

کاربرد مدل‌های رگرسیونی در پیش‌بینی کلاس خاک در بخشی از مناطق ایران مرکزی (منطقه زرنند کرمان)

اعظم جعفری^۱ - شمس اله ایوبی^{۲*} - حسین خادمی^۳

تاریخ دریافت: ۸۹/۱۱/۳

تاریخ پذیرش: ۹۰/۵/۳۰

چکیده

شناسایی رقومی خاک‌ها به‌عنوان ابزاری برای ایجاد اطلاعات مکانی خاک، راه‌حلی برای نیاز رو به افزایش نقشه‌های خاک با تفکیک مکانی بالا را تأمین می‌کند. بنابراین، باید روش‌های جدید به‌منظور به‌دست آوردن اطلاعات مکانی خاک با تفکیک مکانی بالا توسعه پیدا کند. به همین منظور مطالعه‌ای جهت پیش‌بینی کلاس‌های خاک با استفاده از مدل‌های رگرسیونی در منطقه زرنند کرمان طراحی گردید. در این مطالعه، مدل‌های رگرسیونی شامل رگرسیون لاجیستیک چندجمله‌ای و رگرسیون درختی توسعه یافته چندکلاسه برای پیش‌بینی گروه بزرگ خاک به کمک داده‌های سنجش از دور، پارامترهای سرزمین و نقشه ژئومرفولوژی استفاده گردید. کیفیت پیش‌بینی مدل‌ها با شاخص‌های حاصل از آرایه خطا بررسی گردید. نتایج نشان داد در پیش‌بینی همه گروه‌های بزرگ خاک، سطوح ژئومرفیک به‌عنوان یک پیش‌بینی کننده مؤثر محسوب می‌شود. بعد از سطوح ژئومرفیک، پارامترهای سرزمین و شاخص‌های سنجش از دور در پیش‌بینی وارد شدند. در هر دو مدل خلوص نقشه برای همه گروه‌های بزرگ خاک در موقعیت‌های اعتبارسنجی و واسنجی بیشتر از ۰/۶ بود. نتایج نشان داد عملکرد پیش‌بینی برای گروه‌های بزرگ هاپلوچیسید و هاپلوسالید بهتر از گروه‌های بزرگ کلسی جیسید و هاپلوکمبید بود. در بین گروه‌های بزرگ خاک، مقادیر بالای دقت کاربر و قابلیت اطمینان تولیدکننده برای گروه بزرگ هاپلوسالید به‌دست آمد. خاک‌های با قابلیت اطمینان بهتر خاک‌هایی هستند که به شدت تحت تأثیر مشخصات توپوگرافی و ژئومرفولوژی قرار گرفتند (گروه‌های بزرگ هاپلوسالید، تری‌سامنت و هاپلوچیسید) و خاک‌های با قابلیت اطمینان و دقت پیش‌بینی کمتر خاک‌هایی هستند که به سختی تحت تأثیر مشخصات توپوگرافی و ژئومرفولوژی (گروه‌های بزرگ هاپلوکمبید و کلسی جیسید) قرار گرفتند.

واژه‌های کلیدی: نقشه‌برداری رقومی خاک، رگرسیون لاجیستیک چندجمله‌ای، رگرسیون درختی توسعه یافته

مقدمه

اطلاعات مکانی خاک با تفکیک مکانی بالا توسعه پیدا کند (۲۴).

نقشه‌برداری رقومی خاک یک مدل آماری یا عددی را بین متغیرهای محیطی و خصوصیات خاک توسعه می‌دهد. برای مدل-سازی رقومی خاک؛ داده‌های سنجش از دور، ویژگی‌های پستی و بلندی، داده‌های صحرایی، پوشش گیاهی، اقلیم و زمین‌شناسی همراه با دانش کارشناس منجر به استنباط مشخصه خاک در مکان‌های بازدید نشده می‌شود. چنین روش‌هایی اطلاعات مناسب‌تری از خاک و زمین‌نما برای کاربردهای مدل‌سازی تأمین می‌کنند (۲۴). در سال‌های اخیر، به دلیل افزایش منابع داده‌های کمکی یا محیطی، نقشه‌برداری رقومی خاک پیشرفت سریعی نموده است (۸). یکی از تکنیک‌های نقشه‌برداری، رگرسیون لاجیستیک^۴ است که با ایجاد مدل لاگیت^۵

نیاز به اطلاعات خاک در سال‌های اخیر رو به افزایش است. نقشه‌های خاک در مطالعات زیادی از جمله ارزیابی اراضی، برنامه-ریزی استفاده از اراضی و هیدرولوژی کاربرد دارد. اطلاعات رقومی خاک‌ها براساس ابزارهای پردازش جدید و داده‌های رقومی مختلف به‌طور مداوم در حال افزایش است. شناسایی رقومی خاک به‌عنوان ابزاری برای ایجاد اطلاعات مکانی خاک، راه‌حلی برای نیاز رو به افزایش نقشه‌های خاک با تفکیک مکانی بالا را تأمین می‌کند. بنابراین، باید استراتژی‌ها و روش‌های جدید به‌منظور به‌دست آوردن

۱، ۲ و ۳- به ترتیب دانشجوی دکتری، دانشیار و استاد گروه خاکشناسی، دانشکده

کشاورزی، دانشگاه صنعتی اصفهان

(Email: ayoubi@cc.iut.ac.ir

*) نویسنده مسئول:

4- Logistic Regression

5- Logit

رگرسیون لاجیستیک چندجمله‌ای و رگرسیون درختی توسعه یافته تهیه شود. بنابراین، این مطالعه به منظور پاسخگویی به سئوالات زیر طراحی گردید.

- ۱- آیا می‌توان با استفاده از پایگاه داده محدود به نقشه خاک با کیفیت بالا دست پیدا کرد؟
- ۲- کدامیک از فاکتورهای محیطی قدرت پیش‌بینی‌کنندگی بالایی دارند؟
- ۳- کدام روش یا مدل در پیش‌بینی کلاس خاک توانایی بالایی دارد؟

مواد و روش‌ها

تشریح منطقه مطالعاتی

منطقه مورد مطالعه در جنوب شرق ایران، تقریباً در فاصله ۷۰ کیلومتری شهر کرمان بین عرض جغرافیایی $25^{\circ} 30'$ تا 31° درجه شمالی و طول جغرافیایی $50^{\circ} 10'$ تا $56^{\circ} 50'$ درجه شرقی قرار گرفته است. این منطقه مساحتی حدود ۹۰۰۰۰ هکتار را می‌پوشاند و از اطراف توسط رشته کوه‌هایی (از جنس آهک، دولومیت و شیل) که روند شمال‌غربی و جنوب‌شرقی دارند، احاطه شده است. لندفرم‌های اصلی این منطقه شامل مخروطه‌افکنه‌های قدیمی و جدید، سطوح رسی (دق) و تپه‌های شنی می‌باشد (شکل ۱). پسته، مهم‌ترین محصول این منطقه است و باغات پسته روی سطوح رسی یا دق‌ها می‌باشد. متوسط سالانه بارندگی، درجه حرارت و تبخیر و تعرق به ترتیب ۶۱ میلی‌متر، ۱۷ درجه سانتی‌گراد و ۱۷۵۰ میلی‌متر می‌باشد.

طرح نمونه‌برداری و تشریح پروفیل

یک طرح نمونه‌برداری طبقه‌بندی شده^۳ با استفاده از نقشه‌های رقومی زمین‌شناسی، ژئومرفولوژی و توپوگرافی پیاده گردید. در این مطالعه، لایه نمونه‌برداری^۴ منطبق بر تفاوت در لندفرم‌ها (نقشه ژئومرفولوژی)، توپوگرافی (نقشه ارتفاع رقومی) و زمین‌شناسی تعریف گردید، به طوری که در هر طبقه، موقعیت‌های تصادفی نمونه‌برداری انتخاب گردید که اندازه نمونه‌ها متناسب با مساحت طبقه بود. در نهایت، ۱۲۳ خاکرخ حفر، تشریح، نمونه‌برداری، آنالیز و براساس معیارهای تاکسونومی آمریکائی (۲۶) طبقه‌بندی گردید. تشریح پروفیل در منطقه مورد مطالعه هفت گروه بزرگ خاک را نشان داد که شامل هاپلوسالیدز، هاپلوچیپسیدز، هاپلوکلسیدز، کلسی‌چیپسیدز، هاپلوکمبیدز، تری‌سامنتز و تری‌ارتنتز^۵ بود.

احتمال حضور کلاس خاک معینی را پیش‌بینی می‌کند. این مدل، احتمال حضور یک کلاس خاک را به متغیرهای پیش‌بینی‌کننده ارتباط می‌دهد. روش‌های رگرسیون لاجیستیک به طور موفقیت‌آمیزی در علوم خاک و زمینه‌های مربوطه از جمله پیش‌بینی خطرات زمین‌لغزه (۱۷) احتمال حضور کلاس‌های زهکشی خاک (۶) یا ارتباط انواع خاک به ویژگی‌های پستی و بلندی استفاده شده است. بایلی و همکاران (۲) روابط بین سری‌های خاک یک منطقه و تعدادی از متغیرهای همان منطقه (از جمله شیب، جهت شیب و ارتفاع نسبی) را با استفاده از رگرسیون لاجیستیک چندجمله‌ای^۱ به دست آوردند. دابلاگیلو و همکاران (۷) به کمک متغیرهای کمکی (پارامترهای سرزمین) برای تخمین توزیع مکانی کلاس‌های خاک در ایالت وست فولد، جنوب شرق نروژ، از لاجیستیک رگرسیون چندجمله‌ای استفاده کردند. نتایج نشان داد از بین پارامترهای سرزمین، ارتفاع، شیب، جهت و مدت زمان تابش خورشیدی در توزیع مکانی کلاس‌های خاک نقش تعیین‌کننده‌ای دارند. هنگل و همکاران (۱۴) موفقیت رگرسیون لاجیستیک چندجمله‌ای در پیش‌بینی گروه‌های خاک WRB را گزارش دادند و نتایج آنها نشان داد موفقیت مدل شدیداً به همبستگی کلاس‌های خاک با پیش‌بینی‌کننده‌ها بستگی دارد.

یکی دیگر از تکنیک‌های رگرسیون مورد استفاده در نقشه‌برداری خاک، طبقه‌بندی و رگرسیون درختی است (۱۵). پیشرفت‌های آماری اخیر در این مدل‌ها از جمله فرآیند تصادفی، منجر به ایجاد مدل قوی‌تری به نام رگرسیون درختی توسعه یافته^۲ شده است. هدف این مدل‌ها بهبود عملکرد یک مدل تک درخت از طریق برازش مدل‌های متعدد و ترکیب آنها برای پیش‌بینی است. توابع رگرسیون درختی توسعه یافته ممکن است خطی، منحنی یا غیرخطی باشند. شواهد تجربی نشان داده رگرسیون درختی توسعه یافته یکی از بهترین روش‌های مدل‌سازی می‌باشد. باور و کوه‌اوی (۳) و فریدمن و همکاران (۱۰) رگرسیون درختی توسعه یافته و رگرسیون درختی معمولی را روی چهارده پایگاه داده مقایسه کردند و نتیجه گرفتند رگرسیون درختی توسعه یافته بهترین است. گرینند و همکاران (۱۳) توانایی رگرسیون درختی توسعه یافته را برای پیش‌بینی دقیق لنداسکیپ در نواحی بازدید نشده را بررسی کردند. آنها دریافتند توانایی تخمین مدل در موارد برون‌یابی پایین است اما برای درون‌یابی توانایی بالایی دارد. کاربرد رگرسیون درختی توسعه یافته در علوم اکولوژی گزارش شده (۹) اما تاکنون مطالعه‌ای در زمینه پیش‌بینی خصوصیات خاک انجام نشده است.

در این مطالعه سعی شده است یک نقشه خاک قابل قبولی از گروه‌های بزرگ خاک براساس اطلاعات محدود به کمک مدل‌های

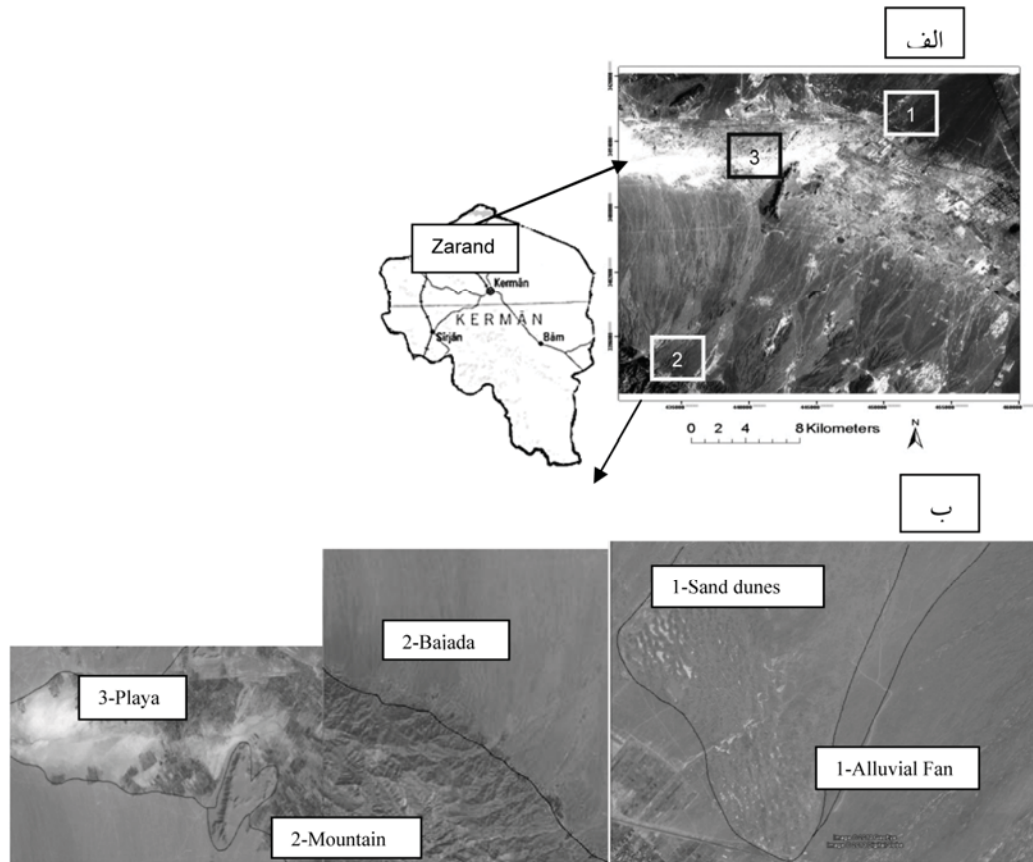
3- Stratified Sampling Scheme

4- Sampling strata

5- Haplosalids, Haplogypsids, Haplocalcids, Calcigypsids, Haplocambids, Torripsamments and Torriorthents

1- Multinomial logistic regression

2- Boosted Regression Tree (BRT)



شکل ۱- منطقه مطالعاتی نزدیک شهر کرمان (الف: تصویر ماهواره‌ای لندست ETM (243 RGB)، ب: تصویر تهیه شده توسط Google Earth)

متغیرهای کمکی

یک سین تصویر ماهواره‌ای ETM+ برای تعیین شاخص‌های گیاهی و شاخص رس و یک مدل رقمی ارتفاع با تفکیک مکانی ۳۰ متر که از وب سایت مدل رقمی ارتفاع جهانی استر (Aster Global Elevation Digital Model) به دست آمد (۲۷) برای تعیین ویژگی‌های پستی و بلندی استفاده گردید.

ویژگی‌های پستی و بلندی حاصل از مدل رقمی ارتفاع و شاخص‌های سنجش از دور در جدول ۱ ارائه گردیده است. مرحله پیش پردازش این پارامترها و شاخص‌ها در نرم افزار SAGA انجام گردید (۱۸). نقشه‌های ویژگی‌های پستی و بلندی، شاخص‌های سنجش از دور و ژئومرفولوژی به سیستم مرجع جهانی یکسانی (WGS 84 UTM 40N) تبدیل شدند. سپس همه مقادیر ویژگی‌های پستی و بلندی، شاخص‌های سنجش از دور و سطوح مختلف ژئومرفولوژی معادل با نقاط نمونه‌برداری، افق‌های مشخصه و گروه‌های بزرگ خاک در یک جدول گردآوری شدند و در نهایت این جدول به همراه نقشه‌های رستری متغیرهای کمکی به نرم افزار R منتقل گردید.

پیش‌بینی کلاس‌های خاک با روش رگرسیون لاجیستیک

چندجمله‌ای

مدل لاجیستیک مورد خاصی از مدل خطی تعمیم یافته^۱ است که به دو صورت قابل اجرا است: مدل لاجیستیک دوتایی و مدل لاجیستیک چندجمله‌ای. در مدل لاجیستیک دوتایی، متغیر وابسته به شکل حضور یا عدم حضور (صفر و یک) است، مثلاً حضور یا عدم حضور افق مشخصه. در حالیکه در مدل لاجیستیک چندجمله‌ای، متغیر وابسته دارای چند کلاس یا طبقه است مثلاً کلاس‌های خاک در یک منطقه. بنابراین در این مطالعه، در اجرای مدل لاجیستیک چند جمله‌ای گروه بزرگ خاک به عنوان متغیر هدف می‌باشد که یک متغیر طبقه‌ای است. بنابراین، متغیر وابسته (گروه بزرگ خاک) دارای بیش از یک سطح است و حل آن به صورت دوتایی^۲ امکان‌پذیر نیست، در این صورت باید یک سطح متغیر وابسته به عنوان طبقه مرجع^۳ انتخاب شود. یک مدل لاجیستیک چندجمله‌ای با طبقه مرجع به شکل زیر نمایش داده می‌شود:

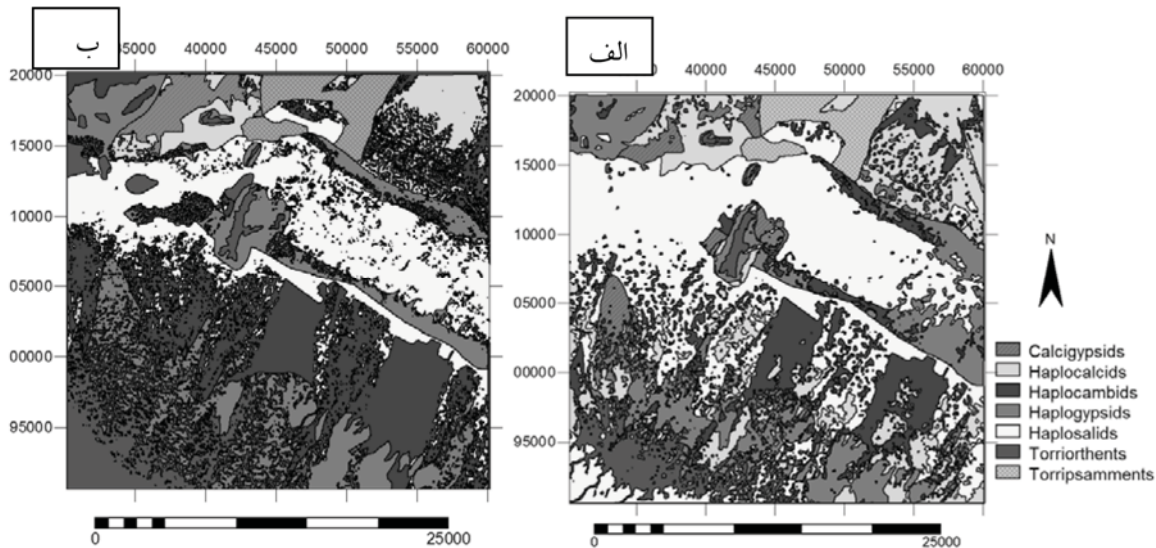
1- Generalized linear model

2- Binary

3- Reference category

جدول ۱- داده‌های محیطی مورد استفاده به‌عنوان متغیرهای پیش‌بینی‌کننده در مدل‌سازی

منبع	نوع متغیر	تشریح	نام	ماهیت فاکتور	مقیاس	داده‌های محیطی
	کمی	Elevation (m)	El	توپوگرافی	۳۰ متر	مدل رقومی ارتفاع (DEM)
		Slope angle (%)	Sl			
		Mean Curvature	MCur			
		Profile Curvature	PrCur			
		Plan Curvature	PlCur			
		Wetness Index	WI			
		Topographic Wetness Index	TWI			
(۱۱)		Multi-resolution Valley Bottom Flatness index	MrVBF			
(۲۳)	کمی	Normalized Difference Vegetation Index	NDVI	انعکاس نور		تصویر لندست (ETM)
(۱۹)		Ratio Vegetation Index	RVI			
(۲۰)		Perpendicular Vegetation Index	PVI			
(۸)		Clay Index	CI			
(۱۶)	کیفی	Hierarchical Four level classification (18 geomorphic surfaces(GS))	GS	ژئومرفولوژی	۱:۱۰۰۰۰۰	نقشه ژئومرفولوژی
	کیفی	Lithological units (9 classes)	Geol	زمین‌شناسی	۱:۱۰۰۰۰۰	نقشه ژئولوژی



شکل ۲- توزیع مکانی گروه‌های بزرگ خاک حاصل از روش‌های مورد استفاده (الف: رگرسیون لاجیستیک چندجمله‌ای و ب: رگرسیون درختی توسعه یافته چندکلاسه)

این مدل مشابه مدل لاجیستیک دو جمله‌ای^۲ است به استثناء اینکه توزیع احتمال متغیر هدف به جای دو جمله‌ای، چندجمله‌ای است و به جای یک معادله، J-1 معادله داریم.

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \alpha_j + \beta_j x_i, j = 1, 2, \dots, J-1 \quad (۱)$$

که α_j یک ثابت، β_j یک برداری^۱ از ضرایب رگرسیون برای $j = 1, 2, \dots, J-1$ و x_i برداری از متغیرهای توصیفی است.

در برازش BRT، دو پارامتر باید مشخص شود: نرخ یادگیری^۶ و پیچیدگی درخت^۷. نرخ یادگیری، سهم هر درخت متوالی را در مدل نهایی تعیین می‌کند. پیچیدگی درخت اثرات اصلی (tree complexity=1) یا اثرات متقابل بین متغیرها (tree complexity=2,3,...) را نشان می‌دهد. در نهایت، نرخ یادگیری و پیچیدگی درخت با هم تعداد کل درختان در مدل نهایی را تعیین می‌کنند (۵ و ۷).

مدل‌های BRT با استفاده از کد BRT تأمین شده توسط الیت و همکاران (۹) و همچنین بسته "gbm" (۲۱) در نرم افزار R برازش داده شدند. برای برازش مدل BRT، پارامترهای قراردادی کد BRT استفاده شد که نرخ یادگیری ۰/۰۱، پیچیدگی درخت ۱ و اعتبارسنجی متقابل^۸ ۱۰ مرتبه بود. اگرچه با سعی و خطا، به دلیل انحراف کمتر، مقدار پارامتر کسر کل (کسری از کل داده‌های موجود) از ۰/۷۵ به ۰/۵ تغییر پیدا کرد. از آنجایی که متغیر وابسته یا متغیر مورد پیش‌بینی، گروه بزرگ خاک است و یک متغیر طبقه‌ای است، روش BRT چندکلاسه^۹ انتخاب گردید.

ارزیابی دقت مدل‌ها و مقایسه آنها

برای مقایسه دقت پیش‌بینی 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 برازش داده شد و سپس پیش‌بینی برای داده‌های اعتبارسنجی انجام گردید. اعتبارسنجی با مقایسه داده‌های مشاهده شده و پیش‌بینی شده صورت گرفت. یک آرایه خطا^{۱۱} (۱۲) برای محاسبه معیارهای دقت استفاده گردید. خلوص نقشه (MP) ارتباط بین همه داده‌های مورد استفاده و داده‌های صحیح طبقه‌بندی شده (ff, tt) را تشریح می‌کند:

$$MapPurity = \frac{tt + ff}{tt + tf + ft + ff} \quad (۴)$$

دقت استفاده کننده (UA) شامل کلاس‌های حضور صحیح پیش‌بینی شده (tt) به مجموع کلاس‌های حضور صحیح پیش‌بینی شده (tt) و کلاس‌های عدم حضور که به غلط جزء کلاس‌های حضور مشاهده شده (واقعی)، پیش‌بینی شدند (tf) می‌باشد:

$$User's Accuracy = \frac{tt}{tt + tf} \quad (۵)$$

قابلیت اطمینان تولید کننده (PR) ارتباط بین همه کلاس‌های

$$\Pr(y_i = j) = \pi_{ij} = \frac{\exp(\beta_j x_i)}{1 + \sum_{j=1}^j \exp(\beta_j x_i)} \quad (۲)$$

بنابراین، احتمال طبقه مرجع:

$$\Pr(y_i = 0) = \pi_{i0} = \frac{1}{1 + \sum_{j=1}^j \exp(\beta_j x_i)} \quad (۳)$$

این روش در R با استفاده از بسته nnet قابل اجرا است. قبل از اجرای مدل، کلاس مرجع باید انتخاب شود. در این مطالعه، کلسی-جیپسیدز براساس پیش فرض مدل، به‌عنوان کلاس مرجع انتخاب شد. اهمیت ضریب رگرسیون هر متغیر پیش‌بینی کننده برای هر متغیر وابسته با استفاده از معیار آماری والد^۱ ارزیابی گردید (۲۲). این معیار در واقع از تقسیم ضریب رگرسیون هر متغیر مستقل به خطای استاندارد آن به دست می‌آید (۸). 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. در انتخاب مدل اقتصادی هر دوی معیار Wald و AIC^۲ (۱) استفاده گردید.

پیش‌بینی کلاس‌های خاک با روش رگرسیون درختی توسعه یافته (BRT)

روش رگرسیون درختی توسعه یافته، ترکیبی از دو تکنیک آماری قوی بوستینگ و رگرسیون درختی است. بوستینگ یک روش مرحله-ای رو به جلو است، که مدل‌های درختی به‌طور تکرارپذیری با زیرمجموعه‌ای از داده‌های آموزشی برازش داده می‌شود. زیرمجموعه‌های داده‌های آموزشی مورد استفاده در هر تکرار از برازش مدل به‌طور تصادفی بدون جایگزینی انتخاب می‌شوند. سهم داده‌های آموزشی مورد استفاده توسط مدل‌ساز با پارامتری به نام کسر کل^۳ (کسری از کل داده‌های موجود) تعیین می‌شود. این فرآیند در واقع فرآیند تصادفی بودن^۴ را به مدل اعمال می‌کند که دقت مدل را افزایش و برازش بیش از حد^۵ را کاهش می‌دهد (۹).

در ابتدا، ۵۰ درخت به‌طور معمولی برازش داده می‌شود. سپس باقی‌مانده‌های حاصل از برازش اولیه با مجموعه دیگری از ۵۰ درخت برازش داده می‌شود، باقی‌مانده‌های حاصله دوباره با ۵۰ درخت دیگر برازش داده می‌شود و این کار ادامه می‌یابد تا این که تابع خطا حداقل شود. در مورد رگرسیون درختی، تابع خطا در واقع انحراف مدل است.

6- Learning rate
7- Tree Complexity
8- Cross Validation
9- Multiclass-BRT
10- Error matrices

1- Wald Statistic
2- Akaike Information Criteria
3- Bag Fraction
4- Stochasticity
5- Overfitting

های پستی و بلندی را روی توزیع گروه‌های خاک توسط روش لاجیستیک چندجمله‌ای نشان دادند. در حقیقت، حضور ویژگی‌های پستی و بلندی به همراه فرآیندهای ژئومرفیک در پیش‌بینی کلاس‌های خاک نشان داد که این فاکتورها برآیند و نتیجه نیروهای مؤثر خاک‌سازی در منطقه مورد مطالعه هستند. نقشه گروه بزرگ خاک حاصل از روش رگرسیون لاجیستیک چندجمله‌ای در شکل ۲ نشان داده شده است.

مدل‌سازی ارتباط خاک-ژئومرفولوژی-سرزمین توسط مدل رگرسیون درختی توسعه یافته

مدل BRT چندکلاس مستقیماً گروه‌های بزرگ خاک را از متغیرهای پیش‌بینی کننده تخمین می‌زند. بهترین مدل برای پیش‌بینی مشابه روش رگرسیون لاجیستیک براساس بیشترین مقدار AUC و کمترین انحراف انتخاب گردید. نقشه ژئومرفولوژی به‌عنوان مهم‌ترین تخمین‌گر برای همه گروه‌های بزرگ خاک تشخیص داده شد، که مجدداً نقش مؤثر فرآیندهای ژئومرفولوژی در توسعه خاک را تأیید می‌کند. این نتیجه همچنین در مورد رگرسیون لاجیستیک چندجمله‌ای نشان داده شد. بنابراین، براساس نتایج به‌دست آمده از مدل‌سازی، نقش انکارناپذیر فرآیندهای ژئومرفولوژی در این مطالعه تأیید می‌شود.

به دلیل فراوانی بالای گروه بزرگ هاپلوسالید و همچنین حضور پوشش گیاهی در پلایا، شاخص‌های سنجش از دور همچون PVI و NDVI و همچنین ویژگی‌های پستی و بلندی مانند MrVBF برای برازش این گروه بزرگ در مدل‌ها وارد شدند. برای دیگر گروه‌های بزرگ خاک عمدتاً ویژگی‌های پستی و بلندی عوامل توجیه‌کننده بودند. نقشه گروه‌های بزرگ خاک حاصل از روش‌های BRT چندکلاس در شکل ۲ نشان داده شده است.

در پیش‌بینی گروه‌های بزرگ خاک توسط این روش‌ها تفاوت‌هایی وجود دارد، به‌ویژه گروه بزرگ تری‌ارتنت تخمین زده شده توسط مدل لاجیستیک چندجمله‌ای به گروه بزرگ هاپلوسالید طبقه‌بندی شدند در حالی که در مدل BRT پیش‌بینی بهتری انجام شده است.

همچنین، باید به تعداد کم نمونه‌های تری‌ارتنت توجه کرد، هرچند که مدل BRT پیش‌بینی بهتری برای این گروه بزرگ به‌ویژه در قسمت شمال‌غربی منطقه مورد مطالعه نشان داد (شکل ۲). همان‌طوری که مورد انتظار بود مدل BRT گروه بزرگ تری‌سامنت را در تپه‌های شنی، هاپلوچیسید را در تپه‌های گچی و هاپلوکلسید را در مخروطه‌افکنه‌ها به خوبی پیش‌بینی کرد (شکل ۲). این نتیجه توسط مدل لاجیستیک نیز به‌دست آمد. بنابراین، مدل BRT پتانسیل بالایی برای پیش‌بینی دارد، به‌ویژه اگر پایگاه داده بزرگی موجود باشد.

صحیح پیش‌بینی شده (tt) و مجموع کلاس‌های صحیح پیش‌بینی شده و کلاس‌های حضور مشاهده شده (واقعی) که به غلط جزء کلاس‌های عدم حضور پیش‌بینی شدند (ft) می‌باشد:

$$\text{Producer's Reliability} = \frac{tt}{tt + ft} \quad (۶)$$

یک نقشه خوب دارای مقادیر خلوص نقشه، دقت استفاده‌کننده و قابلیت اطمینان تولید‌کننده نزدیک به یک است (۵).

نتایج و بحث

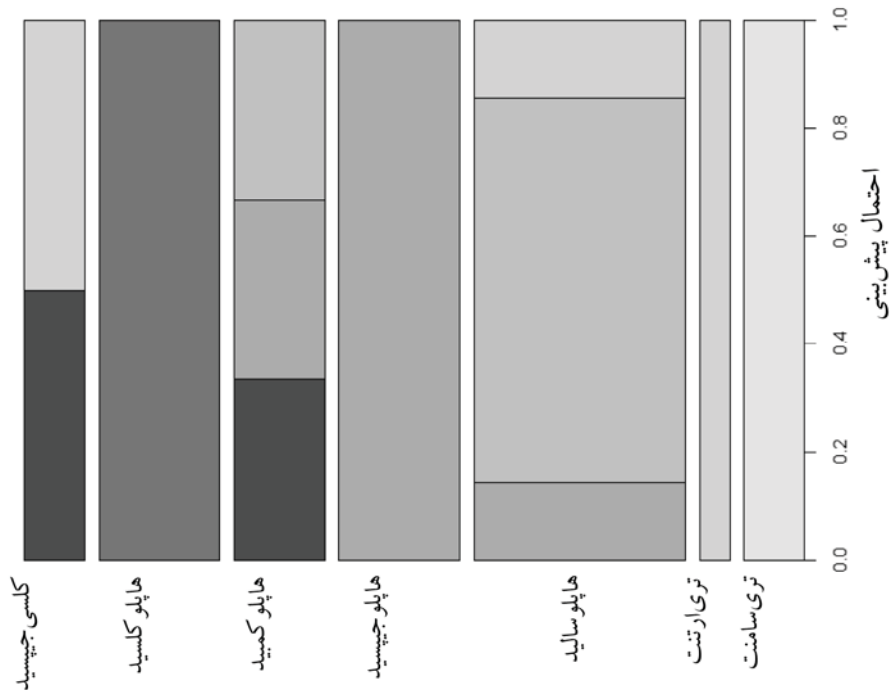
مدل‌سازی ارتباط خاک-ژئومرفولوژی-سرزمین توسط رگرسیون لاجیستیک

نتایج مدل‌سازی توسط روش رگرسیون لاجیستیک و متغیرهای وارد شده در مدل‌سازی جهت پیش‌بینی کلاس‌های خاک در جدول ۲ ارائه شده است. همان‌طوری که نتایج این جدول نشان می‌دهد سطح ژئومرفیک یک پیش‌بینی کننده قدرتمند محسوب می‌شود، در حالی که سطوح دیگر ژئومرفولوژی (زمین‌نما، شکل زمین و سنگ‌شناسی) در پیش‌بینی تأثیری نداشتند. این نتیجه را می‌توان چنین تشریح 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 آخرین و نزدیک‌ترین سطح به شرایط خاک در مقیاس بزرگ می‌باشد. بعد از سطوح ژئومرفیک، ویژگی‌های پستی و بلندی جزء مؤثرترین پارامترها برای پیش‌بینی بودند (جدول ۲) که تأیید کننده این مطلب است که در این ناحیه توپوگرافی یا پستی و بلندی، مهم‌ترین فاکتور خاک‌سازی می‌باشد.

هیچ کدام از شاخص‌های سنجش از دور در پیش‌بینی گروه بزرگ خاک توسط رگرسیون لاجیستیک سهمی نداشتند. بعد از سطح ژئومرفیک، شاخص خیزی و شاخص همواری دره با درجه تفکیک بالا (MrVBF) مهم‌ترین پیش‌بینی کننده بودند (جدول ۳). شاخص همواری دره با درجه تفکیک بالا (MrVBF) در تشخیص سطوح دره‌ای شکل کمک می‌کند و شاخص خیزی درجه رطوبت را نشان می‌دهد (۲۸). این پارامترها فرآیند انتقال رسوب به‌ویژه جریان رسوبات و مواد در حال حرکت را نشان می‌دهند (۲۸). بنابراین، این شاخص‌ها به‌طور غیرمستقیم نیروی رانشی در تشخیص بعضی از گروه‌های بزرگ خاک مورد مطالعه به‌ویژه هاپلوسالید، هاپلوکلسید و کلسی‌چیسید هستند. نقش مؤثر فرآیندهای ژئومرفولوژی توسط اسکول و همکاران (۲۵) در پیش‌بینی کلاس‌های خاک به اثبات رسیده است. همچنین، دبالگیلو و اتزلمولر (۷) تأثیر معنی‌دار ویژگی-

جدول ۲- متغیرهای مورد استفاده در پیش‌بینی گروه بزرگ خاک

متغیرها بر اساس اهمیت	متغیرهای مورد استفاده در مدل	مدل‌های مورد استفاده در پیش‌بینی
GS***, MrVBF***, WI**, EI*	GS+ MrVBF+ EI+ WI	مدل رگرسیون لاجیستیک چندجمله‌ای
GS***, MrVBF***, WI***, PVI**, PrCur**	GS+ MrVBF+ PVI+ WI+ PrCur	مدل رگرسیون درختی توسعه یافته



شکل ۳- احتمال پیش‌بینی گروه‌های بزرگ خاک حاصل از مدل لاجیستیک چندجمله‌ای برای داده‌های اعتبارسنجی

پیش‌بینی نسبتاً خوب با هر دو روش است در مقایسه با مقدار ۰/۷ که برای نقشه‌برداری سنتی خاک، مقدار هدف است (۲۸). همچنین، جدول ۳ نشان می‌دهد که دقت پیش‌بینی برای همه انواع خاک‌ها یکسان نیست. عملکرد ضعیف‌تر برای گروه‌های بزرگ کلسی جیپسید و هاپلوکسید در هر دو روش مشاهده شد. قابلیت اطمینان تولیدکننده برای داده‌های اعتبارسنجی حاصل از مدل لاجیستیک چندجمله‌ای (جدول ۳) را به صورت یک آرایه خطا در جدول ۴ و شکل ۳ نشان داده شده است که می‌توان عملکرد ضعیف را برای گروه‌های بزرگ کلسی جیپسید و هاپلوکسید مشاهده کرد. به عنوان مثال از دو نمونه خاک کلسی جیپسید، یکی به درستی پیش‌بینی شده اما نمونه دیگر به عنوان گروه بزرگ تری‌ارتنت پیش‌بینی شده است (جدول ۴).

همچنین در شکل ۳ ستون مربوط به گروه بزرگ کلسی جیپسید به دو قسمت تقسیم شده که یک قسمت آن همان نمونه صحیح پیش‌بینی شده و دیگری رنگ مربوط به ستون گروه بزرگ تری‌ارتنت را نشان می‌دهد که مطابق جدول ۴ صحیح می‌باشد.

مقایسه مدل‌های پیش‌بینی کننده

خلوص نقشه، دقت استفاده‌کننده و قابلیت اعتبار تولیدکننده برای گروه‌های بزرگ خاک برای مدل‌های مورد بررسی (مدل لاجیستیک و مدل درختی توسعه یافته) در جدول ۳ ارائه شده است. در داده‌های واسنجی، دقت و قابلیت اعتبار مدل‌ها تفاوت معنی‌داری نداشت هرچند که دقت برای گروه بزرگ کلسی جیپسید حاصل از مدل BRT چندکلاسه به طور قابل ملاحظه‌ای بالاتر بود. هدف واسنجی، اطمینان از دقت مدل در شناسایی ورودی‌ها و ارتباط شان است، بنابراین در صورت تشخیص روابط، واسنجی نتایج خوبی را در بر خواهد داشت. به نظر می‌رسد شناسایی خوب روابط در اصل به اندازه و توزیع نمونه در لایه‌ها بستگی دارد. در مکان‌های مختلف این منطقه وسیع، فقط ۵ نمونه کلسی جیپسید وجود داشت، بنابراین دقت پایین مورد انتظار است. جدول ۳ به طور واضح نشان می‌دهد که BRT به دلیل ساختار انعطاف‌پذیر و تکرار اپتیم‌سازی آن، یک تکنیک قدرتمند است.

در هر دو مدل، خلوص نقشه برای همه گروه‌های بزرگ در موقعیت‌های اعتبارسنجی بالاتر از ۰/۶ بود (جدول ۳) که حاکی از

جدول ۳- کیفیت پیش‌بینی مدل‌های مورد استفاده برای گروه‌های بزرگ خاک

گروه بزرگ خاک	داده‌های واسنجی						داده‌های اعتبارسنجی					
	رگرسیون لاجیستیک چندجمله‌ای			رگرسیون درختی توسعه یافته			رگرسیون لاجیستیک چندجمله‌ای			رگرسیون درختی توسعه یافته		
	MP	UA	PR	MP	UA	PR	MP	UA	PR	MP	UA	PR
نقشه رقومی حاصله	۰/۷۸	-	-	۰/۷۳	-	-	۰/۷۴	-	-	۰/۶۵	-	-
کلسی جیپسید		۰/۵	۰/۲	۱	۰/۲	۰/۵	۰/۵	۰/۵	۰	۰	۰	
هاپلوکلسید		۰/۷۱	۰/۶۷	۱	۰/۶۱	۱	۱	۱	۱	۰/۷۵	۰/۷۵	
هاپلوکمبید		۰/۶۹	۰/۵۶	۰/۵۶	۰/۹۴	۰	۰	۰	۰/۲۵	۰/۶۷	۰/۶۷	
هاپلوچیپسید		۰/۶۴	۰/۷۴	۰/۶۰	۰/۷۴	۰/۵۷	۱	۱	۰/۶	۰/۷۵	۰/۷۵	
هاپلوسالید		۰/۶۵	۰/۷۶	۰/۹۵	۰/۷۲	۰/۸۳	۰/۷۱	۰/۷۱	۱	۰/۷۱	۰/۷۱	
تری‌ارتنت		۰/۷۱	۰/۶۳	۰/۶۷	۰/۷۵	۰/۵	۱	۱	۰	۰/۷۵	۰/۷۵	
تری‌سامنت		۱	۱	۱	۱	۱	۱	۱	۱	۱	۱	

MP: Map Purity, UA: User Accuracy, PR: Producer Reliability

جدول ۴- قابلیت اطمینان تولیدکننده حاصل از مدل لاجیستیک چندجمله‌ای برای داده‌های اعتبارسنجی

تری‌سامنت	تری‌ارتنت	هاپلوسالید	هاپلوچیپسید	هاپلوکمبید	هاپلوکلسید	کلسی جیپسید	مشاهده نشده	پیش‌بینی شده
-	-	-	-	۱	۱	۱	کلسی جیپسیدز	-
-	-	-	-	-	۴	-	هاپلوکلسیدز	-
-	-	-	-	-	-	-	هاپلوکمبیدز	-
-	-	۱	۴	۱	-	-	هاپلوچیپسیدز	-
-	-	۵	-	۱	-	-	هاپلوسالیدز	-
-	۱	-	-	-	-	۱	تری‌ارتنتز	-
-	-	-	-	-	-	-	تری‌سامنتز	۲
۲	۱	۷	۴	۳	۴	۲	تعداد کل	۲

البته باید ذکر گردد با این تعداد نمونه کم نمی‌توان به‌طور قطع نتیجه گرفت عملکرد مدل لاجیستیک چند جمله‌ای برای گروه بزرگ هاپلوکلسید ۵۰٪ است چرا که ممکن است این نتیجه به‌صورت تصادفی به‌دست آمده باشد. در جدول ۳، شاخص‌های کیفیت حاصل از لاجیستیک چندجمله‌ای برای داده‌های اعتبارسنجی مربوط به کلسی جیپسید، نتایج بالا را تأیید می‌کند. در مورد گروه بزرگ هاپلوکمبید هم از سه نمونه هیچ کدام به درستی پیش‌بینی نشدند (جدول ۴) در واقع یک نمونه به کلسی جیپسید، یک نمونه هم به هاپلوسالید و نمونه سوم هم به هاپلوچیپسید پیش‌بینی شدند. در شکل ۳ این نتایج تأیید می‌گردد به‌طوری‌که ستون مربوط به گروه بزرگ هاپلوکمبید به سه قسمت تقسیم شده که شامل کلسی جیپسید، هاپلوسالید و هاپلوچیپسید می‌باشند. طبق جدول ۳ شاخص‌های کیفیت حاصل از لاجیستیک چندجمله‌ای برای داده‌های اعتبارسنجی مربوط به هاپلوکمبید صفر است که تأییدکننده نتایج بالا می‌باشد. احتمال دارد پیش‌بینی‌های با درجه بالای عدم قطعیت نتیجه مدل

مفهومی ناقص باشند (در مورد گروه‌های بزرگ کلسی جیپسیدز و هاپلوکمبیدز). البته قضاوت در مورد ضعف و ناتوانی مدل‌ها منطقی نیست چرا که تعداد نسبتاً کم نمونه‌ها (۱۰۰ داده آموزشی و ۲۳ داده اعتبارسنجی) برای چنین منطقه وسیعی، تشخیص رابطه پیش‌بینی دقیق را برای طبقه‌بندی کلاس‌های خاک دچار مشکل می‌کند. علاوه بر این، هیچ خصوصیت مشخصه‌ای برای گروه بزرگ هاپلوکمبید وجود ندارد که بتواند برای طبقه‌بندی کمک کند.

شاید اگر این مطالعه در یک مقیاس تفصیلی‌تر یا حتی با به‌کار بردن پیش‌بینی‌کننده‌های بیشتری انجام می‌شد، پیش‌بینی‌های دقیق‌تری برای گروه بزرگ هاپلوکمبید به‌دست می‌آمد. علاوه بر این، میزان ارب‌بودن^۱ در نمونه‌برداری (فقدان مشخص‌سازی کامل فضای پیش‌بینی‌کننده با توجه به متغیر پاسخ) در عدم قطعیت سهیم است (۳). تعداد نمونه‌ها، سطح منطقه مورد مطالعه و مقیاس نقشه از جمله پارامترهایی هستند که نتایج و عملکرد پیش‌بینی را تحت تأثیر قرار

های BRT برای گروه بزرگ تری‌ارتنت در داده‌های واسنجی بالا بود (شکل ۲) به‌ویژه در قسمت شمال‌غربی که مدل لاجیستیک چندجمله‌ای، گروه بزرگ تری‌ارتنت را به گروه بزرگ هاپلوسالید و هاپلوچیسید طبقه‌بندی کرده است. این نتیجه مجدداً حساسیت BRT را به تعداد نمونه کم نشان می‌دهد. برای گروه بزرگ تری‌سامنت به دلیل ارتباط مستقیم بین این گروه خاک و داده‌های توصیفی (جدول ۲) حتی با تعداد کم نمونه، همه مدل‌ها عملکرد پیش‌بینی بالایی را نشان دادند (جدول ۳).

در بین گروه‌های بزرگ خاک، بهترین پیش‌بینی برای گروه بزرگ هاپلوسالید به‌دست آمد، مقادیر بالای دقت کاربران و قابلیت اطمینان تولیدکننده در جدول ۳ این مسئله را نشان می‌دهد. پیش‌بینی دقیق کلاس هاپلوسالید با توزیع مکانی شاخص‌هایی همچون شاخص خیسی، NDVI و لندفرم پلایا به‌ویژه در مرکز آن همبستگی بالایی دارد. دیلاگیلو و ازلمولر (۷) نشان دادند نواحی با احتمال بالا برای هر گروه بزرگ خاک با لندسکیپ‌های کاملاً شناخته شده منطبق است. پارامترهای سنجش از دور (بازتاب نوری) در پیش‌بینی گروه‌های بزرگ هاپلوسالید، تری‌ارتنت و هاپلوچیسید تعیین‌کننده هستند، نواحی مشترک در نقشه‌های آنها گویای این حقیقت است.

زمانی که پارامترهای حاصل از سرزمین و سنجش از دور و فرآیندهای ژئومرفولوژی همه با هم در پیش‌بینی شرکت کنند، بهترین نتایج پیش‌بینی حاصل می‌شود. به نظر می‌رسد جایی که توزیع پیش‌بینی‌کننده‌ها یکنواخت‌تر باشد و تغییرپذیری کمتری داشته باشد، مدل‌ها سریع‌تر رابطه پیش‌بینی‌کننده و متغیر وابسته را تشخیص می‌دهند و ارتباط برقرار می‌کنند. به‌طور کلی، این روش‌های پیش‌بینی گروه‌های بزرگ خاک (لاجیستیک چندجمله‌ای و BRT چندکلاس) که بیشتر شبیه یک روش جعبه سیاه عمل می‌کنند، نقشه خاک با درجه خلوص بالایی تولید کردند.

نتیجه‌گیری

نتایج حاصل از این مطالعه به‌طور خلاصه بیان می‌شود: نقشه ژئومرفولوژی یک ابزار مهم برای روش‌های نقشه‌برداری رقومی خاک است که به افزایش دقت پیش‌بینی کمک می‌کند. از آنجایی که فرآیندهای ژئومرفولوژی در متن سرزمین عمل می‌کنند، هر دو عامل، به‌ویژه در مناطق خشک توسعه و تکامل خاک را تحت تأثیر قرار می‌دهند.

پیش‌بینی‌کننده‌های انتخاب شده با توجه به شرایط تشکیل خاک در لندسکیپ به تفسیر داده‌ها کمک می‌کنند. زمانی که همه فاکتورهای خاک‌سازی به‌طور همزمان در روش مدل‌سازی استفاده شوند، بهترین پیش‌بینی‌ها حاصل می‌شود.

نقشه گروه بزرگ خاک حاصل از روش رگرسیون درختی توسعه

می‌دهند و همچنین به دلیل طرح نمونه‌برداری مورد استفاده در این مطالعه، یک بایاسی در توزیع انواع خاک‌ها اتفاق می‌افتد که خود می‌تواند دلیلی باشد بر اینکه چرا پیش‌بینی خاک‌های با حضور خیلی محدود دقت کمتری دارد در مقایسه با خاک‌های فراوان‌تر که دقت بیشتری دارد (جدول ۳).

در حالی که همه نتایج برای داده‌های واسنجی (جدول ۳) عملکرد خوب BRT را پیشنهاد کرد، داده‌های اعتبارسنجی نتایج عکسی را برای روش BRT (جدول ۳) نشان داد. بالاترین خلوص نقشه برای پیش‌بینی گروه‌های بزرگ خاک توسط لاجیستیک چندجمله‌ای رقم خورد. هر دو مدل در موقعیت‌های اعتبارسنجی، گروه‌های بزرگ کلسی‌چیسید و هاپلوکامید را ضعیف پیش‌بینی کردند. نتایج ضعیف عمدتاً به دلیل اندازه کم نمونه در منطقه وسیع (۱۰۰ موقعیت آموزشی و ۲۳ موقعیت اعتبارسازی) و نمایش ضعیف انواع خاک‌ها با حضور خیلی محدود در منطقه تحقیقاتی در مقایسه با نمایش بیشتر انواع دیگر خاک (۴) بود. علاوه بر این، اگر معیارهای تشخیص برای بعضی خاک‌ها ناشناخته باشد، شناسایی آنها به‌طور دقیق خیلی مشکل است. این مسئله برای هاپلوکامیدهای مورد مطالعه واقعیت داشت، هرچند که با افزایش داده پیش‌بینی آنها بهبود یافت (جدول ۳).

نتایج نشان داد که حداکثر کاهش در عملکرد پیش‌بینی، مربوط به پیش‌بینی گروه‌های بزرگ کلسی‌چیسید، هاپلوکلسید و هاپلوکامید بود. تعداد این گروه‌های بزرگ خاک در داده‌های واسنجی و اعتبارسازی کمتر بود در حالیکه عملکرد پیش‌بینی برای گروه‌های بزرگ هاپلوسالید و هاپلوچیسید بهتر بود (جدول ۳)، چرا که این گروه‌های بزرگ خاک در داده‌های واسنجی و اعتبارسازی تعداد نمونه بیشتری را شامل می‌شدند. بنابراین، به نظر می‌رسد مدل BRT به تعداد نمونه حساس‌تر است، به‌طوری‌که هرچه داده‌های بیشتری برای برازش مدل استفاده شود، تغییرپذیری کلاس‌های خاک بهتر نشان داده می‌شود و پیش‌بینی کلاس‌های خاک در داده‌های آزمایشی بهبود یافته است.

روش لاجیستیک چندجمله‌ای، گروه بزرگ تری‌ارتنت را بهتر پیش‌بینی کرده است (جدول ۳) هرچند که در این روش هم بعضی از تری‌ارتنت‌ها به هاپلوچیسیدها طبقه‌بندی شده‌اند (جنوب غربی شکل ۲). شکل ۳ پیش‌بینی غلط گروه بزرگ تری‌ارتنت به کلاس‌های دیگر را نشان نمی‌دهد اما جدول ۴ پیش‌بینی غلط کلاس‌های دیگر به گروه بزرگ تری‌ارتنت را به خوبی نشان می‌دهد که باعث دقت کاربر پایین می‌شود. این مسئله با توجه به نتایج جدول ۳ کاملاً مشخص و واضح است. تعداد کم این نوع خاک‌ها و در نتیجه فقدان ارتباط خوب بین پیش‌بینی‌کننده‌ها و این گروه بزرگ خاک، خود عاملی بر قابلیت اطمینان پایین این خاک‌ها می‌باشد. اگرچه نقشه گروه بزرگ خاک حاصل از مدل BRT نسبت به مدل لاجیستیک چندجمله‌ای بیشتر با واقعیت انطباق داشت، عملکرد پیش‌بینی مدل -

یافته با واقعیت سازگارتر بود. نتایج نشان داد مدل رگرسیون درختی توسعه یافته به پایگاه داده کوچک حساس است. تلفیق داده‌های جغرافیایی با نرم افزار R، فرصتی را برای پیش-بینی کلاس‌های خاک فراهم می‌کند و نرم‌افزارهای بر پایه GIS (مثل SAGA) و نرم‌افزارهای آماری قوی (مثل R) به آسانی نقشه-برداران خاک را حمایت می‌کند. انتظار می‌رود روش‌های توسعه یافته آنالیز رقوم سرزمین و تشریح واضح فرآیندهای ژئومورفولوژی-زمین-شناسی-پدولوژی به‌عنوان کلید تکنولوژی در نقشه‌برداری رقومی خاک وارد شود.

منابع

- 1- Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. p. 267-281. In B.N. Petrov and Csaki F. (eds) 2th International Symposium on Information Theory. Akademiai Kiado, Budapest.
- 2- Bailey N., Clements T., Lee, J.T., and Thompson S. 2003. Modelling soil series data to facilitate targeted habitat restoration: a polytomous logistic regression approach. *Journal of Environmental Management*, 67(4): 395-407.
- 3- Bauer E., and Kohavi R. 1999. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36: 105-139.
- 4- Beaudette D.E., and O'Geen A.T. 2008. New digital soil survey products to quantify soil variability over 28 multiple scales. *Kearney Foundation of Soil Science: Understanding and Managing Soil-29 Ecosystem Functions Across Spatial and Temporal Scales. 2006-2011 Mission.*
- 5- Behrens T., Zhu A.X., Schmidt, K., and Scholten T. 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*, 155:175-185.
- 6- Campling P., Gobin, A., and Feyen J. 2002. Logistic modeling to spatially predict the probability of soil drainage classes. *Soil Science Society of American Journal*, 66:1390-1401.
- 7- Debella-Gilo M., and Etzelmuller B. 2009. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vestfold County, Norway. *Catena*, 77:8-18.
- 8- Deventer van A.P., Ward A.D., Gowda, P.H., and Lyon J.G. 1997. Using thematic mapper data to identify contrasting soil plains and tillage practices. *Photogrammetric Engineering and Remote Sensing*, 63(1):87-93.
- 9- Elith J., Leathwick, J.R., and Hastie T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*, 77:802-813.
- 10- Friedman J., Hastie T., and Tibshirani R. 2000. Additive Logistic Regression: A Statistical View of Boosting. *Annals of Statistics*, 28(2):337-407.
- 11- Gallant J.C., and Dowling T.I. 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resource Research*, 39(1):1347.
- 12- Gamberger D., and Smuc T. 2001. Data Mining Server. <http://www.dms.irb.hr/>.
- 13- Grinand C., Arrouays D., Laroche, B., and Martin M.P. 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma*, 143:180-190.
- 14- Hengl T., Toomanian N., Reuter, H., and Malakouti M.J. 2007. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. *Geoderma*, 140:417-427.
- 15- Lagacherie P. 1992. Formalisation des lois de distribution des sols pour automatiser la cartographie pédologique à partir d'un secteur pris comme référence. *Mémoire de Thèse, Université de Montpellier. Institut National de la Recherche Agronomique, France.* p. 175.
- 16- Moameni A., and Zink J.A. 1997. Application of statistical quality control charts and geostatistics to soil quality assessment in a semi-arid environment of south-central Iran. *ITC Journal*, 1997-3/4.
- 17- Ohlmacher G.C., and Davis J.C. 2003. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering Geology*, 69:331-343.
- 18- Olaya V. 2004. A gentle introduction to Saga GIS. User manual, The SAGA User Group, Göttingen.
- 19- Pearson R.L., and Miller L.D. 1972. Remote mapping of standing crop biomass for estimation of the productivity of the short-grass prairie, Pawnee National Grasslands, Colorado. p. 1357-1381. In *Proceedings of the 8th International Symposium on Remote Sensing of Environment*, ERIM International, Inc., Ann Arbor, Michigan, USA.
- 20- Richardson A.J., and Wiegand C.L. 1977. Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, 43(12):1541-1552.
- 21- Ridgeway G. 2009. Package `_gbm`. *R-News*. 08:05:15.
- 22- Rossiter D.G., and Loza A.V. 2010. Analyzing land cover change with logistic regression in R: Technical Note. *International Institute for Geoinformation Science and Earth Observation (ITC), Enschede.* Available at <http://www.itc.nl/personal/rossiter>.
- 23- Rouse J.W., Haas R.H., Schell, J.A., and Deering D.W. 1973. Monitoring vegetation systems in the Great Plains

- with ERTS. p. 309-317. In 3rd Proceeding ERTS Symposium, 10-14 December. 1973. NASA SP-351. Natl. Aeronaut. Space Admin., Washington, DC.
- 24- Scull P. R. 2002. Predictive soil mapping in the Mojave Desert of California. University of California, San Diego State University, PhD thesis.
- 25- Scull P., Franklin J., Chadwick, O.A., and McArthur D. 2003. Predictive soil mapping: a review. *Progress Physical Geography*, 27:171-197.
- 26- Soil Survey Staff. 2010. Keys to Soil Taxonomy, Eleventh edition, NRCS, USDA.
- 27- The Ministry of Economy, Trade and Industry of Japan (METI) and the National Aeronautics and Space Administration (NASA). 2009. Aster Global Digital Elevation Model (Aster GDEM). NASA Official. URL <http://www.gdem.aster.ersdac.or.jp>.
- 28- Wang D., and Laffan S.W. 2009. Characterisation of valleys from DEMs. In Proceedings of the 18th World IMACS Congress and MODSIM International Congress on Modeling and Simulation, 13-17 July. Cairns, Australia. Available at <http://mssanz.org.au/modsim09>.

Application of Regression Models for Prediction of Soil Classes in Some Regions of Central Iran (Zarand district, Kerman Province)

A. Jafari¹- Sh. Ayoubi^{2*}- H. Khademi³

Received:23-1-2011

Accepted:21-8-2011

Abstract

Soil digital survey as tool for soil spatial information provides pathways for producing of high resolution soil maps. Therefore, it should be developed strategic methods for making high resolution soil spatial information. Subsequently, this study was designed for prediction of soil classes by regression models in Zarand region of Kerman. Regression models includes of multinomial logistic regression and multiclass boosted regression tree were used for prediction of soil great groups by relating those with predictors such as remote sensing indices, terrain attributes and geomorphology map. A confusion matrix was used to calculate aspects of map accuracy. The geomorphology map at the fourth level (geomorphic surface) was a powerful predictor unlike the other levels (landscape, landform and lithology). Terrain attributes and finally remote sensing indices after geomorphic surface were imported as predictors in the prediction. The map purity over all soil great groups was above 0.60 in both calibration and validation locations. Poorer performance was observed for Calcigypsid and Haplocambids. Both methods provided good predictions for Haplosalids that shown by high values for users' accuracy and producers' reliability. The results showed soils with better reliability are those highly influenced by topographic and geomorphic characteristics at least in this study area (e.g. Haplosalids, Haplogypsid and Torripsamments) and soils with very low reliability and accuracy of prediction are hardly influenced by the topographic and geomorphic characteristics (e.g. Haplocambids and Calcigypsid).

Keywords: Digital soil mapping, Multinomial logistic regression, Boosted regression tree

1,2,3- PhD Student, Associate Professor and Professor, Department of Soil Science, College of Agriculture, Isfahan University of Technology, Respectively

(* - Corresponding Author Email: ayoubi@cc.iut.ac.ir)