

## نقشه برداری رقومی کلاس های خاک با استفاده از نقشه خاک قدیمی در منطقه خشک جنوب شرق ایران

فریده عباس زاده افشار<sup>۱،۲</sup>، شمس الله ایوبی<sup>۱</sup> و اعظم جعفری<sup>۳</sup>

(تاریخ دریافت: ۱۳۹۴/۸/۲۸؛ تاریخ پذیرش: ۱۳۹۵/۴/۸)

### چکیده

نقشه توزیع مکانی کلاس های خاک برای استفاده مفید و مؤثر از خاک و تصمیم گیری های مدیریتی مهم است. نقشه برداری رقومی خاک (DSM) می تواند توزیع مکانی کمی از کلاس های خاک پیش بینی کند. مؤلفه های کلیدی و اصلی DSM، روش ها و مجموعه ای از متغیرهای کمکی محیطی مورد استفاده برای پیش بینی کلاس های خاک هستند. این مطالعه به منظور تهیه نقشه رقومی گروه های بزرگ خاک با روش رگرسیون لجستیک چندجمله ای با استفاده از دو مجموعه از متغیرهای کمکی شامل: مجموعه (۱) متغیرهای مشتق شده از مدل رقومی ارتفاع، شاخص های سنجش از دور و سطوح ژئومورفیک تفکیک شده در منطقه و مجموعه (۲) متغیرهای مشتق شده از مدل رقومی ارتفاع، شاخص های سنجش از دور، ژئومورفیک تفکیک شده و واحدهای خاک شناسایی شده (نقشه خاک) در بخشی از اراضی شهرستان بم استان کرمان طراحی گردید. یک طرح نمونه برداری طبقه بندی شده تصادفی در منطقه ای به مساحت صد هزار هکتار تعریف شد و در نهایت، ۱۲۶ خاکرخ حفر و تشریح گردید. نتایج ارزیابی دقت مدل رگرسیون لجستیک چندجمله ای با متغیرهای ورودی مختلف نشان داد که با وارد شدن نقشه خاک قدیمی در مدل سازی، شاخص های صحت مدل از جمله صحت کلی و آماره کاپا به ترتیب از ۰/۷۱ و ۰/۶۵ به ۰/۷۹ و ۰/۷۴ افزایش یافت. همچنین نتایج نشان داد که در پیش بینی کلاس خاک، سطح ژئومورفولوژی در بین متغیرهای ورودی دو مجموعه، به عنوان یک متغیر پیش بینی کننده قدرتمند است. به طور کلی نتایج نشان داد که تکنیک های نقشه برداری رقومی می توانند روش سنتی نقشه برداری را ارتقاء بخشند، کاربرد نقشه های تولید شده را افزایش داده و همچنین قابلیت استفاده این نقشه ها را برای شاخه های علمی مختلف را امکان پذیر نماید.

واژه های کلیدی: ماشین یادگیری، رگرسیون لجستیک چندجمله ای، گروه بزرگ خاک، متغیرهای کمکی

۱. گروه علوم خاک، دانشکده کشاورزی، دانشگاه صنعتی اصفهان

۲. گروه علوم خاک، دانشکده کشاورزی، دانشگاه جیرفت

۳. گروه علوم خاک، دانشکده کشاورزی، دانشگاه باهنر کرمان

\*: مسئول مکاتبات: پست الکترونیکی: ayoubi@cc.iut.ac.ir

## مقدمه

نقشه‌های توزیع مکانی کلاس‌های خاک در تصمیم‌گیری‌های مدیریتی لازم و ضروری است و اطلاعات زیادی در اختیار کاربران قرار می‌دهند (۷). نقشه خاک با مقیاس ۱:۱۰۰۰۰۰۰ منبع اصلی اطلاعات خاک در ایران است (۱). این نقشه در اوایل سال ۲۰۰۰ میلادی تهیه و کامل گردید (۱). با توجه به وسعت زیاد کشور ایران و همچنین با گذشت یک دهه از تهیه این نقشه، نیاز به تهیه نقشه‌های با مقیاس منطقه‌ای و صحت و تفکیک بیشتر است. از مشکلات اصلی برای به‌هنگام کردن این نقشه، می‌توان به وسعت بزرگ کشور و همچنین کار میدانی وسیع، که جزء هزینه‌بر این پروژه است، اشاره کرد. نقشه‌برداری رقومی خاک (Digital soil mapping)، یکی از روش‌هایی است، که کار میدانی را کاهش می‌دهد (۶ و ۲۱). نقشه‌برداری رقومی خاک را می‌توان به‌عنوان مولد سیستم‌های اطلاعات مکانی خاک توسط مدل‌های عددی قابل استنباط از تغییرات مکانی و زمانی انواع خاک، خصوصیات خاک حاصل از مشاهدات و دانش خاک و متغیرهای محیطی تعریف کرد (۱۴). نقشه‌برداری رقومی خاک ممکن است مزایای بیشتری نسبت به نقشه‌برداری سنتی خاک داشته باشد؛ به‌عنوان مثال می‌توان به ضبط تغییرات مکانی بهتر مشاهدات و کاهش نیاز به نقاط نمونه‌برداری براساس مقیاس نقشه‌برداری اشاره کرد (۴۳). پایه و اساس نقشه‌برداری رقومی خاک مبتنی بر معادله اسکورپن (Scorpan) می‌باشد. در این معادله ویژگی‌ها یا کلاس‌های خاک با یک‌سری متغیرهای محیطی (متغیرهای کمکی) ارتباط داده می‌شود. به این صورت که در هر موقعیت مشاهده‌ی خاک، یک سری متغیرهای محیطی وجود دارد که همبستگی بالایی با ویژگی‌ها یا کلاس‌های خاک دارند، که یافتن این ارتباط و پیش‌بینی ویژگی‌های خاک یا کلاس‌های خاک در نقاط دیگر، منجر به تهیه نقشه رقومی خاک در تمام منطقه مورد مطالعه می‌گردد. بنابراین جزء اصلی معادله اسکورپن برنامه یادگیری (مدل پیش‌بینی کننده، Predictive model) می‌باشد. برای پیش‌بینی ویژگی‌های خاک یا کلاس خاک مدل‌های مختلفی

مورد بررسی قرار گرفته‌اند که شامل سیستم‌های کارشناسی (۳۳، ۳۸ و ۴۳)، طبقه‌بندی نظارت‌نشده (۱۰ و ۳۷) و ماشین یادگیری (۴، ۷ و ۲۲) هستند.

ماشین یادگیری یک اصطلاح کلی برای مجموعه گسترده‌ای از مدل‌ها برای کشف الگوهای موجود در داده‌ها و پیش‌بینی متغیرهای مورد مطالعه است (۴۱). اگر چه ماشین یادگیری اغلب به پایگاه داده‌های بزرگ نیاز دارد، یک ابزار جذاب برای یادگیری و پیش‌بینی مکانی کلاس‌های خاک، به‌دلیل دانش روابط بین کلاس‌های خاک و متغیرهای کمکی محیطی است (۱۳). انواع مختلفی از ماشین یادگیری برای نقشه‌برداری رقومی خاک استفاده شده است، از جمله می‌توان به رگرسیون لجستیک (۱۷، ۱۹، ۲۱ و ۲۶) درخت تصمیم (۲۲ و ۳۲) و شبکه عصبی مصنوعی (۵، ۱۸ و ۲۷) اشاره کرد.

دو روش معمول در پیش‌بینی کلاس‌های خاک با استفاده از ماشین یادگیری به کار رفته‌اند. اولین روش، جستجو برای پیدا و استخراج کردن روابط خاک-زمین‌نما از رقومی کردن نقشه‌های پلی‌گونی خاک هنگامی که موقعیت صحیح مشاهدات خاک مشخص نیست (۵، ۱۲ و ۳۱). روش دوم، جستجو برای پیدا و استخراج کردن روابط خاک-زمین‌نما از مشاهدات خاک که از نمونه‌برداری در نقاط مشخص به‌دست آمدند (۳، ۱۷، ۱۸، ۲۰ و ۳۵). انتخاب روش تا حد زیادی به در دسترس بودن مشاهدات بدون‌های خاک با موقعیت مشخص بستگی دارد. مطالعات اندکی در ارتباط با اسفاده از روش‌های DSM برای داده‌های طبقه‌ای از قبیل نوع خاک و یا کلاس‌های خاک، به‌ویژه هنگامی که روابط خاک-زمین‌نما نتیجه مشاهدات بدون‌های خاک باشند، انجام شده است. در ایران، نقشه‌برداری رقومی خاک توسط محققین متعددی مانند اسفندیارپور و همکاران (۱۰)، جعفری و همکاران (۱۸ و ۱۹)، تقی‌زاده و همکاران (۳۵) و پهلوان‌راد و همکاران (۲۹) به‌ترتیب در منطقه بروجن استان چهارمحال و بختیاری، زرنده استان کرمان، اردکان استان یزد و گرگان استان گلستان برای تهیه نقشه‌های گروه بزرگ خاک به‌کار گرفته شده است. همچنین کمپن و همکاران (۲۰)،

درجه و ۹ دقیقه شمالی و طول‌های جغرافیایی ۵۸ درجه و ۴ دقیقه تا ۵۸ درجه و ۲۸ دقیقه شرقی قرار گرفته است (شکل ۱). این منطقه، مساحتی حدود ۱۰۰۰۰۰ هکتار را می‌پوشاند. زمین‌نماهای اصلی این منطقه شامل مخروط افکنه، پلایا، باهادا، تپه و کوه می‌باشند. این منطقه توسط کوه‌های آهکی و آتش‌فشانی از شمال غرب تا جنوب شرق احاطه شده است. سطح زیر کشت محصولات کشاورزی این شهرستان ۶۵ تا ۷۰ هزار هکتار می‌باشد که مهم‌ترین محصولات باغی این شهرستان به ترتیب سطح زیر کشت، شامل خرما، مرکبات، گردو و زیتون می‌باشند. میانگین بارندگی سالیانه این منطقه ۶۴ میلی‌متر، متوسط درجه حرارت آن ۲۲ درجه سلسیوس و میانگین سالیانه تبخیر با تشتک تبخیر کلاس A در آن، ۳۰۰۰ میلی‌متر است. این منطقه براساس سیستم طبقه‌بندی آمریکا (۳۴) دارای رژیم رطوبتی اریدیک و رژیم حرارتی هایپرترمیک است.

#### تهیه نقشه ژئومرفولوژی منطقه

بر روی عکس‌های هوایی ۱/۴۰۰۰۰ منطقه روکش طلق مانندی کشیده شد و خطوط سطوح ژئومرفیک روی آنها ترسیم گردید. برای تفکیک واحدهای اراضی از سیستم استریوسکوپ برای تشخیص تفاوت‌ها استفاده شد. در تفکیک واحدها از فهم پدولوژیکی استفاده شد و براساس دانش ژئومرفولوژی و شناخت فرآیندهای تکامل واحدهای اراضی، مرز واحدها ترسیم شده و با ارتباط دادن آنها با تکامل خاک‌ها همراه با تأثیر دادن فاکتورهای زمین‌شناسی، توپوگرافی و پوشش گیاهی یا کاربری اراضی، سطوح ژئومرفیک یکنواختی ترسیم گردید. سپس عکس‌های هوایی و روکش مربوطه اسکن شده و وارد محیط سامانه جغرافیایی شد. در مراحل بعد، عکس‌ها وارد شده به سامانه جغرافیایی زمین مرجع شدند. با رقومی کردن مرزهای ترسیم شده بر روکش عکس‌ها، نقشه ژئومرفولوژی تهیه گردید (۳۶).

یک طرح نمونه‌برداری طبقه‌بندی شده تصادفی (Stratified sampling scheme) با استفاده از نقشه‌های رقومی زمین‌شناسی، ژئومرفولوژی و توپوگرافی پیاده گردید. در

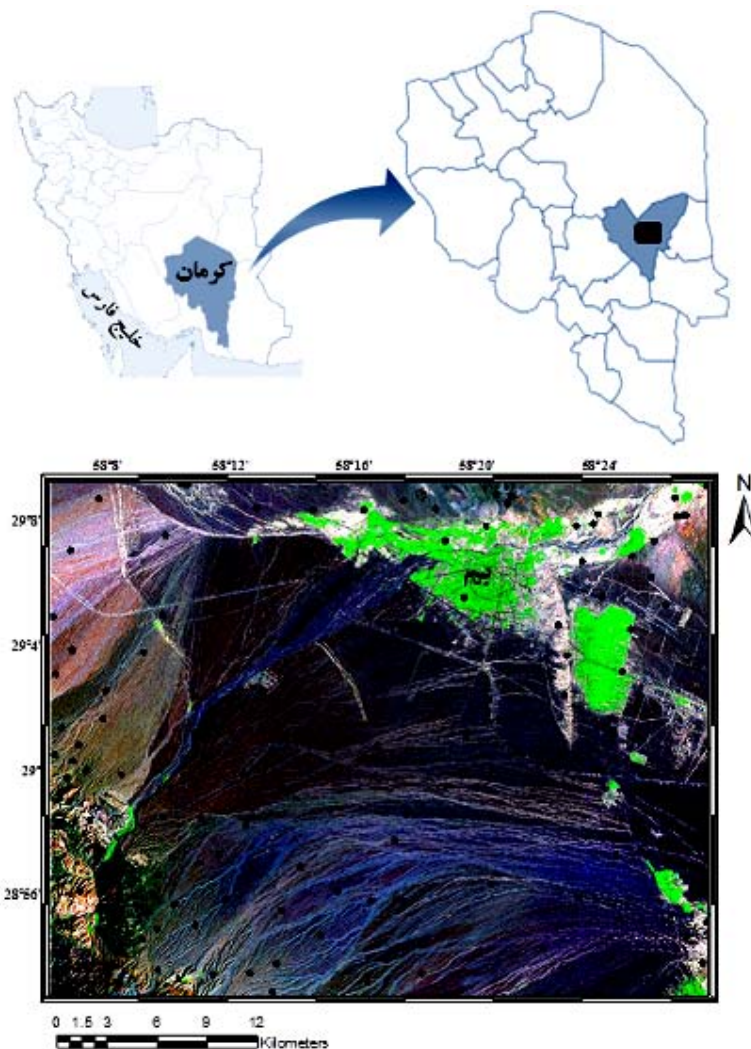
یانگ و همکاران (۴۲) و پهلوان راد و همکاران (۲۹) از تکنیک‌های ماشین یادگیری برای به‌هنگام کردن نقشه‌های خاک موجود منطقه مورد مطالعه استفاده کردند.

در حال حاضر از هر زمان دیگری به اطلاعات و نقشه‌های خاک بیشتر و با دقت و صحت بیشتر نیاز است. نگرانی درباره تأمین مواد غذایی و تغییرات آب و هوایی جهانی، کشاورزی و راه‌های مدیریت خاک را تغییر داده است. در نتیجه نیاز به اطلاعات خاک بیشتر احساس می‌شود. همچنین روش‌های سنتی برای تهیه نقشه ویژگی‌های خاک، براساس مشاهدات و آنالیز نقطه‌ای خاک هستند، که این روش‌ها مقرون به صرفه زمانی و اقتصادی برای زمین‌نماهای بزرگ نیستند. نه تنها این روش‌ها بسیار وقت‌گیر و پرهزینه‌اند (به دلیل هزینه‌های نیروی کار)، همچنین می‌توانند تهدیدی جدی برای هدررفت منابع خاک باشند. به دلیل این محدودیت‌ها، روش‌هایی سریع و ارزان، برای تکمیل تکنیک‌های سنتی نقشه‌برداری خاک مورد نیاز است. با توجه به نیاز کشور به نقشه‌های به‌هنگام خاک، این تحقیق با هدف کاربرد روش رگرسیون لجستیک چندجمله‌ای با استفاده از دو گروه از متغیرهای کمکی که شامل مجموعه (۱) متغیرهای مشتق شده از مدل رقومی ارتفاع، شاخص‌های سنجش از دور استخراج شده از تصاویر ماهواره‌ای و سطوح ژئومرفیک تفکیک شده در منطقه و مجموعه (۲) متغیرهای مشتق شده از مدل رقومی ارتفاع، شاخص‌های سنجش از دور، سطوح ژئومرفیک تفکیک شده و نقشه خاک قدیمی منطقه (کلاس‌های خاک) در بخشی از اراضی شهرستان بزم استان کرمان طراحی گردید. مقایسه کارایی این متغیرها برای پیش‌بینی گروه‌های بزرگ خاک با استفاده از معیارهای ارزیابی صحت مدل، از اهداف دیگر این تحقیق بود.

#### مواد و روش‌ها

##### توصیف جغرافیایی منطقه مورد مطالعه

منطقه مورد مطالعه در جنوب شرقی ایران و در شرق استان کرمان بین عرض‌های جغرافیایی ۲۸ درجه و ۵۲ دقیقه تا ۲۹



شکل ۱. موقعیت منطقه مورد مطالعه و نقاط نمونه برداری شده در شهرستان بم، استان کرمان

کلسی جیپسیدز، کلسی آرچیدز، پتروکلسیدز و توری ارتنتز هستند.

از لایه‌های ژنتیکی تمامی خاک‌ها و بعد از هوا خشک شدن و عبور از الک ۲ میلی متری، برخی ویژگی‌های فیزیکی و شیمیایی در ۴۶۰ نمونه خاک به شرح زیر اندازه‌گیری شد: بافت خاک به روش پی‌پت (۱۱)، اسیدیته و هدایت الکتریکی نمونه‌های خاک در عصاره ۲/۵ : ۱ خاک به آب به وسیله pH متر و هدایت‌سنج، کربنات کلسیم معادل خاک به روش تیتراسیون برگشتی (۲۸)، درصد کربن آلی خاک به روش اکسایش تر (۴۰)، درصد گچ به روش خشک کردن در آن

این پژوهش، لایه نمونه برداری منطبق بر تفاوت در اشکال اراضی (نقشه ژئومرفولوژی)، توپوگرافی (نقشه ارتفاع رومی) و زمین‌شناسی طراحی شد. به‌طوریکه در هر طبقه، موقعیت‌های نمونه برداری انتخاب گردید که اندازه نمونه‌ها متناسب با مساحت طبقه بود. در نهایت ۱۲۰ خاک‌خ حفر (شکل ۱)، تشریح، نمونه برداری، آنالیز و براساس معیارهای رده‌بندی آمریکایی طبقه‌بندی شد (۳۴). با تشریح خاک‌های حفر شده براساس سیستم طبقه‌بندی آمریکایی در منطقه مورد مطالعه هشت گروه بزرگ خاک شناسایی شد که شامل هیلوسالیدز، هیلوکلسیدز، هیلوجیپسیدز، هیلوکمیدز،

جدول ۱. داده‌های محیطی مورد استفاده به‌عنوان متغیرهای پیش‌بینی کننده در مدل‌سازی

| نوع متغیر | نام   | راه‌نما | ماهیت فاکتور | مقیاس     | داده‌های محیطی         |
|-----------|---|---------|--------------|-----------|------------------------|
| کمی       | Elevation (m)                                 | EL      | توپوگرافی    |           | مدل رقومی ارتفاع (DEM) |
| کمی       | Slope angle (%)                               | SL      |              |           |                        |
| کمی       | Mean Curvature                                | MCur    |              |           |                        |
| کمی       | Profile Curvature                             | PrCur   |              |           |                        |
| کمی       | Plan Curvature                                | PlCur   |              |           |                        |
| کمی       | Topographic Wetness Index                     | TWI     |              |           |                        |
| کمی       | Multi-resolution Valley Bottom Flatness index | MrVBF   |              |           |                        |
| کمی       | Stream Power Index                            | SPI     |              |           |                        |
| کمی       | Catchment Slope                               | CS      |              |           |                        |
| کمی       | Catchment Area                                | CA      |              |           |                        |
| کمی       | Normalized Difference Vegetation Index        | NDVI    | انعکاس نور   |           | تصویر لندست (ETM+)     |
| کمی       | Ratio Vegetation Index                        | RVI     |              |           |                        |
| کمی       | Perpendicular Vegetation Index                | PVI     |              |           |                        |
| کمی       | Clay Index                                    | CI      |              |           |                        |
| کمی       | Salinity Index                                | SI      |              |           |                        |
| کمی       | Gypsum Index                                  | GI      |              |           |                        |
| کمی       | Soil & Vegetation Index                       | SAI     |              |           |                        |
| کمی       | Carbonate Index                               | CaI     |              |           |                        |
| کیفی      | 17 geomorphic surface                         | GS      | ژئومرفولوژی  | ۱:۱۰۰۰۰۰  | نقشه ژئومرفولوژی       |
| کیفی      | 30 Geology                                    | Geol    | زمین‌شناسی   | ۱:۱۰۰۰۰۰  | نقشه ژئولوژی           |
| کیفی      | 4 Soil unit                                   | SU      | کلاس خاک     | ۱:۱۰۰۰۰۰۰ | نقشه خاک               |

ارائه شده، متغیرهای کمی در محیط سامانه جغرافیایی SAGA محاسبه و استخراج شد. روش استخراج همه پارامترهای مزبور را، هنگل و همکاران تشریح کردند (۱۶). شایان ذکر است که تصاویر ماهواره‌ای و مدل رقومی ارتفاع استفاده شده دارای ابعاد ۳۰×۳۰ متر بودند. شاخص‌های سنجش از دور و اجزاء سرزمین، نقشه ژئومرفولوژی و نقشه خاک در نرم افزار SAGA به سیستم مرجع جهانی یکسانی (WGS 84 UTM 40) تبدیل شدند.

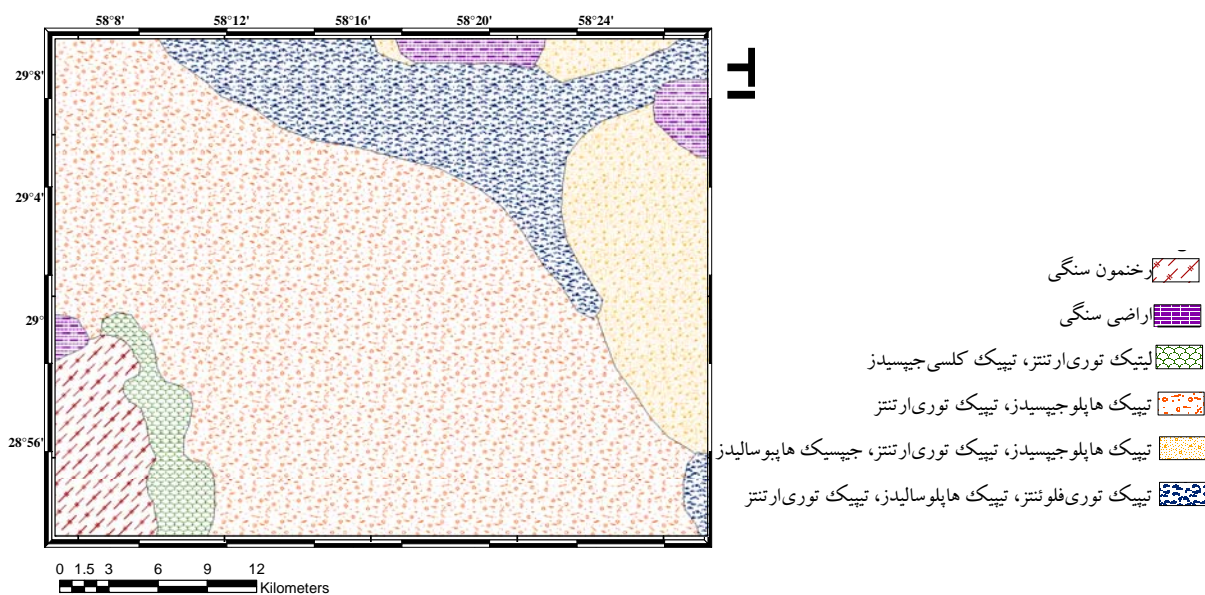
#### آماده‌سازی داده‌ها برای مدل‌سازی

هدف از ماشین یادگیری پیدا کردن یک تقریب مفید از تابعی است که اساس آن پیش‌بینی رابطه بین متغیرهای کمی ورودی

(۲۸)، نسبت جذب سدیم پس از اندازه‌گیری غلظت کاتیون‌های سدیم، کلسیم و منیزیم در عصاره اشباع محاسبه شد و ظرفیت تبادل کاتیونی به روش استات سدیم در اسیدپته ۷ (۳۰) تعیین شد.

#### متغیرهای کمی

از یک مدل رقومی ارتفاع که از وب سایت مدل رقومی ارتفاع جهانی استر (Aster Global Digital Elevation Model) به‌دست آمد، برای تعیین اجزاء سرزمین و یک سین تصویر ماهواره‌ای ETM+ برای تعیین شاخص‌های گیاهی، شاخص رس و شاخص شوری استفاده گردید. اجزاء سرزمین حاصل از مدل رقومی ارتفاع و شاخص‌های سنجش از دور در جدول (۱)



شکل ۲. نقشه خاک قدیمی با مقیاس ۱:۱۰۰۰۰۰۰ منطقه مورد مطالعه (۱)

و نتایج مورد نظر است (۱۵). در این مطالعه متغیرهای ورودی کمکی شامل شاخص‌های سنجش از دور و اجزاء سرزمین، نقشه ژئومرفولوژی و نقشه خاک و متغیر هدف، گروه‌های بزرگ خاک بودند. این متغیرهای کمکی به دو مجموعه تبدیل شدند و به‌عنوان متغیر ورودی برای پیش‌بینی گروه‌های بزرگ خاک استفاده شدند. این مجموعه‌ها شامل: مجموعه ۱، متغیرهای مشتق شده از مدل رقومی ارتفاع، شاخص‌های سنجش از دور و سطوح ژئومرفیک تفکیک شده در منطقه (نقشه ژئومرفولوژی) و مجموعه ۲، متغیرهای مشتق شده از مدل رقومی ارتفاع، شاخص‌های سنجش از دور، سطوح ژئومرفیک تفکیک شده و واحدهای خاک شناسایی شده در نقشه خاک قدیمی منطقه مورد مطالعه بودند. در شکل (۲) نقشه خاک قدیمی منطقه مورد مطالعه ارائه شده است. گروه‌های بزرگ خاک طبقه‌بندی شده در این نقشه که به‌صورت ترکیبی (Associate) است، شامل، توری ارتنتز، هپلو جیپسیدز، هپلوسالیدز و توری فلونتز هستند (۱).

به‌عنوان متغیرهای مستقل در مدل اسکورپن شناخته می‌شوند. اطلاعات خاک و متغیرهای محیطی به ترتیب به فرمت sgrd و csv تبدیل و برای انجام فرآیند مدل‌سازی در نرم‌افزار مربوطه آماده شد.

### مدل‌سازی خاک

مدل رگرسیون لجستیک، جزء مدل‌های خطی تعمیم یافته است که می‌تواند برای توصیف رابطه چندین متغیر مستقل (X) با یک متغیر وابسته دو یا چند حالتی (Y) مورد استفاده قرار گیرد. در مدل‌سازی متغیر پاسخ دو حالتی، سه مسئله خاص در مقایسه با رگرسیون خطی رخ می‌دهد: نرمال نبودن خطاها، ثابت نبودن واریانس خطاها و محدودیت تابع پاسخ. سه تابع برای مدل‌سازی متغیر پاسخ دو حالتی وجود دارد که هر سه، جزء مدل‌های خطی تعمیم یافته هستند و می‌توانند در کاربردهای مختلف مورد استفاده قرار گیرند: لاجیت (Logistic)، پروبیت و لوگ-لوگ مکمل (۲۳).

مدل لجستیک به دو صورت: لجستیک دوتایی ولجستیک چند جمله‌ای قابل اجرا است. در این مطالعه، از مدل رگرسیون لجستیک چند جمله‌ای برای پیش‌بینی گروه بزرگ‌های خاک (متغیر هدف) استفاده شد. متغیر وابسته (گروه بزرگ خاک)

شده صورت گرفت. یک آرایه خطا برای محاسبه معیارهای صحت استفاده گردید. صحت کلی ارتباط بین همه داده‌های مورد استفاده و داده‌های طبقه‌بندی شده (ff,tt) را تشریح می‌کند:

$$\text{Overall accuracy} = \frac{tt + ff}{tt + ft + tf + ff} \quad (4)$$

صحت کاربر شامل کلاس‌های حضور صحیح پیش‌بینی شده (tt) به مجموع کلاس‌های حضور صحیح پیش‌بینی شده و کلاس‌های عدم حضور که به غلط جزء کلاس‌های حضور مشاهده شده، پیش‌بینی شدند (tf) می‌باشد:

$$\text{users' accuracy} = \frac{tt}{tt + tf} \quad (5)$$

قابلیت اطمینان تولید کننده ارتباط بین همه کلاس‌های صحیح پیش‌بینی شده و مجموع کلاس‌های صحیح پیش‌بینی شده کلاس‌های حضور مشاهده شده که به غلط جزء کلاس‌های عدم حضور پیش‌بینی شدند (ft) می‌باشد:

$$\text{producers' reliability} = \frac{tt}{tt + ft} \quad (6)$$

آماره کاپا یک شاخص قوی است که نسبت احتمال حضور یا عدم حضور کلاس، که به درستی به وسیله مدل پیش‌بینی شدند را محاسبه می‌کند. بنابراین آماره کاپا همیشه کمی کمتر از خلوص نقشه است.

$$\text{kappa index} = \frac{\text{observed accuracy} - \text{chance agreement}}{1 - \text{chance agreement}} \quad (7)$$

که در آن، دقت مشاهده به‌عنوان مجموع ارزش قطر در ماتریس و شانس خطا شامل مجموع ارزش غیر قطر ماتریس است. آماره کاپا بالاتر نشان دهنده عملکرد مدل بالا است. دامنه تغییرات آماره کاپا بین صفر تا یک است. آماره کاپا بالاتر از ۰/۸، ۰/۴-۰/۸ و کمتر از ۰/۴ به ترتیب نشان دهنده توافق قوی، متوسط و ضعیف هستند (۱۹).

درجه برای تفاوت بین کلاس واقعی و احتمال تخمین‌زده شده از کلاس واقعی را محاسبه می‌کند:

کلاس داده‌های تست،  $F_{ij}$  برآورد احتمال تعداد مشاهدات  $n_i$  متعلق به کلاس  $i$  و  $F_{ij}$  یک متغیر شاخص، از جمله  $E_{ij}=1$  اگر  $n_i$  یک گروه بزرگ خاک و ۰ دیگر گروه‌های بزرگ خاک است.

دارای بیش از یک سطح است و حل آن به‌صورت دوتایی امکان پذیر نیست، در این صورت باید یک سطح متغیر وابسته به‌عنوان طبقه مرجع انتخاب شود. یک مدل لجستیک چند جمله‌ای با طبقه مرجع به شکل زیر نمایش داده می‌شود:

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \alpha_j + \beta_j x_i, \quad j = 1, 2, \dots, J-1 \quad (1)$$

که  $\alpha_j$  یک ثابت،  $\beta_j$  یک برداری از ضرایب رگرسیون برای  $j = 1, 2, \dots, J-1$  و  $x_i$  برداری از متغیرهای توصیفی است.

این مدل شبیه مدل لجستیک دو جمله‌ای است به استثناء اینکه توزیع احتمال متغیر هدف به جای دو جمله‌ای، چند جمله‌ای است و به جای یک معادله،  $J-1$  معادله داریم.

$$P(y_i = j) = \pi_{ij} \frac{\exp(\beta_j x_i)}{1 + \sum_{j=1}^j \exp(\beta_j x_i)} \quad (2)$$

بنابراین، احتمال طبقه مرجع:

$$P(y_i = 0) = \pi_{i0} = \frac{1}{1 + \sum_{j=1}^j \exp(\beta_j x_i)} \quad (3)$$

این مدل در نرم‌افزار R با استفاده از بسته nnet قابل اجرا است. اهمیت ضریب رگرسیون هر متغیر پیش‌بینی کننده برای هر متغیر با استفاده از معیار آماری والد ارزیابی گردید. این معیار در واقع از تقسیم ضریب رگرسیون هر متغیر مستقل به خطای استاندارد آن به دست می‌آید. بنابراین، ما می‌توانیم اهمیت هر پیش‌بینی کننده را استنباط کنیم، بدین معنی که هر چه مقدار مطلق این پارامتر بزرگتر باشد، اهمیت متغیر پیش‌بینی کننده بیشتر می‌شود (۱۹).

## بررسی عملکرد مدل‌های پیش‌بینی کننده با دو مجموعه

### متغیر ورودی

برای بررسی صحت و دقت مدل‌های مورد استفاده، داده‌ها به‌طور تصادفی به داده‌های آموزشی و اعتبارسنجی تقسیم شدند. داده‌های آموزشی، ۸۰ درصد و داده‌های اعتبارسنجی، ۲۰ درصد کل داده‌ها را شامل شدند. هر مدل با داده‌های آموزشی برآورد داده شد و سپس پیش‌بینی برای داده‌های اعتبارسنجی انجام گردید. اعتبارسنجی با مقایسه داده‌های مشاهده شده و پیش‌بینی

**جدول ۲. متغیرهای وارد شده در پیش‌بینی گروه‌های بزرگ خاک در برازش مدل رگرسیون لجستیک چندجمله‌ای**

| مجموعه متغیرهای ورودی | متغیرهای وارد شده در مدل*                     |
|-----------------------|---|
| مجموعه ۱              | GS, MrVBF, PVI, EL, RVI, GI, SL, SVI, CS, CaI |
| مجموعه ۲              | GS, MrVBF, EL, SU, PVI, RVI, CI, SVI, CaI     |

\* علائم اختصاری ارائه شده در جدول به طور کامل در جدول ۱ تعریف شده‌اند

نقشه خاک قدیمی موجود در منطقه مورد مطالعه است، بعد از انجام مدل‌سازی، علاوه بر سطح ژئومرفیک و پارامترهای سرزمین و شاخص‌های سنجش از دور، نقشه خاک موجود با مقیاس ۱:۱۰۰۰۰۰۰ نیز وارد مدل شد. در این مجموعه از بین پارامترهای سرزمین، شاخص همواری دره با درجه تفکیک بالا (MrVBF) و ارتفاع (EL) در پیش‌بینی گروه‌های بزرگ خاک نقش داشتند. نقش مؤثر فرآیندهای ژئومورفولوژیکی در توسعه و تکامل خاک و همبستگی نزدیک ژئومورفولوژی و خاک در بسیاری از مطالعات خاک-ژئومورفولوژی نشان داده شده است (۹، ۱۹، ۲۹، ۳۵). آکسوی و همکاران (۲) گزارش کردند ترکیب اجزای سرزمین و داده‌های سنجش از دور، ابزاری قدرتمند در نقشه‌برداری خاک است. ویژگی‌های مختلف مدل رقومی ارتفاع هم به لحاظ منطقی و ریاضی و هم از نظر تجربی، دارای رابطه‌ی نزدیکی با ویژگی‌های محیطی و خاک هستند. استفاده از این ویژگی‌ها در شناسایی خاک می‌تواند تا حد زیادی بر دقت کار بیفزاید و از صرف هزینه و زمان بکاهد. دبلاگیلو و اتزلمولر (۹)، جعفری و همکاران (۱۸) و وایسی و لاگچری (۳۹) تأثیر معنی‌دار ویژگی‌های توپوگرافی را در پیش‌بینی کلاس‌ها و ویژگی‌های خاک با استفاده از روش رگرسیون لجستیک چندجمله‌ای را نشان دادند. همچنین می‌توان بیان کرد که حضور ویژگی‌های پستی و بلندی به‌همراه فرآیندهای ژئومرفیک در پیش‌بینی گروه‌های بزرگ خاک نشان داد که این فاکتورها برآیند نیروهای مؤثر بر تشکیل خاک در منطقه مورد مطالعه هستند. پهلوان راد و همکاران (۲۹) نیز از نقشه‌برداری رقومی خاک برای به‌هنگام کردن نقشه خاک موجود در منطقه گرگان استان گلستان استفاده کردند. نتایج آنها نشان داد که

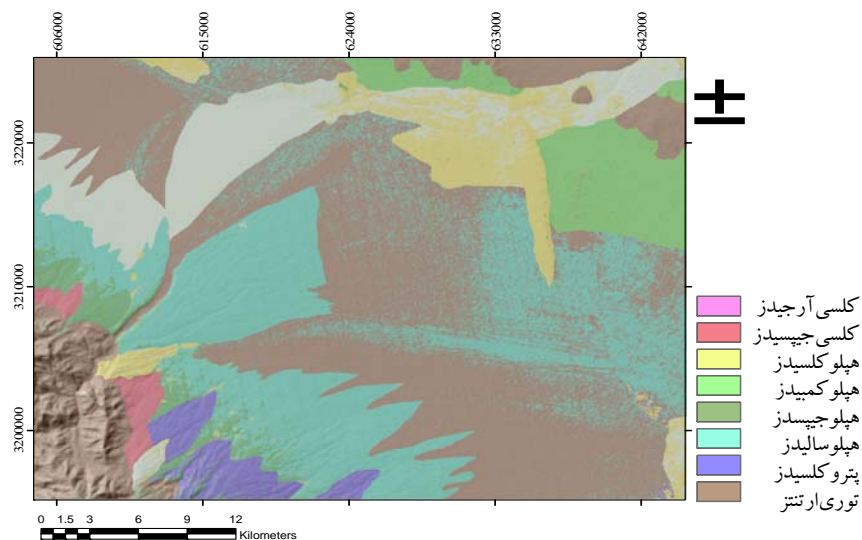
درجه برای بین محدوده صفر و دو است، درجه برای پایین‌تر نشان‌دهنده عملکرد بهتر مدل است (۷). یک نقشه خوب دارای مقادیر صحت کلی، صحت کاربر، قابلیت اطمینان تولید کننده و آماره کاپا نزدیک به یک و درجه برای پایین است.

### نتایج و بحث

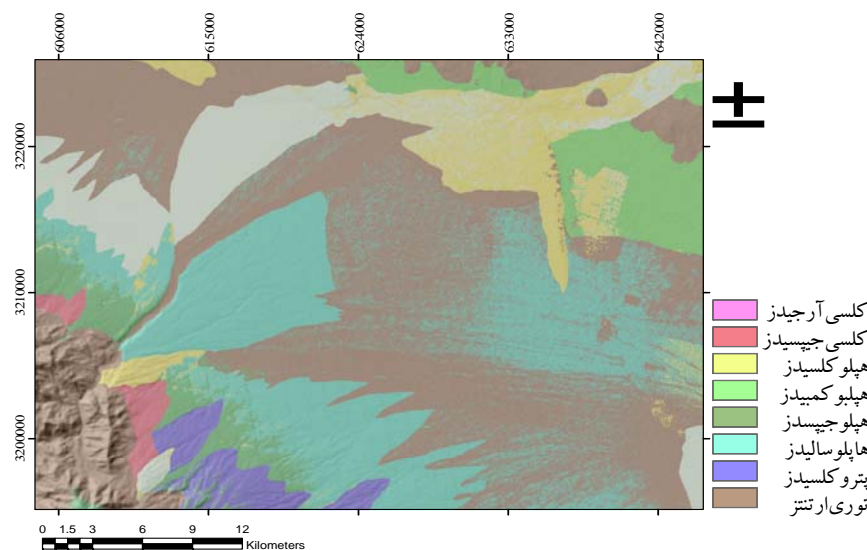
نتایج مدل‌سازی توسط رگرسیون لجستیک چندجمله‌ای و دو مجموعه از متغیرهای ورودی جهت پیش‌بینی گروه‌های بزرگ خاک در جدول (۲) ارائه شده است. در بین متغیرهای ورودی دو مجموعه، سطح ژئومرفیک اولین متغیری بود که وارد مدل شد و به‌عنوان یک متغیر پیش‌بینی کننده قدرتمند انتخاب شد. علت این نتیجه را می‌توان چنین بیان کرد که سطوح ژئومرفیک در طول یک دوره زمین‌شناسی همراه با شرایط تشکیل خاک نزدیک به فرآیندهای خاک‌سازی کنونی در مناطق خشک تشکیل شدند، چرا که سطوح ژئومرفیک نزدیک‌ترین سطح به شرایط خاک در مقیاس بزرگ می‌باشد. نتایج جعفری و همکاران (۱۹) و پهلوان راد و همکاران (۲۹) نیز نشان داد که نقشه ژئومورفولوژی، متغیری مهم برای پیش‌بینی گروه‌های بزرگ خاک می‌باشد.

در مجموعه ۱ علاوه بر سطح ژئومرفیک، برخی از متغیرهای مشتق شده از مدل رقومی ارتفاع و شاخص‌های سنجش از دور وارد مدل شدند. این نتایج تأیید کننده این است که توپوگرافی و پستی و بلندی و همچنین پوشش گیاهی از جمله فاکتورهای مؤثر در خاک‌سازی در منطقه مورد مطالعه هستند. مجموعه ۲ که شامل همه متغیرهای ورودی به همراه





شکل ۳. توزیع مکانی گروه‌های بزرگ خاک با استفاده از مدل رگرسیون لجستیک چند جمله‌ای با متغیرهای ورودی مجموعه ۱



شکل ۴. توزیع مکانی گروه‌های بزرگ خاک با استفاده از مدل رگرسیون لجستیک چند جمله‌ای متغیرهای ورودی مجموعه ۲

مورد مطالعه ارائه شده است. در نقشه‌های پیش‌بینی شده، هشت گروه بزرگ موجود در منطقه مورد مطالعه تشخیص داده شد. نقشه خاک قدیمی موجود در منطقه فقط چهار نوع خاک به صورت ترکیبی را نشان می‌دهد. نتایج حاصل از ماشین یادگیری و نقشه‌های خروجی با هر دو مجموعه متغیرهای ورودی نشان داد که روش نقشه‌برداری رقومی باعث ارتقاء

علاوه بر نقشه خاک موجود، شاخص SAVI و نقشه ژئومرفولوژی از متغیرهای مهم در پیش‌بینی کلاس‌های خاک بودند. همچنین نقشه رقومی تهیه شده ۱۳/۴ درصد دقیق‌تر، نسبت به نقشه خاک موجود بود.

در شکل (۳) و (۴) نقشه‌های خروجی مدل رگرسیون چندجمله‌ای با استفاده از متغیرهای ورودی مختلف در منطقه

جدول ۳. کیفیت پیش‌بینی مدل‌های مورد استفاده برای گروه‌های بزرگ خاک در داده‌های اعتبارسنجی

| PR   | UA   | Brier score | Kappa index | MP           | متغیرهای ورودی | گروه‌های بزرگ خاک |
|------|------|-------------|-------------|--------------|----------------|-------------------|
| ۱    | ۱    | ۰/۵۷        | ۰/۶۵        | ۰/۷۱         | مجموعه ۱       | هپلوسالیدز        |
| ۰/۵۰ | ۰/۸۳ |             |             | هپلوچیپسیدز  |                |                   |
| ۰/۲۵ | ۰/۲۵ |             |             | هپلوکلسیدز   |                |                   |
| ۰/۴۰ | ۰/۵۰ |             |             | هپلوکمیدز    |                |                   |
| ۱    | ۱    |             |             | کلسی‌چیپسیدز |                |                   |
| ۱    | ۱    |             |             | کلسی‌آرجیدز  |                |                   |
| ۱    | ۱    |             |             | پتروکلسیدز   |                |                   |
| ۱    | ۰/۶۱ |             |             | توری‌ارتنتز  |                |                   |
| ۱    | ۱    | ۰/۵۳        | ۰/۷۴        | ۰/۷۹         | مجموعه ۲       | هپلوسالیدز        |
| ۰/۵۵ | ۰/۸۳ |             |             | هپلوچیپسیدز  |                |                   |
| ۰/۵۰ | ۰/۵۰ |             |             | هپلوکلسیدز   |                |                   |
| ۰/۶۰ | ۰/۷۵ |             |             | هپلوکمیدز    |                |                   |
| ۱    | ۱    |             |             | کلسی‌چیپسیدز |                |                   |
| ۱    | ۱    |             |             | کلسی‌آرجیدز  |                |                   |
| ۱    | ۱    |             |             | پتروکلسیدز   |                |                   |
| ۱    | ۰/۶۹ |             |             | توری‌ارتنتز  |                |                   |

مشاهدات، خاک‌های این سطح در گروه‌های بزرگ هپلوکلسیدز و هپلوکمیدز طبقه‌بندی شدند. اما در نقشه خاک قدیمی در سطح پلایا گروه‌های بزرگ توری‌فلوئنتز، هپلوسالیدز و توری‌ارتنتز شناسایی شدند.

در جدول ۳ نتایج ارزیابی صحت مدل رگرسیون لجستیک چندجمله‌ای با متغیرهای ورودی مختلف ارائه شده است. صحت کلی برای گروه‌های بزرگ خاک با استفاده از این مدل با مجموعه ۱ و ۲ به ترتیب ۰/۷۱ و ۰/۷۹ محاسبه شد. در مدل‌های برازش شده صحت کلی برای همه گروه‌های بزرگ خاک در داده‌های اعتبارسنجی بالاتر از ۰/۷ بود که حاکی از پیش‌بینی نسبتاً خوب این مدل‌ها بود. همچنین آماره کاپا برای هر دو متغیرهای ورودی بین ۰/۸-۰/۴ محاسبه شد، که نشان‌دهنده این است که دقت مدل از شرایط متوسط برخوردار هستند.

پایین‌ترین کیفیت پیش‌بینی مدل برای هر دو مجموعه

نقشه قدیمی خاک در منطقه مورد مطالعه شده و خاک‌های منطقه با دقت و جزئیات بیشتری تفکیک و شناسایی شدند. آنچه که از شکل‌های (۲) و (۳) منتج می‌شود این است که هر دو مدل گروه بزرگ هپلوسالیدز در پدیمت، کلسی‌چیپسیدز، کلسی‌آرجیدز و پتروکلسیدز در باهادای بریده و باهادا را به خوبی پیش‌بینی کردند. با توجه به مقیاس نقشه خاک قدیمی و مقیاس مطالعه، گروه بزرگ‌های کلسی‌آرجیدز و پتروکلسیدز در این نقشه شناسایی نشدند.

با اطلاعات تصویر ماهواره‌ای منطقه و تفسیر عکس‌های هوایی، جهت حرکت آبراهه‌ها و شیب منطقه به طرف زمین ریخت پلایا است و باعث تجمع رسوبات ریزدانه در این سطح شده است. در منطقه مورد مطالعه، بیشتر باغ‌های خرما در سطح ژئومرفیک پلایا قرار دارند. آبیاری غرقابی باغ‌های خرما، باعث شستشو و خارج شدن نمک و گچ از این خاک‌ها شده و افق‌های کمبیک و کلسیک تشکیل شده‌اند. براساس این

تأثیری در صحت پیش‌بینی این گروه بزرگ خاک نداشت. در بین گروه‌های بزرگ خاک در هر دو مدل، بهترین پیش‌بینی برای گروه بزرگ هیلوسالیدز، کلسی‌جیسیدز، کلسی‌آرجیدز و پتروکلسیدز به دست آمد، مقادیر بالای صحت کاربر و قابلیت اطمینان تولید کننده در جدول (۳) این مسئله را نشان می‌دهد. دبالگیلو و ازلمولر نشان دادند نواحی با احتمال بالا برای هر گروه بزرگ خاک با زمین‌نماهای کاملاً شناخته شده، منطبق است (۹).

شاخص صحت و قابلیت اطمینان برای داده‌های اعتبارسنجی برای گروه‌های بزرگ مختلف خاک در هر دو مجموعه ورودی متفاوت بود. با این حال، صحت برای پیش‌بینی گروه‌های بزرگ خاک به دست آمده از مدل با مجموعه ۲ بالاتر از مدل با مجموعه ۱ محاسبه شد. نتایج ارائه شده در جدول (۳) تأیید میکنند که مقادیر صحت کاربر و قابلیت اطمینان مدل با مجموعه ۱ و مجموعه ۲، دامنه تغییرات بین ۰/۲۵-۱ و ۰/۵۰-۱ داشتند. صحت کاربر و قابلیت اطمینان تهیه‌کننده به ما اجازه می‌دهد که سطح بیش‌برآورد و کم‌برآورد در پیش‌بینی گروه‌های بزرگ خاک را تخمین بزنیم (۲۴). صحت کاربر برای گروه‌های بزرگ (۰/۸۳) هیلوجیسیدز و (۰/۵۰) هیلوکمبیدز در هر دو مجموعه همواره بیشتر از قابلیت اطمینان تهیه‌کننده اندازه‌گیری شد، که این نشان می‌دهد مدل‌ها این گروه‌های بزرگ را در منطقه مورد مطالعه کم‌برآورد کرده‌اند. گروه بزرگ توری‌ارتتت، با قابلیت اطمینان بیشتر از صحت کاربر واضح است که پیش‌بینی آن دارای بیش‌برآورد بوده است.

نتایج حاصل از پیش‌بینی گروه‌های بزرگ خاک توسط مدل رگرسیون لجستیک با دو مجموعه ورودی نشان داد که مدل با استفاده از داده‌های ورودی مجموعه ۲ باعث افزایش صحت پیش‌بینی شده است. صحت کلی و آماره کاپا بالاتر و درجه برای پایین مجموعه ۲ این مسئله را تأیید می‌کند (جدول ۳). همچنین برای افزایش صحت گروه‌های بزرگ خاک که دارای دقت کمی هستند می‌توان با افزایش تعداد نقاط نمونه‌برداری،

متغیرهای ورودی برای گروه بزرگ هیلوکلسیدز (جدول ۳) مشاهده شد. این نتایج در تطابق با نتایج به دست آمده توسط جعفری و همکاران (۱۹)، تقی‌زاده و همکاران (۳۵) می‌باشد. در مناطق خشک به دلیل فرسایش و رسوب شدیدی که رخ می‌دهد، تمایز خاک بسیار دشوار و حتی مدل‌های قدرتمند دقت زیادی را ارائه نمی‌دهند. البته قضاوت در مورد ضعف و ناتوانی مدل‌ها منطقی نیست چرا که تعداد کم نمونه‌ها (۸۸ داده آموزشی و ۳۸ داده اعتبارسنجی) برای چنین منطقه وسیعی، تشخیص رابطه پیش‌بینی دقیق را برای طبقه‌بندی کلاس‌های خاک دچار مشکل می‌کند. علاوه بر این، اگر معیارهای تشخیص برای بعضی خاک‌ها ناشناخته باشد، شناسایی آنها به طور دقیق خیلی مشکل است. همچنین نتایج حاصل از جدول (۳) نشان داد که مدل با متغیرهای ورودی مجموعه ۲ باعث افزایش دقت گروه بزرگ هیلوکلسیدز از صحت کاربر و قابلیت اطمینان تولیدکننده ۰/۲۵ به ۰/۵۰ شد. جعفری و همکاران (۱۹) و پهلوان راد و همکاران (۲۹) نیز به چنین نتیجه‌ای دست یافتند که کلاس‌های خاکی که نمونه کمتری داشتند، دارای خطای پیش‌بینی بیشتری بودند.

عملکرد مدل لجستیک چند جمله‌ای با مجموعه ۱ برای گروه بزرگ هیلوکمبیدز ۰/۵۰ محاسبه شد، چرا که ممکن است این نتیجه به صورت تصادفی به دست آمده باشد. در جدول (۳)، شاخص‌های کیفیت حاصل از مدل برای داده‌های اعتبارسنجی مربوط به این گروه بزرگ خاک، نتایج بالا را تأیید میکند. هیچ خصوصیت مشخصه‌ای برای تشخیص و تفکیک هیلوکمبیدز وجود ندارد که بتواند مدل از آن به عنوان معیار تفکیک کننده استفاده کند. بنابراین احتمال دارد خلوص پایین پیش‌بینی، نتیجه ناقص بودن مدل مفهومی باشد. نتایج حاصل از مدل با متغیرهای ورودی مجموعه ۲ نیز همچنین باعث افزایش صحت پیش‌بینی این گروه بزرگ خاک در منطقه مورد مطالعه شد. برای گروه بزرگ هیلوجیسیدز مقدار صحت کاربر و قابلیت اطمینان تولیدکننده در هر دو مجموعه متغیرهای ورودی تقریباً مشابه اندازه‌گیری شد. اضافه شدن نقشه قدیمی منطقه

### نتیجه گیری

در این تحقیق از روش رگرسیون لجستیک چندجمله‌ای با دو مجموعه متغیرهای ورودی (مجموعه ۱ و ۲) برای پیش‌بینی گروه‌های بزرگ خاک در قسمتی از اراضی شهرستان بم استان کرمان استفاده شد. در بین متغیرهای ورودی، سطوح ژئومرفیک مهمترین و قدرتمندترین متغیر برای پیش‌بینی در هر دو مجموعه بود. نتایج ارزیابی صحت مدل رگرسیون لجستیک چندجمله‌ای با متغیرهای ورودی مختلف در داده‌های اعتبارسنجی نشان داد که، مجموعه ۲ با وارد شدن نقشه خاک قدیمی در مدل‌سازی شاخص‌های صحت مدل از جمله صحت کلی و آماره کاپا به ترتیب از ۰/۷۱ و ۰/۶۵ به ۰/۷۹ و ۰/۷۴ افزایش یافت و درجه برابری از ۰/۵۷ به ۰/۵۳ کاهش یافت. همچنین استفاده از ابزارهای مطالعاتی با ماهیت پیوسته مانند تصاویر ماهواره‌ای و یا مدل رقومی ارتفاع می‌تواند در شناسایی تغییرپذیری تدریجی پدیده‌های طبیعی مانند خاک، نقش مؤثری داشته باشند. بنابراین، استفاده از مدل رقومی ارتفاع و ویژگی‌های آن به همراه اطلاعات دیگر مانند تصاویر ماهواره‌ای، نقشه زمین‌شناسی و نقشه‌های خاک قدیمی به منظور ارتقای کیفی این نقشه‌ها و بهنگام نمودن آنها پیشنهاد می‌گردد. به‌طور کلی نتایج نشان داد که تکنیک‌های ماشین یادگیری از جمله رگرسیون لجستیک چندجمله‌ای می‌توانند فرآیند نقشه‌برداری خاک‌ها را در گستره‌ای وسیع و متشکل از هر گونه عوارض طبیعی به یکباره انجام داده، روش سنتی را ارتقاء بخشیده و قابلیت استفاده آنها را برای شاخه‌های علمی مختلف را امکان‌پذیر نماید.

صحت گروه‌های بزرگ خاک و در نتیجه صحت نقشه برازش‌شده را افزایش داد. استفاده از سایر متغیرهای محیطی که در این مطالعه استفاده نگردید ممکن است سبب افزایش صحت نقشه برازش‌شده گردد.

کمپن و همکاران بیان کردند که داده‌های خاک موجود در ترکیب با متغیرهای کمکی محیطی برای به‌هنگام کردن نقشه خاک مفید هستند (۲۱). کاتن و همکاران از تکنیک نقشه‌برداری خاک برای پیش‌بینی مدل‌های توسعه‌یافته در یک منطقه مرجع که نزدیک منطقه‌ای با روابط خاک-زمین‌نمای مشابه استفاده کردند (۸). نتایج آنها نشان داد که مدل قادر به تشخیص سه شکل اصلی از زمین‌نماهای منطقه بودند. کلاس‌های خاکی که به متغیرهای کمکی همبستگی بیشتری داشتند بهتر پیش‌بینی شدند. نقشه خاک تولیدشده، قابلیت ۴۶/۱۲٪ و دقت ۲۱/۰۶ را نشان داد. لمرسیر و همکاران در مقیاس منطقه‌ای، ویژگی‌های خاک را توسط مدل خاک-زمین‌نما با استفاده از روش ماشین یادگیری برون‌یابی کردند (۲۵). نتایج آنها نشان داد که این روش رویکرد مؤثر و قوی در پیش‌بینی ویژگی‌های خاک با استفاده از اطلاعات خاک موجود و متغیرهای محیطی کمکی دارد. همچنین بروگراد و همکاران برای برآورد کلاس‌های خاک از مدل‌سازی متعدد و مجموعه متغیرهای کمکی متفاوتی استفاده کردند (۷). نتایج آنها نشان داد که پیش‌بینی‌های مکانی دقیق‌ترین مدل‌ها معمولاً با روابط خاک-زمین‌نماهای مورد انتظار در توافق بودند. همچنین کمترین عدم قطعیت پیش‌بینی مکانی در مناطق نسبتاً پست در منطقه مورد مطالعه برآورد شد.

### منابع مورد استفاده

۱. بنایی، م. ح. ۱۳۸۰. نقشه منابع و استعدادهای خاک‌های ایران. مؤسسه تحقیقات آب و خاک ایران، تهران.
2. Aksoy, E., G. Ozsoy and M. Sabri Dirim. 2009. Soil mapping approach in GIS using Landsat satellite imagery and DEM data. *Afr. J. Agr. Res.* 4(11): 1295-1302.
3. Barthold, F. K., M. Wiesmeier, L. Breuer, H. G. Frede, J. Wu and F. B. Blank. 2013. Land use and climate control the spatial distribution of soil types in the grasslands of Inner Mongolia. *J. Arid Environ.* 88: 194-205.
4. Behrens, T. and T. Scholten. 2006. A comparison of data-mining techniques in predictive soil mapping. PP. 353-617. *In: Lagacherie, P. McBratney, A. B. and Voltz, M. (Eds.), Digital Soil Mapping: An Introductory Perspective.* Elsevier, Amsterdam.

5. Behrens, T., H. Förster, T. Scholten, U. Steinrücken, E. D. Spies and M. Goldschmitt. 2005. Digital soil mapping using artificial neural networks. *J. Plant Nut. Soil Sci.* 168: 21-33.
6. Boruvka, L., L. Pavlu, R. Vasat, V. Penizek and O. Drabek. 2008. Delineating acidified soils in the Jizera Mountains region using fuzzy classification. PP. 303–309. *In: Hartemink, A.E. McBratney, A. and Mendonça-Santos, M.L. (Eds.), Digital Soil Mapping With Limited Data.* Springer, Netherlands.
7. Brungard, C. W., J. L. Boettinger, M. C. Duniway, S. A. Wills, T. C. Edwards Jr. 2015. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma* 239-240: 68-83.
8. Caten A. T., S. Ricardo, D. Dalmolin, F. A. Pedron and M. L. Mendonça Santos. 2011. Extrapolation of soil-landscape relationship from a reference area. *Cienc. Rural.* 41: 812-816.
9. Debella-Gilo, M. and B. Etzelmuller. 2009. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vest fold County, Norway. *Catena.* 77: 8-18.
10. Esfandiarpour Borujeni, I., J. Mohammadi, M. H. Salehi, N. Toomanian and R. M. Poch. 2010. Assessing geopedological soil mapping approach by statistical and geostatistical methods: A case study in the Borujen region, Central Iran. *Catena.* 82: 1-14.
11. Gee, G. W. and J. W. Bauder. 1986. Particle-size analysis. PP. 372-382. *In: Klute, A. (Ed.), Methods of Soil Analysis. Part 1.* 2nd, Agronomy Monograph No. 9. ASA and SSSA, Madison, WI.
12. Grinand, C., D. Arrouays, B. Laroche and M.P. Martin. 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma.* 143: 180-190.
13. Grunwald, S. 2006. *Environmental Soil–Landscape Modeling: Geographic Information Technologies and Pedometrics.* CRC/Taylor & Francis, Boca Raton.
14. Hartemink, A. E., A. McBratney and M. d. L. Mendonca-Santos. 2008. *Digital soil mapping with limited data.* Springer.
15. Hastie, T., R. Tibshirani and J. H. Friedman. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, New York.
16. Hengl, T., G. Heuvelink and A. Stein. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma.* 120(1-2): 75-93.
17. Hengl, T., N. Toomanian, H. I. Reuter and M. J. Malakouti. 2007. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. *Geoderma.* 140: 417-427 .
18. Jafari, A., S. Ayoubi, H. Khademi, P. A. Finke and N. Toomanian. 2013. Selection of a taxonomic level for soil mapping using diversity and map purity indices: a case study from an Iranian arid region. *Geomorphology.* 201: 86-97.
19. Jafari, A., P. A. Finke, J. Van deWauw, S. Ayoubi and H. Khademi. 2012. Spatial prediction of USDA- great soil groups in the arid Zarand region, Iran: comparing logistic regression approaches to predict diagnostic horizons and soil types. *Eur. J. Soil Sci.* 63: 284-298.
20. Kempen, B., D. J. Brus and G. B. M. Heuvelink. 2012. Soil type mapping using the generalized linear geostatistical model: a case study in a Dutch cultivated peatland. *Geoderma.* 189: 540-553.
21. Kempen, B., D. J. Brus, G. B. M. Heuvelink and J. J. Stoorvogel. 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma* 151: 311-326.
22. Kim, J., S. Grunwald, R. G. Rivero and R. Robbins. 2012. Multi-scale modeling of soil series using remote sensing in a wetland ecosystem. *Soil Sci. Soc. Am. J.* 76: 2327-2341.
23. Kutner, M. H., C. J. Nachtsiem, J. Neter and W. Li W. 2004. *Applied linear statistical models,* 4th ed, New York.
24. Lacoste, M., B. Lemerrier and C. Walter. 2011. Regional mapping of soil parent material by machine learning based on point data. *Geomorphology* 133: 90-99.
25. Lemerrier, B., M. Lacoste, M. Loum and C. Walter. 2012. Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach. *Geoderma* 171-172: 75-84.
26. Marchetti, A., C. Piccini, S. Santucci, I. Chiuchiarelli and R. Francaviglia. 2011. Simulation of soil types in Teramo province (Central Italy) with terrain parameters and remote sensing data. *Catena* 85: 267-273.
27. Moonjun, R., A. Farshad, D. P. Shrestha and C. Vaiphasa. 2010. Artificial neural network and decision tree in predictive soil mapping of Hoi Num Rin sub-watershed, Thailand. PP: 151-164. *In: Boettinger, J. L. Howell, D. W. Moore, A. C. Hartemink, A. E. and Kienast-Brown, S. (Eds.), Digital Soil Mapping: Bridging Research,*

- Environmental Application, and Operation. Springer, Dordrecht.
28. Nelson, R. E. 1982. Carbonate and gypsum. PP: 181-196. *In*: Page, A.L. Miller, R.H. and Keeny, R. (Eds.), Methods of soil analysis, Part II-chemical and microbiological properties, Madison, W.I.
  29. Pahlavan Rad, M. R., N. Toomanian, F. Khormali, C. W. Brungard, C. B. Komaki and P. Bogaert. 2014. Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. *Geoderma* 232-234: 97-106.
  30. Rhoades, J. D. 1982. Soluble salts. PP. 933-951. *In*: Page, A.L. Methods of Soil Analysis, Part II 2nd, Agronomy Monograph.ASA, Madison, WI.
  31. Schoeneberger, P. J., D. A. Wysocki, E. C. Benham and W.D. Broderson (Eds.), 2003. Field Book for Describing and Sampling Soils, Version 2.0. Natural Resources Conservation Service. National Soil Survey Center, Lincoln.
  32. Scull, P., J. Franklin and O. A. Chadwick. 2005. The application of classification tree analysis to soil type prediction in a desert landscape. *Eco. Model.* 181: 1-15.
  33. Smith, C. A. S., B. Daneshfar and G. Frank. 2012. Use of weights of evidence statistics to define inference rules to disaggregate soil survey maps. PP: 215-220. *In*: Minasny, B. Malone, B.P. and McBratney, A. (Eds.), Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping. CRC Press, Sydney.
  34. Soil Survey Staff. 2010. Keys to Soil Taxonomy, 11th end. United States Department of Agriculture, Washington, 337p.
  35. Taghizadeh-Mehrjardi, R., B. Minasny, A. B. McBratney, J. Triantafilis, F. Sarmadian and N. Toomanian. 2012. Digital soil mapping of soil classes using decision trees in central Iran. *Digital Soil Assessments and Beyond - Proceedings of the 5th Global Workshop on Digital Soil Mapping*, 24, July, Sydney, Australia.
  36. Toomanian, T., A. Jalalian, H. Khademi, M. Karimian Eghbal and A. Papritz. 2006. Pedodiversity and pedogenesis in Zayandeh-rud Valley, Central Iran. *Geomorphology* 81: 376-393.
  37. Triantafilis, J., N. Y. Earl and I. D. Gibbs. 2012. Digital soil-classmapping across the Edgeroi district using numerical clustering and gamma-ray spectrometry data. PP: 187-191. *In*: Minasny, B. Malone, B.P. and McBratney, A. (Eds.), Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping. CRC Press, Sydney .
  38. Van Zijl, G. M., P. A. L., le Roux and A.B. Smith. 2012. Rapid soil mapping under restrictive conditions in Tete, Mozambique. PP: 335-339. *In*: Minasny, B. Malone, B.P. and McBratney, A. (Eds.), Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping. CRC Press, Sydney.
  39. Vaysse, K. and P. Lagacherie. 2015. Evaluating Digital Soil Mapping approaches for mapping Global Soil Map soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Regional* 4: 20-30.
  40. Walkly A. and I. A. Black. 1934. An examination of digestion method for determining soil organic matter and a proposed modification of the chromic acid titration. *Soil Sci.* 37: 29-38.
  41. Witten, I. H., E. Frank and M. A. Hall. 2011. Data mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Burlington.
  42. Yang, L., Y. Jiaob, S. Fahmy, A. X. Zhu, S. Hanne, J. E. Burtf and F. Qig. 2010. Updating Conventional Soil Maps through Digital Soil Mapping. *Soil Sci. Soc. Am. J.* 75(3): 1044-1053.
  43. Zhu, A. X., B. Hudson, J. Burt, K. Lubich and D. Simonson. 2001. Soil mapping using GIS, expert knowledge and fuzzy logic. *Soil Sci. Soc. Am. J.* 65: 1463-1472.

## Digital Soil Mapping of Soil Classes using Conventional Soil Maps in the Arid Region Southeastern Iran

F. Abbaszadeh Afshar<sup>1,2</sup>, S. Ayoubi<sup>1\*</sup> and A. Jafari<sup>3</sup>

(Received: Nov. 18-2015 ; Accepted: June 28-2016)

### Abstract

Mapping the spatial distribution of soil taxonomic classes is important for useful and effective use of soil and management decisions. Digital soil mapping (DSM) may have advantages over conventional soil mapping approaches as it may better capture observed spatial variability and reduce the need to aggregate soil types. A key component of any DSM activity is the method used to define the relationship between soil observations and environmental covariates. This study aims to compare multiple logistic regression models and covariate sets for predicting soil taxonomic classes in Bam district, Kerman province. The environmental covariates derived from digital elevation models, Landsat imagery, geomorphology map and soil unit map that were divided into two different sets: (1) variables derived from digital elevation models, remote sensing and geomorphology map, (2) variables derived from digital elevation model, remote sensing, geomorphology map and the soil map. Stratified sampling schemes were defined in 100000 hectares, and 126 soil profiles were excavated and described. The results of accuracy model showed that data set 2 increased accuracy of model including overall accuracy, kappa index, user accuracy and reliability of the producer. The results showed that the multiple logistic regression model can promote traditional soil mapping and it can be used to large group of other scientific fields.

**Keywords:** Auxiliary variables, Machine learning, Multiple logistic regression, Soil great groups

---

1. Dept. of Soil Sci., Faculty of Agric., Isf. Univ. of technol., Isfahan, Iran.

2. Dept. of Soil Sci., Faculty of Agric., Univ. of Jiroft, Jiroft, Iran.

3. Dept. of Soil Sci., Faculty of Agric., Shahid Bahonar Univ. of Kerman, Kerman, Iran.

\*: Corresponding Author, Email: ayoubi@cc.iut.ac.ir