

## Comparison of Different Data Mining Methods for Digital Mapping of Soil Particle-size Fractions in Lands of Semnan Plain

ALI ASGHAR ZOLFAGHARI<sup>\*1</sup>, MOHAMMADREZA YAZDANI<sup>1</sup>, MAHIN KHOSRAVI<sup>1</sup>, SEYED MEHDI MAHMOUDI<sup>2</sup>

1. Department of Management of Arid Areas, Faculty of Desertification, University of Semnan, Semnan, Iran.

2. Department of Statistics, Faculty of Basic Sciences, University of Semnan, Semnan, Iran.

(Received: Aug. 13, 2019- Revised: Sep. 27, 2019- Accepted: Oct. 5, 2019)

### ABSTRACT

Knowledge about the spatial distribution of particle-size fractions in different areas is required for various land management applications and resources, modeling, and monitoring practices. In recent years, with the advancement of data mining methods and the availability of cheap data from satellite imagery, digital soil mapping methods have been developed to predict the spatial distribution of primary soil particles. The objective of this study was to conduct a spatial prediction of particle-size fractions such as clay, sand and silt using digital soil mapping in agricultural lands in Semnan. To achieve these goals, a total of 84 soil samples were collected from 0 to 20 cm of soil surface. Also, the environmental variables were obtained from OLI Satellite Landsat to make dependence with soil particles. In this study a linear model such as Partial Least Squares (PLS) and two non-linear models, including Random Forest (RF) and Stochastic Gradient Boosting Machin (GBM) were used for spatial prediction of particle-size fractions. The models were calibrated and validated by the 10-fold cross-validation methods. Three statistics, such as Root Mean Squared Error (RMSE), Coefficient of determination ( $R^2$ ), and Mean Absolute Error (MAE) were used to determine the performance of the investigated models. Values of RMSE,  $R^2$ , and MAE statics of RF model for prediction of sand, silt and clay were (15.6, 0.35, 12.62), (11.49, 0.33, and 9.34), and (8.42, 0.28, and 5.9), respectively. These results indicated that the most accurate model for the prediction of particle-size fractions was the RF model. Also, the results showed that the most important environmental covariates for predicting particle-size fractions were band 10 (B10), band 5 (B5), and the gypsum index (GI). This indicated that the variables containing the near-infrared and infrared thermal waves had a major contribution to explaining the spatial patterns of particle-size fractions.

**Keywords:** Agricultural lands, Data mining, Digital soil mapping.

## مقایسه روش‌های مختلف داده‌کاوی برای نقشه‌برداری رقومی ذرات اولیه خاک در اراضی دشت سمنان

علی اصغر ذوالفقاری<sup>۱\*</sup>، محمدرضا یزدانی<sup>۱</sup>، مهین خسروی<sup>۱</sup>، سید مهدی محمودی<sup>۲</sup>

۱. گروه مدیریت مناطق خشک، دانشکده کویرشناسی، دانشگاه سمنان، سمنان، ایران

۲. گروه آمار، دانشکده علوم پایه، دانشگاه سمنان، سمنان، ایران

(تاریخ دریافت: ۱۳۹۸/۵/۲۲ - تاریخ بازنگری: ۱۳۹۸/۷/۵ - تاریخ تصویب: ۱۳۹۸/۷/۱۳)

### چکیده

آگاهی از نحوه توزیع فضایی اندازه ذرات اولیه خاک برای مدیریت اراضی، مدیریت منابع، اجرای برنامه‌ها و مدل‌سازی دیگر خصوصیات خاک ضروری است. در سال‌های اخیر با پیشرفت‌های روش‌های داده‌کاوی و با در اختیار بودن داده‌های ارزان قیمت حاصل از تصاویر ماهواره‌ای، روش‌های نقشه‌برداری رقومی خاک برای پیش‌بینی توزیع فضایی ذرات اولیه خاک به وفور مورد استفاده قرار گرفته است. لذا هدف این تحقیق، پیش‌بینی فضایی ذرات اولیه خاک از قبیل رس، شن و سیلت با استفاده از نقشه‌برداری رقومی خاک در اراضی کشاورزی دشت سمنان می‌باشد. همچنین بررسی کارایی سه روش داده‌کاوی برای تهیه نقشه رقومی ذرات خاک از دیگر اهداف این مطالعه است. برای رسیدن به این اهداف، مجموع ۸۴ نمونه خاک از عمق ۰ تا ۲۰ سانتی‌متر جمع‌آوری شدند. همچنین متغیرهای محیطی با استفاده از تصاویر سنجنده OLI ماهواره لندست برای برقراری ارتباط با ذرات خاک استخراج گردید. در این مطالعه از مدل خطی حداقل مربعات جزئی (PLS) و دو مدل غیرخطی شامل جنگل تصادفی (RF)، ماشین تقویت‌شده گرادین (GBM) جهت ارتباط میان متغیرهای محیطی و ذرات اولیه خاک استفاده شد. مدل‌های مورد مطالعه با استفاده از روش اعتبارسنجی متقابل و ارزیابی شدند. به منظور بررسی کارایی مدل‌های مختلف داده‌کاوی از آماره‌های میانگین ریشه‌ی مربعات خطا (RMSE)، ضریب تبیین ( $R^2$ ) و میانگین قدر مطلق خطا (MAE) استفاده شد. بر اساس نتایج RMSE،  $R^2$  و MAE مدل RF با مقادیر این آماره‌ها به ترتیب برای ذرات شن (۱۵/۶۰، ۰/۳۵ و ۱۲/۶۲)، سیلت (۱۱/۴۹، ۰/۳۳ و ۹/۳۴) و برای ذرات رس (۸/۴۲، ۰/۲۸ و ۵/۹) بودند، این نتایج نشان داد که مدل RF به نسبت مدل‌های PLS و GBM دارای کارایی و دقت بیشتری در پهنه‌بندی ذرات اولیه خاک است. نتایج نشان داد که مهم‌ترین متغیرهای محیطی، برای پیش‌بینی ذرات اولیه خاک باندهای ۱۰، ۵ و شاخص گج (GI) می‌باشند. بنابراین، متغیرهایی که دارای طیف مادون‌قرمز نزدیک و مادون‌قرمز حرارتی بودند، سهم عمده‌ای در توصیف مکانی ذرات خاک را بر عهده داشتند.

واژه‌های کلیدی: اراضی کشاورزی، داده‌کاوی، نقشه‌برداری رقومی.

### مقدمه

نقشه‌های فضایی خصوصیات خاک با استفاده از روش‌های معمول وقت‌گیر و گران است (Forkuor et al., 2017). بنابراین استفاده از روش‌های جدید داده‌کاوی برای پیش‌بینی مکانی خصوصیات خاک با استفاده از متغیرهای کمکی استخراج‌شده از تصاویر ماهواره‌ای می‌تواند مشکلات ذکر شده را برطرف نماید.

در سال‌های اخیر، روش‌های نقشه‌برداری رقومی با برقراری ارتباط بین متغیرهای محیطی و خصوصیات خاک برای تهیه نقشه‌های مکانی، توسط محققین مختلفی ارائه و استفاده شده است. به‌عنوان مثال Scudiero et al. (2014) با استفاده از داده‌های تصاویر ماهواره‌ای به این نتیجه رسیدند که از میان متغیرهای مختلف محیطی داده‌های چندین ساله تصاویر ماهواره لندست یک شاخص مفید برای توصیف تنوع فضایی شوری خاک مناسب می‌باشند. مطالعات دیگر نیز از داده‌های سنجش از دور (RS) برای

بافت خاک یکی از مهم‌ترین خصوصیات خاک است که به‌عنوان درصد نسبی مقادیر رس، شن و سیلت تعریف می‌شود. بافت خاک عامل مهم و تأثیرگذار بر رفتار فیزیکی و شیمیایی خاک از قبیل ظرفیت نگهداری آب، ظرفیت تبادل کاتیونی، حاصلخیزی و باروری خاک و زهکشی خاک است (Makabe et al., 2009). به همین علت نقشه‌های رقومی خاک به‌طور گسترده‌ای برای ارزیابی توزیع فضایی ذرات اولیه خاک در مناطق مختلف Minasny and Hartemink (2011)؛ Taghizadeh-Mehrjardi et al. (2016a)؛ Forkuor et al. (2017) مورد استفاده قرار گرفته‌اند. روش‌های نقشه‌برداری سنتی به‌ندرت اطلاعات مربوط به توزیع فضایی اندازه ذرات خاک را بر اساس قدرت تفکیک مکانی در حد یک پیکسل را ارائه می‌دهد (McBratney et al., 2003). علاوه بر این، تهیه

al. (2017) انجام شد، این محققین نشان دادند که مدل جنگل تصادفی از توانایی بالایی برای پیش‌بینی خصوصیات خاک در مناطق خشک و نیمه‌خشک برخوردار است.

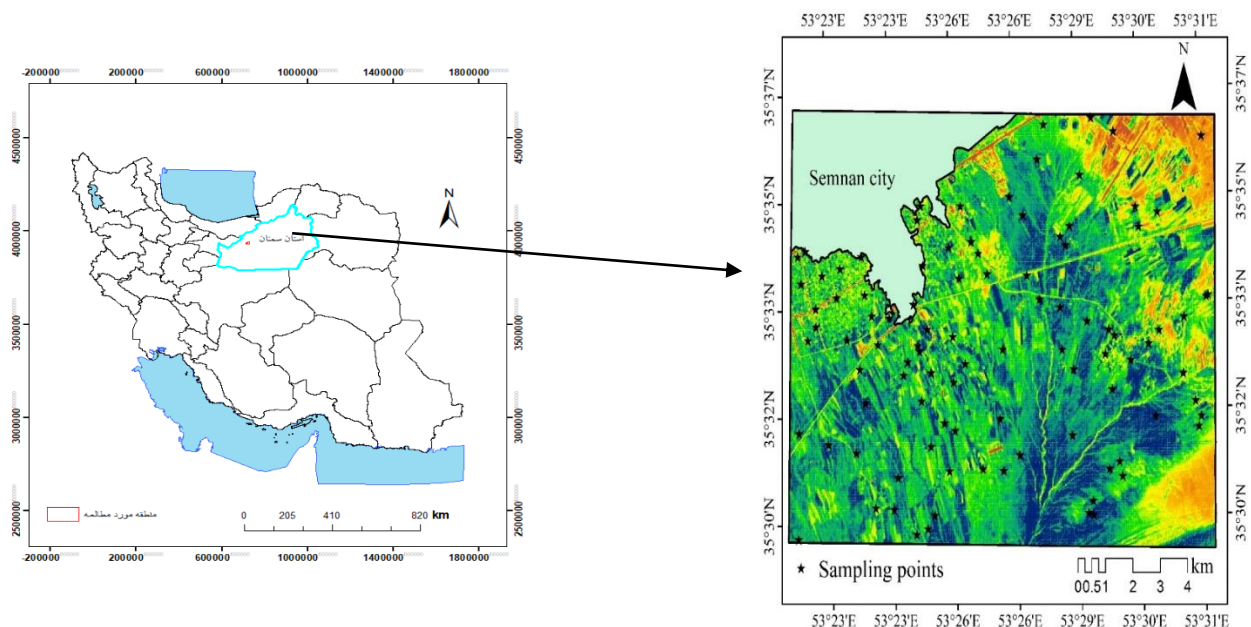
اما مطالعات اندکی به‌منظور مقایسه روش‌های مختلف داده‌کاوی را برای مدل‌سازی مکانی توزیع اندازه ذرات خاک در اراضی کشاورزی مناطق خشک کشور انجام شده است. لذا هدف اصلی از این مطالعه، پیش‌بینی پراکنش فضایی ذرات اولیه خاک و تهیه نقشه‌های رقومی هر یک از ذرات خاک با استفاده از مناسب‌ترین روش داده‌کاوی در اراضی کشاورزی دشت سمنان است. همچنین تعیین مؤثرترین متغیر محیطی در توزیع فضایی ذرات اولیه خاک به‌منظور تفسیر متغیرها و اثر آن‌ها در دقت روش‌های داده‌کاوی نیز از دیگر اهداف این مطالعه است.

### مواد و روش‌ها

#### منطقه مورد مطالعه و نقاط نمونه‌برداری

منطقه مورد مطالعه در اراضی کشاورزی و مرتعی دشت سمنان واقع شده و وسعتی برابر با ۱۲۶۳۰ هکتار از اراضی دشت سمنان را شامل می‌شود. شکل (۱) موقعیت منطقه مورد مطالعه و نقاط نمونه‌برداری را نشان می‌دهد. کاربری منطقه در اراضی کشاورزی عمدتاً گندم‌زار است که به‌صورت آبی کشت می‌شود همچنین در عرصه منابع طبیعی نیز کاربری منطقه مراتع با پوشش ضعیف می‌باشند. میانگین دما و بارندگی سالانه منطقه مورد مطالعه به ترتیب ۱۸/۵ درجه سلسیوس و ۱۳۸ میلی‌متر است.

تهیه نقشه‌های مکانی خصوصیتی از قبیل درصد شن، سیلت، رس و کربن آلی خاک (SOC) توسط (Taghizadeh- (2016b) و Forkuor et al. (2017)؛ Mehrjardi et al. (2014) استفاده کرده‌اند. تکنیک‌های مدل‌سازی در نقشه‌برداری رقومی ذرات اولیه خاک عمدتاً به‌منظور ارتباط بین ذرات اولیه خاک و متغیرهای محیطی از قبیل داده‌های سنجش از دور به کار گرفته می‌شوند. رابطه بین متغیرهای هدف (مانند شن، رس و سیلت) و متغیرهای محیطی می‌تواند یک رابطه خطی تک متغیره، یک رابطه خطی چند متغیره، یک توابع بسیار پیچیده غیرخطی و یا یک الگوریتم پیچیده مانند الگوریتم جنگل تصادفی باشد. مزیت تکنیک‌های مدل‌سازی بر اساس الگوریتم‌ها و توابع اما و اگرها (if- then) این است که این الگوریتم‌ها توانایی تشخیص روابط غیرخطی دارند و بنابراین برای نشان دادن روابط بین خصوصیات خاک و متغیرهای محیطی مناسب می‌باشند (Taghizadeh-Mehrjardi et al., 2016a). روش‌های یادگیری ماشین مانند رگرسیون خطی، رگرسیون‌های غیرخطی و رگرسیون‌های درختی از قبیل جنگل تصادفی به‌وسیله (Hengl et al. (2015)؛ Vaysse and Lagacherie., (2015)؛ Forkuor et al. (2017) برای پیش‌بینی خصوصیات خاک و متغیرهای محیطی بکار گرفته شده است. نتایج بدست آمده توسط این محققین نشان داد که روش‌های غیرخطی مانند جنگل تصادفی از توانایی بیشتری برای شبیه‌سازی خصوصیات خاک نسبت به مدل‌های خطی مانند حداقل مربعات جزئی برخوردار هستند. به‌عنوان مثال در مطالعه‌ای که توسط Zeraatpisheh et



شکل ۱. موقعیت منطقه مورد مطالعه و توزیع نقاط نمونه‌برداری با استفاده از روش هایپرکیوب لاتین

اولیه خاک استفاده شد (Gee and Bauder., 1986). به این منظور خاک‌ها ابتدا با هوا خشک و نرم شده، و سپس از غربال دو میلی-متری عبور داده شدند. حذف مواد آلی در نمونه‌هایی که بیش از ۲ درصد ماده آلی داشتند با استفاده از محلول ۳۰ درصد آب‌اکسیژنه (H<sub>2</sub>O<sub>2</sub>) صورت گرفت. سپس جرم معینی از هر یک از این نمونه‌ها با محلول پراکنش (هگزامتا فسفات سدیم ۰.۵٪) تیمار شدند. اعداد قرائت شده روی هیدرومتر متناسب با حجم مایع جابجا شده خواهد بود. بنابراین پس از ۴۰ ثانیه قرائت اول هیدرومتر برای تعیین درصد شن خاک و قرائت ۶ ساعت برای تعیین مقدار رس خاک مورد استفاده قرار گرفت.

#### متغیرهای محیطی و انتخاب داده‌ها

در این مطالعه، برای توسعه مدل‌های داده‌کاوی، داده‌های سنجش‌ازدور (RS) به عنوان پیش‌بینی کننده در نظر گرفته شدند. داده‌های RS از تصاویر بدون ابر سنجنده OLI ماهواره لندست ۸ در ماه مه ۲۰۱۵ به دست آمدند. متغیرهای محیطی شامل سه باند مرئی (باندهای ۲، ۳ و ۴)، مادون قرمز نزدیک (باند ۵)، مادون قرمز کوتاه (باندهای ۶ و ۷)، مادون قرمز بلند (باند ۱۰) و شاخص‌هایی که در جدول (۱) ذکر شده‌اند با وضوح فضایی ۳۰ متر به ۳۰ متر مورد استفاده قرار گرفتند. در نهایت، ۱۸ متغیر محیطی از تصاویر OLI بدست آمده به عنوان متغیرهای کمکی مورد استفاده قرار گرفت شد (جدول ۱).

در این مطالعه برای تعیین نقاط نمونه‌برداری از روش هایپرکیوب لاتین<sup>۱</sup> استفاده شد. این روش نمونه‌برداری به عنوان یک طرح نمونه‌برداری برای نقشه‌برداری رقومی خاک در جایی که هیچ نمونه خاک برداشت‌شده قبلی موجود نیست (فقط اطلاعات کمکی موجود هستند) مورد استفاده قرار می‌گیرد (Minasny and McBratney, 2006). برای اجرا و تعیین نقاط نمونه‌برداری با استفاده از این روش، احتیاج به متغیرهای محیطی است. این متغیرهای محیطی می‌توانند از تصاویر ماهواره‌ای و یا مشتقات حاصل از نقشه رقومی ارتفاع باشند. در این مطالعه باندها و شاخص‌های تصویر ماهواره‌ای لندست ۸ که بیشترین ارتباط را به توزیع ذرات خاک دارند، به عنوان متغیرهای محیطی در تعیین نقاط نمونه‌برداری با استفاده از هایپرکیوب لاتین استفاده شد. مطالعات گذشته نشان داد که شاخص‌های رس<sup>۲</sup>، کربنات<sup>۳</sup>، روشنایی<sup>۴</sup>، شاخص شدت<sup>۵</sup>، پوشش گیاهی نرمال شده<sup>۶</sup> و شوری نرمال شده<sup>۷</sup> و باند ۷ (B7) با خصوصیات توزیع ذرات خاک مرتبط است (Taghizadeh-Mehrjardi *et al.*, 2016a). لذا در این مطالعه از خصوصیات مذکور به عنوان متغیرهای ورودی برای تعیین نقاط نمونه‌برداری استفاده شد. در نهایت با استفاده از روش مذکور تعداد ۸۴ نقطه انتخاب و نمونه‌برداری از عمق ۲۰-۰ سانتی‌متر خاک انجام شد.

در این مطالعه از روش هیدرومتر برای تعیین درصد ذرات

جدول ۱. متغیرهای محیطی مورد استفاده برای پیش‌بینی ذرات اولیه خاک

ردیف	شرح	متغیرهای محیطی
۱	باند ۲ (آبی مرئی) (عرض باند = ۰,۴۵۰-۰,۵۱۵ μm)	باند ۲ ماهواره لندست (B2)
۲	باند ۳ (سبز مرئی) (عرض باند = ۰,۵۲۵-۰,۶۰۰ μm)	باند ۳ ماهواره لندست (B3)
۳	باند ۴ (قرمز مرئی) (عرض باند = ۰,۶۳۰-۰,۶۸۰ μm)	باند ۴ ماهواره لندست (B4)
۴	باند ۵ (مادون قرمز نزدیک) (عرض باند = ۰,۸۴۵-۰,۸۸۵ μm)	باند ۵ ماهواره لندست (B5)
۵	باند ۶ (طول موج کوتاه مادون قرمز-۱) (عرض باند = ۱,۵۶۰-۱,۶۶۰ μm)	باند ۶ ماهواره لندست (B6)
۶	باند ۷ (طول موج کوتاه مادون قرمز ۲) (عرض باند = ۲,۱۰-۲,۳۰ μm)	باند ۷ ماهواره لندست (B7)
۷	باند ۱۰ (طول موج بلند مادون قرمز) (عرض باند = ۱۰,۳۰-۱۱,۳۰ μm)	باند ۱۰ ماهواره لندست (B10)
۸	(مادون قرمز نزدیک - قرمز) / (مادون قرمز نزدیک + قرمز)	شاخص پوشش گیاهی نرمال شده (NDVI)
۹	(مادون قرمز نزدیک / قرمز)	شاخص پوشش گیاهی (RVI)
۱۰	(کوتاه موج IR-1 / موج کوتاه IR-2)	شاخص رس (CI)
۱۱	(قرمز) / (سبز)	شاخص کربنات (CrI)
۱۲	((قرمز) ۲ + (مادون قرمز نزدیک) ۲) ۰,۵	شاخص روشنایی (BI)
۱۳	(مادون قرمز کوتاه IR-1) / (مادون قرمز کوتاه IR-1 + مادون قرمز نزدیک)	شاخص گچ (GI)
۱۴	((سبز) ۲ + (قرمز) ۲) ۰,۵	شاخص شدت (IN)
۱۵	(IR-2 موج کوتاه IR-1 - موج کوتاه) / (کوتاه موج IR-1 + IR-2 موج کوتاه)	شوری اختلاف طبیعی (NDSI)
۱۶	(قرمز - به مادون قرمز نزدیک) / (قرمز + مادون قرمز نزدیک)	شاخص شوری (SI)

6 - normalized difference vegetation index(NDVI)

7 - normalized difference salinity index(NDSI)

1 - Latin hypercube

2 - clay index(CI)

3 - carbonate index(CrI)

4 - brightness index(BI)

5 - intensity index (IN)

جدید به مدل قبلی اضافه شده و این الگوریتم با توجه به توجه به نظر کاربران تکرار می‌شود.

جنگل تصادفی (RF) یک الگوریتم ترکیبی است که از چندین درخت تصمیم برای الگوریتم خود استفاده می‌کند. در واقع مجموعه‌ای از درخت‌های تصمیم، با هم یک جنگل را تولید می‌کنند و این جنگل می‌تواند نتایج بهتری را (نسبت به یک درخت) اتخاذ نماید. در روش RF مجموعه داده‌ها را به  $n$  نشان‌دهنده تعداد درختان است تقسیم می‌کنند سپس به هر زیرمجموعه از داده‌ها یک درخت تصمیم ساخته می‌شود Hastie (2005). نتیجه نهایی با استفاده از میانگین‌گیری از تمامی درختان به دست خواهد آمد. روش RF به تنظیمات زیاد نیاز ندارد (Kuhn and Johnson, 2013). در مطالعه حاضر، تعداد درختان به ۱۰۰۰ و حداقل تعداد گره هر درخت برابر با پنج در نظر گرفته شد.

در این مطالعه برای انتخاب بهترین مدل در پیش‌بینی ذرات اولیه خاک در مرحله آموزش مدل، ۸۰ درصد داده‌ها برای آموزش مدل و ۲۰ درصد برای آزمون مدل مورد استفاده قرار گرفتند. در نهایت پس از تعیین بهترین مدل تمام داده‌ها برای تهیه نقشه پراکنش ذرات اولیه خاک مورد استفاده قرار گرفته شد.

#### تعیین ارتباط بین متغیرها محیطی

جهت تعیین ارتباط بین متغیرهای محیطی و ذرات اولیه خاک از تابع ارزیابی اهمیت استفاده شد. سپس تأثیرگذارترین متغیرها در هر مدل داده کاوی تعیین شد. در این توابع ارزیابی اهمیت متغیرها بیشتر به عملکرد مدل‌ها مرتبط می‌شوند و قادر به ترکیب ساختار همبستگی بین پیش‌بینی‌کننده‌ها در محاسبه اهمیت می‌باشند. اندازه‌گیری اهمیت متغیرها از موضوعات بسیار مهم در مدل‌های داده کاوی است (Genuer et al., 2010) که به ما در تعیین متغیرهای تأثیرگذار محیطی کمک می‌کند (Nauman and Thompson., 2014) و همچنین برای تفسیر متغیرها و اثر آنها در دقت مدل کاربرد دارد (Genuer et al., 2010). در این مطالعه تمام مدل‌سازی و تعیین اهمیت متغیرها با استفاده از بسته «Caret» در محیط RStudio (R Development Core Team, 2015) انجام شد.

#### ارزیابی کارایی مدل‌ها

جهت ارزیابی کارایی مدل‌های داده کاوی در برآورد ذرات اولیه خاک از آماره‌های ضریب تبیین<sup>۶</sup> ( $R^2$ ) (رابطه ۲)، ریشه دوم

#### روش‌های داده کاوی

در این مطالعه به منظور تعیین توزیع مکانی ذرات اولیه خاک از مدل خطی حداقل مربعات جزئی<sup>۱</sup> (PLS) و دو مدل یادگیری غیرخطی شامل جنگل تصادفی<sup>۲</sup> (RF) و روش ماشین‌گرادیان تقویت‌شده<sup>۳</sup> (GBM) استفاده شد. در انتها مدلی که دارای بیشترین دقت در برآورد ذرات اولیه خاک بود، برای تهیه نقشه نهایی توزیع مکانی ذرات خاک مورد استفاده قرار گرفت.

مدل حداقل مربعات جزئی (PLS)، یک روش آماری نظارت شده است که ارتباط بین مؤلفه‌های اصلی<sup>۴</sup> و متغیر هدف را با استفاده از معادله رگرسیون خطی نشان می‌دهد. این روش نسخه نظارت‌شده رگرسیون چندمتغیره خطی مؤلفه‌های اصلی است. این روش زمانی مورد استفاده قرار می‌گیرد که تعداد متغیرهای ورودی زیاد بوده و همچنین وجود همبستگی بالا بین متغیرهای ورودی (متغیرهای محیطی) وجود داشته باشد. روش PLS متغیرهای ورودی را به گونه‌ای تغییر می‌دهد که متغیرهای جدید دارای کمترین همبستگی با یکدیگر بوده اما دارای بیشترین همبستگی با متغیر هدف (Y) داشته باشند.

در نهایت در روش PLS رابطه خطی بین مؤلفه‌های اصلی و متغیر هدف (Y) با استفاده از رابطه (۱) که نشان‌دهنده یک معادله رگرسیون خطی چند متغیره می‌باشد، برقرار می‌کند.

$$Y = XB + \varepsilon \quad (\text{رابطه ۱})$$

روش ماشین‌گرادیان تقویتی (GBM) یک روش داده کاوی پارامتری است که می‌تواند روابط غیرخطی و خطی را اداره کند (Myles et al., 2004). روش تقویتی یا Boosting، ابتدا با استفاده از رگرسیون درختی مقادیر متغیر هدف (Y) پیش‌بینی می‌شود، معمولاً رگرسیون درختی منفرد برای پیش‌بینی متغیرهای غیرخطی مانند خصوصیات خاک الگوریتم‌های ضعیفی محسوب می‌شوند. سپس در روش تقویتی مشکلات رگرسیون‌های خطی ضعیف تقویت‌شده و نتایج پیش‌بینی بهبود می‌یابد. برای اجرای ماشین‌گرادیان تقویتی این مراحل انجام می‌گیرد. انتخاب یک تابع ضرر<sup>۵</sup> (به عنوان مثال تابع مربعات خطا در رگرسیون) و مدل‌های آموزش ضعیف (درخت تصمیم)، در مرحله بعد الگوریتم با استفاده حداقل کردن تابع ضرر مدل‌های آموزش را بهینه می‌کند. در مرحله بعد، باقیمانده‌ها (تفاوت بین مقادیر اندازه‌گیری شده و برآورد شده) محاسبه شده و یک مدل بر روی باقیمانده برآورد داده می‌شود به گونه‌ای که باقیمانده‌ها حداقل شوند. سپس مدل

5. Loos functions

6. Coefficient of determination

1. Partial Least Squares (PLS)

2. Random forest (RF)

3. Stochastic Gradient Boosting Machin (GBM)

4. principle component

میانگین مربعات خطا ( $RMSE^1$ ) (رابطه ۳) و میانگین قدر مطلق خطا ( $MAE^2$ ) (رابطه ۴) استفاده می‌گردد که روابط ریاضی آن‌ها در زیر آمده است.

$$R^2 = \frac{|\sum_{i=1}^n (Q_i - \bar{Q}_i)(P_i - \bar{P}_i)|^2}{\sum_{i=1}^n (Q_i - \bar{Q}_i)^2 \sum_{i=1}^n (P_i - \bar{P}_i)^2} \quad (\text{رابطه ۲})$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - Q_i)^2}{n}} \quad (\text{رابطه ۳})$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - Q_i| \quad (\text{رابطه ۴})$$

$P_i$  و  $O_i$  به ترتیب مقادیر مشاهداتی و تخمین زده شده و  $n$  تعداد داده‌ها می‌باشد.

آماره‌های  $MAE$  و  $RMSE$  (معادلات ۳ و ۴) هر چه به صفر نزدیک‌تر باشند نشان‌دهنده دقت بیشتر است. مدل‌های درون‌یابی که دارای حداقل خطا و حداکثر ضریب تبیین است را به‌عنوان مناسب‌ترین مدل برای تهیه نقشه مکانی توزیع اندازه ذرات خاک انتخاب شد.

## نتایج و بحث

### توصیف آماری داده‌های اندازه‌گیری شده

نتایج مقادیر آماری داده‌های اندازه‌گیری شده برای ذرات خاک در جدول (۲) نشان داده شده است. مقدار رس خاک در منطقه مورد مطالعه بین ۱ تا ۳۹ درصد، سیلت در محدوده‌ی ۶ تا ۸۰ درصد و شن در محدوده‌ی ۵ تا ۹۳ درصد متغیر بودند. ضریب تغییرات نشان داد که میزان رس و شن به ترتیب دارای کمترین و بیشترین تغییرات هستند. ضریب تغییرات رس و سیلت برابر با ۱۰ درصد و ضریب تغییرات شن در منطقه مورد مطالعه برای با ۴۶ درصد بود.

جدول ۲. مقادیر آماری داده‌های اندازه‌گیری شده برای ذرات بافت خاک

حداقل	حداکثر	میانگین	انحراف معیار ذرات خاک	ضریب تغییرات (درصد)	
۱	۳۹	۱۶/۹۷	۸/۴۱	۱۰/۱	رس (%)
۶	۸۰	۴۴/۳۹	۱۳/۳۰	۱۰/۱	سیلت (%)
۵	۹۳	۳۸/۶۳	۱۷/۹۴	۴۶	شن (%)

### بررسی همبستگی بین درصد ذرات بافت خاک و خصوصیات باندها و شاخص‌های تصاویر ماهواره‌ای

به منظور شناسایی روابط بین درصد اندازه ذرات خاک با باندها و شاخص‌های سنجنده OLI ماهواره لندست، همبستگی بین این ذرات و ۱۸ متغیر مورد مطالعه تعیین شد. نتایج همبستگی بین ذرات خاک با متغیرها مورد مطالعه در جدول (۳) نشان داده شده است. نتایج نشان داد که ضریب همبستگی شن خاک با باند ۵ منفی و معنی‌دار ( $-0/۳۱^*$ ) بود در حالی که همبستگی بین شن

و باند ۱۰ مثبت و معنی‌دار ( $0/۳۸^*$ ) به‌دست آمد. این ضریب برای ذرات رس خاک با هیچ یک از باندها دارای رابطه معنی‌داری نبود. مقدار ضریب همبستگی سیلت خاک با باندهای ۵ و ۱۰ به ترتیب برابر با ( $0/۳۳^*$  و  $0/۳۶^*$ ) به‌دست آمد. همبستگی بین ذرات خاک با باندهای ۵ و ۱۰ قوی و با باند ۴ ضعیف تا خیلی ضعیف مشاهده گردید. نتایج ارتباط بین ذرات خاک با شاخص‌ها نشان داد که ضریب همبستگی شن خاک با شاخص روشنایی منفی و معنی‌دار ( $-0/۳۴^*$ ) می‌باشد در حالی که همبستگی بین شن و شاخص گچ مثبت و معنی‌دار ( $0/۲۸^*$ ) به‌دست آمد. این ضریب برای ذرات رس خاک با هیچ یک از شاخص‌ها رابطه‌ی معنی‌داری را نشان نداد. ضریب همبستگی برای سیلت خاک با شاخص‌های روشنایی و گچ به ترتیب اهمیت با مقادیر همبستگی حدود ( $0/۲۵^*$ ،  $0/۲۵^*$ ) مثبت و با شاخص گچ منفی بود. بیشترین مقدار ضریب همبستگی بین شاخص‌ها، مربوط به شاخص گچ با ذره شن خاک با همبستگی حدود ( $0/۲۸^*$ ) و کمترین مقدار همبستگی مربوط به شاخص شدت است که با رس خاک همبستگی حدود ( $0/۱۱$ ) را نشان داد. (Summers et al. (2011)؛ Forkuor et al. (2017) نتایج مشابهی را برای خصوصیات خاک در محدوده مادون‌قرمز بدست آوردند نتایج به‌دست آمده نشان داد که از روی میزان انعکاس در محدوده مادون‌قرمز نزدیک، مادون‌قرمز و قرمز می‌توان به همبستگی بین متغیرهای محیطی و خصوصیات خاک پی برد. همچنین، نتایج این مطالعه نشان داد که در میان ذرات خاک، مقادیر همبستگی میان ذرات شن و باندهای تصاویر ماهواره‌ای قوی بوده، در حالی این ضریب برای ذرات رس خاک به‌صورت منفی و ضعیف بوده است. این می‌تواند به دلیل وجود ظرفیت رطوبتی بالاتر ذرات رس نسبت به شن و بازتابش طیفی کمتر رس نسبت به بازتابش شن باشد که در نهایت باعث ایجاد رابطه معکوس بین باندها و ذرات رس و رابطه مستقیم باندها با ذرات شن شده است. هر چند بر طبق نتایج Bellinaso et al. (2010) مشخص شده است که هر چه اندازه ذرات کوچک‌تر باشد خاک نرم‌تر و میزان بازتابش بیشتر است. نتایج متضاد ما و افزایش شدت بازتابش در بافت‌های درشت‌تر را می‌توان به وجود ترکیبات خاک‌های شنی نسبت داد. نتایج این مطالعه نشان داد که طول موج‌های در محدوده مادون‌قرمز با درصد ذرات اولیه خاک ارتباط بیشتری را نشان می‌دهند، این نتایج با Curcio et al. (2013) که گزارش دادند که در محدوده مادون‌قرمز نزدیک و مادون‌قرمز کوتاه برای توزیع اندازه ذرات خاک مناسب است، همخوانی دارد.

جدول ۳. ضریب همبستگی بین ذرات اولیه خاک با باندها سنجنده OLI ماهواره لندست

ذرات اولیه خاک	باند ۱۰	باند ۷	باند ۶	باند ۵	باند ۴	باند ۳	باند ۲	باند ۱
شن	۰/۳۷*	-۰/۰۴۷	-۰/۰۷۳	۰/۳۰۹*	-۰/۰۷۹	-۰/۰۶۶	-۰/۰۶۸	-۰/۰۵۴
سیلت	-۰/۳۷*	۰/۰۱۶	۰/۰۶۲	۰/۳۳*	۰/۰۴۳	۰/۰۴۰	۰/۰۳۹	۰/۰۱۶
رس	-۰/۱۹۷	۰/۰۹۷	۰/۰۵۹	۰/۱۲۹	۰/۱۲۴	۰/۱۰۳	۰/۰۸۳	۰/۰۶۸

\* نشان دهنده معنی داری در سطح ۰/۰۵

جدول ۳. همبستگی بین ذرات اولیه خاک با شاخص های حاصل از سنجنده OLI ماهواره لندست

ذرات اولیه خاک	SI	RVI	NDVI	NDSI	IN	GI	CrI	CI	BI
شن	۰/۰۹۰	-۰/۰۸۴	-۰/۰۹۰	-۰/۰۱۳	-۰/۰۸۶	۰/۲۸	-۰/۱۰۳	-۰/۰۱۵	-۰/۲۵۳
سیلت	-۰/۱۶۷	۰/۱۵۷	۰/۱۶۷	۰/۱۰۸	۰/۰۴۳	-۰/۲۵	۰/۰۳۷	۰/۱۰۸	۰/۲۵
رس	۰/۰۷۳	-۰/۰۷۴	-۰/۰۷۳	-۰/۱۴۳	۰/۱۱۶	-۰/۱۲۸	۰/۱۶۰	-۰/۱۳۹	۰/۱۴۰

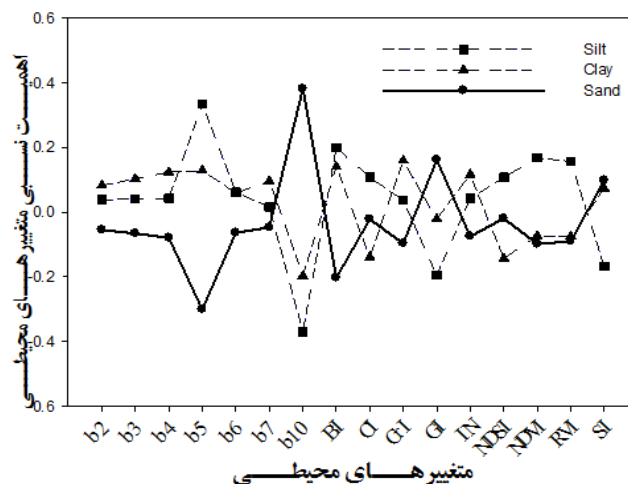
\* معنی داری در سطح ۰/۰۵ (توضیحات و نحوه محاسبه هر یک از شاخص ها در جدول (۱) ارائه شده است)

بیشترین توزیع فضایی میزان رس و سیلت را توضیح می دهد. بسیاری از مطالعات گزارش دادند شاخص های پوشش گیاهی اثرات قابل توجهی بر ذرات بافت خاک داشتند و منعکس کننده تغییرات ذرات بافت خاک بودند. در مطالعه ای که توسط Mahmoudabadi *et al.* (2017) انجام شد، گزارش دادند که با توجه به رابطه قوی بین خواص خاک و پوشش گیاهی، شاخص های پوشش گیاهی می توانند تغییرات خاک را منعکس کنند.

نتایج کلی این مطالعه نشان داد که میان متغیرهای محیطی مستخرج از تصویر ماهواره ای و ذرات اولیه خاک در اراضی کشاورزی همبستگی چندان بالایی وجود دارد، اما این متغیرها در پیش بینی نحوه توزیع ذرات بافت خاک مؤثر هستند. لذا این نتایج با نتایج بدست آمده توسط Khanamani *et al.* (2011) و Forkuor *et al.* (2016a) و Taghizadeh-Mehrjardi *et al.* (2017) مطابقت دارد.

### تعیین مهمترین متغیرهای محیطی در پیش بینی فضایی ذرات اولیه خاک

در مدل های RF افزایش میانگین خطای یک درخت، بیشترین اهمیت یک متغیر داده شده را نشان می دهد. در مدل خطی متغیرهای مهم بر اساس برخی شاخص ها مانند پراکندگی و واریانس متغیر تعیین می شوند (Genuer *et al.*, 2010). نتایج این مطالعه نشان داد که اهمیت نسبی متغیرهای محیطی در پیش بینی توزیع ذرات شن، سیلت و رس خاک با یکدیگر متفاوت بوده است. اهمیت نسبی متغیرها برای مدل منتخب در شکل (۲) ارائه شده است. باند ۱۰، باند ۵، شاخص GI، شاخص SI مؤثرترین متغیرها در پیش بینی محتوای شن خاک بودند. باندها و شاخص ها شامل باندهای ۱۰ و ۵ و شاخص های BI, SI، مهم ترین متغیرهای محیطی که مسئول پیش بینی محتوای رس و سیلت خاک بودند. علاوه بر این، شاخص های NDVI, RVI که پوشش گیاهی را نشان می دهند



شکل (۲) نمودار اهمیت نسبی باندها و شاخص های مختلف سنجنده (OLI) برای ذرات خاک (توضیحات و نحوه محاسبه هر یک از شاخص ها در جدول (۱) ارائه شده است)

جدول ۴. بهترین متغیرهای محیطی مهم برای پیش‌بینی ذرات شن، سیلت و رس خاک

متغیرهای مهم به ترتیب اهمیت					ذرات اولیه خاک
BI	SI	GI	B5	B10	شن
RVI	NDVI	BI	B5	B10	سیلت
RVI	NDVI	CrI	BI	B10	رس

توضیحات و نحوه محاسبه هر یک از شاخص‌ها در جدول (۱) ارائه شده است.

### بررسی کارایی مدل‌ها

در این مطالعه از اعتبارسنجی متقابل برای واسنجی مدل‌های استفاده شد. در این روش داده‌ها به ۱۰ قسمت مساوی تقسیم شدند. در هر بار اجرای برنامه، ۹۰ درصد داده‌ها برای آموزش و ۱۰ درصد داده‌ها برای آزمون استفاده شدند و در نهایت میانگین پارامتر مدل در ۱۰ بار اجرای برنامه به‌عنوان بهترین مدل برای اعتبارسنجی و پهنه‌بندی رقومی ذرات خاک استفاده شد.

نتایج مقایسه خطاهای پیش‌بینی مدل PLS نشان داد که ریشه مربعات خطا به ترتیب برای ذرات رس، شن و سیلت (۹/۲۱ و ۱۶/۱۸، ۱۲/۷۲) بود. همچنین قدرمطلق خطا (MAE) مدل PLS برای ذرات شن، سیلت و رس به ترتیب برابر با ۱۳/۳۹، ۹/۹۹ و ۷/۵۸ درصد می‌باشد. همچنین، نتایج ضریب تبیین نشان داد که تغییرات متغیر شن نسبت به سایر ذرات خاک بیشتر تحت تأثیر متغیرهای مستقل قرار داشته است. به عبارت دیگر باندها و شاخص‌های حاصل از تصاویر ماهواره‌ای، متغیرهای مناسبی برای برآورد ذرات شن خاک می‌باشند. مقادیر ضریب تبیین بر اساس مدل PLS برای ذرات رس، شن و سیلت به ترتیب برابر با ۰/۱۷، ۰/۳۰ و ۰/۲۴ بدست آمد (جدول ۵).

نتایج ریشه مربعات خطا در مدل RF برای ذرات رس، شن و سیلت به ترتیب برابر با ۸/۴۲ و ۱۵/۶۸، ۱۱/۴۹ بدست آمد این نتایج نشان داد که در پیش‌بینی ذرات خاک کمترین و بیشترین مقدار RMSE به ترتیب مربوط به رس و شن خاک است. نتایج مدل RF بر اساس آماره ضریب تبیین برای ذرات رس، شن و سیلت به ترتیب برابر ۰/۲۸، ۰/۳۵ و ۰/۳۳ است که نشان می‌دهد که مدل RF به نسبت مدل PLS دارای دقت بیشتری در برآورد ذرات اولیه خاک است (جدول ۵). این نتایج با نتایج Ryan et al. (2000)؛ Florinsky et al. (2002)؛ Malone et al. (2009) شباهت دارد. نتایج اعتبارسنجی مدل‌ها نشان داد که در بین مدل‌های مورد مطالعه مدل‌های RF، GBM بیش‌ترین دقت پیش‌بینی را برای برآورد شن دارا هستند (جدول ۵). هر دو RF و GBM عملکرد مشابهی را برای پیش‌بینی محتوای شن نشان دادند ( $R^2=0/35$ ) (جدول ۵). اگرچه بر اساس آماره  $R^2$  مدل RF به نسبت مدل GBM در برآورد سیلت ۸ درصد دقت بالاتری را

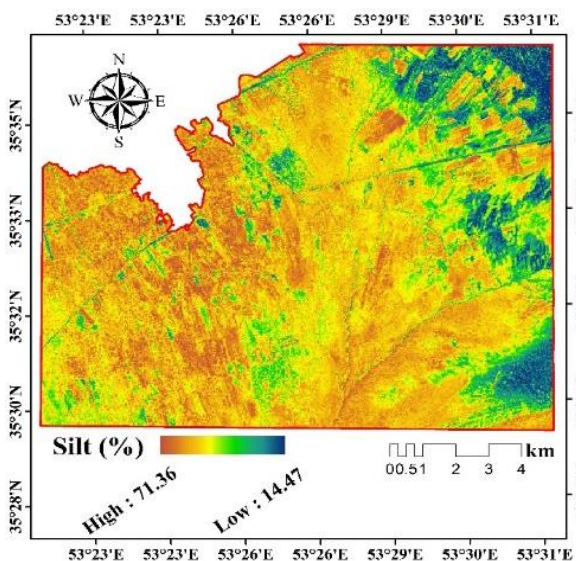
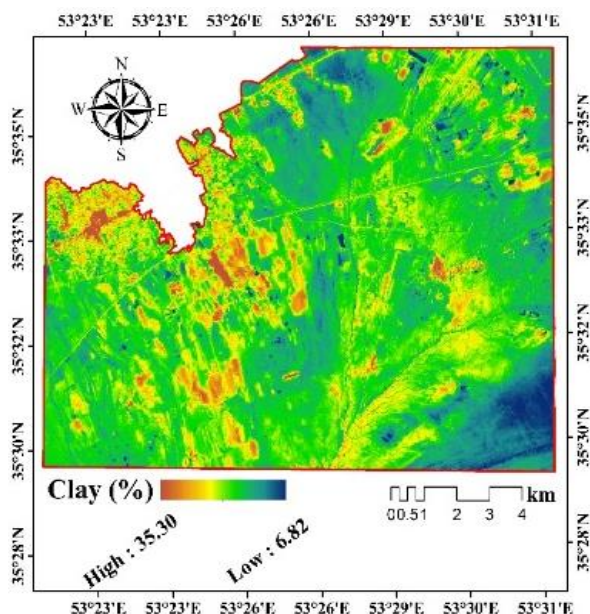
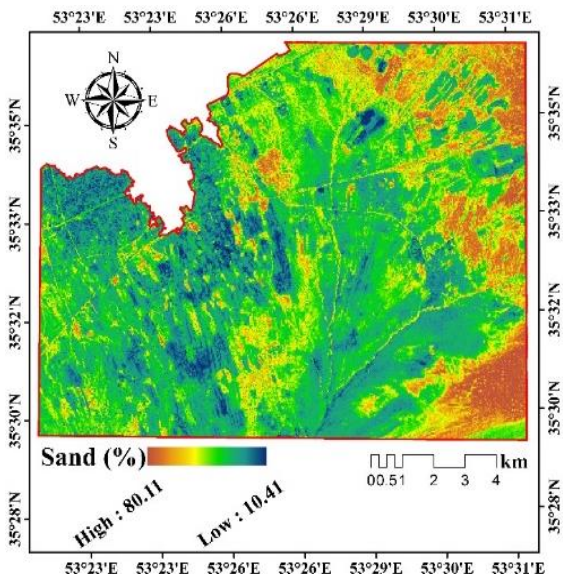
دارد. در میان مدل‌های داده‌کاوی، برای پیش‌بینی ذرات رس و سیلت بیشترین دقت را به ترتیب مدل‌های RF، GBM، PLS داشتند (جدول ۵).

همچنین، مقایسه مدل‌های غیرخطی RF و GBM نشان داد که مدل‌های پیچیده‌تر به‌عنوان مثال RF بهتر از مدل ساده‌تر GBM واقعیات را نشان می‌دهد. زیرا مدل‌های ساده‌تر از مقادیر گسسته در گره‌های ترمینال برای تقسیم درخت استفاده می‌کند و به لحاظ تغییرپذیری، عدم قطعیت بیشتری را نشان می‌دهند، بنابراین، در مدل GBM تعمیم روابط پایه‌ای بین ذرات خاک و پیش‌بینی‌های آن‌ها دقت کمتری را نسبت به مدل RF دارا هستند. همچنین مقایسه مدل‌های غیرخطی RF و مدل خطی PLS نشان داد که مدل RF نسبت به مدل PLS از کارایی بهتری برخوردار بوده است دلیل عملکرد بهتر مدل RF نسبت به مدل PLS را می‌توان به غیرخطی بودن ارتباط بین ذرات اولیه خاک و متغیرهای محیطی دانست. در این شرایط مدل‌های خطی قادر به توصیف ارتباط بین متغیرهای ورودی و متغیرهای هدف نمی‌باشند و برای برقراری ارتباط بین این متغیرها یک مدل غیرخطی نتایج بهتری را نشان می‌دهد. در این مطالعه بهترین مدل داده‌کاوی بر اساس مقادیر ضریب تبیین ( $R^2$ ) برای پیش‌بینی توزیع فضایی هریک از ذرات خاک انتخاب شد. مدل RF برای پیش‌بینی ذرات رس، سیلت و شن خاک دارای بیشترین  $R^2$  بود که این نتایج با نتایج (Hengl et al. 2015)؛ Tayebi et al. (2018) مبنی بر برتری مدل RF نسبت به مدل PLS برای پیش‌بینی ذرات خاک مطابقت دارد.

### پیش‌بینی فضایی

شکل (۳)، توزیع فضایی ذرات رس، سیلت و شن خاک را نشان می‌دهد. تجزیه و تحلیل ضریب همبستگی بین نقشه‌های رس، سیلت و شن خاک نشان داد که همبستگی منفی قوی بین شن و سیلت خاک ( $R=-0/89$ ) وجود داشت. همچنین ارتباط معکوسی بین نقشه‌های شن و رس خاک ( $R=-0/47$ ) مشاهده شد، در حالی که همبستگی بین نقشه سیلت و رس خاک به مثبت ( $R=0/39$ ) به‌دست آمد. بالاترین مقدار رس و مقدار سیلت





شکل ۳. نقشه‌های توزیع فضایی شن، رس و سیلت خاک

در گوشه جنوب غربی (شکل ۳) مشاهده می‌شود. بالاترین مقادیر شن در شمال غربی و جنوب شرقی منطقه مطالعه مشاهده شد که این اراضی مربوط به تپه‌های ماسه‌ای می‌باشند که در کنار اراضی کشاورزی قرار دارند. بنابراین مقادیر بالای شن در این اراضی کاملاً قابل قبول و منطقی است. با توجه به نقشه رس، بیشترین مقدار رس در نواحی مرکزی منطقه مورد مطالعه مشاهده شد. در حالی که تپه ماسه‌ای در شمال شرقی و جنوب شرقی دارای کمترین میزان رس بودند. مطالعات مشابهی نشان داد که امکان استفاده از داده‌های تصاویر ماهواره‌ای برای تهیه نقشه‌های رقوم ذرات اولیه خاک وجود دارد به‌عنوان مثال Liu *et al.* (2013) از روش‌های زمین‌آمار با کمک داده‌های سنجش از راه دور برای نقشه‌برداری رقوم ذرات خاک استفاده کردند. همچنین اخیراً Taghizadeh-Mehrjardi *et al.* (2016a) نیز از تصاویر ماهواره‌ای و شاخص‌های حاصل از تصویر به همراه متغیرهای حاصل نقشه رقوم ارتفاع (DEM) برای پیش‌بینی توزیع مکانی ذرات خاک استفاده کردند. نتایج آن‌ها نشان داد که استفاده از متغیرهای استخراج‌شده از نقشه DEM سبب افزایش کارایی روش‌های نقشه‌برداری رقوم خواهد شد. بیشتر اراضی منطقه مورد مطالعه، اراضی کشاورزی می‌باشند از طرف دیگر منطقه مورد مطالعه مسطح و تفاوت معنی‌داری بین کمترین و بیشترین ارتفاع در منطقه مورد مطالعه دیده نشد. به طوری که تفاوت بین بیشترین و کمترین ارتفاع منطقه مورد مطالعه کمتر از ۲۵۰ متر است به همین علت در این مطالعه از متغیرها حاصل از DEM به‌عنوان متغیرهای ورودی استفاده نشد.

جدول ۵. مقادیر ریشه مربعات خطا، ضرایب تبیین و میانگین قدر مطلق خطا در مدل‌های PLS, GBM, RF برای ذرات اولیه خاک

آمارهای مورد استفاده				
ذرات خاک	مدل	RMSE	R <sup>2</sup>	MAE
شن	RF	۱۵/۶۸	۰/۳۵	۱۲/۶۲
	GBM	۱۵/۶۰	۰/۳۵	۱۲/۵۴
	PLS	۱۶/۱۸	۰/۳۰	۱۳/۳۹
سیلت	RF	۱۱/۴۹	۰/۳۳	۹/۳۴
	GBM	۱۱/۷۱	۰/۳۱	۹/۴۹
	PLS	۱۲/۷۲	۰/۲۴	۹/۹۹
رس	RF	۸/۴۲	۰/۲۸	۵/۹
	GBM	۸/۳۰	۰/۲۰	۶/۷۳
	PLS	۹/۲۱	۰/۱۷	۷/۵۸

(RMSE, R<sup>2</sup> و MAE به ترتیب نشان‌دهنده ریشه میانگین مربعات خطا، ضریب تبیین و میانگین قدر مطلق خطا می‌باشند)

## نتیجه‌گیری

متغیرهای حاصل از تصویر برای پهنه‌بندی ذرات خاک در اراضی کشاورزی دارای دقت خیلی بالایی نمی‌باشند اما الگوی پراکنش شن، سیلت و رس را به خوبی پیش‌بینی می‌کنند. این مطالعه نشان داد که ارتباط قوی بین متغیرهای محیطی حاصل از تصاویر ماهواره‌ای و ذرات اولیه خاک در اراضی کشاورزی وجود ندارد. در این مطالعه از متغیرهای محیطی که از داده‌های حاصل از یک تصویر ماهواره‌ای به‌عنوان متغیر کمکی برای تهیه نقشه رقوم ذرات اولیه خاک استفاده شد. استفاده از یک تصویر می‌تواند در اراضی کشاورزی با خطاهایی همراه باشد. به‌عنوان مثال آیش بودن قطعاتی از اراضی کشاورزی در تاریخی که تصویربرداری انجام شده می‌تواند سبب گمراه شدن محقق شده و نتایج دقت مدل‌سازی را کاهش دهند. بنابراین برای مطالعات آینده پیشنهاد می‌شود که از شاخص‌های آماری از قبیل میانه و یا میانگین سری زمانی تصاویر برای مدل‌سازی ذرات اولیه خاک و دیگر خصوصیات خاک استفاده شود.

نتایج مقایسه مدل‌های جنگل تصادفی (RF)، حداقل مربعات جزئی (PLS) و روش ماشین تقویت‌شده گرادیان (GBM) برای پیش‌بینی ذرات شن، سیلت و رس خاک نشان داد که مدل RF بهترین مدل برای پیش‌بینی شن، رس و سیلت خاک می‌باشد. نتایج همچنین نشان داد که مهم‌ترین متغیرهای محیطی، باندهای ۱۰، ۵ و شاخص گج (GI) سنجنده OLI می‌باشند. در این مطالعه همبستگی منفی قوی بین نقشه‌های شن و سیلت خاک ( $R = -0/89$ ) بدست آمد. همچنین همبستگی بین نقشه‌های شن و رس خاک برابر با  $-0/47$  و همبستگی بین نقشه‌های سیلت و رس خاک برابر با  $0/39$  تعیین شد. بالاترین مقادیر شن در شمال‌غربی و جنوب‌شرقی منطقه مطالعه مشاهده شد که این اراضی مربوط به تپه‌های ماسه‌ای می‌باشند که در کنار اراضی کشاورزی قرار دارند. بنابراین مقادیر بالای شن در این اراضی کاملاً قابل قبول و منطقی می‌باشد. لذا این نتایج نشان می‌دهد که گرچه استفاده از

## REFERENCES

- Bellinaso, H., Demattê, J. A. M., and Romeiro, S. A. (2010). Soil spectral library and its use in soil classification. *Revista Brasileira de Ciência do Solo*, 34(3): 861-870.
- Curcio D., Ciralo G., D'Asaro F., and Minacapillia M. (2013). Prediction of soil texture distributions using VNIR SWIR reflectance spectroscopy. *Procedia Environmental Sciences*, 19:494 – 503.
- Florinsky, I. V., Eilers, R. G., Manning, G. R., and Fuller, L. G. (2002). Prediction of soil properties by digital terrain modelling. *Environmental Modelling & Software*, 17(3): 295-311.
- Forkuor, G., Hounkpatin, O. K., Welp, G., & Thiel, M. (2017). High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. *PloS one*, 12(1), e0170478.
- Gee, G.W., and Bauder, J.W. (1986). Particle- size analysis, In: Klute, A., et al. (Ed.), Methods of soil analysis. Part1, Physical and mineralogical methods, seconded. ASA, Inc., Madison, WI, pp. 383-411.
- Genuer, R., Poggi, J. M., and Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14): 2225-2236.
- Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2): 83-85.
- Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., and Tondoh, J. E. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PloS one*, 10(6), e0125814.
- Khanamani, A., Jafari, R., Jafari, A., Sangoony, H., and Shahbazi, A. (2011). Evaluation of soil status using RS and GIS technology (Case study: Segzi plain). *Journal of Applied RS & GIS Techniques in Natural Resource Science*, 2(3): 25-37.
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). New York: Springer.
- Liu, Z. P., Shao, M. A., and Wang, Y. Q. (2013). Large-scale spatial interpolation of soil pH across the Loess Plateau, China. *Environmental Earth Sciences*, 69(8): 2731-2741.
- Mahmoudabadi, E., Karimi, A., Haghnia, G. H., and Sepehr, A. (2017). Digital soil mapping using remote sensing indices, terrain attributes, and vegetation features in the rangelands of northeastern Iran. *Environmental monitoring and assessment*, 189(10): 500.
- Malone, B. P., McBratney, A. B., Minasny, B., and Laslett, G. M. (2009). Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma*, 154(1-2): 138-152.
- Makabe, S., Kakuda, K. I., Sasaki, Y., Ando, T., Fujii, H., and Ando, H. (2009). Relationship between mineral composition or soil texture and available silicon in alluvial paddy soils on the Shounai Plain, Japan. *Soil science and plant nutrition*, 55(2): 300-308.
- McBratney, A. B., Santos, M. M., and Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1-2): 3-52.
- Minasny, B., and Hartemink, A. E. (2011). Predicting soil properties in the tropics. *Earth-Science Reviews*, 106(1-2): 52-62.
- Minasny, B., and McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information.

- Computers & geosciences*. 32(9): 1378-1388.
- Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., and Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*. 18(6): 275-285.
- Nauman, T. W., and Thompson, J. A. (2014). Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. *Geoderma*. 213, 385-399.
- R Development Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna Austria. <http://www.Rproject.org/>.
- Ryan, P. J., McKenzie, N. J., O'Connell, D., Loughhead, A. N., Leppert, P. M., Jacquier, D., and Ashton, L. (2000). Integrating forest soils information across scales: spatial prediction of soil properties under Australian forests. *Forest Ecology and Management*. 138(1-3): 139-157.
- Scudiero, E., Skaggs, T. H., and Corwin, D. L. (2014). Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. *Geoderma Regional*. 2: 82-90.
- Summers, D., Lewis, M., Ostendorf, B., and Chittleborough, D. (2011). Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. *Ecological Indicators*. 11(1): 123-131.
- Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., and Malone, B. P. (2014). Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma*. 213: 15-28.
- Taghizadeh-mehrjardi, R., Toomanian, N., Khavaninzadeh, A. R., Jafari, A., and Triantafyllis, J. (2016a). 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(6): 707-725.
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., and Kerry, R. (2016b). Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma*. 266: 98-110.
- Tayebi, M., Naderi, M., Mohammadi, J., & Zadeh, M. H. (2018). Comparing different statistical models and pre-processing techniques for estimation of soil particles using VNIR/SWIR spectrum. *Journal of Water and Soil*. 32(1): 73-85. (In Farsi).
- Vaysse, K., and Lagacherie, P. (2015). Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Regional*. 4: 20-30.
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., and Finke, P. (2017). Comparing the efficiency of digital and conventional soil mapping to predict soil types in a semi-arid region in Iran. *Geomorphology*. 285: 186-204.