

شناسایی رنگ و بافت خاک با استفاده از الگوریتم هوش مصنوعی و تصاویر ماهواره‌ای

لاله پرویز*

تاریخ پذیرش: ۱۳۹۹/۱۱/۰۴

تاریخ دریافت: ۱۳۹۹/۰۹/۱۷

چکیده

نیاز به اطلاعات دقیق و با کیفیت خاک به دلیل کاربرد در برنامه‌ریزی‌های کشاورزی در حال افزایش است. هدف این تحقیق برآورد رنگ و بافت خاک با استفاده از اطلاعات تصاویر ماهواره‌ای به‌عنوان داده‌های ورودی رگرسیون بردار پشتیبان و رگرسیون درختی می‌باشد. در پژوهش حاضر، شاخص‌های ماهواره‌ای TVX ، TCI ، $MSAVI$ ، $MPDI$ ، $SAVI$ ، PDI ، $NDVI$ و VCI استفاده شد. آزمون دانکن در سطح احتمال پنج درصد حاکی از وجود تغییرات معنی‌دار زمانی بین شاخص‌ها است. تفاوت معنی‌دار بین میانگین شاخص‌ها از نظر تنوع بافت خاکی براساس آزمون دانکن در سطح احتمال پنج درصد وجود نداشت. آماره‌های خطای $RMSE$ ، $RRMSE$ ، $AMAPE$ و MSE در مورد شن از رگرسیون درختی به رگرسیون بردار پشتیبان به ترتیب $۱۵/۴۳$ ، $۱۳/۳۳$ ، $۱۶/۴۱$ و $۲۸/۷$ درصد کاهش داشتند. تعیین بافت خاک با مثلث بافت خاک در دوره صحت‌سنجی حاکی از تطابق نوع بافت خاک مشاهداتی و رگرسیون بردار پشتیبان بود. با در نظر گرفتن اجزاء بافت خاک و مولفه‌های رنگ خاک آماره RPD از رگرسیون درختی به رگرسیون بردار پشتیبان به میزان $۱۲/۴۳$ درصد افزایش داشت که بیانگر کارایی رگرسیون بردار پشتیبان در برابر رگرسیون درختی است. درصد کاهش آماره‌های $RMSE$ ، $RRMSE$ و MSE از رگرسیون خطی چندگانه به رگرسیون بردار پشتیبان در هیو به ترتیب $۷۶/۸۸$ ، $۷۷/۴$ و $۹۴/۶$ و در رگرسیون درختی به ترتیب $۷۲/۱۵$ ، $۷۲/۵۸$ و $۹۲/۹۲$ بیانگر عملکرد بهتر دو مدل رگرسیونی نسبت به رگرسیون خطی چندگانه بود. براساس تحلیل از جهات گوناگون رگرسیون بردار پشتیبان نسبت به رگرسیون درختی عملکرد بهتری در تعیین رنگ و بافت خاک داشت.

واژه‌های کلیدی: شاخص‌های ماهواره‌ای، رگرسیون، آماره‌های خطا، RPD

پرویز ل. ۱۴۰۰. شناسایی رنگ و بافت خاک با استفاده از الگوریتم هوش مصنوعی و تصاویر ماهواره‌ای. تحقیقات کاربردی خاک. جلد ۹، شماره ۴. صفحه: ۸۸-۱۰۱.

۱- دانشیار دانشکده کشاورزی دانشگاه شهید مدنی آذربایجان
*پست الکترونیک: Jaleh_parviz@yahoo.com

مقدمه

خاک نقش مهمی در محیط‌زیست و زندگی انسان دارد. در این میان آگاهی از ویژگی‌های خاک مانند بافت و رنگ خاک نقش مهمی در مطالعه وضعیت خاک‌ها دارد. ویژگی‌های خاک در مهندسی کشاورزی مانند بهبود خاک، مدیریت زهکشی، فرسایش خاک و آبیاری تأثیر دارد. آگاهی در مورد تغییرات ویژگی‌های خاک می‌تواند اطلاعات ارزشمندی را برای طراحی منطقی‌تر و برنامه ریزی مدیریتی به‌ویژه در مناطق کشاورزی فراهم کند. اقلیم، زیواگان^۱ و تاریخ زمین‌شناسی از عوامل مهم تأثیرگذار بر ویژگی‌های شیمیایی و فیزیکی خاک در مقیاس‌های بزرگ‌تر و فعالیت‌های انسانی و توپوگرافی عوامل اصلی کنترل‌کننده ویژگی‌های خاک در مقیاس های کوچک‌تر هستند (Wu et al., 2018).

بافت خاک به‌عنوان یک عامل تأثیرگذار مهم در بسیاری از ویژگی‌ها و رفتارهای فیزیکی و شیمیایی خاک نظیر حاصلخیزی خاک، ظرفیت تبادل کاتیونی، ذخیره آب، زهکشی داخلی در نظر گرفته می‌شود (Ding et al., 2020). هم‌چنین رنگ خاک نیز به‌عنوان شاخص قوی در توصیف سایر ویژگی‌های خاک مانند آهن و مواد آلی بشمار می‌آید (Stiglitz et al., 2017). تعیین بافت خاک و نقشه مربوط به توزیع آن نقش مهمی در برنامه‌ریزی کاربری اراضی، رژیم آبی و مطالعات حفاظت خاک دارد. با این حال تجزیه و تحلیل معمول بافت خاک مانند روش‌های هیدرومتری و پیپت وقت‌گیر و گران بوده و امروزه برای تعداد زیاد نمونه و نقشه‌های با قدرت تفکیک مکانی بالا مناسب نیست (Coblinski et al., 2020). تعیین رنگ خاک با دفترچه مانسل دارای یکسری محدودیت‌هایی از جمله حساسیت به کاربر، شرایط محیطی را دارد (Stiglitz et al., 2016)، بنابراین استفاده از تکنیک سنجش از دور در تخمین ویژگی‌های خاک در حال گسترش است (Coblinski et al., 2020) که می‌تواند به تغییرات مکانی ذرات معدنی خاک با استفاده از تصاویر ماهواره ترا و سنجنده MODIS با قدرت تفکیک مکانی ۵۰۰ متر اشاره کرد (Asadzadeh et al., 2019). در تعیین رنگ خاک امکان بکارگیری باندهای مختلف تصاویر ماهواره‌ای Landsat در دشت هرات در استان یزد (خاک‌های آهکی، نمکی، سنگلاخی، ماسه‌ای) مورد

بررسی قرار گرفت. نتایج نشان دادند که بهترین باندهای هیو شامل باندهای ۵، ۱۰، ۷ و بهترین باندهای ولیو شامل باندهای ۶، ۱۱، ۱۰ و بهترین باندهای کروما شامل ۷، ۱ و ۱۱ است (Hasanzadeh et al., 2019).

برای ایجاد توابع انتقالی برای برآورد هدایت آبی اشباع و پارامتر عکس طول درشت موئینگی در ۶۰ نقطه از دشت آزادگان در منطقه شهرکرد از رگرسیون خطی چندگانه و درختی استفاده شد. معیارهای ارزیابی عملکرد مدل یعنی میانگین خطا (ME)، خطای نسبی (RE)، جذر میانگین مربعات خطا (RMSE) و ضریب همبستگی پیرسون بیانگر عملکرد بهتر رگرسیون درختی در برابر رگرسیون خطی چندگانه در برآورد پارامترهای بیان‌شده بودند. با بکارگیری رگرسیون درختی هدایت آبی اشباع بیش‌تر و با رگرسیون خطی چندگانه کم‌تر نسبت به مقدار واقعی تخمین زده شد (Dehghani Baniani et al., 2012). استفاده از برخی ویژگی‌های خاک مانند بافت خاک، pH، ماده آلی و چگالی ظاهری حاصل از داده‌های ۱۰۶ نمونه خاک بانک اطلاعاتی UNSODA و بکارگیری روش‌های رگرسیون درختی و خطی، ظرفیت تبادل کاتیونی تخمین زده شد. مدل رگرسیونی براساس ضرایب ماده آلی و درصد رس که بیش‌ترین همبستگی را با ظرفیت تبادل کاتیونی خاک داشتند، توسعه یافت. براساس معیارهای ارزیابی عملکرد مدل، مدل رگرسیون درختی در دو بخش صحت‌سنجی و اعتبارسنجی دارای کارایی بالایی بود (Ostovari et al., 2015). با تصاویر ماهواره‌ای Landsat-8 و Sentinel-2 به ارزیابی توزیع مکانی رنگ خاک سطحی در اردن شمالی پرداخته شد. نتایج نشان‌دهنده شباهت قابل توجه طبقه‌بندی ماشین بردار پشتیبان از Landsat-8 و Sentinel-2 در منطقه و رابطه قوی از توزیع رنگ قرمز خاک، میانگین بارش سالانه و جنبه‌های ژئومورفولوژی بود (Sahwan et al., 2018). مساله قابل توجه ارتقاء رابطه بین باندهای انعکاسی و مؤلفه‌های رنگ خاک (هیو، ولیو و کروما) به‌جای استفاده از رگرسیون خطی ساده است به‌عنوان نمونه در تحقیق متیکالی (Mattikalli, 1997) ضریب همبستگی از ۰/۵ (رگرسیون خطی ساده) به بیش‌تر از ۰/۸ (روش انتقال یافته چرخشی بهینه) در مدل‌سازی رنگ خاک ارتقاء یافت. مقایسه‌ای بین عملکرد ماشین بردار پشتیبان، شبکه عصبی مصنوعی و

در استان آذربایجان شرقی با استفاده از سنجنده MOIDS به کار برده شدند. متغیرهای ورودی شامل شاخص‌ها و باندهای حاصل از تصاویر ماهواره‌ای بودند. مدل‌سازی با شبکه عصبی مصنوعی انجام گرفت. MPDI و باندهای حرارتی به عنوان شاخص‌های دقیق در تعیین رنگ خاک معرفی شدند (Parviz, 2020).

هدف این تحقیق مقایسه عملکرد رگرسیون بردار پشتیبان و رگرسیون درختی در تعیین رنگ و بافت خاک است. ورودی‌های مدل شامل باندها و شاخص‌های حاصل از تصاویر ماهواره‌ای سنجنده MODIS است. با آزمون دانکن تغییرات شاخص‌ها و باندها از نظر زمانی و نوع بافت خاک مورد بررسی قرار می‌گیرد. همچنین تأثیر انتخاب دقیق پارامترهای شاخص‌های ماهواره‌ای هم مورد ارزیابی قرار گرفت.

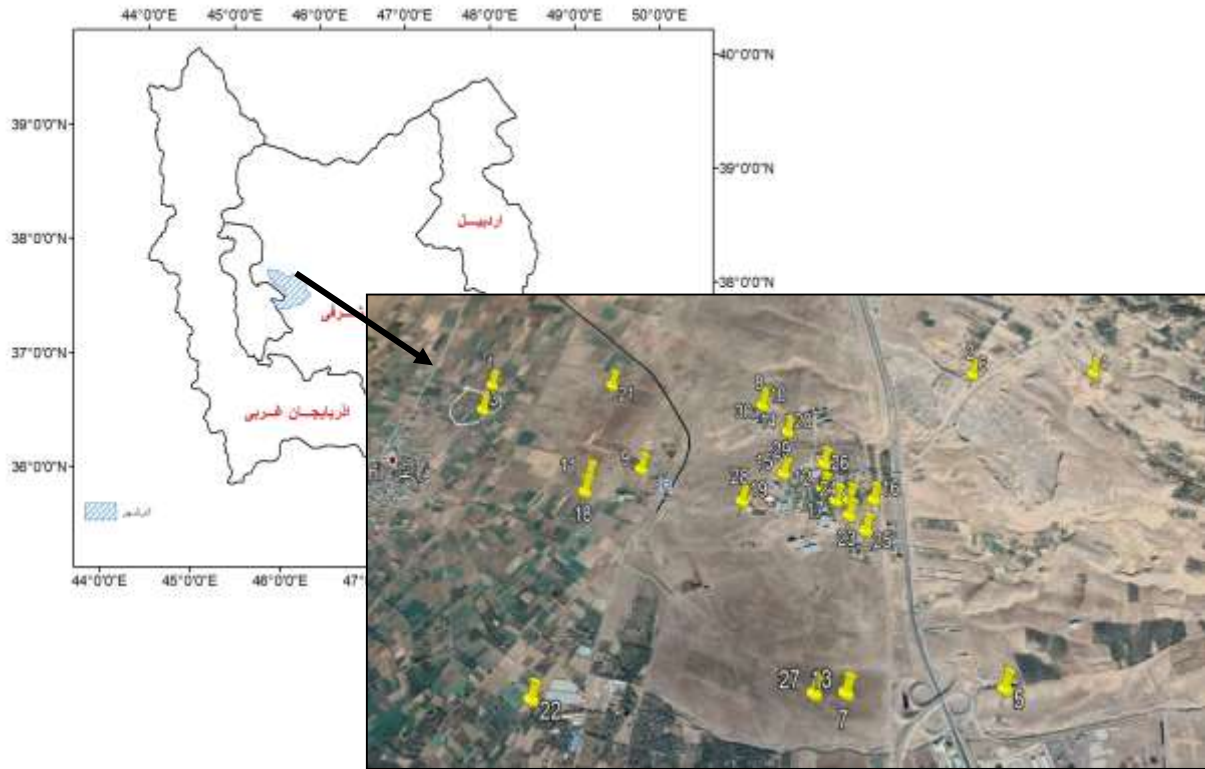
مواد و روش‌ها

منطقه مورد مطالعه

داده‌های جمع‌آوری‌شده برای تجزیه و تحلیل بافت و رنگ خاک مربوط به زمین‌های دانشگاه شهید مدنی آذربایجان (۳۰ نمونه) واقع در ۳۵ کیلومتری جاده تبریز- آذرشهر بود که موقعیت نقاط نمونه‌برداری شده در شکل ۱ نشان داده شده است. دوره زمانی تصاویر سنجنده MODIS مربوط به بهار، تابستان و پاییز (۲۰۱۵ تا ۲۰۱۷) بود. بافت خاک منطقه حاصل از نمونه‌برداری تصادفی شامل دو نوع بافت خاک شن و شن لومی بود. براساس درون‌یابی داده‌های دما و بارش ایستگاه‌های تبریز و مراغه، اقلیم منطقه نمونه‌برداری در بازه زمانی ۱۹۸۷-۲۰۱۷ براساس اقلیم‌نمای دومارتن نیمه‌خشک و با استفاده از سیستم طبقه‌بندی کوپن استپی برآورد شد.

در این تحقیق شاخص‌های حاصل از تصاویر ماهواره‌ای به‌عنوان داده‌های ورودی رگرسیون بردار پشتیبان و درختی مورد استفاده قرار گرفت که در ادامه در مورد شاخص‌های به‌کار گرفته‌شده و مدل‌های رگرسیونی توضیح داده شده است.

درخت طبقه‌بندی در تعیین کلاس‌های بافت خاک براساس پارامترهای مربوط به عوارض زمین در جنوب‌غربی چین انجام گرفت. توابع کرنل مورد استفاده در ماشین بردار پشتیبان چندجمله‌ای و شعاعی بودند. ماشین بردار پشتیبان در برابر شبکه عصبی مصنوعی و درخت طبقه‌بندی دارای عملکرد بهتری بود. افزون بر این، ماشین بردار پشتیبان با تابع کرنل چند جمله‌ای نسبت به پایه شعاعی بهتر عمل می‌کند. ارتفاع، شاخص طبقه‌بندی زمین برای مناطق پست و طول مسیر جریان مهم‌ترین شاخص‌های زمینی بودند که بر تغییر در کلاس بافت خاک در منطقه مورد مطالعه تأثیرگذار بودند (Wu *et al.*, 2018). با استفاده از ماشین بردار پشتیبان چندکلاسه در منطقه‌ای در هند به طبقه‌بندی بافت خاک پرداخته شد. بدین منظور پردازش تصاویر خاک برای تولید یک سیستم طبقه‌بندی خاک رقومی انجام گرفت. درصدهای شن، رس و سیلت با استفاده از آزمایش هیدرومتری بدست آمدند. از مقادیر هیو و ولیو در برآورد بردار ویژگی استفاده شد. به طور متوسط روش پیشنهادی ۹۱/۳۷ درصد دقت را برای تمام نمونه‌های خاک نشان داد (Barman & Choudhury, 2020). مدل‌های جنگل تصادفی و ماشین بردار پشتیبان برای ارزیابی متغیرهای ورودی مختلف در پیش‌بینی و تهیه نقشه‌های کلاس‌های زهکشی خاک در بخشی از چهارمحال و بختیاری توسعه داده شدند. متغیرهای ورودی شامل ویژگی‌های توپوگرافی استخراج شده از DEM، شاخص‌های پوشش گیاهی سنجش از دور، توصیف‌کننده‌های رنگ خاک حاصل از طیف‌سنج بازتابی (ولیو-کروما) بودند. کروما به همراه NDVI و SAVI مؤثرترین پیش‌بینی‌کننده زهکشی خاک بودند. بهترین عملکرد مدل‌ها زمانی به‌دست آمد که صفات توپوگرافی، SAVI، کروما و ولیو به عنوان متغیرهای ورودی برای پیش‌بینی کلاس‌های زهکشی خاک در نظر گرفته شد. دقت کلی پیش‌بینی و ضریب کاپا برای طبقه‌بندی زهکشی به ترتیب ۰/۸۳ و ۰/۷۳ برای جنگل تصادفی و ۰/۸۶ و ۰/۷۴ برای ماشین بردار پشتیبان بودند (Asgari *et al.*, 2020). رگرسیون گام به گام، تحلیل مولفه‌های اصلی و روش تابع حساسیت برای تعیین متغیرهای ورودی مؤثر در تعیین رنگ خاک



شکل ۱- موقعیت منطقه مورد مطالعه و نقاط نمونه برداری شده خاک
Figure 1. Location of case study and soil sampling points

$$SAVI = \frac{(1 + L)(\rho_{nir} - \rho_{red})}{\rho_{nir} + \rho_{red} + L} \quad (5)$$

$$(6)$$

$MSAVI = 0.5[2\rho_{nir} + 1 - \sqrt{(2\rho_{nir} + 1)^2 - 8(\rho_{nir} - \rho_{red})}]$
 شاخص تفاضل نرمال شده پوشش گیاهی،
 شاخص عمودی خشکسالی، $MPDI$: شاخص
 عمودی خشکسالی اصلاح شده، VCI : شاخص شرایط
 پوشش گیاهی، $SAVI$: شاخص گیاهی تعدیل کننده
 اثرات خاک، $MSAVI$: شاخص گیاهی تعدیل کننده
 اثرات خاک اصلاح شده، ρ_{nir} : مقادیر بازتاب در باند مادون
 قرمز نزدیک، ρ_{red} : مقادیر بازتاب در باند قرمز، M : شیب
 خط خاک، f_v : کسر پوشش گیاهی، $NDVI_{max}$: $NDVI$
 سطحی با ۱۰۰ درصد پوشش گیاهی و شاخص سطح
 برگ بی نهایت و $f_v = 1$ ، $NDVI_{min}$: $NDVI$ خاک لخت
 و $f_v = 0$ ، L : فاکتور تصحیح اثرات خاک.

شاخص‌های حاصل از تصاویر ماهواره‌ای

شاخص‌های حاصل از تصاویر ماهواره‌ای شامل ترکیب باندهای مختلف (باندهای مرئی، مادون قرمز نزدیک و غیره) براساس رابطه ریاضی می‌باشند. فرمول شاخص‌های مورد استفاده براساس دسته‌بندی باندها در روابط ۱-۱۳ آورده شده است.

شاخص‌هایی براساس بازتاب طیفی

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \quad (1)$$

$$VCI = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \quad (2)$$

$$PDI = \frac{1}{\sqrt{M^2 + 2}} (R_{red} + MR_{nir}) \quad (3)$$

$$MPDI = \frac{R_{red} + MR_{NIR} - f_v(R_{v,red} + MR_{v,NIR})}{(1 - f_v)\sqrt{M^2 + 1}} \quad (4)$$

$$f_v = 1 - \left(\frac{NDVI_{max} - NDVI}{NDVI_{max} - NDVI_{min}} \right)^{0.6175}$$

4Vegetation Condition Index
 5Soil Adjustment Vegetation Index
 6Modified Soil Adjustment Vegetation Index

1Normalized Difference Vegetation Index
 2Perpendicular Drought Index
 3Modified Perpendicular Drought Index

رگرسیون بردار پشتیبان

مدل‌های ماشین بردار پشتیبان با وجود آن که مبانی ریاضی آن‌ها در دهه ۱۹۶۰ مشابه مفهوم شبکه‌های عصبی تنظیم شده است ولی هنوز جزء مفاهیم نسبتاً جدیدی معرفی می‌شوند. ماشین بردار پشتیبان یک روش یادگیری نظارت شده است که براساس مجموعه‌ای از نمونه‌های آموزشی ساخته شده است و مدلی را ایجاد می‌کند که می‌تواند نمونه‌های طبقه‌بندی شده را به دسته‌های جداگانه تقسیم کند. ماشین بردار پشتیبان یک طبقه‌بندی غیراحتمالی براساس تفکیک نمونه‌ها به کلاس‌های مجزا با تعریف ابرصفحه که دسته‌ها را جدا می‌کند، می‌باشد. اصل اساسی طبقه‌بندی ماشین بردار پشتیبان تبدیل فضای ورودی اصلی به فضای دیگر با ابعاد بالاتر است. افزایش ابعاد فضای داده امکان یافتن راه‌حل خطی برای جدا سازی داده‌ها که در فضای اصلی داده وجود ندارد را امکان‌پذیر می‌کند. این فرآیند را می‌توان در نمونه‌ای از فضای دو بعدی با داده‌های باینری شامل نمونه‌های مثبت و منفی آموزشی که توزیع آن‌ها اجازه جداسازی خطی کلاس‌ها را نمی‌دهد، نشان داد (شکل ۲-ا). فضای دو بعدی اولیه جایی که هر بردار با دو ویژگی تعریف می‌شود می‌تواند با اضافه کردن ویژگی سوم براساس تبدیل ویژگی‌های اولیه به فضای سه بعدی تبدیل شود. چنین تبدیلی امکان جابجایی نقاط داده را در امتداد محور جدید فراهم می‌کند تا این که صفحه‌ای پیدا شود و کلاس‌های داده‌ها از هم جدا شوند (شکل ۲-ب). تبدیل بیان شده در مدل ماشین بردار پشتیبان با کرنل براساس اصول متفاوت همانند خطی، چند جمله‌ای، سیگموئید انجام می‌گیرد. رگرسیون بردار پشتیبان ایده اصلی ماشین بردار پشتیبان را برای نگاشت داده‌ها به فضای با ابعاد بالا از طریق نگاشت غیرخطی و انجام رگرسیون خطی در این فضا را حفظ می‌کند. تابع کرنل داده‌ها را به فضای با ابعاد بالا تبدیل می‌کند که جداسازی خطی را ممکن می‌سازد (Langhammer & Česák, 2016).

شاخص‌های حرارتی

$$TCI = \frac{LST_{max} - LST}{LST_{max} - LST_{min}} \quad (7)$$

TCI : شاخص شرایط دمایی، LST_{max} : بیشینه دمای سطح زمین، LST_{min} : کمینه دمای سطح زمین، LST : دمای سطح زمین.

شاخص‌های ترکیبی

$$TVX = \frac{LST}{NDVI} \quad (8)$$

$$VHI = \alpha VCI + (1 - \alpha) TCI \quad (9)$$

$$VHI = \frac{f_v}{100} VCI + \frac{100 - f_v}{100} TCI \quad (10)$$

$$NVWSI = \frac{VWSI - VWSI_{min}}{VWSI_{max} - VWSI_{min}}, \quad (11)$$

$$VWSI = \frac{NDVI}{LST} \quad (12)$$

$$RVWSI = \frac{100 \times NDVI}{LST - 100} \quad (12)$$

$$MVWSI = \frac{RNDVI}{RLST^2}, \quad (13)$$

$$RNDVI = \frac{NDVI}{NDVI}$$

$$RLST = \frac{LST}{LST}$$

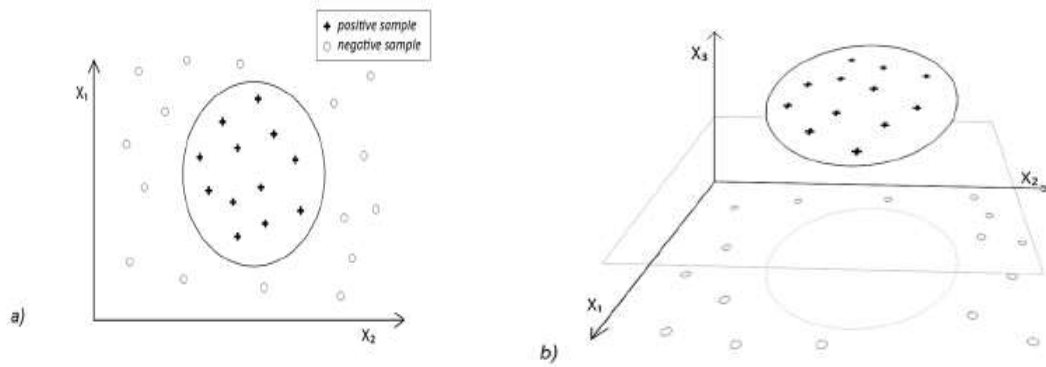
TVX ^۱: شاخص پوشش گیاهی-دمایی، VHI : شاخص شادابی پوشش گیاهی، $NVWSI$: شاخص تأمین آب پوشش گیاهی نرمال شده، $VWSI$: شاخص تأمین آب پوشش گیاهی، $RVWSI$: شاخص تأمین آب پوشش گیاهی اصلاح شده، $MVWSI$: شاخص تأمین آب پوشش گیاهی تعدیل شده، LST : دمای سطح زمین، $NDVI$: شاخص تفاضل نرمال شده پوشش گیاهی، \overline{LST} : میانگین دمای سطح زمین، \overline{NDVI} : متوسط شاخص تفاضل نرمال شده پوشش گیاهی، α ، $1 - \alpha$: به ترتیب نمایانگر سهم نسبی شاخص‌های VCI و TCI در مقادیر شاخص VHI ، TCI : شاخص شرایط دمایی، VCI : شاخص شرایط پوشش گیاهی (Menxin & Houquan, 2016; Shahabfar et al., 2012).

3 Normalized Vegetation Supply Water Index

4 Temperature Condition Index

1 Temperature Vegetation Index

2 Vegetation Health Index



شکل ۲- اصل یافتن راه حل خطی طبقه‌بندی با افزایش ابعاد (a) نمونه‌ای از داده‌های آموزش با مقادیر مثبت و منفی جایی که تفکیک خطی امکان‌پذیر نیست (b) فضای داده با ابعاد افزایش یافته - امکان پیدا کردن صفحه‌ای برای جدا کردن داده‌ها (Langhammer & Česák, 2016).

Figure 2. Principle of finding the linear solution of classification by augmenting the dimensionality: (a) sample of training data with positive and negative values, where linear separation is not possible (b); data space with augmented dimensionality, enabling to find a plane, separating the data (Langhammer & Česák, 2016).

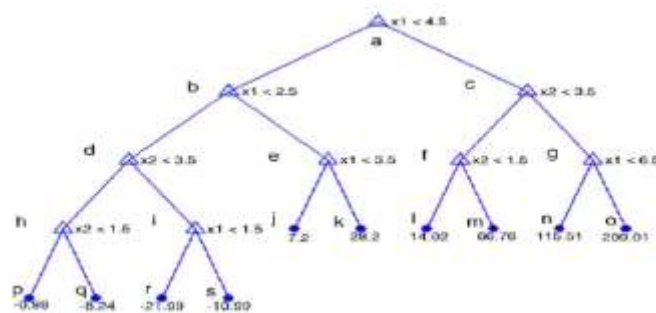
فرزندی با توجه به متغیرهای ورودی نمونه ساخته می‌شود. هر تقسیم یک بررسی در مورد متغیرهای ورودی است و پاسخ بلی و نه به ترتیب به زیرمجموعه‌های (نسل) چپ و راست منتهی می‌شود. شکل ۳ رگرسیون درختی تابع

$$y = x_1 x_2 (x_1 - 1) + e^{-x_2} - 4x_2, x_1, x_2 \in [0, 7]$$

است.

رگرسیون درختی

رگرسیون درختی به عنوان یک الگوریتم ناپارامتری، یک طبقه‌بندی کننده خودکار است. برای نمونه‌های آموزشی شامل n مورد $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ که x متغیر مستقل در یک فرم به صورت بردار m بعدی و y متغیر پاسخ می‌باشند و رگرسیون درختی یک طبقه‌بندی ساختاری درختی (باینری) را تشکیل می‌دهد. این درخت با تقسیم‌های مکرر زیر مجموعه‌ها به دو زیرشاخه



شکل ۳- رگرسیون درختی تابع $y = x_1 x_2 (x_1 - 1) + e^{-x_2} - 4x_2, x_1, x_2 \in [0, 7]$ (Yang & Suergen, 2006)

Figure 3. Regression tree of $y = x_1 x_2 (x_1 - 1) + e^{-x_2} - 4x_2, x_1, x_2 \in [0, 7]$ (Yang & Suergen, 2006)

نمونه اصلی است مانند گره a. هر گره برگ دارای یک مقدار خروجی و یک قانون (بصورت اگر... سپس بیان می‌شود) است مانند گره m که قانون به این صورت نوشته می‌شود:

$$\text{if } 1.5 \leq x_2 \leq 3.5 \text{ and } x_1 > 4.5, \text{ then } y = 66.76$$

از آنجائیکه الگوریتم درخت رگرسیون تنها با مقادیر گسسته مرتبط است ابتدا تابع به ۶۴ بردار ورودی و خروجی گسسته شد. به هر زیرمجموعه گره گفته می‌شود مانند گره a, b, ..., s. اگر تقسیم گره ND_1 را به دو گروه انجام گیرد ND_2 و ND_3 ، بنابراین ND_1 والدین ND_2 و ND_3 است و ND_2 هم‌نژاد ND_3 است. گره برگ گره‌ای است بدون تقسیم بیش‌تر مانند گره k. گره ریشه دسته

می‌باشند. شاخص شادابی پوشش گیاهی به دو طریق محاسبه شد: الف) با در نظر گرفتن ضرایب α برابر با ۰/۵ (معادله VHI - 0.59). ب) تعیین ضرایب براساس کسر پوشش گیاهی (رابطه ۱۰). بافت خاک در نقاط نمونه برداری شده که با آزمایش هیدرومتری به دست آمد، حاکی از دو نوع بافت خاک شنی و شن لومی در منطقه بود. تغییرات داده‌های ورودی از نظر بافت و زمان مورد بررسی قرار گرفت به اینصورت که تفاوت میانگین داده‌های ورودی بین دو نوع بافت خاک و از نظر سه زمان ارزیابی شد که نتایج در جدول ۱ آورده شده است.

با توجه به جدول ۱ تغییرات میانگین داده‌های ورودی از نظر بافت خاک کم است به طوری که به غیر از باند ۳۱ و ۳۲، تغییرات میانگین داده‌های ورودی بین دو بافت خاک معنی‌دار نمی‌باشد. دلیل این مساله را می‌توان در نزدیکی دو نوع بافت خاک دانست؛ چون تغییرات زیادی از نظر نوع بافت خاک در منطقه وجود ندارد و درصد‌های اجزاء بافت اختلاف محسوسی ندارند. بین سه فصل در میانگین شاخص‌های باند ۲، باند ۳، شاخص عمودی خشکسالی و شاخص عمودی خشکسالی اصلاح شده تفاوت معنی‌داری وجود دارد. در باند ۱ و ۴ بین میانگین باندها در بهار و پاییز با تابستان تفاوت وجود دارد. در باند ۳۱، باند ۳۲ و شاخص‌های VCI، SAVI، MSAVI، NDVI، NVSWI، TVX و VHI بین میانگین در بهار و تابستان با پاییز تفاوت وجود دارد. در شاخص شادابی پوشش گیاهی حاصل از ضرایب ۰/۵ تفاوت معنی‌داری بین سه گروه زمانی وجود ندارد. مساله قابل توجه تفاوت تغییرات میانگین‌ها در سه گروه زمانی بین شاخص شادابی پوشش گیاهی حاصل از ضرایب ۰/۵ و شاخص شادابی پوشش گیاهی حاصل از کسر پوشش گیاهی است. در بررسی دیگر ضریب همبستگی بین اجزاء بافت خاک (درصد شن، رس، سیلت) و مولفه‌های رنگ خاک (هیو، ولیو و کروما) با باندها و شاخص‌های استخراجی تصاویر ماهواره‌ای مورد ارزیابی قرار گرفت که نتایج در شکل ۴ آورده شده است.

برای گره k که شامل N مورد است: $(x_{k1}, y_{k1}), \dots, (x_{kn}, y_{kn})$ پراکندگی داده‌ها براساس انحراف معیار اندازه گرفته می‌شود.

$$DEV(k) = \sqrt{\left(\sum_{i=1}^N (y_{ki} - \bar{y}_i)^2\right) / (N-1)} \quad (14)$$

به منظور یافتن بهترین متغیر تقسیم شده و بهترین مقدار تقسیم برای این متغیر، الگوریتم RT تمام متغیرهای تقسیم ممکن و هم‌چنین تمام مقادیر ممکن متغیر مورد استفاده برای تقسیم گره را بررسی می‌کند. فرض کنید که یک تقسیم می‌تواند گره K را به k_L و k_R تقسیم کند.

$$f = DEV(k) - DEV(k_L) - DEV(k_R) \quad (15)$$

بهترین تقسیم گره معادله ۱۵ را بیشینه می‌کند (Yang & Stenzel, 2006).

آماره‌های خطا جهت ارزیابی عملکرد مدل

برای مقایسه عملکرد رگرسیون درختی و رگرسیون بردار پشتیبان از برخی آماره‌های خطا استفاده شد. آماره‌های مورد استفاده شامل MAE^۱ (مجموع قدر مطلق اختلاف بین داده‌های مشاهداتی و شبیه‌سازی تقسیم بر تعداد داده)، MSE^۲ (مجموع توان دو اختلاف بین داده‌های مشاهداتی و شبیه‌سازی تقسیم بر تعداد)، RMSE^۳ (جذر مجموع توان دو اختلاف بین داده‌های مشاهداتی و شبیه‌سازی تقسیم بر تعداد)، RRMSE^۴ (تقسیم RMSE بر میانگین داده‌های مشاهداتی)، AMAPE^۵ (محاسبه مجموع قدرمطلق اختلاف بین داده‌های مشاهداتی و شبیه‌سازی تقسیم بر میانگین داده‌های مشاهداتی و در نهایت تقسیم بر تعداد)، RPD^۶ (تقسیم انحراف معیار بر RMSE) (Agussabti et al., 2020; Carmen Bas et al., 2017).

نتایج و بحث

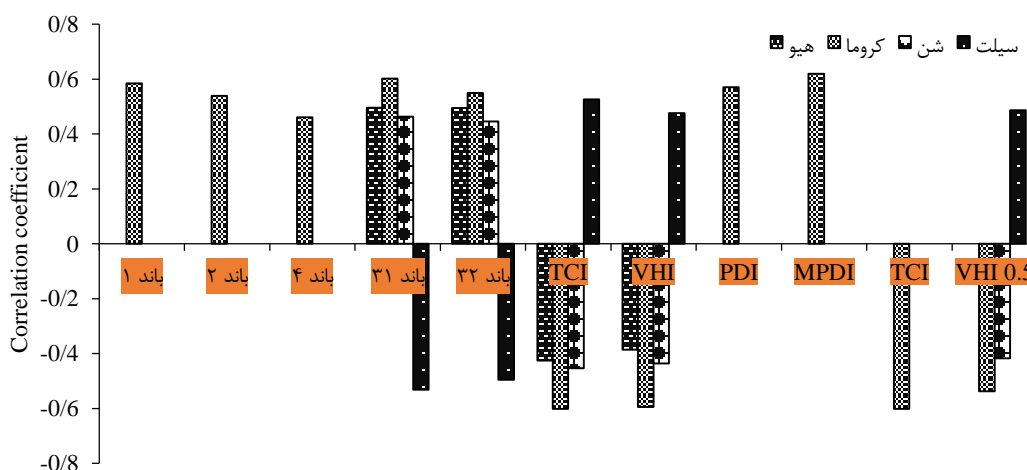
داده‌های ورودی مدل‌های رگرسیونی برای تعیین رنگ و بافت خاک شامل باندها (باند ۱، ۲، ۳، ۴، ۳۱ و ۳۲) و شاخص‌های (NDVI، PDI، SAVI، MPDI، TVX، TCI، MSAVI، VHI، NVWSI) حاصل از تصاویر ماهواره‌ای

4Relative Root Mean Square Error
5Adapted Mean Absolute Percentage Error
6Residual Predictive Deviation

1Mean Absolute Error
2Mean Square Error
3Root Mean Square Error

جدول ۱- نتایج آزمون دانکن داده‌های ورودی در مقایسه از نظر زمانی و تغییرات بافت خاک در سطح احتمال پنج درصد
Table 1. Result of Duncan's test of input variables basis on time and soil texture variation at the 5% probability level

Input variables	loamy sand	Sand	Autumn	Summer	Spring
Band1	A	a	b	a	B
Band2	a	a	c	a	B
Band3	a	a	b	a	C
Band4	a	a	b	a	B
Band31	b	a	b	A	A
Band32	b	a	b	a	A
NDVI	a	a	b	a	A
PDI	a	a	c	a	B
SAVI	a	a	b	a	A
MPDI	a	a	c	a	B
MSAVI	a	a	b	a	A
TCI	a	a	a	b	b
TVX	a	a	a	b	b
VHI -0.5	a	a	a	a	a
VHI	a	a	a	b	a
NVSWI	a	a	b	a	a
RVWSI	a	a	b	a	a
MVWSI	a	a	b	a	a
VCI	a	a	b	a	a
LST	a	a	b	a	a



شکل ۴- ضرایب همبستگی بافت و رنگ خاک با اطلاعات تصاویر ماهواره‌ای

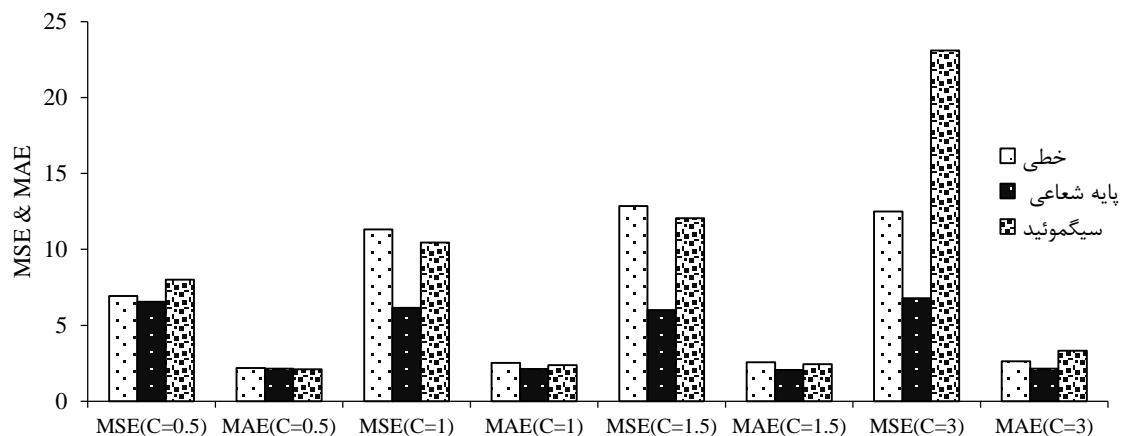
Figure 4. Correlation coefficient of soil texture and color with information of satellite images

۳۲، شاخص شرایط دمایی، شاخص شادابی پوشش گیاهی (هر دو روش) و شاخص پوشش گیاهی-دمایی بود. باند ۳۱ در تمامی اجزاء و مؤلفه‌ها به غیر از کروما دارای بیشینه مقدار ضریب همبستگی معنی‌دار بود. در کروما شاخص عمودی خشکسالی اصلاح شده دارای بیشینه مقدار ضریب همبستگی معنی‌دار بود. داده‌های ورودی که در تمامی اجزاء و مؤلفه‌ها دارای ضریب همبستگی معنی‌دار بوده و عبارتند از باند ۳۱، باند ۳۲، شاخص شرایط دمایی و

ضرایب همبستگی معنی‌دار باندها و شاخص‌ها با هیو شامل باند ۳۱، باند ۳۲، شاخص شرایط دمایی و شاخص شادابی پوشش گیاهی بود. در کروما شامل باند ۱، باند ۲، باند ۴، باند ۳۱، باند ۳۲، شاخص عمودی خشکسالی، شاخص عمودی خشکسالی اصلاح شده، شاخص شرایط دمایی، شاخص شادابی پوشش گیاهی (هر دو روش) و شاخص پوشش گیاهی-دمایی بود. در شن شامل باند ۳۱، باند ۳۲، شاخص شرایط دمایی و شاخص شادابی پوشش گیاهی (هر دو روش) بود. در سیلت شامل باند ۳۱، باند

تحلیل حساسیت انجام گرفته در رگرسیون بردار پشتیبان Epsilon در مورد تابع کرنل و پارامتر تنظیم کننده مدل (C) انجام گرفت که نتایج تحلیل حساسیت در مورد شن در شکل ۵ آورده شده است.

شاخص شادابی پوشش گیاهی (حاصل از کسر پوشش گیاهی) بودند. توابع کرنل مورداستفاده در رگرسیون بردار پشتیبان Epsilon شامل خطی، پایه شعاعی و سیگموئید می باشند.

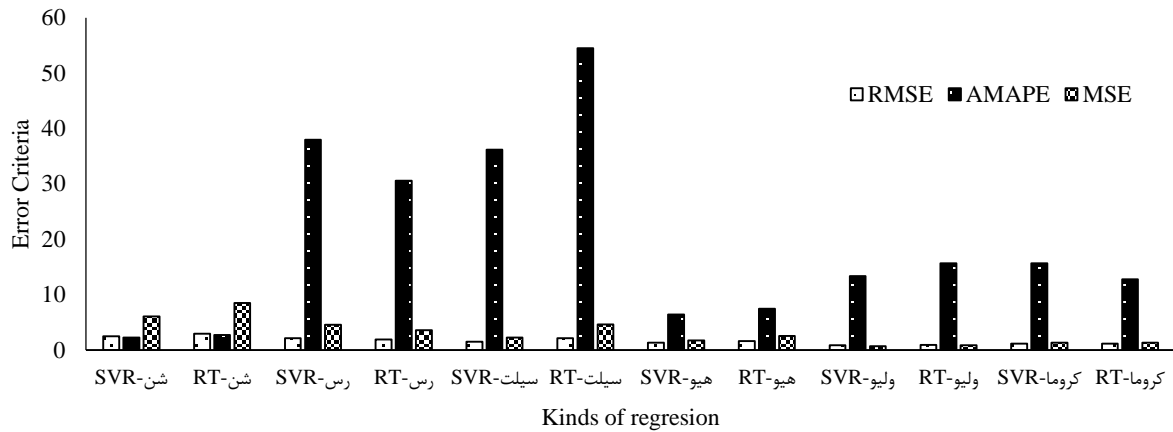


شکل ۵- تحلیل حساسیت رگرسیون بردار پشتیبان (شن)
Figure 5. Sensitivity analysis of support vector regression (sand)

کمینه تعداد برگها (پارامتر ۲) انجام گرفت به عنوان نمونه در تحلیل حساسیت شن، تغییرات MSE و MAE از ۱۱/۵۶ - ۲/۷۵ (پارامتر ۱ برابر با ۱۱ و پارامتر ۲ برابر با ۲) به ۸/۵ - ۲/۴۷ (پارامتر ۱ برابر با ۳ و پارامتر ۲ برابر با ۷) تغییر پیدا کرد.

بعد از انجام تحلیل حساسیت مدل های رگرسیونی، اقدام به برآورد اجزاء بافت خاک و مؤلفه های رنگ خاک با رگرسیون بردار پشتیبان (SVR) و رگرسیون درختی (RT) شد. برای مقایسه دو نوع رگرسیون از برخی آماره های خطا استفاده شد که نتایج در شکل ۶ آورده شده است.

بر اساس شکل ۵ کمینه خطا در پارامتر تنظیم کننده برابر با ۱/۵ و تابع کرنل پایه شعاعی است. به همین ترتیب تابع کرنل و پارامتر تنظیم کننده بهینه به این صورت می باشد: رس (پارامتر تنظیم کننده = ۳ تابع کرنل برابر با سیگموئید (پارامتر تنظیم کننده = ۴/۵۷، MSE=۱/۷، MAE=۱/۷)، سیلت (پارامتر تنظیم کننده = ۰/۵ تابع کرنل برابر با پایه شعاعی MSE=۲/۳۲، MAE=۱/۲۴)، هیو (پارامتر تنظیم کننده = ۰/۵ تابع کرنل برابر با پایه شعاعی MSE=۱/۷۴، MAE=۱/۱۸)، ولیو (پارامتر تنظیم کننده = ۰/۵ تابع کرنل برابر با سیگموئید MSE=۰/۶۷، MAE=۰/۷)، کروما (پارامتر تنظیم کننده = ۰/۵ تابع کرنل برابر با خطی MSE=۱/۳۱، MAE=۱) در مورد رگرسیون درختی تحلیل حساسیت در مورد کمینه تعداد گرهها برای تقسیم (پارامتر ۱) و



شکل ۶- مقایسه عملکرد رگرسیون بردار پشتیبان و درختی در برآورد رنگ و بافت خاک

Figure 6. Comparison the performance of support vector regression and tree regression for soil color and texture determination

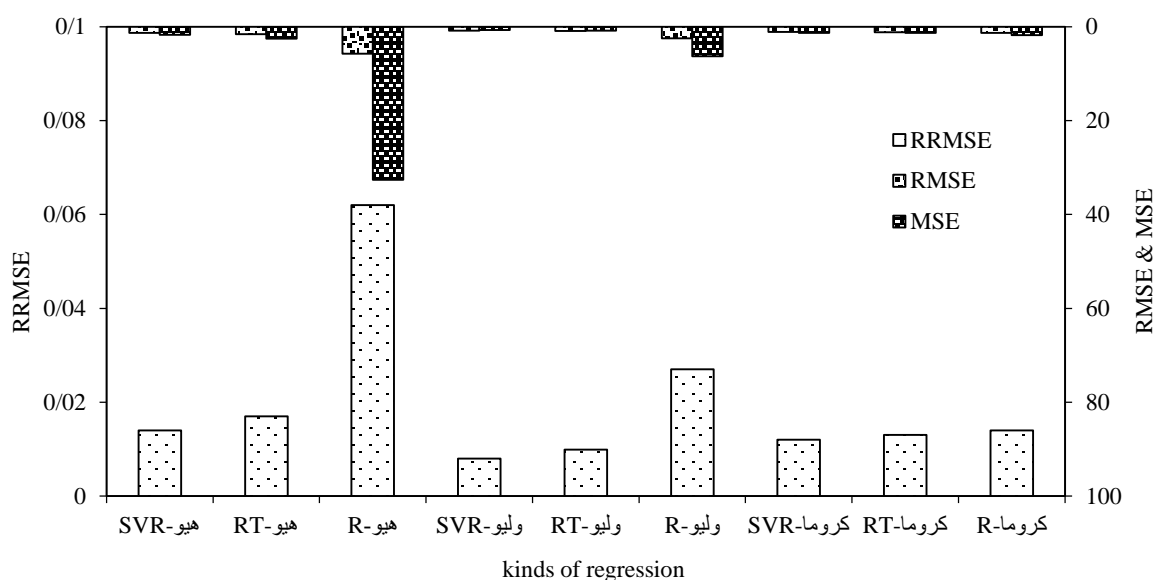
در مورد تمام موارد رنگ و بافت خاک محاسبه شد. میزان آماره RPD از رگرسیون درختی به رگرسیون بردار پشتیبان ۱۳/۴۳ درصد افزایش نشان داد که حاکی از عملکرد بهتر رگرسیون بردار پشتیبان است. در بررسی دیگر در مورد مقایسه عملکرد رگرسیون درختی و رگرسیون بردار پشتیبان به تعیین بافت خاک (با اطلاعات اجزای بافت خاک حاصل از دو نوع رگرسیون) با استفاده از مثلث بافت خاک در دوره صحت سنجی پرداخته شد که نتایج در جدول ۲ آورده شده است. همان طور که از جدول ۲ مشخص است در دوره صحت سنجی در نقاط نمونه برداری شده بجز یک مورد (۳۷/۸-۴۵/۹۳) که بافت آن شن لومی است، در بقیه نقاط بافت خاک شن می-باشد. این روند در داده‌های حاصل از رگرسیون بردار پشتیبان حفظ شده است. هم‌چنین دوره‌ای از صحت سنجی که در آن بیشینه مقدار شن، رس و سیلت اتفاق افتاده برای داده‌های مشاهداتی، داده‌های حاصل از رگرسیون بردار پشتیبان و رگرسیون درختی محاسبه شد. بین داده‌های مشاهداتی و داده‌های حاصل از رگرسیون درختی در سه جزء خاک مورد مشابهی در نقطه بیشینه مشاهده نشد، ولی در داده‌های حاصل از رگرسیون بردار پشتیبان در مورد رس در نقطه بیشینه دوره صحت سنجی (نقطه اول) با داده‌های مشاهداتی مشابهت وجود داشت. در حالت دیگر مقایسه‌ای بین عملکرد رگرسیون درختی، رگرسیون بردار پشتیبان و رگرسیون خطی چندگانه (R) در تعیین رنگ خاک انجام گرفت که نتایج در شکل ۷ آورده شده است.

با توجه به شکل ۶، میزان آماره‌های خطا در برآورد شن، سیلت، هیو و ولیو با استفاده از رگرسیون بردار پشتیبان دارای خطای کمی بود. در برآورد رس رگرسیون درختی دارای عملکرد بهتری بود (در مورد رس در RRMSE بین رگرسیون بردار پشتیبان و رگرسیون درختی تفاوت قابل ملاحظه‌ای نبود) و در مورد کروما اختلاف دو نوع رگرسیون کم بود. میزان کاهش RMSE، RRMSE، AMAPE و MSE در شن از رگرسیون درختی به رگرسیون بردار پشتیبان به ترتیب ۱۶/۴۱، ۱۳/۳۳، ۱۵/۴۳ و ۲۸/۷ درصد، در سیلت به ترتیب ۳۰/۴۳، ۳۰/۳۷، ۳۳/۶۳ و ۵۱/۴۱ درصد، در هیو به ترتیب ۱۶/۹۸، ۱۷/۶۴، ۱۴/۶۱ و ۳۱/۴۹ درصد، در ولیو به ترتیب ۱۰/۸۶، ۱۹/۱۹، ۱۴/۷۳ و ۲۰/۲۳ درصد بود. میزان کاهش RMSE، RRMSE و MSE در کروما از رگرسیون درختی به رگرسیون بردار پشتیبان به ترتیب ۰/۸۶، ۷/۶۹، ۱۴/۷۳ و ۰/۷۶ درصد بود. در تحقیق وو و همکاران (Wu 2018) عملکرد بهتر ماشین بردار پشتیبان در برابر شبکه عصبی مصنوعی و درخت طبقه بندی به اثبات رسید. در تحقیق شافری و رامل (Shafri & Ramle, 2009) در رابطه با مقایسه بین طبقه بندی تصاویر با ماشین بردار پشتیبان و درخت تصمیم، ماشین بردار پشتیبان عملکرد بهتری داشت. در تحقیق چوبین (Choubin et al., 2019) در زمینه پیش بینی سیلاب نیز ماشین بردار پشتیبان نسبت به رگرسیون درختی نتایج قابل قبول تری داشت.

برای بررسی بیش تر در مورد مقایسه عملکرد رگرسیون درختی و رگرسیون بردار پشتیبان، متوسط آماره RPD

جدول ۲- بافت خاک (مشاهداتی - شبیه‌سازی) در دوره صحت‌سنجی
Table 2. Soil texture (observation-simulation) in validation period

Validation Period	Observation	Support vector regression	Tree regression
1	Loamy Sand	Loamy Sand	Sand
2	Sand	Sand	Sand
3	Sand	Sand	Sand
4	Sand	Sand	Sand
5	Sand	Sand	Sand
6	Sand	Sand	Sand
7	Sand	Sand	Sand
8	Sand	Sand	Sand
9	Sand	Sand	Sand



شکل ۷- مقایسه عملکرد سه نوع رگرسیون در برآورد رنگ خاک

Figure 7. Comparison the performance of three kinds of regression for soil color determination

با توجه به شکل ۷ میزان آماره‌های خطا در رگرسیون ساده افزایش داشته است، به طوری که درصد کاهش RRMSE، RMSE و MSE از رگرسیون ساده به رگرسیون بردار پشتیبان در هیو به ترتیب ۷۶/۸۸، ۷۷/۴ و ۹۴/۶۶ و در رگرسیون درختی به ترتیب ۷۲/۱۵، ۷۲/۵۸ و ۹۴/۶۶ بود. درصد کاهش RRMSE، RMSE و MSE از رگرسیون ساده به رگرسیون بردار پشتیبان در ولیو به ترتیب ۶۷/۱۲، ۷۰/۳۷ و ۸۹/۳۴ و در رگرسیون درختی به ترتیب ۶۳/۲، ۶۳/۳۳ و ۸۶/۶۴ بود. درصد کاهش RRMSE، RMSE و MSE از رگرسیون ساده به رگرسیون بردار پشتیبان در کروما به ترتیب ۱۳/۶۳، ۱۴/۲۸ و ۲۶/۱۳ و در رگرسیون درختی به ترتیب ۱۲/۸۷

در تحقیق دهقانی بانیانی و همکاران (Dehghni Baniani *et al.*, 2012) رگرسیون درختی نسبت به رگرسیون خطی چندگانه در برآورد هدایت آبی اشباع و پارامتر عکس موئینگی دارای عملکرد بهتری بود. بررسی درصدهای کاهش یافته در تمام اجزاء رنگ خاک حاکی از آن است که درصدهای کاهش یافته با رگرسیون بردار پشتیبان بیش‌تر از رگرسیون درختی است و این مساله عملکرد بهتر رگرسیون بردار پشتیبان را نسبت به رگرسیون درختی نشان می‌دهد.

نتیجه گیری کلی

حالی رسیدن به راه حل سراسری در ماشین بردار پشتیبان تضمین شده است. الگوریتم یادگیری ماشین به طور خودکار در مورد ساختار مدل تصمیم می گیرد. رگرسیون بردار پشتیبان خطای بین پیچیدگی مدل و کیفیت برازش داده های آموزشی را کنترل می کند که پارامتر تنظیم کننده مدل در این مورد تأثیر دارد. انعطاف پذیری ماشین بردار پشتیبان مربوط به تابع کرنل است که به طور ضمنی منجر به نگاشت داده ها به فضای ابعادی بالا می شود. یک راه حل خطی در فضای ابعادی بالا با راه حل غیر خطی در فضای ابعادی کم مطابقت دارد که این مسأله ماشین بردار پشتیبان را در به کارگیری برای سری های زمانی مختلف کارا تر می کند. تغییر تابع کرنل امکان پیدا کردن ابرصفحه های برای طبقه بندی غیرخطی را ممکن می سازد. استفاده از توابع کرنل، رگرسیون بردار پشتیبان را برای تقریب های خطی و غیرخطی قابل استفاده می کند و از عملکرد خوبی در نتیجه استفاده از بردارهای پشتیبانی برای پیش بینی عدم وجود کمینه های محلی به دلیل خاصیت تحدب تابع هدف و قیود آن برخوردار است. با توجه به کارایی بالای رگرسیون بردار پشتیبان نسبت به رگرسیون درختی و اهمیت تعیین دقیق پارامترهای رگرسیون بردار پشتیبان یکی از پیشنهادات این تحقیق تلفیق یکی از الگوریتم های بهینه سازی مانند الگوریتم ژنتیک PSO با رگرسیون بردار پشتیبان است.

برآورد ویژگی های خاک مانند بافت، رنگ خاک در مقیاس بزرگ در مطالعات حفاظت خاک، کاربری اراضی از اهمیت چشمگیری برخوردار است. از موارد مهم در زمینه مدل سازی اطلاعات حاصل از تصاویر ماهواره ای و ویژگی های خاک، نوع مدل مورد استفاده می باشد. در این راستا عملکرد رگرسیون بردار پشتیبان و رگرسیون درختی مورد مقایسه قرار گرفت. نوع اطلاعات حاصل از تصاویر ماهواره ای هم در زمینه تخمین رنگ و بافت خاک دارای اهمیت است به طوری که باند ۳۱، باند ۳۲، شاخص شرایط دمایی و شاخص شادابی پوشش گیاهی در تمام اجزاء و مولفه های بافت و رنگ خاک ضریب همبستگی معنی داری داشتند. البته این مساله قابل تامل است که تعیین دقیق پارامترهای شاخص های ماهواره ای دارای اهمیت است مانند شاخص شادابی پوشش گیاهی. متوسط آماره RPD در مورد تمام اجزاء بافت و مولفه های رنگ خاک حاکی از عملکرد بهتر رگرسیون بردار پشتیبان بود. ماشین بردار پشتیبان بهترین تکنیک در تشخیص الگوی داده ها می باشد. ماشین بردار پشتیبان براساس کمینه سازی ریسک ساختاری می باشد که سعی در کمینه سازی حد بالای خطا با ایجاد تعادل بین خطای آموزشی و ظرفیت ماشین است. مدل هایی مانند شبکه عصبی ممکن است در نقاط بهینه موضعی گیر بیفتند در

References

- Agussabti R., Purwana S., and Agus A.M. 2020. Data analysis on near infrared spectroscopy as a part of technology adoption for cocoa farmer in Aceh Province, Indonesia, Data in brief 29 105251, 1-9.
- Asadzadeh F., Khosraviaqdam K., Yaghmaeian Mahabadi N., and Ramezanpour H. 2019. Spatial variation of mineral particles of the soil using remote sensing data and geostatistics to the soil texture interpolation. *Journal of Water and Soil*, 32(6): 1207-1222.
- Asgari N., Ayoubi SH., Dematte J.A.M., Jafari A., Safanelli J.L., and Silveira D.D. 2020. Digital mapping of soil drainage using remote sensing, DEM and soil color in a semiarid region of Central Iran. *Geoderma Regional*, 22(e00302): 1-10.
- Barman U., and Choudhury R.D. 2020. Soil texture classification using multi class support vector machine. *Information Processing in Agriculture*, 7: 318-332.
- Carmen Bas M., Ortiz J., Ballesteros L., and Martorell S. 2017. Forecasting 7BE concentrations in surface air using time series analysis. *Atmospheric Environment*, 155: 154-161.
- Choubin B., Moradi E., Golshan M., Adamowski J., Sajedi-Hosseini F., and Mosavi A. 2019. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of the Total Environment*, 651: 2087-2096.

- Coblinski J.A., Giasson E., Dematte J.A.M., Carnieletto A., Costa J.J.F., and Vašát R. 2020. Prediction of soil texture classes through different wavelength regions of reflectance spectroscopy at various soil depths. *Catena*, 189(104485): 1-12.
- Dehghani Banian S., Ghorbani Dashtaki SH., Mohammadi J., Khodaverdilo H., and Khalil Moghadam B. 2012. Comparing the performance of multiple linear regression and tree to predict saturated hydraulic conductivity and the inverse of macroscopic capillary length. *Iranian Water Research Journal*, 5(9): 193-204. (In Persian)
- Ding X., Zhao Zh., Yang Q., Chen L., Tian Q., Li X., and Meng F.R. 2020. Model prediction of depth-specific soil texture distributions with artificial neural network: A case study in Yunfu, a typical area of Udults Zone, South. *Computers and Electronics in Agriculture China*, 169(105217): 1-13.
- Hasanzadeh A., Zakeryan S., Mansouri Daneshvar M.R. 2019. Preparing soil color map using Landsat satellite imagery. *Geographical Sciences Journal*, 15(31): 23-33.
- Langhammer J., and Česák J., 2016. Applicability of a nu-support vector regression model for the completion of missing data in hydrological time series. *Water*, 8(12): 560-575.
- Mattikalli N.M. 1997. Soil color modeling for the visible and Near-Infrared bands of landsat sensors using laboratory spectral measurements. *Remote Sensing Environment*, 59: 14-28.
- Menxin W., and Houquan L. 2016. A modified vegetation water supply index (MVWSI) and its application in drought monitoring over Sichuan and Chongqing, China. *Journal of Integrative Agriculture*, 15(9): 2132–2141.
- Ostovari Y., Asgari K., and Motaghian H.R. 2015. Assessment of tree and multiple linear regressions in estimation of cation exchange capacity. *Journal of Water and Soil*, 29(3): 683-694.
- Parviz L. 2020. Performance evaluation of remote sensing data with machine learning technique to determine soil color. *Polish Journal of Soil Science*, LIII/1: 97-116.
- Sahwan W., Lucke B., Kappas M., and Baumler R. 2018. Assessing the spatial variability of soil surface colors in northern Jordan using satellite data from Landsat-8 and Sentinel-2. *European Journal of Remote Sensing*, 51(1): 850-862.
- Shafri H.Z.M., and Ramle F.S.H. 2009. A comparison of support vector machine and decision tree classification using satellite data of Langkawi Island. *Information Technology Journal*, 8(1): 64-70.
- Shahabfar A., Ghulam A., and Eitzinger J. 2012. Drought monitoring in Iran using the perpendicular drought indices. *International Journal of Applied Earth Observation and Geoinformation*, 18: 119–127.
- Stiglitz R., Mikhailova E., Post CH., Schlautman M., Sharp J., Pargas R., Glover B., and Mooney J. 2017. Soil color sensor data collection using a GPS-enabled smartphone application. *Geoderma*, 296: 108–114.
- Wu W., Li A.D., He X.H., Ma R., Liu H.B., and 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 J., and Stenzel J. 2006. Short-term load forecasting with increment regression tree. *Electric Power Systems Research*, 76: 880–888.

Soil Texture and Color Identification Using Artificial Intelligence Algorithm and Satellite Images

Laleh Parviz^{1*}

(Received: December 2020 Accepted: January 2021)

Abstract

The demand for quality and low-cost soil information is growing because of necessity in land use planning and precision agriculture. The aim of this study is soil texture and color estimation using satellite images information as input variables of support vector regression and tree regression. NDVI, PDI, SAVI, MPDI, MSAVI, TCI, TVX, VHI, NVWSI, RVWSI, MVWSI, VCI are satellite indices which related to the region in Azarshahr (East Azarbaijan). Duncan's test at the 5% probability level indicates significant time differences between the indices. There was no significant difference among the average of indices in terms of soil texture diversity. The error criteria RMSE, RRMSE, MAPE and MSE decreasing regard to sand from tree regression to support vector regression was 15.43, 13.33, 16.41 and 28.7%, respectively. Determination of soil texture with soil texture triangle in the validation period indicated the agreement of soil texture between observation and support vector regression. Considering soil texture and color components, RPD statistic increased from tree regression to support vector regression by 12.43%, which indicates the efficiency of support vector regression against tree regression. RMSE, RRMSE and MSE decreasing from multiple linear regression to support vector regression in hue were 76.88, 77.4 and 94.6%, respectively and for tree regression were 72.15, 72.58 and 92.92%, respectively, which is indicative of better performance of two regression models relative to simple regression. Based on various aspects of analysis, support vector regression had better performance for soil color and texture determination than tree regression.

Keywords: Satellite indices, Regression, Error Criteria, RPD

Parviz L. 2022. Soil texture and color identification using artificial intelligence algorithm and satellite images. *Applied Soil Research*, 9(4): 88-101.

1. Associate Professor, Faculty of Agriculture, Azarbaijan Shahid Madani University

* Corresponding Author Email: laleh_parviz@yahoo.com