه ای با هدف ارزيابی تأثير پيش پردازش پارامترهای ورودی حاصل از تصاویر ماهواره ای به شبکه عصبی مصنوعی در تعيين بافت خاک بيان کرد حالت تکی باندهای ماهواره ای همبستگي بالایی با اجزا بافت خاک دارند. Liao و همکاران (2013) در نتایج مشابه بيان داشتند که باندهای ۱ تا ۵ و باند ۷ ماهواره ی Landsat ETM بیش ترين تأثير را در پيش بينی اجزا بافت سطحی دارد.

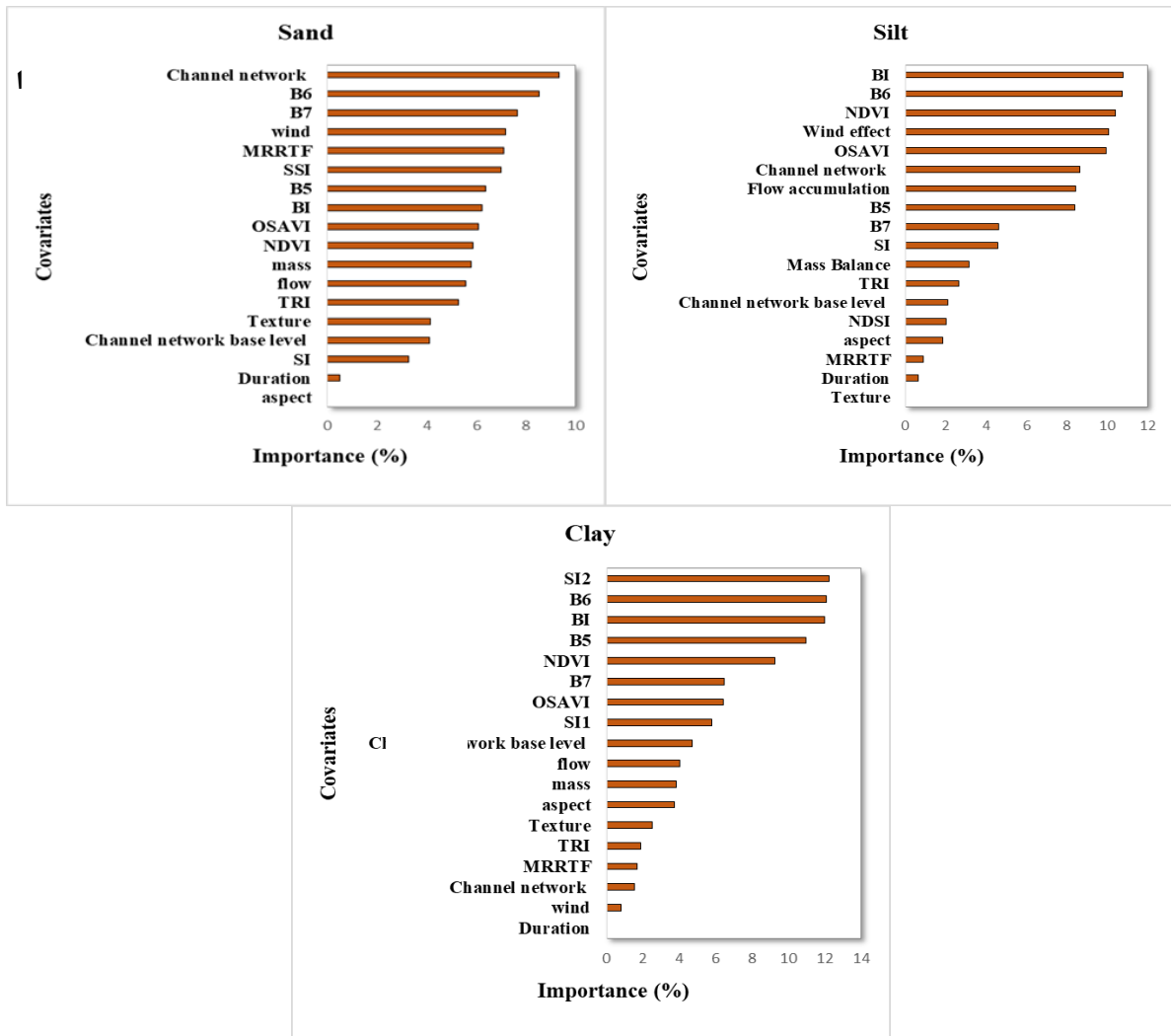


شکل ۵: نقشه های تغییرات مکانی مهم ترین متغیرهای محیطی پيش بينی کننده ی تخمين اجزای بافت خاک در منطقه مورد مطالعه

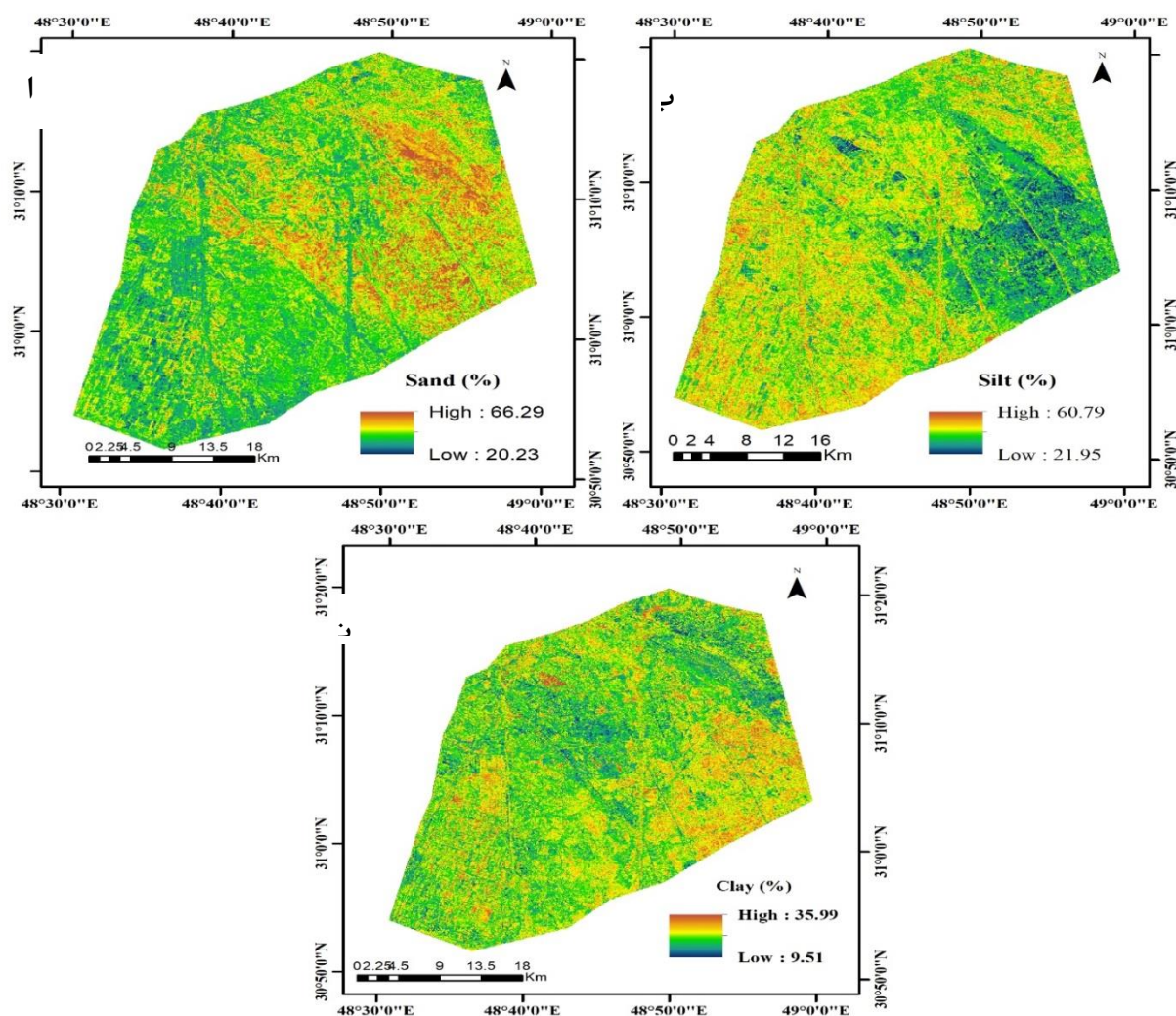
پيش بينی مکانی بافت خاک

شکل (۷) نقشه نهایی تغییرات مکانی اجزای بافت خاک با استفاده از مدل RF را در منطقه ی مطالعاتی نشان می دهد. بر اين اساس، بیش ترين تخمين میزان شن در بخش های مرکزی و شرقی منطقه مورد مطالعه بود. (شکل ۷ الف) اين مناطق در نزديکی شبکه آبراهه ای و کانال های موجود در منطقه قرار دارند. مقادير بالای شن را می توان در نتیجه نشست سريع ماسه از حالت تعليق در هنگام وقوع سيلاب های

فصلی در منطقه محتمل دانست. بر اساس نتایج اهمیت نسبی (شکل ۶ الف) پراکنش مکانی شن بیشتر متاثر از دو فاکتور محیطی شبکه آبراهه‌ای و باند ۶ ماهواره لندست بود که الگوی تغییرات مکانی این دو شاخص محیطی به‌خوبی با پراکنش مکانی شن همخوانی دارد (شکل ۵ ت و ب). همچنین بر اساس شکل (۷ ب) میزان سیلت در قسمت‌های مرکزی، شرقی و جنوب شرقی منطقه مورد مطالعه کم پیش‌بینی شده است، مقادیر بالاتری از سیلت در بخش‌های جنوبی، جنوب غربی و در اراضی با کاربری کشاورزی یافت شد. شکل‌های (۵ پ و ب) الگوی تغییرات مکانی شاخص روشنایی (BI) و باند ۶ ماهواره لندست را نشان می‌دهد که بیانگر این موضوع است که پراکنش مکانی سیلت هم‌راستا با تغییرات این دو شاخص محیطی در منطقه است. براساس نقشه‌ی رس، قسمت‌های شمال شرقی و شرق منطقه دارای کمترین میزان رس هستند که با توجه به ارتفاع بیشتر این ناحیه نسبت به کل منطقه مورد مطالعه قابل توصیف می‌باشد. همچنین در مراتع و اراضی پست فاقد کاربری (اراضی بایر) قسمت جنوب شرقی منطقه مورد مطالعه که حوضه آبخیز دارای ارتفاع کمتری است، بیشترین میزان رس دیده می‌شود (شکل ۷ پ). دو فاکتور محیطی شاخص شوری (SI) و باند ۶ بر اساس نتایج اهمیت نسبی، مهمترین فاکتورهای محیطی در پیش‌بینی مکانی رس بودند (شکل ۶ ب). الگوی تغییرات مکانی این دو شاخص محیطی با نقشه پراکنش مکانی رس در منطقه مطابقت دارد (شکل‌های ۵ الف و ب). به‌طورکلی با توجه به نقشه‌های خروجی می‌توان گفت که مدل RF در پیش‌بینی مکانی تمامی اجزا بافت خاک با شرایط طبیعی و فیزیوگرافی منطقه همگنی مناسبی داشته و توانسته است خصوصیات فیزیکی خاک محدوده مطالعاتی را با دقت و صحت بالایی مدل‌سازی کند.



شکل ۶. اهمیت نسبی متغیرهای محیطی در پیش‌بینی اجزا بافت خاک



شکل ۷: نقشه‌های رقمی اجزا بافت خاک با استفاده از مدل RF

نتیجه‌گیری

هدف از این مطالعه توسعه رویکردی برای تخمین و تهیه نقشه بافت خاک سطحی، با استفاده از داده‌های سنجش از دور و مدل رقمی ارتفاع بود. نتایج ما نشان داد هر دو مدل SVM و RF دارای کارایی لازم جهت تخمین بافت سطحی در منطقه هستند اما الگوریتم جنگل تصادفی با توجه به میزان دقت بیشتر و خطای کمتر دارای عملکرد بهتری از ماشین بردار پشتیبان است. همچنین با توجه به اینکه الگوریتم RF می‌تواند با حجم زیادی از داده‌ها برای یک زمان پردازش کامپیوتری معین دقت بالایی بدهد، مدل بهتری برای پیش‌بینی بافت خاک است. همچنین، نتایج نشان داد که شاخص‌های شاخص شوری، روشنایی و شبکه آبراه‌های، به همراه باند ۶ ماهواره لندست بیشترین تاثیر را در پیش‌بینی اجزا بافت سطحی دارند. به طور کلی نتایج این مطالعه نشان داد استفاده از نقشه‌برداری رقمی می‌تواند فرآیند نقشه‌برداری خاک را بهبود بخشد و سرعت عمل و کارایی نقشه‌ها را در دسترسی به اطلاعات و داده‌ها جهت استفاده کارشناسان خاک و آب افزایش دهد. بنابراین پیشنهاد می‌شود از روش RF برای نقشه‌برداری از ویژگی‌های خاک در سایر مناطق خشک ایران نیز استفاده گردد.

سیاس‌گذاری

هزینه‌های اجرای این پژوهش توسط معاونت پژوهشی، فناوری و ارتباط با جامعه دانشگاه شهید چمران اهواز (SCU.AS1400.50) و SCU.AS1400.365 پرداخت شده است.



REFERENCE

- Abyat A, Azhdari A, Jodaki M, Darvishi Khatoni J., (2017). Study and separation of Quaternary sedimentary environments in Khuzestan plain, *Iranian Journal of Advanced Applied Geology*, 7(3), 49-64 (In Persian with English Abstract).
- Bagheri Bodaghabadi, M., Antonio Martinez-Casasnovas, J., Salehi, M. H., Mohammadi, J., Esfandiarpour Borujeni, I., Toomanian, N., & Gandomkar, A.,(2015). Digital soil mapping using artificial neural networks and terrain-related attributes. *Pedosphere*, 25, 580–591.
- Bannai, M.H., (1998). Soil moisture and temperature map, *Soil and Water Research Institute of Iran*.
- Bousbih, S., Zribi, M., Pelletier, C., Gorrab, A., Lili-Chabaane, Z., Baghdadi, N., Ben Aissa, N., & Mougenot, B., (2019). Soil texture estimation using radar and optical data from sentinel-1 and sentinel-2. *Remote Sensing*, 11, 1520.
- Brown, D.J., & Shepherd K.D., Walsh M.G., Mays M.D., and Reinsch, T.G., (2006). Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma*, 132(3-4), 273-290.
- Camera, C., Zomeni, Z., Noller, J. S., Zissimos, A. M., Christoforou, I. C., & Bruggeman, A., (2017). A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. *Geoderma*, 285, 35–49.
- Castaldi, F.; Palombo, A.; Santini, F.; Pascucci, S.; Pignatti, S.; Casa, R., (2016). Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. *Remote Sensing. Environment*, 179, 54–65.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Bohner, J., (2015). System for Automated Geoscientific analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* 8, 1991–2007.
- Dharumarajan S, Hegde R., (2022). Digital mapping of soil texture classes using Random Forest classification algorithm. *Soil Use Manage* 38:135–149.
- Douaoui, A.E.K., Nicolas, H., Walter, C., (2006). Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma* 134, 217–230.
- Foody, G.M., Cutler, M.E., McMorro, J., Pelz, D., Tangki, H., Boyd, D.S., Douglas, I., (2001). Mapping the biomass of Bornean tropical rain forest from remotely sensed data. *Glob. Ecol. Biogeogr.* 10, 379–387.
- Friedman, J.H., Meulman, J.J., (2003). Multiple additive regression trees with application in epidemiology. *Stat. Med.* 22 (9), 1365–1381.
- Gomez, C., Dharumarajan, S., Féret, J. B., Lagacherie, P., Ruiz, L., & Sekhar, M., (2019). Use of Sentinel-2 time-series images for classification and uncertainty analysis of inherent biophysical property: Case of soil texture mapping. *Remote Sensing*, 11, 565.
- Gomez, C.; Adeline, K.; Bacha, S.; Driessen, B.; Gorretta, N.; Lagacherie, P.; Roger, J.M.; Briottet, X., (2018). Sensitivity of clay content prediction to spectral configuration of VNIR/SWIR imaging data, from multispectral to hyperspectral scenarios. *Remote Sensing. Environment*, 204, 18–30.
- Grinand, C., Arrouays, D., Laroche, B., Martin, M.P., (2008). Extrapolating regional soil land-scapes from an existing soil map: sampling intensity, validation procedures and integration of spatial context. *Geoderma* 143, 180–190.
- Hengl, T., de Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotic, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B., (2017). Soil grids 250m: Global gridded soil information based on machine learning. *PLoS One*, 12, e0169748.
- Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., and Schmidt, M. G., (2016): An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping, *Geoderma*, 265, 62-77.
- Khaledian, Y., & Miller, B.A., (2020). Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, 81, 401–418.
- Khan, N.M., Rastokuev, V.V., Sato, Y., Shiozawa, S., (2005). Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agric. Water Manag.* 77, 96–109.
- Li, Q., Meng, Q., Cai, J., Yoshino H., and Mochida, A., (2009). Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*. 86: 2249-2256
- Liao, K., Xu, S., Wu, J., & Zhu, Q., (2013). Spatial estimation of surface soil texture using remote sensing data. *Soil Science Plant Nutrient*, 59(4), 488–500.

- McBratney, A. B., Mendonça-Santos, M. L., & Minasny, B., (2003). On digital soil mapping. *Geoderma*, 117, 3–52.
- Mehrabi-Gohari, E., Matinfar, H. R., Jafari, A., Taghizadeh-Mehrjardi, R., & Triantafilis, J., (2019). The spatial prediction of soil texture fractions in arid regions of Iran. *Soil System*, 3(4), 65.
- Momtaz, H. R., A. A. Jafarzadah, H. Torabi, S. Oustan, A. Samadi, N. Davatgar and R. J. Gilkes., (2009). An assessment of the variation in soil properties within and between landforms in Amol region, Iran. *Geoderma* 149: 10-18.
- Mountrakis, G.; Im, J.; Ogole, C., (2011). Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sensing*, 66, 247–259.
- Mousavi, S. R., Sarmadian, F., & Rahmani, A., (2020). Modeling and Prediction of Soil Classes Using Boosting Regression Tree and Random Forests Machine Learning Algorithms in Some Part of Qazvin Plain. *Iranian Journal of Soil and Water Research*, 50(10), 2525-2538 (In Persian).
- Okun, O., & Priisalu, H.; (2007). Random forest for gene expression based cancer classification: Overlooked issues. In J. Marti et al., (Eds.). Pattern recognition and image analysis: *Third Iberian Conference*, IbPRIA 2007, Girona, Spain. 6–8 June. Springer, Berlin. p. 483–490
- Pahlavan-Rad, M.R., Akbarimoghaddam, A., (2018). Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). *Catena* 160, 275–281
- Parviz, L.; (2017). Evaluation the Preprocessing Effect of Satellite Images Input Parameters in to Artificial Neural Network for Soil Texture Determination. *Iranian Journal of Applied Soil Research*, 5(2), 66-80 (In Persian).
- Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., & Thompson, J.; (2018). Soil property and class maps of the conterminous United States at 100-meter spatial resolution. *Soil Science Society of America Journal*, 82(1), 186–201.
- Rondeaux, Geneviève, et al.; (1996). Optimization of Soil-Adjusted Vegetation Indices. *Remote Sensing of Environment*, vol. 55, no. 2, 95–107.
- Taalab, K., Corstanje, R., Zawadzka, J., Mayr, T., Whelan, M. J., Hannam, J. A., & Creamer, R.; (2015). On the application of Bayesian networks in digital soil mapping. *Geoderma*, 259, 134–148.
- Taghizadeh-Mehrjardi, R., Sarmadian, F., Minasny, B., Triantafilis, J., & Omid, M.; (2014). Digital mapping of soil classes using decision tree and auxiliary data in the Ardakan region, Iran. *Arid Land Research and Management*, 28, 147–168.
- Taghizadeh-Mehrjardi, R., Toomanian, N., Khavaninzadeh, A. R., Jafari, A., & Triantafilis, J.; (2016). Predicting and mapping of soil particle-size fractions with adaptive neuro-fuzzy inference and ant colony optimization in central Iran. *European Journal of Soil Science*, 67, 707–725.
- Taghizadeh-Mehrjardi, R., Mahdianpari, M., Mohammadimanesh, F., Behrens, T., Toomanian, N., Scholten, T., & Schmidt, K.; (2020). Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in central Iran. *Geoderma*, 376, 114552.
- Vagen, T.G., Leigh, A., Winowiecki, L.A., Tondoh, J.E., Desta, L.T., Gumbrecht, T., (2016). Mapping of soil properties and land degradation risk in Africa using MODIS re-flectance. *Geoderma* 263, 216–225
- Vapnik, V.N. The Nature of Statistical Learning Theory; *Springer-Verlag: Berlin/Heidelberg, Germany*, (1995); Volume 8, ISBN 0387945598.
- Wang, B., Waters, C., Origill, S., Gray, J., Cowie, A., Clark, A., Liu, D.L., (2018). High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Science of the Total Environment*. 630, 367–378.
- Wilding, L.; (1985). Spatial variability: its documentation, accommodation and implication to soil surveys. pp. 166-194.
- Wu, W., Li, A.-D., He, X.-H., Ma, R., Liu, H.-B., & Lv, J.-K.; (2018). A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. *Computers and Electronics in Agriculture*, 144, 86–93.
- Yang, R.M., Zhang, G.L., Liu, F., Lu, Y.Y., Yang, F., Yang, F., Yang, M., Zhao, Y.G., & Li, D.C.; (2016). Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological Indices*, 60, 870–878.
- Zhang, M., & Shi, W.; (2019). Systematic comparison of five machine-learning methods in classification and interpolation of soil particle size fractions using different transformed data. *Hydrology and Earth System Sciences Discussion*; 2018-584
- Zhang, Y., Sui, B., Shen, H., and Ouyang, L.; (2019). Mapping stocks of soil total nitrogen using remote sensing data, A comparison of random forest models with different predictors. *Computers and Electronics*



in Agriculture, 160: 23-30

Zolfaghari, A. A., Yazdani, M. R., Khosravi. M. and Mahmoudi S. M.; (2019). Comparison of Different Data Mining Methods for Digital Mapping of Soil Particle-size Fractions in Lands of Semnan Plain. *Iranian Journal of Soil and Water Research*, 51(2), 375-385 (In Persian).