

نقشه برداری رقومی افق‌های مشخصه و گروه‌های بزرگ خاک در منطقه زرنند کرمان

اعظم جعفری^۱، حسین خادمی^{۲*} و شمس الله ایوبی^۲

(تاریخ دریافت: ۱۳۹۰/۴/۲۵؛ تاریخ پذیرش: ۱۳۹۰/۱۱/۸)

چکیده

نقشه برداری رقومی خاک شامل پیش‌بینی مکانی خاک‌ها و ویژگی‌های آنها براساس ارتباط با متغیرهای کمکی می‌باشد. این مطالعه به منظور نقشه برداری رقومی خاک با استفاده از مدل‌های رگرسیونی لاجیستیک دوتایی و درختی توسعه یافته در منطقه زرنند کرمان طراحی گردید. یک طرح نمونه برداری طبقه بندی شده در منطقه‌ای به مساحت نود هزار هکتار تعریف شد و در نهایت، ۱۲۳ خاک‌رخ حفر و تشریح گردید. در هر دو مدل، ابتدا احتمال حضور افق‌های مشخصه از طریق ارتباط دادن با متغیرهای کمکی، تخمین زده شد و سپس حضور کلاس‌های مختلف خاک پیش‌بینی گردید. متغیرهای کمکی شامل نقشه ژئومورفولوژی، اجزای سرزمین و شاخص‌های سنجش از دور بودند. نتایج نشان داد که نقشه ژئومورفولوژی، یک ابزار مهم در روش‌های نقشه برداری رقومی خاک است که به افزایش دقت پیش‌بینی کمک می‌کند. پس از سطوح ژئومرفیک، اجزای سرزمین به عنوان پارامترهای کمکی مؤثر شناخته شدند. هر دو مدل، احتمال بالای حضور افق سالیک را در سیمای اراضی پلایا، افق جیپسیک را در تپه‌های گچی و افق کلسیک را در مخروط‌افکنه‌ها پیش‌بینی کردند. در بین کلاس‌های خاک مورد پیش‌بینی، کلاس "کلسی جیپسیدز"، دقت و قابلیت اعتبار تخمین پایینی را نشان داد، در صورتی که تخمین کلاس‌های "هپلوسالیدز" و "هپلو جیپسیدز" از دقت بالایی برخوردار بود.

واژه‌های کلیدی: رگرسیون لاجیستیک دوتایی، رگرسیون درختی توسعه یافته، افق‌های مشخصه خاک، کلاس خاک، رده بندی خاک آمريکايي

۱. گروه خاک‌شناسی، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان

۲. گروه خاک‌شناسی، دانشکده کشاورزی، دانشگاه صنعتی اصفهان

*: مسئول مکاتبات، پست الکترونیکی: hkhademi@cc.iut.ac.ir

مقدمه

نقشه برداری کلاس های خاک، یکی از تکنیک هایی است که برای تعیین تغییرات مکانی خاک استفاده شده است (۲۱). کلاس های خاک از سامانه های طبقه بندی محلی، ملی یا جهانی خاک (مانند طبقه بندی عددی (۷)، سامانه طبقه بندی خاک استرالیا (۱۹)، یا سامانه مرجع جهانی (۲۰)) به دست می آیند. بعضی از خاک شناسان، نقشه کلاس خاک را ترجیح می دهند؛ چرا که آنها با مفهوم کلاس آشنا هستند و می توانند ویژگی های خاک ها را از آنها استنباط کنند.

در هر موقعیت مشاهده خاک، یک سری متغیرهای محیطی وجود دارد که هم بستگی بالایی با کلاس های خاک دارند که یافتن این ارتباط و پیش بینی کلاس های خاک در نقاط دیگر، پایه و اساس نقشه برداری رقومی خاک است. مکرراتنی و همکاران (۲۳) ارتباط کلاس های خاک و متغیرهای محیطی را به صورت زیر بیان کردند:

$$Sc = f(s, c, o, r, p, a, n) \quad [1]$$

که در این معادله، Sc: کلاس خاک؛ s: اطلاعات مربوط به خاک که از نقشه خاک، پایگاه داده یا دانش کارشناس به دست می آیند؛ c: اقلیم؛ o: موجودات زنده؛ r: پستی و بلندی یا توپوگرافی؛ p: مواد مادری؛ a: زمان و n: موقعیت مکانی می باشد. تابع f، یک طبقه بندی نظارت شده یا برنامه یادگیری نظارت شده را نشان می دهد. تابع f برای پیش بینی کلاس خاک می تواند رگرسیون لاجیستیک (۸)، آنالیز تشخیص خطی (۶)، شبکه های عصبی (۳۲)، یا طبقه بندی درختی یا تصمیم گیری درختی (۲۲) باشد.

کمپلینگ و همکاران (۸) رگرسیون لاجیستیک را برای پیش بینی مکانی کلاس های زه کشی خاک با کمک اجزای سرزمین و شاخص های پوشش گیاهی به کار بردند. آنها گزارش کردند عملکرد مدل های چندمتغیره حاصل از ترکیب متغیرهای توصیفی بهتر از مدل های تک متغیره است. هنگل و همکاران (۱۶) رگرسیون لاجیستیک چند جمله ای (Multinomial Logistic Regression) را برای پیش بینی کلاس

خاک استفاده کردند. نتایج آنها نشان داد در صورتی که هم بستگی بین متغیرهای کمکی و کلاس های خاک ضعیف باشد، این مدل پیش بینی ضعیفی را به دنبال خواهد داشت. دابلاگیلو و اتزلمولر (۹) در مطالعه ای در ارتباط با پیش بینی مکانی کلاس های خاک براساس آنالیز سرزمین و به کمک رگرسیون لاجیستیک چند جمله ای گزارش کردند که برای کلاس هایی که متأثر از پستی و بلندی منطقه نیستند، پیش بینی ضعیف می باشد. در زمینه استفاده از رگرسیون لاجیستیک دوتایی و رگرسیون درختی توسعه یافته برای پیش بینی کلاس خاک تاکنون مطالعه ای انجام نشده است. بنابراین، مطالعه حاضر با هدف به کارگیری این روش ها در پیش بینی احتمال حضور افق های مشخصه و کلاس های خاک و ارزیابی توان تخمین آنها در منطقه زرنند استان کرمان، طراحی گردید.

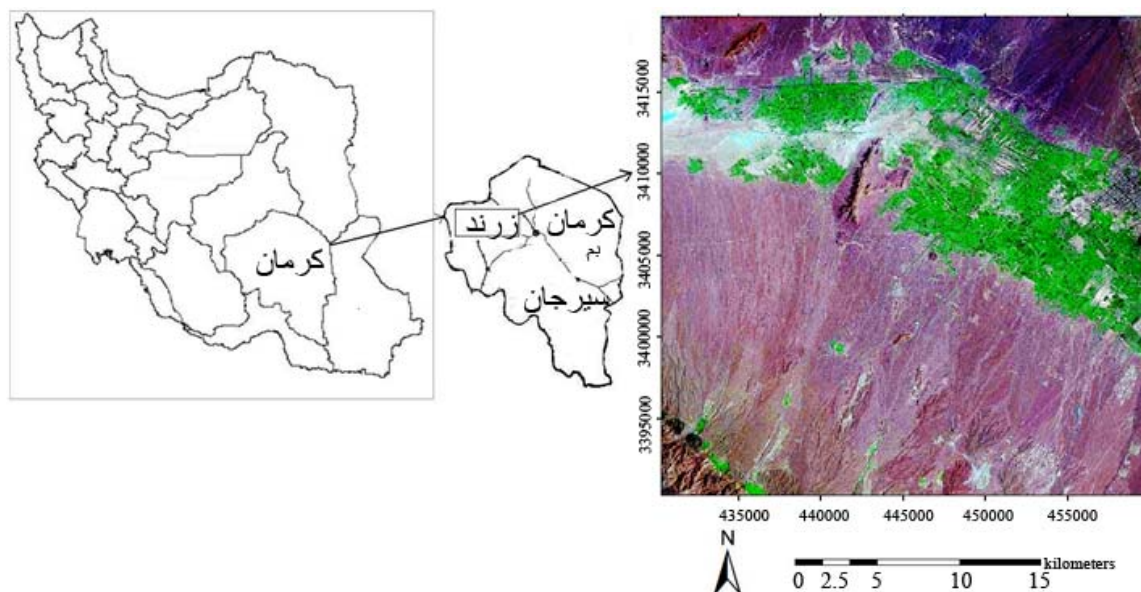
مواد و روش ها

۱. تشریح منطقه مطالعاتی

منطقه مورد مطالعه در جنوب شرق ایران، در فاصله حدود ۷۰ کیلومتری شهر کرمان بین عرض های جغرافیایی ۲۵° ۳۰' تا ۲۶° ۰۰' شرقی و طول های جغرافیایی ۵۶° ۱۰' تا ۵۶° ۵۰' شمالی قرار گرفته است. این منطقه، مساحتی حدود ۹۰۰۰۰ هکتار را می پوشاند. سیمای اراضی اصلی این منطقه شامل مخروط افکنه، پلایا، باهادا، تپه های گچی و تپه های شنی می باشند (شکل ۱). متوسط سالانه بارندگی، درجه حرارت، تبخیر و تعرق و ارتفاع به ترتیب ۶۱ میلی متر، ۱۷ درجه سانتی گراد، ۱۷۵۰ میلی متر و ۱۶۵۰ متر می باشد (۱).

۲. طرح نمونه برداری

یک طرح نمونه برداری طبقه بندی شده (Stratified Sampling Scheme) با استفاده از نقشه های رقومی زمین شناسی، ژئومرفولوژی و توپوگرافی پیاده گردید. در این مطالعه، لایه نمونه برداری (Sampling Stratum) منطبق بر تفاوت در اشکال اراضی (نقشه ژئومرفولوژی)، توپوگرافی (نقشه رقومی ارتفاع)



شکل ۱. موقعیت جغرافیایی منطقه مطالعاتی در حوالی شهر زرنند کرمان (تصویر ماهواره‌ای لندست ETM+ (RGB: 243)

توسعه‌یافته (Boosted Regression Tree) در نرم‌افزار R مورد استفاده قرار گرفتند.

۴. پیش‌بینی افق‌های مشخصه و کلاس‌های خاک با روش رگرسیون لاجیستیک دوتایی

ابتدا احتمال حضور افق‌های مشخصه تخمین زده شد و سپس با ترکیب حضور یا عدم حضور افق‌های مشخصه، کلاس خاک سلول به سلول تعیین گردید. بنابراین، مرحله اول برای پیش‌بینی کلاس خاک توسط لاجیستیک دوتایی و هم‌چنین رگرسیون درختی توسعه‌یافته؛ تخمین حضور یا عدم حضور افق‌های مشخصه است. از این‌رو متغیر هدف، دارای توزیع دوتایی (Binary) است. در چنین موقعیتی، رگرسیون لاجیستیک دوتایی، یک روش پیش‌بینی مناسب است (۱۷). برای اجرای این مدل در نرم‌افزار R باید توزیع متغیر مشخص شود که در این مورد توزیع دوتایی است. مدل لاجیستیک به‌صورت زیر تعریف می‌گردد:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{i=1}^n \beta_i * x_i \quad [2]$$

که در این مدل، p احتمال حضور یک افق مشخصه، β_i ضریب رگرسیون، x_i متغیر مستقل و n تعداد متغیرهای مستقل است. در

زمین‌شناسی تعریف گردید. در هر طبقه، موقعیت‌های تصادفی نمونه‌برداری به‌گونه‌ای انتخاب شد که تعداد نمونه‌ها متناسب با مساحت طبقه باشد. در نهایت، ۱۲۳ خاک‌رخ حفر، تشریح، نمونه‌برداری، آنالیز و براساس معیارهای سامانه رده‌بندی آمریکایی (۳۰) طبقه‌بندی گردید.

۳. متغیرهای کمکی

یک مدل رقومی ارتفاع با تفکیک مکانی ۳۰ متر که از وب‌سایت مدل رقومی ارتفاع جهانی استر (Aster Global Elevation Digital Model) به‌دست آمد (۳۱)، برای تعیین اجزای سرزمین (Terrain attributes) و یک سین تصویر ماهواره‌ای ETM+ برای تعیین شاخص‌های گیاهی استفاده گردید. اجزای سرزمین حاصل از مدل رقومی ارتفاع و شاخص‌های سنجش از دور در جدول ۱ ارائه گردیده است. نقشه‌های اجزای سرزمین، شاخص‌های سنجش از دور و ژئومرفولوژی در نرم‌افزار SAGA (۲۵) زمین‌مرجع شدند. سپس، این متغیرهای کمکی برای پیش‌بینی افق‌های مشخصه و کلاس‌های خاک با استفاده از رگرسیون لاجیستیک دوتایی (Binary Logistic Regression) و رگرسیون درختی

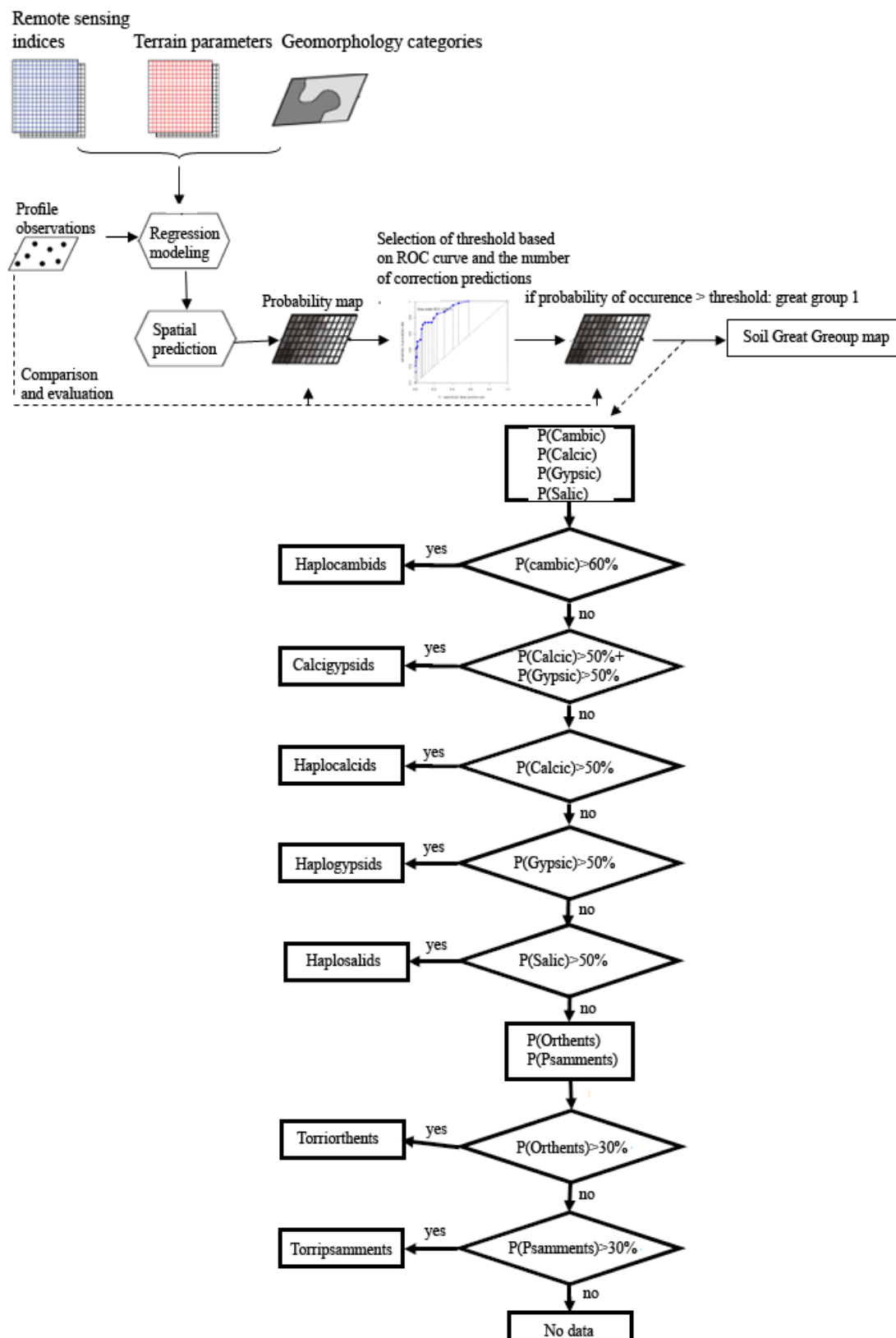
جدول ۱. داده‌های محیطی مورد استفاده به‌عنوان پیش‌بینی‌کننده در مدل‌سازی

داده‌های محیطی	مقیاس	ماهیت فاکتور	نام	تشریح	نوع متغیر	منبع
مدل رقومی ارتفاع (DEM)	۳۰ متر	توپوگرافی	El	ارتفاع (متر)	کمی	
			SI	زاویه شیب (درصد)		
			MCur	متوسط انحنا		
			PrCur	انحنای خاک‌رخ		
			PlCur	انحنای صفحه‌ای		
			WI	شاخص خیزی		
			TWI	شاخص خیزی توپوگرافی		
			MrVBF	(Multi-resolution Valley Bottom Flatness index) شاخص همواری دره با درجه تفکیک بالا		(۱۳)
تصویر لندست (ETM+)		انعکاس نور	NDVI	شاخص گیاهی تفاضلی نرمال شده	کمی	(۲۹)
			RVI	شاخص گیاهی نسبی		(۲۶)
			PVI	شاخص گیاهی عمودی		(۲۷)
			CI	شاخص رس		(۹)
نقشه ژئومرفولوژی	۱:۱۰۰۰۰۰	ژئومرفولوژی	GS	سطوح ژئومرفیک (۱۸ سطح)	کیفی	(۲۴)
نقشه ژئولوژی	۱:۱۰۰۰۰۰	زمین‌شناسی	Geol	واحدهای زمین‌شناسی (۹ واحد)	کیفی	

احتمال‌های پیش‌بینی شده تعیین گردید. بنابراین، حدود آستانه بایستی انتخاب می‌شدند. مقادیر حدود آستانه هر نقشه شاخص، براساس منحنی ROC (Receiver Operating Characteristic) یا مشخصه عامل گیرنده انتخاب گردیدند. سپس، سطح زیر منحنی (AUC) (Area Under Curve) (۲۶) به‌عنوان شاخصی از کیفیت کلی نقشه تخمین زده شد. AUC مساوی یک، نشان‌دهنده قدرت بالای روش در تشخیص حضور و عدم حضور می‌باشد و AUC مساوی ۰/۵ نشان می‌دهد که روش مورد استفاده، هیچ‌گونه قدرت تمایزی در تشخیص حضور و عدم حضور ندارد و در واقع، تشخیص به‌صورت تصادفی است.

موارد عدم حضور افق مشخصه، خاک‌ها در رده انتی‌سول طبقه‌بندی گردیدند که در مرحله پیش‌بینی افق مشخصه یا متغیر شاخص از نام زیر رده این خاک‌ها استفاده شد (مانند Orthents یا Psamments). برای انتخاب مدل از بین مدل‌های برازش داده شده، شاخص AIC (Akaike Information Criterion) به‌کار برده شد (۳). در نهایت، مدل با کمترین AIC و انحراف باقی‌مانده انتخاب گردید (۵).

به منظور ترکیب نقشه‌های حاصل از پیش‌بینی افق‌های مشخصه برای ایجاد گروه‌های بزرگ خاک، یک درخت تصمیم‌گیری ایجاد گردید (۳۱). در شکل ۲ مراحل ایجاد درخت تصمیم‌گیری نمایش داده شده است. حضور افق‌های مشخصه براساس مقادیر بزرگ‌تر از حدود آستانه در



شکل ۲. روندنمای مراحل پیش‌بینی گروه‌های بزرگ خاک با استفاده از احتمال حضور افق‌های مشخصه، به کمک مشاهدات خاک‌رخ و اطلاعات کمی

۵. پیش‌بینی افق‌های مشخصه و کلاس‌های خاک با روش رگرسیون درختی توسعه یافته (BRT)

روش رگرسیون درختی توسعه یافته از مجموعه ماشین‌های یادگیری، ترکیبی از دو تکنیک آماری قوی بوستینگ و رگرسیون درختی است (۲). بوستینگ یک روش مرحله‌ای رو به جلو است که در آن، مدل‌های درختی به‌صورت تکرارپذیر با زیرمجموعه‌ای از داده‌های آموزشی برازش داده می‌شوند. اولین مرحله در برازش یک درخت بوستینگ، تعیین تعداد درختان در توالی‌ها، مقدار یادگیری، اندازه درختان و سهم داده‌های آموزشی می‌باشد. حدود ۰/۴ تا ۰/۶ سهم داده‌ها به‌عنوان داده‌های آموزشی انتخاب می‌شوند. سهم داده‌های آموزشی مورد استفاده توسط مدل‌ساز با پارامتری به نام "Bag fraction" تعیین می‌شود. این عمل، فرآیند تصادفی بودن (Stochasticity) را به مدل اعمال می‌کند که دقت مدل را افزایش و برازش بیش از حد (Overfitting) را کاهش می‌دهد (۱۲).

در برازش BRT، دو پارامتر مقدار یادگیری (Learning Rate) و پیچیدگی درخت (Tree Complexity) برای تعیین تعداد کل درختان باید مشخص شود (۱۱ و ۱۲). مقدار 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. مدل‌های BRT با استفاده از کد BRT نوشته شده توسط الیت و همکاران (۱۲) و هم‌چنین بسته "gbm" (Generalized Boosted Regression Models) (۲۸) موجود در نرم‌افزار R برازش داده شدند.

در پیش‌بینی گروه بزرگ خاک (متغیر هدف) به‌وسیله BRT لاجیستیک (Logistic-BRT)، ابتدا احتمال حضور افق‌های مشخصه تعیین شد و سپس نقشه‌های مختلف برای طبقه‌بندی به انواع خاک، با همدیگر ترکیب گردیدند. برای ترکیب نقشه‌های شاخص، یک درخت تصمیم‌گیری باید تعریف شود (شکل ۲).

۶. 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 شد. اعتبارسنجی با مقایسه داده‌های مشاهده شده و پیش‌بینی شده انجام گرفت. خلوص نقشه (MP, Map Purity) ارتباط بین همه داده‌های مورد استفاده و داده‌های صحیح طبقه‌بندی شده (ff, tt) را تشریح می‌کند:

$$\text{Map Purity} = \frac{tt + ff}{tt + tf + ft + ff} \quad [3]$$

صحت کاربر (UA, User's Accuracy) شامل کلاس‌های حضور صحیح پیش‌بینی شده (tt) به مجموع کلاس‌های حضور صحیح پیش‌بینی شده (tt) و کلاس‌های عدم حضور که به غلط جزو کلاس‌های حضور مشاهده شده (واقعی) پیش‌بینی شدند (tf)، می‌باشد:

$$\text{User's Accuracy} = \frac{tt}{tt + tf} \quad [4]$$

قابلیت اطمینان تولیدکننده (PR, Producer's Reliability) بیانگر ارتباط بین همه کلاس‌های صحیح پیش‌بینی شده (tt) و مجموع کلاس‌های صحیح پیش‌بینی شده و کلاس‌های حضور مشاهده شده (واقعی) که به غلط جزو کلاس‌های عدم حضور پیش‌بینی شدند (ft)، می‌باشد:

$$\text{Producer's Reliability} = \frac{tt}{tt + ft} \quad [5]$$

یک نقشه خوب، دارای مقادیر خلوص نقشه، صحت کاربر و قابلیت اطمینان تولیدکننده نزدیک به عدد یک است (۲).

نتایج و بحث

۱. مدلسازی ارتباط خاک - سرزمین توسط رگرسیون لاجیستیک دوتایی

جدول ۲ متغیرهای کمکی مورد استفاده در پیش‌بینی افق‌های مشخصه توسط مدل رگرسیون لاجیستیک دوتایی را نشان

جدول ۲. متغیرهای مورد استفاده در پیش‌بینی افق‌های مشخصه خاک در برازش مدل‌های لاجستیک دوتایی

متغیرها براساس اهمیت	متغیرهای مورد استفاده در مدل	افق مشخصه یا کلاس خاک
GS ^{***} , NDVI ^{**} , MrVBF ^{**} , PVI [*]	GS+ NDVI+ MrVBF+ CI+PVI	سالیک
GS ^{***} , WI ^{**} , PICur ^{**} , PVI [*]	GS+ PICur+ WI+ PVI	چیپسیک
GS ^{***} , PVI [*]	GS+ PVI	کلسیک
EI [*] , SI [*] , PVI ^{**} , MrVBF ^{**} , TWI [*]	EI+ SI+ PVI+ TWI+ MrVBF	کمبیک
GS ^{***}	GS	سامنتز
EI ^{***} , GS [*]	EI+ GS	ارتنتز

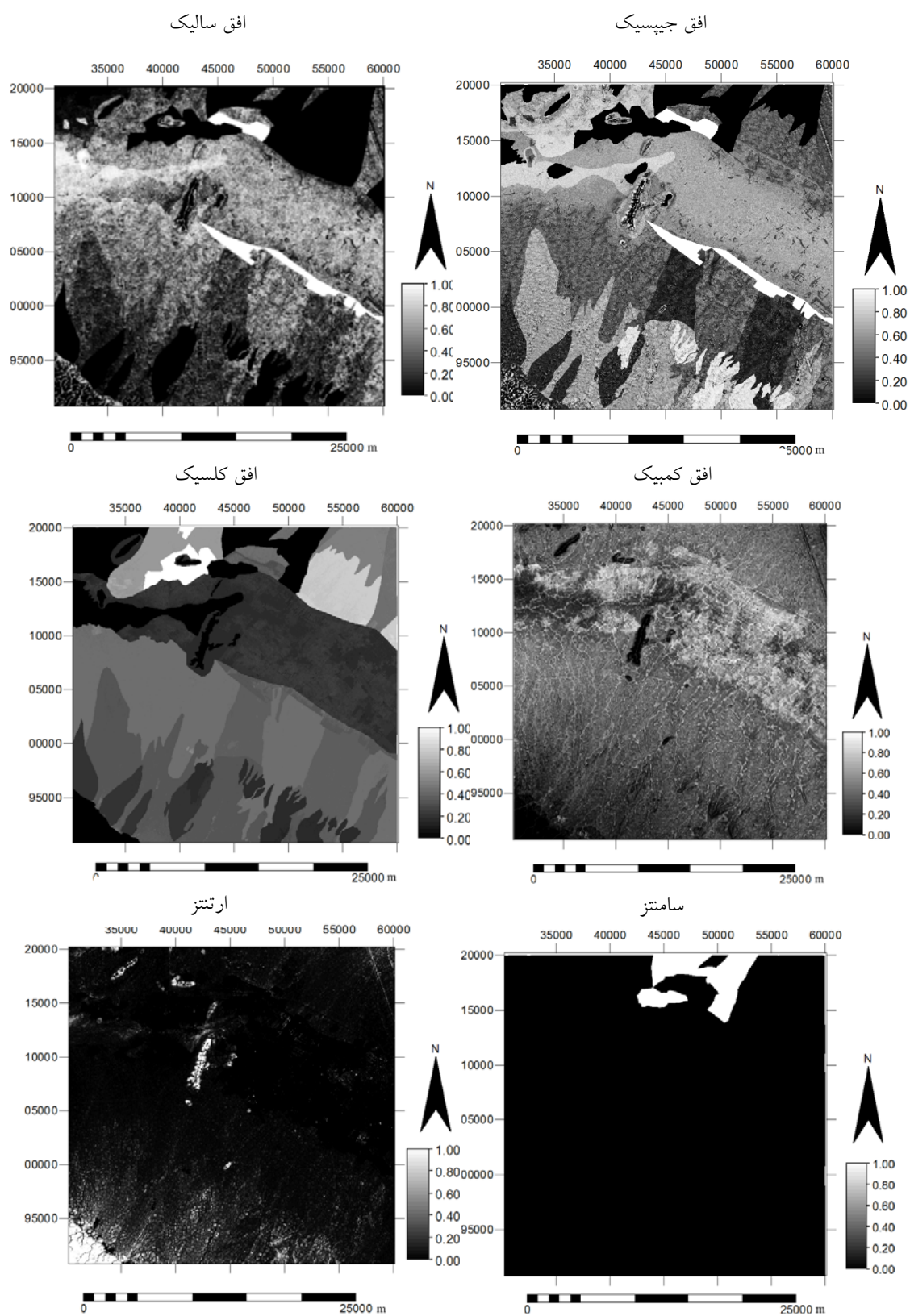
***: $P < 0.001$, **: $P < 0.01$, *: $P < 0.05$

می‌باشد. نقشه شاخص افق سالیک، احتمال بالای حضور این افق مشخصه را در پلایا نشان می‌دهد. مقایسه این نقشه با تصویر ماهواره‌ای منطقه، هم‌بستگی مکانی بالای شاخص‌های سنجش از دور (مانند NDVI) با حضور افق مشخصه سالیک را نشان می‌دهد. هم‌چنین، بالاترین احتمال حضور افق مشخصه سالیک در غربی‌ترین قسمت پلایا با بیشترین درجه شوری مشاهده گردید. پوشش گیاهی ضعیف در این ناحیه، مؤید احتمال بالای حضور افق سالیک است (شکل ۱). بیشترین احتمال حضور افق چیپسیک در تپه‌های گچی که شامل ۵۰ تا ۷۰ درصد گچ هستند (جدول ویژگی‌های فیزیکی و شیمیایی خاک ارایه نشده است) مشاهده شد (شکل ۳). براساس یافته‌های صحرایی، احتمال می‌رود که این تپه‌ها منبع اصلی گچ در قسمت‌های پایین‌تر منطقه باشند.

افق کلسیک عمده‌تاً در اشکال اراضی مخروط‌افکنه و باهادا پیش‌بینی گردید (شکل ۳) که با مشاهدات صحرایی مطابقت دارد. البته حضور افق کلسیک در این اشکال اراضی به پایداری خاک در موقعیت نمونه‌برداری بستگی دارد، چرا که در شیب‌های جانبی، افق کلسیک مشاهده نگردید و در موقعیت‌های باثبات شیب، احتمال حضور افق کلسیک افزایش می‌یابد؛ از این رو برای دربرگرفتن چنین مشاهداتی، فرآیندها و سطوح ژئومرفیک بایستی دقیق‌تر و جزئی‌تر تفکیک شوند و به دنبال آن، تعریف معیارهای ژئومرفولوژی باید بهتر صورت گیرد

می‌دهد. سطح ژئومرفیک به‌عنوان یک پیش‌بینی‌کننده قدرتمند در پیش‌بینی افق‌های مشخصه به استثنای افق کمبیک مشخص گردید. بعد از سطوح ژئومرفیک، اجزای سرزمین از مؤثرترین پارامترها در پیش‌بینی افق‌های مشخصه بودند (جدول ۲). این موضوع می‌تواند مؤید این مطلب باشد که در این منطقه، پستی و بلندی از مهم‌ترین فاکتورهای خاک‌سازی می‌باشد. در الگوریتم‌های پیشنهادی توسط دوبوس و همکاران (۱۱)، مدل رقومی ارتفاع و اجزای سرزمین حاصل از آن برای طبقه‌بندی خاک انتخاب می‌شود. شاخص‌های سنجش از دور (از جمله شاخص گیاهی تفاضلی نرمال شده (NDVI) و شاخص گیاهی عمودی (PVI)) هم جزو متغیرهای کمکی، قابلیت مناسبی در پیش‌بینی کلاس‌های خاک دارند. نقش مؤثر و معنی‌دار آنها در پیش‌بینی افق مشخصه سالیک، نشان‌دهنده این موضوع است (جدول ۲). آکسوی و همکاران (۴) گزارش کردند ترکیب اجزای سرزمین و داده‌های سنجش از دور، ابزاری قدرتمند در نقشه‌برداری خاک است.

نقشه‌های احتمال حضور افق‌های مشخصه یا کلاس خاک حاصل از لاجستیک دوتایی در شکل ۳ نشان داده شده‌اند. در این نقشه‌ها، احتمال حضور یک افق مشخصه یا کلاس خاک در یک سلول معین به‌وسیله مقادیر بین صفر و یک نشان داده شده است، که عدد صفر، نشان‌دهنده عدم حضور و عدد یک (نواحی سفید)، نشانگر حضور یک افق مشخصه یا کلاس خاک



شکل ۳. نقشه‌های توزیع احتمال افق‌های مشخصه یا گروه بزرگ خاک حاصل از رگرسیون لاجیستیک دوتایی

که مستلزم افزایش مقیاس مطالعه می‌باشد. ناگفته پیداست که، افزایش مقیاس مطالعه منجر به افزایش تعداد نمونه‌ها و در نتیجه افزایش هزینه می‌شود.

از آنجا که کلاس خاک "سامنتز" فقط در تپه‌های شنی مشاهده می‌شود، پیش‌بینی آن به این سطح ژئومرفیک محدود می‌شود (جدول ۲). بنابراین، به دلیل رابطه مستقیم سطح ژئومرفیک تپه‌های شنی و این کلاس خاک، پیش‌بینی با دقت بالایی انجام پذیرفت. نقشه‌های پیش‌بینی کلاس‌های خاک "ارتنتز" و "سامنتز" کاملاً منعکس‌کننده حضور سطوح سنگی فرسایش‌یافته و رسوبات سخت نشده می‌باشند.

۲. مدل‌سازی ارتباط خاک - سرزمین توسط مدل رگرسیون درختی توسعه یافته

برای مدل‌های BRT، یک شاخص تأثیر نسبی (Relative Influence) محاسبه شد که سهم هر متغیر را در پیش‌بینی نشان می‌دهد و معادل با مجموع طول شاخه برای هر متغیر در رگرسیون درختی می‌باشد. برای مدل BRT برازش داده شده برای افق سالیک، پنج متغیر کمکی مؤثر شامل سطح ژئومرفیک (۳۹/۴ درصد)، MrVBF (۱۷/۵ درصد)، شاخص رطوبت (۸/۶ درصد)، شاخص رس (۷/۷۹ درصد) و شیب (۵/۵ درصد) شناخته شد (شکل ۴). در بین متغیرهای پیش‌بینی‌کننده، سطح ژئومرفیک یک تخمینگر مهم برای همه افق‌های مشخصه تشخیص داده شد (جدول ۳). نقش مؤثر فرآیندهای ژئومرفولوژیکی در توسعه و تکامل خاک و هم‌بستگی نزدیک ژئومرفولوژی و خاک در بسیاری از مطالعات خاک-ژئومرفولوژی نشان داده شده است (۹).

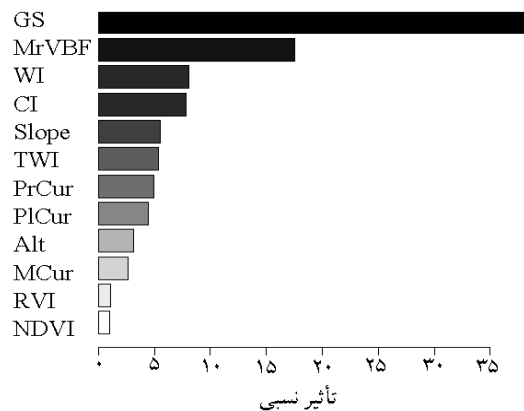
به‌طور مشابه، مدل BRT برای افق‌های مشخصه دیگر برازش داده شد. مقادیر AUC و متغیرهای انتخاب‌شده در پیش‌بینی افق‌های مشخصه توسط مدل‌های BRT در جدول ۳ نشان داده شده است. در مدل BRT افق سالیک، بعد از سطح ژئومرفیک، متغیرهای پیش‌بینی‌کننده MrVBF و WI، جزو مؤثرترین پیش‌بینی‌کننده‌ها بودند (جدول ۳). متغیرهای

MrVBF و WI به‌طور غیرمستقیم، نیروهای پیش‌برنده را در تشکیل افق سالیک نشان می‌دهند (۱۴). شکل ۵ نشان می‌دهد که افزایش مقادیر این متغیرها منطبق بر افزایش احتمال حضور افق سالیک می‌باشد. اجزای سرزمین در پیش‌بینی افق‌های مشخصه توسط مدل BRT، نقش مهم و معنی‌داری را ایفا کردند در حالی که هیچ‌کدام از شاخص‌های سنجش از دور در مدل وارد نشدند. نقش مؤثر اجزای سرزمین در مدل‌سازی خاک - سیمای اراضی و نتایج مناسب پیش‌بینی در مطالعه بهرنز و همکاران (۴) به اثبات رسیده است. هم‌چنین، گیاسن و همکاران (۱۵) توزیع مکانی خاک را با استفاده از رگرسیون لاجیستیک چندگانه در جنوب برزیل مطالعه کردند. نتایج آنها نشان داد اجزای سرزمین شامل ارتفاع، شاخص خیزی و جریان تجمعی از جمله متغیرهای کمکی هستند که توزیع خاک را به خوبی تشریح می‌نمایند.

نقشه‌های شاخص حضور افق‌های مشخصه حاصل از BRT لاجیستیک، مشابه روش رگرسیون لاجیستیک دوتایی به‌دست آمدند. روش BRT هم، احتمال بالای حضور افق سالیک را در پلایا، افق جیپسیک را در تپه‌های گچی، افق کلسیک را در مخروط‌افکنه‌ها، سامنتز را در تپه‌های شنی و ارتنتز را در کوه‌ها پیش‌بینی کرد.

۳. پیش‌بینی مکانی کلاس‌های خاک

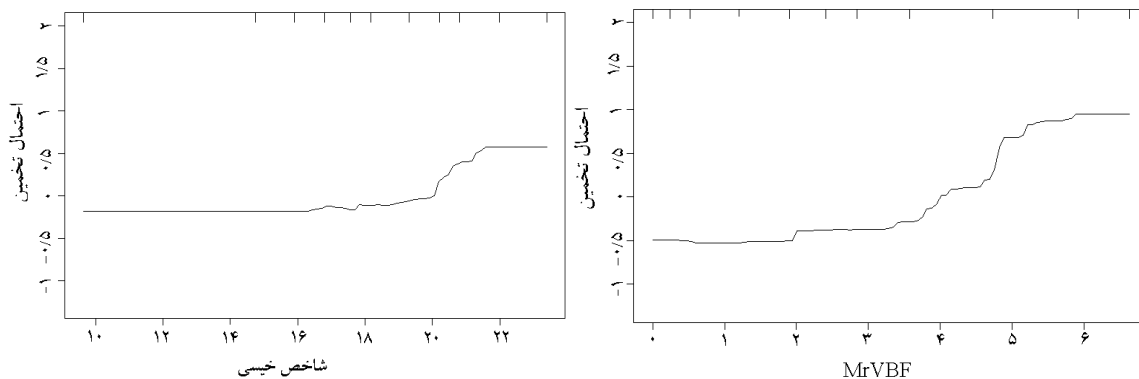
تشریح خاک‌رخ‌ها در منطقه مورد مطالعه هفت گروه بزرگ خاک را نشان داد که شامل هپلوسالیدز (Haplosalids)، هپلوچیپسیدز (Haplogypsid)، هپلوکلسیدز (Haplocalcids)، کلسی‌چیپسیدز (Calcigypsid)، هپلوکمبیدز (Haplocambids)، توری‌سامنتز (Torripsamments) و توری‌ارتنتز (Torriorthents) بود. پیش‌بینی گروه‌های بزرگ خاک از ترکیب احتمال حضور افق‌های مشخصه (حاصل از مرحله قبل) انجام گردید. نقشه‌های گروه بزرگ خاک حاصل از مدل‌های BRT لاجیستیک و رگرسیون دوتایی در شکل ۶ نشان داده شده‌اند. آنچه که از این شکل منتج می‌شود این است که هر دو



شکل ۴. تأثیر نسبی متغیرهای مدل، بسته به سهم هر متغیر در کاهش انحراف کل مدل برای پیش‌بینی افق سالیکی

جدول ۳. متغیرهای انتخاب شده و سطح زیر منحنی ROC در برازش مدل‌های BRT

افق مشخصه یا کلاس خاک	متغیرهای مورد استفاده در مدل	سطح زیر منحنی ROC
سالیکی	GS+ MrVBF+ WI+ CI+ SI+ TWI+ PlCur	۰/۹
جیپسیک	GS+ WI+ CI+ TWI+ PrCur	۰/۸
کلسیک	GS+ WI+ CI+ PrCur	۰/۹
کمبیک	GS+ TWI+ SI+ EI+ PlCur	۰/۸
سامنتز	GS+ MCur	۱
ارتنتز	GS+ SI+ MCur+ PlCur+ EI	۰/۹

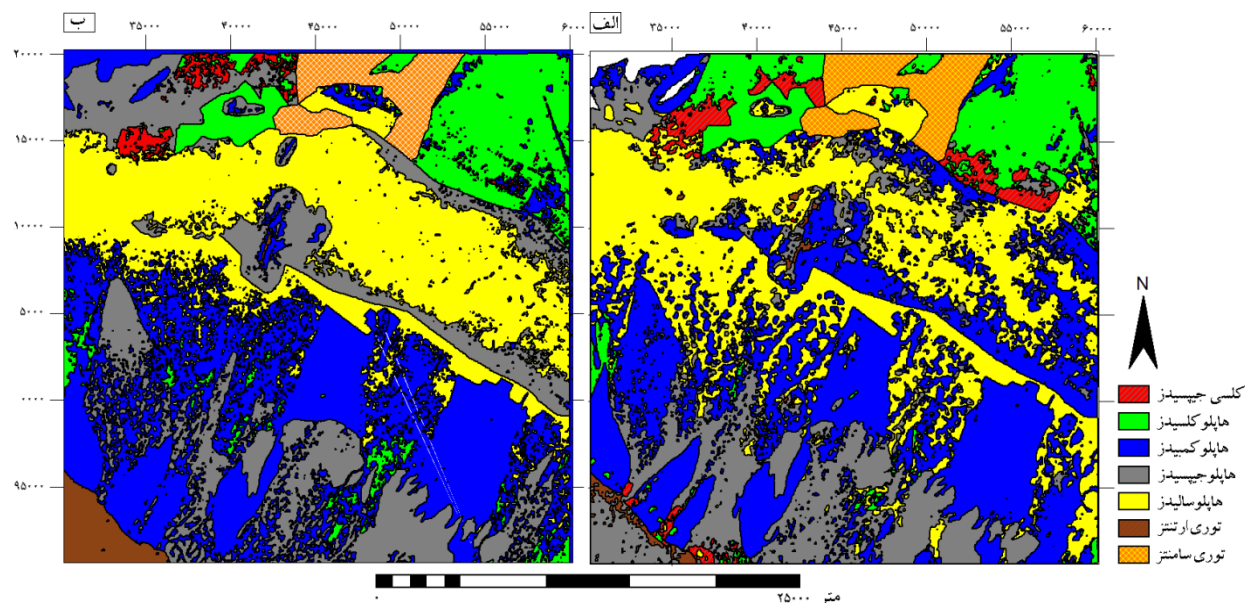


شکل ۵. احتمال حضور افق سالیکی توسط مدل BRT به‌عنوان تابعی از MrVBF و شاخص خیزی

۴. ارزیابی مدل‌های پیش‌بینی‌کننده

جدول ۴ کیفیت پیش‌بینی افق‌های مشخصه حاصل از رگرسیون درختی توسعه‌یافته و رگرسیون لاجیستیک دوتایی را برای داده‌های اعتبارسنجی و واسنجی نشان می‌دهد. خلوص جزئی

مدل، گروه بزرگ "توری‌سامنتز" را در تپه‌های شنی، "هیلوجیپسیدز" را در تپه‌های گچی، "هیلوکلسیدز" را در مخروط‌افکنه‌ها و "هیلوسالیدز" را در پلایا به خوبی پیش‌بینی کردند.



شکل ۶. توزیع مکانی گروه‌های بزرگ خاک پیش‌بینی‌شده توسط روش‌های رگرسیون لاجیستیک دوتایی (الف) و رگرسیون درختی توسعه‌یافته لاجیستیک (ب)

شود. با توجه به حضور کم گروه بزرگ "توری‌ارتنتز" در داده‌های واسنجی، احتمالاً مدل به‌طور مناسب و به خوبی آموزش ندیده و در نتیجه پیش‌بینی ضعیفی را به‌دنبال داشته است. در مورد مدل لاجیستیک دوتایی، دقت پیش‌بینی برای همه افق‌های مشخصه قابل قبول بود چرا که مدل توانسته ارتباط نسبتاً قوی بین پیش‌بینی‌کننده‌ها و افق‌های مشخصه برقرار کند. از طرفی، خلوص نسبی برای هر افق مشخصه در موقعیت‌های اعتبارسنجی بیش از ۷۰٪ بود. این مطلب، انتخاب شایسته متغیرهای پیش‌بینی‌کننده و پتانسیل بالای مدل‌های تخمینگر را نشان می‌دهد. دقت پیش‌بینی برای داده‌های واسنجی بهتر از داده‌های اعتبارسنجی می‌باشد که احتمالاً به تعداد کم داده‌های اعتبارسنجی مربوط می‌شود (جدول ۴).

خلوص نقشه، صحت کاربر و قابلیت اطمینان تولیدکننده برای گروه‌های بزرگ خاک، برای دو مدل (مدل لاجیستیک دوتایی و مدل درختی توسعه یافته) در جدول‌های ۵ و ۶ ارایه شده است. در داده‌های واسنجی خلوص نقشه در هر دو مدل، تفاوت معنی‌داری نداشت و در هر دو مدل، خلوص بالای ۶۰ درصد به‌دست آمد. خلوص (درجه یک‌نواختی) بالا، نتیجه

حاصل از مدل BRT برای همه افق‌های مشخصه در داده‌های واسنجی بیشتر از مدل لاجیستیک دوتایی بود. این نتیجه قابل انتظار است چرا که ساختار سامانه‌های درختی توسعه‌یافته، انعطاف‌پذیرتر و سازگارتر از مدل‌های رگرسیون لاجیستیک می‌باشد. از طرفی، بهترین مدل درختی توسعه‌یافته با تکرار آموزش به‌دست می‌آید که هر مرحله به مقدار انحراف کمتری منجر می‌شود. در فرآیند واسنجی، قسمت اعظم داده‌ها مورد استفاده قرار می‌گیرند. بنابراین منطقی است که مدل، توانایی تشخیص رابطه بین متغیرهای کمکی و متغیر پاسخ را به‌دست آورد و تخمین بهتری را ارایه نماید. در داده‌های اعتبارسنجی، تفاوت قابل ملاحظه‌ای بین رگرسیون درختی توسعه‌یافته و لاجیستیک دوتایی وجود ندارد؛ به استثنای افق کمبیک که مدل لاجیستیک دوتایی، خلوص جزئی بهتری را نسبت به BRT نشان داد (جدول ۴).

با نگاهی به جدول ۴ به نظر می‌رسد که بحث در مورد خلوص گروه بزرگ "توری‌ارتنتز" منصفانه نباشد، زیرا حضور کم این گروه بزرگ در داده‌های اعتبارسنجی، ممکن است که باعث شده باشد پیش‌بینی به‌صورت شانسی و تصادفی انجام

جدول ۴. کیفیت پیش‌بینی مدل‌های رگرسیون درختی توسعه‌یافته و لاجیستیک دوتایی برای افق‌های مشخصه یا کلاس خاک

مجموعه داده	مدل	افق سالیکی	افق جیپسیک	افق کلسیک	افق کمبیک	ارتتیز	سامنتز
واسنجی	رگرسیون درختی	۲۵/۳۱	۴۳/۴۶	۲۲/۳۰	۳۳/۳۸	۷/۸	۵/۵
	توسعه یافته	۰/۸	۰/۹	۰/۷	۰/۸	۰/۸	۱
	نسبت خاک‌رخ‌های صحیح پیش‌بینی شده خلوص جزیی						
اعتبارسنجی	رگرسیون دوتایی	۲۱/۳۱	۴۰/۴۶	۲۰/۳۰	۲۷/۳۸	۵/۸	۵/۵
	لاجیستیک	۰/۷	۰/۸	۰/۶	۰/۷	۰/۶	۱
	نسبت خاک‌رخ‌های صحیح پیش‌بینی شده خلوص جزیی						
اعتبارسنجی	رگرسیون درختی	۵/۷	۱۲/۱۲	۵/۷	۴/۹	۰/۱	۲/۲
	توسعه یافته	۰/۷	۱	۰/۷	۰/۴	۰	۱
	نسبت خاک‌رخ‌های صحیح پیش‌بینی شده خلوص جزیی						
اعتبارسنجی	رگرسیون دوتایی	۶/۷	۱۲/۱۲	۵/۷	۸/۹	۱/۱	۲/۲
	لاجیستیک	۰/۸	۱	۰/۷	۰/۸	۱	۱
	نسبت خاک‌رخ‌های صحیح پیش‌بینی شده خلوص جزیی						

تشخیص روابط بین متغیرها در مدل‌سازی می‌باشد. با این درجه خلوص بالا، به نظر می‌رسد که مدل‌های مورد استفاده، قابلیت کاربرد در تخمین یک متغیر طبقه‌ای (مانند کلاس خاک) را داشته باشند. به‌طور کلی، کیفیت پیش‌بینی مدل‌های مورد استفاده برای داده‌های اعتبارسنجی نسبت به واسنجی کمتر می‌باشد (جدول ۵) که به تعداد کم نمونه‌ها مربوط می‌شود. برای مدل لاجیستیک دوتایی، خلوص نقشه برای همه گروه‌های بزرگ در موقعیت‌های اعتبارسنجی، بالاتر از ۰/۶ بود (جدول ۶)، در مقایسه با مقدار ۰/۷، که مقدار هدف برای نقشه‌برداری سنتی خاک است. گیاسن و همکاران (۱۵) با استفاده از رگرسیون لاجیستیک چندگانه برای نقشه‌برداری رقومی خاک خلوص بالاتر از ۵۰ درصد را به‌دست آوردند. ایلز و همکاران (۱۸) نقشه رقومی خاک با دقت بیش از ۶۰ درصد را با استفاده از روش طبقه‌بندی درختی به‌دست آوردند.

هم‌چنین، جدول‌های ۵ و ۶ نشان می‌دهند که دقت پیش‌بینی برای همه انواع خاک‌ها یکسان نیست. نتایج نشان می‌دهند که ماکزیمم کاهش در عملکرد پیش‌بینی هر دو مدل، مربوط به

پیش‌بینی گروه‌های بزرگ "کلسی جیپسیدز"، "هپلوکلسیدز" و "هپلوکمبیدز" است؛ چرا که تعداد این گروه‌های بزرگ خاک در داده‌های واسنجی و اعتبارسازی کمتر بوده است. در حالی که عملکرد پیش‌بینی برای گروه‌های بزرگ "هپلوسالیدز" و "هپلوچیپسیدز"، تفاوت قابل ملاحظه‌ای با این گروه‌های بزرگ خاک داشت (جدول ۶)، که احتمالاً به تعداد نمونه بیشتر این گروه‌های بزرگ خاک در داده‌های واسنجی و اعتبارسازی مربوط می‌شود. بنابراین، به نظر می‌رسد که مدل‌ها به تعداد نمونه حساس هستند؛ به‌طوری‌که هر چه داده‌های بیشتری برای برازش مدل استفاده شوند، تغییرپذیری کلاس‌های خاک بهتر نشان داده می‌شود و پیش‌بینی کلاس‌های خاک در داده‌های آزمایشی بهتر است. بنابراین، اندازه نمونه، نقش اساسی را در فرآیند ارزیابی دقت ایفا می‌کند (۱۳).

در بین کلاس‌های خاک مورد پیش‌بینی، کلاس "کلسی جیپسیدز"، دقت و قابلیت اطمینان تخمین پایینی را نشان داد (جدول ۵)؛ در صورتی که تخمین کلاس‌های "هپلوسالیدز" و "هپلوچیپسیدز" در هر دو مدل از دقت بالایی برخوردار بود.

جدول ۵. کیفیت پیش‌بینی مدل‌های مورد استفاده برای گروه‌های بزرگ خاک در داده‌های واسنجی

گروه بزرگ خاک			رگرسیون لاجیستیک دوتایی			رگرسیون درختی توسعه یافته		
			MP	UA	PR	MP	UA	PR
کلسی جیپسیدز			۰/۳	۰/۴	۰/۲		۰/۵	۰/۲
هپلوکلسیدز			۰/۷	۰/۶	۰/۸		۰/۷	۰/۸
هپلوکمبیدز			۰/۴	۰/۵	۰/۶		۰/۶	۰/۶
هپلو جیپسیدز			۰/۶	۰/۷	۰/۷	۰/۷	۰/۶	۰/۷
هپلوسالیدز			۰/۷	۰/۶	۰/۸		۰/۶	۰/۸
توری‌ارتتز			۰/۸	۰/۷	۰/۷		۰/۷	۰/۷
توری‌سامتتز			۱	۱	۱		۱	۱

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

جدول ۶. کیفیت پیش‌بینی مدل‌های مورد استفاده برای گروه‌های بزرگ خاک در داده‌های اعتبارسنجی

گروه بزرگ خاک			رگرسیون لاجیستیک دوتایی			رگرسیون درختی توسعه یافته		
			MP	UA	PR	MP	UA	PR
کلسی جیپسیدز			۰	۰	۰		۰	۰
هپلوکلسیدز			۰/۳	۱	۰/۳		۱	۰
هپلوکمبیدز			۰	۰	۰		۰	۰
هپلو جیپسیدز			۰/۶	۱	۰/۶	۰/۵	۰/۴	۱
هپلوسالیدز			۰/۹	۰/۸	۰/۷		۰/۷	۰/۷
توری‌ارتتز			۱	۰/۵	۰		۰	۰
توری‌سامتتز			۱	۱	۱		۱	۱

خصوصیت مشخصه‌ای وجود ندارد که بتواند برای آموزش مدل و در نهایت پیش‌بینی کمک کند (جدول ۶). هنگل و همکاران (۱۶) با استفاده از رگرسیون لاجیستیک چندجمله‌ای، دقت پیش‌بینی ۴ درصد را برای "کمبی‌سولز" به‌دست آوردند در حالی که دقت پیش‌بینی برای جیپسی‌سولز، ۷۱ درصد و کلسی‌سولز، ۷۴ درصد حاصل شد.

قابلیت اطمینان پایین حاصل از پیش‌بینی گروه بزرگ "توری‌ارتتتز" به‌وسیله BRT لاجیستیک در داده‌های اعتبارسنجی کاملاً در شکل ۶ مشخص است، جایی که اکثر خاک‌های "توری‌ارتتتز" به‌صورت "هپلوکمبیدز" طبقه‌بندی شده‌اند (قسمت شمال غربی شکل ۶). با مقایسه نقشه‌های احتمال حضور افق کمبیک و گروه بزرگ "توری‌ارتتتز" (شکل ۶)، در می‌یابیم که احتمالاً گروه‌های بزرگ "توری‌ارتتتز" حاصل از روش لاجیستیک تحت درخت تصمیم‌گیری، اغلب به‌صورت

دقت تخمین پایین گروه بزرگ "کلسی جیپسیدز" احتمالاً به‌دلیل تعداد نمونه کم این کلاس خاک در منطقه مورد مطالعه است. شناسایی خوب روابط متغیرهای پیش‌بینی‌کننده و هدف، در اصل به اندازه و توزیع نمونه در لایه‌ها بستگی دارد. در این منطقه وسیع، فقط پنج نمونه "کلسی جیپسیدز" وجود داشت. بنابراین، دقت پایین مورد انتظار است؛ چرا که مدل نتوانسته است رابطه این کلاس را با متغیرهای محیطی برقرار کند و تشخیص مقادیر حدود آستانه برای طبقه‌بندی کلاس‌های خاک را دچار مشکل می‌کند و در نتیجه، مدل آموزش‌دیده ضعیف عمل می‌کند. بالطبع، نتایج ضعیف آموزش، نتایج ضعیف اعتبارسنجی را به‌دنبال خواهد داشت و این در مورد گروه بزرگ "کلسی جیپسیدز" کاملاً صادق است. از طرفی، در مورد گروه بزرگ "هپلوکمبیدز" احتمال دارد که خلوص و دقت پایین پیش‌بینی، نتیجه ناقص بودن مدل مفهومی باشد؛ چرا که هیچ

"هپلوکمبیدز" پیش‌بینی شده‌اند (شکل ۶). بنابراین، خطای حاصل از روش‌های لاجیستیک تا حدودی از درخت تصمیم‌گیری ناشی می‌شود که حضور افق‌های مشخصه را به گروه‌های بزرگ خاک ترجمه می‌کند.

در بین گروه‌های بزرگ خاک، بهترین پیش‌بینی برای "هپلوسالیدز" در هر دو مدل به‌دست آمد. مقادیر بالای صحت کاربر و قابلیت اطمینان تولیدکننده در جدول‌های ۵ و ۶ مؤید این مطلب است. دِلاگیلو و اِترلمولر (۹) نشان دادند که نواحی با احتمال پیش‌بینی بالا برای گروه بزرگ خاک با فرآیندهای هر سیمای اراضی منطبق هستند. بنابراین، تشخیص متغیرهای توصیفی و داشتن ارتباط منطقی با متغیر هدف از اهمیت بالایی برخوردار است. با افزایش تعداد نمونه و استفاده از متغیرهای توصیفی مناسب، تشخیص ارتباط بین متغیرهای توصیفی و متغیر هدف توسط مدل‌ها امکان‌پذیر می‌شود و در نتیجه، سهم و تأثیر شانس به‌طور خودبخود در پیش‌بینی کاهش می‌یابد. به نظر می‌رسد که خطای پیش‌بینی در این روش‌ها، در درجه اول، ناشی از عدم وجود مدل مفهومی مناسب به‌دلیل عدم حضور پایگاه داده کافی و در درجه دوم ناشی از درخت تصمیم‌گیری

است که حضور افق‌های مشخصه را به گروه‌های بزرگ خاک ترجمه می‌کند.

نتیجه‌گیری

نتایج پیش‌بینی کلاس‌های خاک در منطقه خشک زرنند استان کرمان به‌وسیله تکنیک‌های نقشه‌برداری رقومی نشان داد که نقشه ژئومرفولوژی، یک ابزار مهم برای روش‌های نقشه‌برداری رقومی خاک است که به افزایش دقت پیش‌بینی کمک می‌کند. فرآیندهای ژئومرفولوژی به همراه پارامترهای سرزمین، توسعه و تکامل خاک را به‌ویژه در مناطق خشک تحت تأثیر قرار می‌دهند. حضور اجزای سرزمین به همراه فرآیندهای ژئومرفیک در پیش‌بینی کلاس‌های خاک نشان داد که این فاکتورها برآیند و نتیجه نیروهای مؤثر خاکسازي در منطقه مورد مطالعه هستند. زمانی که همه فاکتورهای خاکسازي به‌طور هم‌زمان در روش مدل‌سازی استفاده شوند، بهترین پیش‌بینی‌ها حاصل می‌شود. در صورت شناخت کامل منطقه مورد بررسی و انتخاب مناسب متغیرهای کمکی، مدل‌های رگرسیون لاجیستیک، پتانسیل بالایی در پیش‌بینی کلاس‌های خاک خواهند داشت.

منابع مورد استفاده

۱. سازمان هواشناسی کشور. ۱۳۸۷. آمار هواشناسی شهرستان زرنند، استان کرمان. پایگاه اینترنتی <http://www.irimo.ir>.
2. Abeare, S. 2009. Comparison of boosted regression tree, GLM and GAM performance in the standardization of yellowfin tuna catch-rate data from the Gulf of Mexico longline fishery. MSc. Thesis, Louisiana State University.
3. Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. PP. 267-281. In: B.N. Petrov and F. Csaki (Eds.), 2nd Intl. Symposium on Information Theory, Akademia, Kiado, Budapest.
4. Aksoy, E., G. Ozsoy and M. Sabri Dirim. 2009. Soil mapping approach in GIS using Landsat satellite imagery and DEM data. African J. Agric. Res. 4(11): 1295-1302.
5. Behrens, T., A. X. Zhu, K. Schmidt and T. Scholten. 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. Geoderma 155: 175-185.
6. Bell, J. C., R. L. Cunningham and M. W. Havens. 1992. Calibration and validation of a soil-landscape model for predicting soil drainage class. Soil Sci. Soc. Amer. J. 56: 1860-1866.
7. Bragato, G. 2004. Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain. Geoderma 118: 1-16.
8. Campling, P., A. Gobin and J. Feyen. 2002. Logistic modeling to spatially predict the probability of soil drainage classes. Soil Sci. Soc. Amer. J. 66: 1390-1401.
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 Vestfold County, Norway. Catena 77: 8-18.
10. Deventer, van A. P., A. D. Ward, P. H. Gowda and J. G. Lyon. 1997. Using thematic mapper data to identify contrasting soil plains and tillage practices. Photo. Eng. Remote Sens. 63(1): 87-93.

11. Dobos, E., L. Montanarella, T. Negre and E. Micheli. 2001. A regional scale soil mapping approach using integrated AVHRR and DEM data. *Intl. J. Appl. Earth Obs.* 3(1): 30-42.
12. Elith, J., J. R. Leathwick and T. Hastie. 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77: 802-813.
13. Foody, G. M. 2002. Status of land covers classification accuracy assessment. *Remote Sens. Environ.* 80: 185-201.
14. Gallant, J. C. and T. I. Dowling. 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Res. Res.* 39(12): 1347-1359.
15. Giassen, E., R. T. Clarke, A. V. I. Junior, G. H. Merten and C. G. Tornquist. 2006. Digital soil mapping using multiple logistic regression on terrain parameters in Southern Brazil. *Sci. Agric.* 63: 262-268.
16. 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.
17. Hosmer, D. W. and S. Lemeshow. 2000. *Applied Logistic Regression*. John Wiley & Sons, Inc., New York.
18. Illes, C., G. Kovacs and B. Heil. 2011. Comparing and evaluating digital soil mapping methods in a Hungarian forest reserve. *Can. J. Soil Sci.* 91: 615-626.
19. Isbell, R. F. 1996. *The Australian Soil Classification*. CSIRO Pub. Co., Collingwood, Victoria.
20. IUSS Working Group WRB. 2006. *World Reference Base for Soil Resources, A framework for International Classification, Correlation and Communication*. World Soil Resources Reports No. 103. FAO, Rome.
21. Lagacherie, P. 2005. An algorithm for fuzzy pattern matching to allocate soil individuals to pre-existing soil classes. *Geoderma* 128: 274-288.
22. Lagacherie, P. and S. Holmes. 1997. Addressing geographical data errors in a classification tree soil unit prediction. *Intl. J. Geog. Inform. Sci.* 11: 183-198.
23. McBratney, A. B., M. L. Mendonça-Santos and B. Minasny. 2003. On digital soil mapping. *Geoderma* 117: 3-52.
24. Moameni, A. and J. A. Zink. 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, CD Rom.
25. Olaya, V. 2004. A gentle introduction to Saga GIS user's manual, The SAGA User Group, Göttingen.
26. Pearson, R. L. and L. D. Miller. 1972. Remote mapping of standing crop biomass for estimation of the productivity of the short-grass Prairie, Pawnee National Grasslands, Colorado. PP. 1357-1381 *In: Proceedings of the 8th International Symposium on Remote Sensing of Environment*, ERIM International Inc., Ann Arbor, Michigan, USA.
27. Richardson, A. J. and C. L. Wiegand. 1977. Distinguishing vegetation from soil background information. *Photo. Eng. Remote Sens.* 43(12): 1541-1552.
28. Ridgeway, G. 2009. Package gbm. *R-News*. 08:05:15. 8.
29. Rouse, J. W., R. H. Hass, J. A. Schell and D. W. Deering. 1973. Monitoring vegetation systems in the Great Plains with ERTS. PP. 309-317. *In: 3rd ERTS Symp.*, Washington, DC.
30. Soil Survey Staff. 2010. *Keys to Soil Taxonomy*. 11th ed., NRCS, USDA.
31. 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. <http://www.gdem.aster.ersdac.or.jp>.
32. Zhu, A. X. 2000. Mapping soil landscape as spatial continua: the neural network approach. *Water Res. Res.* 36: 663-677.