

Soil Distribution Pattern Analysis in a Low Relief Area Using Decision Trees Algorithm

MARZIYE ZANDI BAGHCHE-MARYAM^{1*}, PARVIZ SHEKAARI²

1. Senior Expert, Department of Soil Science and Engineering, Faculty of Agriculture, Razi University, Kermanshah, Iran

2. Assistant Professor, Department of Soil Science and Engineering, Faculty of Agriculture, Razi University, Kermanshah, Iran

(Received: Jan. 20, 2018- Revised: Sep. 11, 2018- Accepted: Sep. 29, 2018)

ABSTRACT

Digital soil mapping (DSM) can be defined as a production of spatial soil information. Decision tree (DT) algorithm is one of the most popular machine learning methods which was applied in several recent DSM studies. This study was carried out to evaluate the capability of DT in mapping soils in Miandarband region with area of 50,000 ha in Kermanshah province. The C5.0 decision tree algorithm (with and without boosting meta-algorithm) used to establish spatial relationships between known soil taxonomic classes and environmental variables. Using simple systematic sampling, 78 pedons were studied and 6 great groups and 14 subgroups of Soil Taxonomy (ST) were identified. Thirty environmental items were derived from a digital elevation model (DEM) file and a landsat-8 OLI/TIRS (July/Tir 1394) image of the area. Predictions made by C5.0 algorithm showed OA values of 73 percent for great group and subgroup, while comparable values for Kappa Index were 0.61 and 0.63, respectively. Combination of boosting meta-algorithm with C5.0 increased OA values for ST categories 0.80 and 0.76 and Kappa Index values to 72 percent and 66 percent. Results showed a considerable capability for DT in recognition of soil pattern over the study area and the topographic variables seems to be most important. Also, analysis of the produced maps, compared with the observed soil pattern during the field survey, revealed a reasonable agreement of decision tree algorithm predictions with reality.

Keyword: Digital Soil Mapping, C5.0 algorithm, Boosting, Environmental Covariates, Miandaband Plain.

تحلیل الگوی پراکنش خاک‌ها در یک منطقه هموار با استفاده از الگوریتم درخت تصمیم‌گیری

مرضیه زندی باغچه‌مریم^{۱*}، پرویز شکاری^۲

۱. کارشناسی ارشد گروه علوم و مهندسی خاک-دانشکده کشاورزی-دانشگاه رازی-کرمانشاه-ایران

۲. استادیار، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه رازی، کرمانشاه، ایران

(تاریخ دریافت: ۱۳۹۶/۱۰/۳۰ - تاریخ بازنگری: ۱۳۹۷/۶/۲۰ - تاریخ تصویب: ۱۳۹۷/۷/۷)

چکیده

نقشه‌برداری رقومی خاک را می‌توان تولید اطلاعات مکانی خاک تعریف کرد. یکی از روش‌های محبوب که اخیراً در چندین مورد از مطالعات نقشه‌برداری رقومی خاک به کار رفته، درخت تصمیم‌گیری است. پژوهش حاضر به منظور ارزیابی قابلیت درخت تصمیم‌گیری در نقشه‌برداری خاک‌ها در منطقه میان‌دربند با مساحت ۵۰۰۰۰ هکتار در استان کرمانشاه انجام شد. الگوریتم C5.0 (با و بدون متالگوریتم بوستینگ) برای ایجاد روابط مکانی بین کلاس‌های خاک و متغیرهای محیطی مورد استفاده قرار گرفت. بر پایه نمونه‌برداری سیستماتیک ۷۸ خاک‌رخ مورد مطالعه قرار گرفت و ۶ گروه بزرگ و ۱۴ زیرگروه شناسایی شد. ۳۰ متغیر محیطی از مدل رقومی ارتفاع و تصویر سنجنده OLI/TIRS ماهواره لندست ۸ مربوط به تاریخ تیرماه ۱۳۹۴ مشتق شد. صحت عمومی برای گروه بزرگ و زیرگروه برابر با ۷۳ درصد به دست آمد در حالی که مقادیر متناظر برای نمایه کاپا به ترتیب ۰/۶۱ و ۰/۶۳ بود. ترکیب متالگوریتم بوستینگ با C5.0 مقادیر صحت عمومی را به ترتیب به ۸۰ درصد و ۷۶ درصد و مقادیر نمایه کاپا را به ۰/۷۲ و ۰/۶۶ افزایش داد. نتایج توانایی قابل توجهی را برای درخت تصمیم‌گیری در باز شناخت الگوی خاک در منطقه مورد مطالعه نشان داد و متغیرهای توپوگرافی از سایر متغیرهای محیطی پر اهمیت‌تر به نظر می‌رسید. همچنین، بررسی نقشه‌های تولیدشده از طریق مقایسه با الگوی خاک مشاهده‌شده در خلال بررسی زمین، نشان‌گر تطابق پذیرفتنی پیش‌بینی‌های الگوریتم درخت تصمیم‌گیری با واقعیت بود.

واژه‌های کلیدی: نقشه‌برداری رقومی خاک، الگوریتم C5.0، بوستینگ، متغیرهای محیطی، دشت میان‌دربند.

مقدمه

اهمیت شناسایی خاک از نقش پایه‌ای آن در تصمیم‌گیری‌های مربوط به ارزیابی زیست‌بوم‌ها، مدیریت سرزمین و برنامه‌ریزی پروژه‌های آب‌شناختی، صنعتی، ساخت‌وسازها، کشاورزی و ... سرچشمه می‌گیرد (Hash, 2008). از جمله اینکه ارزیابی خاک و سرزمین بر اساس نقشه خاک و واحدهای نقشه امکان‌پذیر می‌شود (Daigle et al., 2005). تولید داده‌های دقیق برای اطلاع‌رسانی توصیه‌های مدیریتی با هدف افزایش بهره‌وری در کشاورزی و امنیت غذایی برای کشورهای در حال توسعه که در آنها تولید ناخالص ملی به بخش کشاورزی وابستگی دارد بسیار مهم است. (Cook et al., 2008; Msanya et al., 2002). مدل‌های زیست محیطی، اجتماعی و اقتصادی به ویژگی‌های خاک برای برآورد و پیش‌بینی تغییرات در شرایط زندگی و آینده آدمی نیازمند هستند. تغییرپذیری خاک امری تصادفی نیست و نتیجه‌ی تغییر عوامل خاک‌سازی از نقطه‌ای به نقطه دیگر است. در واقع تفاوت عوامل خاک‌سازی در مناطق مختلف فرآیندهای

خاک‌سازی را متأثر ساخته و موجبات تغییر خاک را فراهم می‌کند. این فاکتورها می‌توانند در پیش‌بینی کلاس خاک و تهیه نقشه الگوی پراکنش خاک‌ها به کار رود. اولین گام برای مدیریت خاک تهیه نقشه پراکنش جغرافیایی خاک‌ها در سراسر زمین با نمایش سطح قابل‌اعتمادی از جزئیات مفید است (Nauman, 2009). برای پاسخ به سؤالات مربوط به پویایی شاخص‌های کلیدی سلامت زیست‌بوم در زمان و مکان به مدل‌ها و داده‌هایی نیاز داریم که از صحت و اعتبار کافی برخوردار باشند. این امر به‌ویژه در شرایط حاضر مهم است زیرا انسان به‌طور روزافزونی در حال تغییر محیط‌زیست و خاک می‌باشد. در نتیجه نیازمند مدل‌ها و نقشه‌هایی هستیم که اثرات انسان بر خاک را مدیریت کند (Grunwald et al., 2011). اما اطلاعات موجود و نقشه‌های خاک سنتی یا در مقیاس مناسب وجود ندارند و یا کیفیت داده‌های آنها سؤال برانگیز است. از طرفی هم قابلیت به‌روزرسانی محدود داشته و یا ندارند. در نتیجه ایجاد نقشه‌های جدید که کاستی‌های گذشته را نداشته باشد و توانایی آشکار کردن پراکنش مکانی و تغییرات خاک را دارا باشد ضروری است.

* نویسنده مسئول: mzandi8867@gmail.com

عملاً مرجع اطلاعات ارزشمند مورداستفاده توسط مدیران و تصمیم‌گیران منابع است (Elnaggar, 2007).

با پیشرفت‌های اخیر در زمینه فن‌آوری اطلاعات مقادیر زیادی از اطلاعات به سرعت در همه زمینه‌ها ایجاد شد و علوم خاک نیز از این قاعده مستثنی نیست. چالش درک، دسته‌بندی و استفاده‌ی بهینه از این اطلاعات منجر به ایجاد و توسعه ابزارهای جدید مانند داده‌کاوی و رایان‌آموزی شده است. هم‌زمان با آن توسعه‌ی ابزارهای سیستم اطلاعات جغرافیایی، سنجش‌ازدور و مدل‌های رقومی ارتفاع با درجه تفکیک مکانی بالا منجر به پیشرفت و یک حرکت روبه‌جلو در دانش نقشه‌برداری خاک شده است (McBratney *et al.*, 2003; Malone, 2013).

(McBratney *et al.*, 2003) یک چارچوب کلی به نام تابع پیش‌بینی مکانی خاک با خطاهای همبسته مکانی (SSPFe) به‌عنوان یک روش برای نقشه‌برداری رقومی خاک تعریف کردند. این تابع یک مدل برای توصیف کمی روابط خاک و متغیرهای محیطی تشکیل‌دهنده‌ی خاک است. از این روابط به‌عنوان یک تابع برای پیش‌بینی مکانی کلاس یا ویژگی‌های خاک به هدف تولید نقشه رقومی خاک استفاده می‌شود. این تئوری برای اولین بار توسط مطالعات واسیلی داکوچائف در سال ۱۸۸۳ درباره‌ی پراکنش خاک و ارزیابی اراضی در استپ‌های روسیه ارائه شد. بعدها هانس‌ینی در سال ۱۹۴۱ مدل داکوچائف و نقشه‌برداری توسط فاکتورهای محیطی را تعمیم داد و آن را در کتاب مشهور خود به‌طور کامل شرح داد (Hash, 2008).

تئوری عامل حالت ینی به‌صورت زیر است:

$$S = f(cl, o, r, p, t) \quad (\text{رابطه ۱})$$

که در آن:

(S) کلاس خاک، (cl) اقلیم، (o) موجودات زنده، (r) پستی‌وبلندی، (p) مواد مادری، (t) زمان را نشان می‌دهند. بر اساس تئوری عامل حالت ینی، خاک به‌عنوان یک متغیر وابسته تابعی از متغیرهای مستقل مواد مادری، اقلیم، موجودات زنده و پستی‌وبلندی در طول زمان است. این معادله خاک را تابعی از عوامل خاک‌سازی در یک مکان معین می‌داند اما پیوستگی خاک را که به‌موجب آن، خاک در هر مکان وابسته به موقعیت جغرافیایی و ویژگی‌های خاک واقع در همسایگی خود است را در نظر نمی‌گیرد. تئوری عامل حالت اساس نقشه-برداری سنتی خاک است و نقشه‌های کلروپت یا پلیگونی را

نقشه‌های پلیگونی بر اساس دیدگاه گسسته استوار است. این دیدگاه منجر به ایجاد نقشه‌های چند ضلعی شده به‌طوری‌که خاک‌ها را درون واحد نقشه یکنواخت در نظر می‌گیرد. درواقع نتایج مشاهدات خاکرخ شاهد به کل نقشه تعمیم داده شده و تأثیر عوامل محیطی بر تغییرات تدریجی خاک در نظر گرفته نمی‌شود، بلکه تغییرات تنها در محل مرزهای واحد نقشه رخ می‌دهد. آشکار است که مرزهای کشیده شده بر روی نقشه ممکن است واقعاً نشان‌دهنده‌ی مرزهای بین واحدهای نقشه نباشد (Mark and Csillag, 1989). بر این اساس نتیجه-گیری یکسانی برای بسیاری از ویژگی‌های فیزیکی و شیمیایی مؤثر در مدل‌سازی‌های محیط زیست و مدیریت خاک وجود ندارد. درواقع در نقشه‌های سنتی الگوهای پراکنش مکانی خاک در درون مرزهای گسسته‌ی واحد نقشه محصور شده‌اند که به معنی همگن بودن کلاس خاک در هر واحد نقشه است (Elnaggar, 2007).

علاوه بر آن مدل مفهومی خاک-سرزمین در مناطق نقشه‌برداری شده مشخص نیستند چون اولاً مفهومی هستند و ثانیاً بیشتر اطلاعات آنها مستند نشده است. در نتیجه امکان انتقال دانش و تجربه‌ی کسب‌شده به نسل بعد امکان‌پذیر نیست. بدین منظور نقشه‌برداری رقومی خاک (DSM) در پاسخ به نقد نقشه‌برداری سنتی و به‌عنوان جایگزین بهتر تکنیک‌های سنتی نقشه‌برداری توسعه یافت. بر همگان روشن است که نقشه‌برداری وابسته به فهم روابط میان خصوصیات خاک و عوامل محیطی در یک زمین‌نمای پیچیده است. نقشه‌برداری رقومی را می‌توان توسعه‌ی یک مدل کمی یا آماری از ارتباط متغیرهای محیطی و ویژگی‌های خاک تعریف کرد که با اعمال یک پایگاه داده‌های جغرافیایی منجر به ایجاد نقشه می‌شوند (Franklin, 1995; McBratney *et al.*, 2000). نقشه‌برداری رقومی خاک (DSM) بیانگر یک مجموعه از محاسبات رایانه‌ای به‌منظور پیش‌بینی پراکنش مکانی خاک‌ها در یک زمین‌چهر است که به‌عنوان ابزاری برای تولید اطلاعات مکانی داده‌های خاک و یا رفع نیاز به نقشه‌های با صحت و دقت معتبر و قدرت تفکیک مکانی بالا توسعه‌یافته است (Brungard *et al.*, 2015).

نقشه‌برداری رقومی خاک (DSM) توان بسیار زیادی برای برقراری رابطه میان متغیرهای محیطی و ویژگی‌های خاک برای بهبود دقت و صحت مکانی شناسایی خاک همراه با کاهش هزینه، زمان و افزایش بهره‌وری خاک دارد (Bui *et al.*, 2006; Cole and Boettinger *et al.*, 2006; Henderson *et al.*, 2005). نیروی رانش اصلی نقشه‌برداری رقومی خاک (DSM) کمک به مدیران و برنامه‌ریزان است زیرا نقشه‌های رقومی خاک اکنون

ایجاد می‌کند که برای بسیاری از برنامه‌های کاربردی کارایی دارند. اما برای کسانی که نیاز به گزارش‌های کمی و همین‌طور نمایش مکانی واقعی و درست دارند کافی نمی‌باشد و نیاز به نقشه‌هایی هست که توان ارائه داده‌های خاک را داشته باشند. سرانجام (McBratney *et al.*, 2003) مفاهیم و روش‌های DSM را به صورت گسترده مورد بررسی قرار دادند و نهایتاً مدل اسکورپن را به عنوان چارچوب نقشه‌برداری رقومی خاک ارائه دادند.

مدل اسکورپن به صورت زیر می‌باشد:

$$S = f(s, c, o, r, p, a, n) + e \quad (\text{رابطه ۲})$$

$$S = f(Q) + e \quad (\text{رابطه ۳})$$

که در آن:

(S) کلاس یا ویژگی خاک، (s) ویژگی‌های اندازه‌گیری شده خاک در یک نقطه، (c) ویژگی‌های آب و هوایی خاک در یک نقطه، (o) موجودات زنده شامل پوشش گیاهی طبیعی، جانوران و یا فعالیت‌های انسانی، (r) ویژگی‌های توپوگرافی و زمین‌شناسی، (p) مواد مادری و سنگ‌شناسی، (a) عامل زمان و (n) مکان یا موقعیت جغرافیایی یا همبستگی مکانی، (e) تغییرات مکانی تصادفی همبسته است.

f(Q): یک تابع کمی است که S (کلاس یا ویژگی خاک)

را به فاکتورهای اسکورپن پیوند می‌دهد (Lagacherie and McBratney, 2006).

انگیزه‌ی پیشرفت سریع و اجرای پروژه‌های نقشه‌برداری رقومی خاک در سراسر جهان، ایجاد زیرساخت داده‌های مکانی با هدف آسان‌سازی جمع‌آوری، نگهداری، انتشار و استفاده از اطلاعات مکانی بوده است (Malone, 2013). علوم خاک از طریق اطلاعات موجود در داده‌های خاک به طور بالقوه‌ای موجب توسعه‌ی زیرساخت داده‌های خاک به صورت منطقه‌ای، قاره‌ای و جهانی شده است (Malone, 2013). در یک مرور کلی، Heung *et al.*, 2016 با مقایسه روش‌های رایان‌آموزی برای اهداف طبقه‌بندی در نقشه‌برداری رقومی خاک مجموعه‌ای از ده الگوریتم را برای پیش‌بینی در سطح رده و گروه بزرگ خاک در پائین دره فریزر در کلمبیا به کار بردند. آنها از ۲۰ متغیر محیطی ارائه‌شده از توپوگرافی، آب‌وهوا و پوشش گیاهی منطقه مورد مطالعه برای مدل‌سازی استفاده کردند. آنها نتیجه گرفتند که روش k نزدیک‌ترین همسایگی و ماشین‌های بردار پشتیبان مبتنی بر توابع پایه شعاعی جزء دقیق‌ترین روش‌ها بود و به دنبال آن مدل‌های مبتنی بر درخت دقت بالاتری داشتند. Nelson and Odeh (2009) در نقشه‌برداری رقومی خاک با استفاده از مشاهدات خاکرخ یک بررسی مرسوم، به مقایسه

کارایی الگوریتم ژنتیک^۱ و درخت تصمیم‌گیری پرداختند. آنها نقشه رقومی کلاس خاک را برای حوضه ناموئی در استرالیا به مساحت ۴۲۰۰۰ کیلومترمربع تهیه کردند و نتیجه گرفتند که الگوریتم ژنتیک کارایی بالاتری نسبت به درخت تصمیم‌گیری داشته است. Subburayalu *et al.* (2014) برای جداسازی سری خاک‌ها در نقشه شناسایی خاک در شهرستان مونرو در جنوب شرقی اوهایو به این نتیجه رسیدند که خوشه‌بندی مبتنی بر درخت تصمیم‌گیری دارای دقت بالاتری بود. (Lu *et al.* 2016) برای پیش‌بینی درجه ضخامت خاک در یک حوضه کوهستانی از روش خوشه‌بندی فازی برای طبقه‌بندی زمین‌چهر و از الگوریتم C5.0 برای پیش‌بینی و به دست آوردن رابطه متغیرهای محیطی و خاک استفاده کردند و صحت عمومی و نمایه سازگاری کاپا به ترتیب برابر ۷۴/۲ و ۰/۶۵ به دست آمد. در ایران شمار پژوهش‌های مرتبط با نقشه‌برداری رقومی خاک بسیار کم بوده که تنها در سال‌های اخیر انجام شده‌اند. از جمله Jafari *et al.* (2013) در مطالعه‌ای به منظور نقشه‌برداری رقومی خاک در منطقه‌ای به مساحت ۹۰ هزار هکتار از مدل‌های رگرسیونی لجستیک دودویی و رگرسیون درختی در منطقه زرد کرمان بهره بردند. (pahlavan rad *et al.* 2015a) برای پیش-بینی کلاس‌های خاک در منطقه‌ای با وسعت ۸۵۰۰۰ هکتار از درخت تصمیم‌گیری تصادفی در استان گلستان بهره گرفتند. (Nabiollahi *et al.* 2014) برای نقشه‌برداری رقومی بافت خاک در منطقه‌ای با وسعت ۲۰۰۰۰ هکتار در استان کردستان از روش رگرسیون درختی و شبکه عصبی مصنوعی استفاده کردند. (Taghizadeh-Mehrjardi *et al.* 2015) به منظور تهیه نقشه رقومی بافت خاک در منطقه‌ای به مساحت ۷۲۰۰۰ هکتار در اردکان استان یزد اقدام به مقایسه درخت تصمیم‌گیری و شبکه عصبی مصنوعی کردند. به این منظور بر اساس روش نمونه-برداری ابر مربع لاتین مشروط مکان ۱۷۳ خاکرخ برای نمونه-برداری تعیین شد. متغیرهای محیطی مورد استفاده شامل اجزاء سرزمین، داده‌های تصاویر Landsat7 ETM و نقشه سطوح ژئومورفولوژی بود. نتایج نشان داد که مدل درخت تصمیم‌گیری نسبت به شبکه عصبی ۷ درصد دقیق‌تر بوده و تفسیر آسان‌تری هم دارند.

در کل نتایج نشان داد که درخت تصمیم‌گیری به طور موفقیت‌آمیزی در پیش‌بینی نقشه‌های خاک عمل کرده است. در استان کرمانشاه نقشه‌های خاک، یا سنتی هستند و یا به طور

شامل چهار واحد فیزیوگرافی مخروط افکنه‌های آبرفتی سنگریزه‌دار، دشت‌های دامنه‌ای، دشت‌های آبرفتی رودخانه‌ای و دشت پست است (Banaii, 1977). این دشت با ارتفاع متوسط ۱۳۵۷/۷ متر از سطح دریا دارای آب‌وهوای معتدل با زمستان-های سرد است. میانگین سالیانه دما و بارش به ترتیب برابر با ۱۴° و ۴۶۲/۶ میلی‌متر و رژیم رطوبتی و دمایی خاک منطقه به ترتیب زیریک و ترمیک است (Banaii, 1977). گونه‌های گیاهی چیره در منطقه شامل شیرین‌بیان و خارزرد است که نشان-دهنده‌ی حاصلخیزی فراوان است. در دامنه‌ها و مخروط افکنه‌ها که حاصلخیزی کمتری دارند خارشتر و تلخ‌بیان وجود دارد. کشاورزی در منطقه شامل کشت آبی و دیم است. محصولات منطقه عبارت از گندم و جو، حبوبات، صیفی‌جات، سبزیجات، چغندرقد، ذرت، شبدر، یونجه، باغات میوه و باغات غیر مثمر است. در شکل (۲) نقشه باغات و زمین‌های کشاورزی و همچنین مسیر رودخانه رازآور و در شکل (۳) نقشه زمین‌شناسی منطقه مطالعاتی موجود است. دشت میان‌دریبد در اکثر قسمت‌ها توسط کوهستان احاطه شده است. همچنین بخشی از این ارتفاعات از قسمت شمال غربی وارد پهنه خاکی دشت شده و قسمت بالایی پهنه خاکی منطقه را به دو قسمت تقسیم کرده است. علاوه بر آن در قسمت‌های میانی نیز ارتفاعات سلطان کوه قرار دارد. تأثیر این ارتفاعات بر متغیرهای محیطی حاکم در منطقه و عملکرد آنها و همچنین همواربودن منطقه و تغییرپذیری کم در متغیرهای توپوگرافی و عدم تنوع در زمین‌شناسی پهنه خاکی موجب شد که برای استخراج روابط خاک-سرزمین در منطقه مطالعاتی علاوه بر پهنه خاکی قسمت کوهستانی منطقه نیز وارد جریان مدل‌سازی شود. به امید اینکه مدلی واقعی‌تر از روابط خاک-سرزمین در منطقه مطالعاتی برای پیش‌بینی پراکنش خاک‌ها حاصل شود.

کامل تهیه نشده‌اند. دشت میان‌دریبد جزء قطب‌های مهم کشاورزی استان است که دارای نقشه سنتی مربوط به سال ۱۳۶۷ است. با توجه به اینکه قسمت عمده زمین‌های این دشت زیر کاربری کشاورزی و باغداری است به‌منظور مدیریت بهینه، کاربری صحیح اراضی و حفظ منابع آب‌و خاک حاصلخیز دشت که لازمه‌ی توسعه‌ی پایدار هستند، نیاز به نقشه‌های خاک با دقت بالا است که اطلاعات جامع و قابل‌اعتمادی را در اختیار مدیران و برنامه‌ریزان قرار دهد. در این راستا، با انجام این پروژه تلاش بر آن است که با دستیابی به نقشه رقوم‌ی خاک‌ها اطلاعات دقیق‌تری از منطقه مورد مطالعه به دست آید. با توجه به اینکه کاربرد روش‌های نوین نقشه‌برداری خاک موجب کسب آگاهی بیشتر در مورد خاک، ویژگی‌های آن و آشکار شدن الگوی پراکنش خاک‌ها می‌شود پیگیری این‌گونه مطالعات موجب توانمند شدن تصمیم‌گیری‌ها در برنامه‌ریزی و مدیریت بهینه سرزمین خواهد شد.

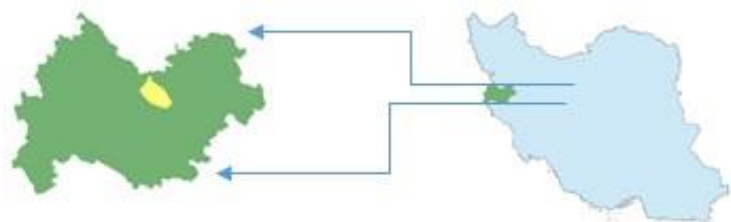
هدف از پژوهش حاضر عبارت است از: (۱) تهیه نقشه رقوم‌ی از خاک‌ها در سطح گروه بزرگ و زیرگروه. (۲) مقایسه عملکرد الگوریتم C5.0 با و بدون روش بوستینگ برای نقشه-برداری رقوم‌ی خاک. (۳) تعیین مهم‌ترین متغیرهای محیطی مؤثر بر تمایز خاک‌ها در سطح منطقه.

مواد و روش‌ها

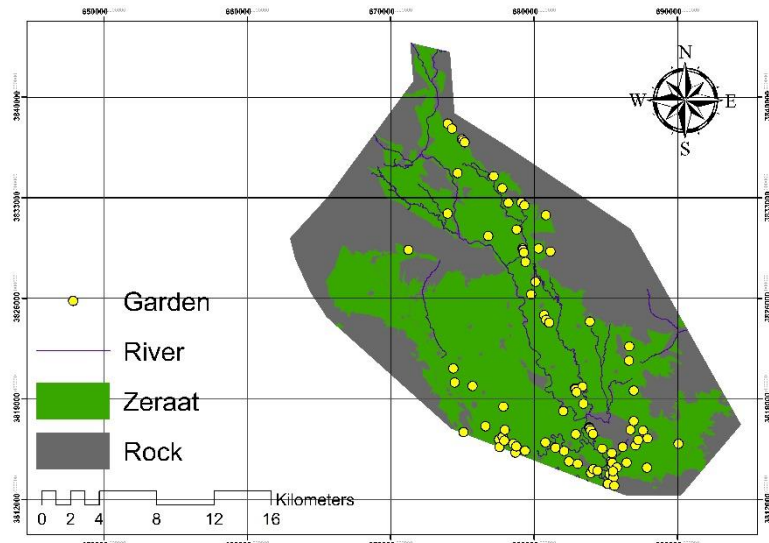
منطقه مطالعاتی

منطقه مورد مطالعه (شکل ۱) دشت میان‌دریبد واقع در استان کرمانشاه است که مساحت حدود ۵۰۰۰۰ هکتار را در برمی-گیرد. این منطقه در شمال غربی شهر کرمانشاه بین طول جغرافیایی ۴۵° ۴۶' تا ۴۷° ۰۵' و عرض جغرافیایی ۲۳' ۳۴° تا ۳۵' ۳۴° واقع شده است. زمین‌چهره‌های^۱ موجود در منطقه

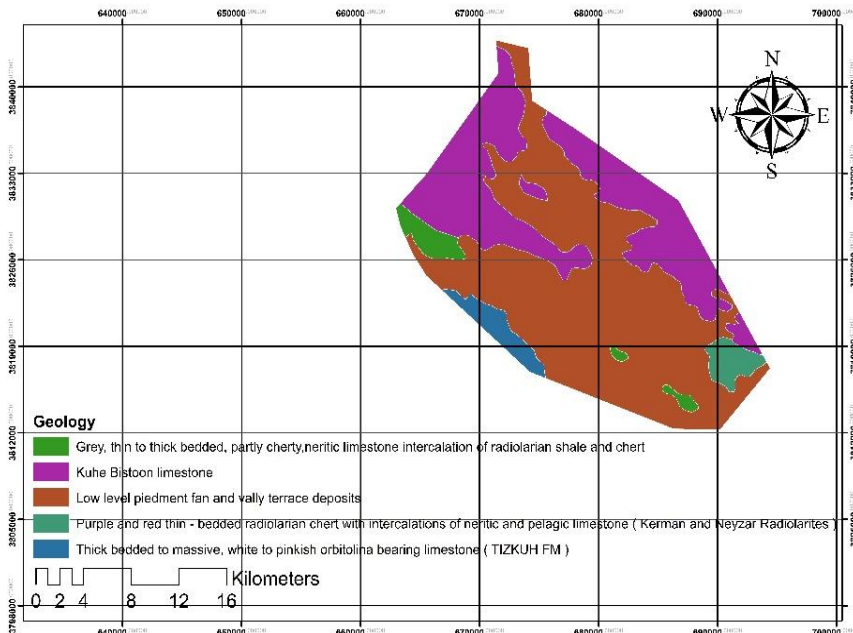
1. Land Landscape



شکل ۱. موقعیت منطقه مطالعاتی



شکل ۲. نقشه کاربری اراضی و مسیر رودخانه دشت میان دربند



شکل ۳. نقشه زمین‌شناسی دشت میان دربند

مطالعات صحرائی

بررسی و تشریح خاک‌ها از دیدگاه ریخت‌شناختی بر اساس راهنمای شناسایی خاک در صحرا (Schoeneberger *et al.*, 2012) انجام شد. پس از تشریح خاک، نمونه‌های خاک تمام افق‌های شناسایی‌شده برای انجام تجزیه و تحلیل فیزیکی و شیمیایی کدگذاری و به آزمایشگاه منتقل شد.

تجزیه‌های آزمایشگاهی

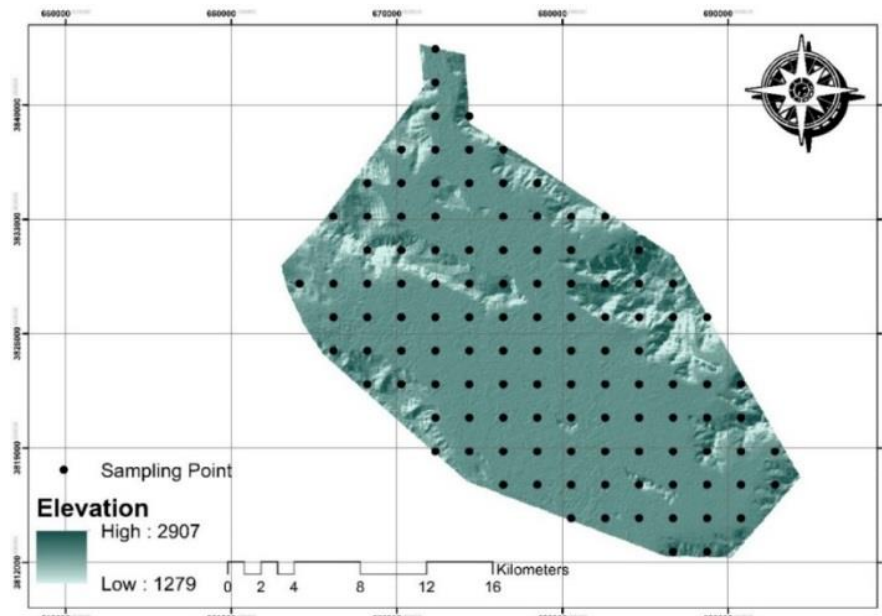
به منظور تجزیه و تحلیل فیزیکی و شیمیایی، نمونه‌های خاک از الک ۲ میلی‌متری عبور داده شد. ویژگی‌های خاک شامل واکنش خاک، بافت خاک، هدایت الکتریکی، ماده آلی خاک، کربنات کلسیم معادل و Mg^{2+} ، Ca^{2+} ، K^+ ، Na^+ طبق

طرح نمونه‌برداری

به منظور تعیین مکان نقاط نمونه‌برداری خاک و استخراج متغیرهای محیطی با استفاده از روش نمونه‌برداری سیستماتیک در محیط نرم‌افزار R یک شبکه منظم به ابعاد ۲×۲ کیلومتر طراحی و بر اساس آن مکان ۱۲۰ نقطه در مساحت ۵۰ هزار هکتار شناسایی شد (شکل ۴). مکان نقاط تعیین‌شده توسط GPS (با دقت برآورد ± 3 متر) پیمایش شد. با توجه به محدودیت‌های موجود در منطقه ۵ نقطه از طرح نمونه‌برداری حذف گردید. تعداد ۷۸ نقطه مشاهداتی در پهنه خاکی و ۳۷ نقطه در کوهستان قرار گرفت. در محل هر نقطه یک خاک‌خاک تا عمق حداقل ۱۲۰ سانتی‌متر حفر شد به جز نقاطی که سنگ‌بستر یا عوامل محدودکننده مانع شد.

و تجزيه‌هاى آزمائشگاهى خاکرخ‌ها به روش (Soil Taxonomy, 2014) رده‌بندى شد.

دستورالعمل‌هاى استاندارد (Klute, 1986; Sparks *et al.*, 1996) انجام شد. سپس بر اساس نتايج ريخت‌شناختى در صحرا



شکل ۴- طرح نمونه‌برداری برای حفر پدانها

(تجزیه و تحلیل زمین) به صورت نقطه‌ای از محل حفر خاکرخ، به نرم‌افزار SAGA 3.0.0 منتقل شد. تمام متغیرهای محیطی به صورت لایه‌های رستری که بیشتر برای نمایش تغییرات مکانی پیوسته مناسب است مورد استفاده قرار گرفت.

تجزیه‌های آماری

درخت تصمیم‌گیری

درخت تصمیم‌گیری یکی از روش‌های محبوب و امیدوارکننده در گروه روش‌های رایان‌آموزی است که تکنیک‌های موفقی برای پیش‌بینی و تبیین رابطه میان برخی اندازه‌گیری‌ها و یک متغیر هدف می‌باشد. در این مطالعه برای استخراج ارتباط خاک-سرزمین و پیش‌بینی الگوی پراکنش خاک‌ها از الگوریتم C5.0 با و بدون روش بوستینگ استفاده شده است. درخت تصمیم‌گیری از بین متغیرهای محیطی، متغیری را برای تشخیص کلاس خاک بر می‌گزیند که دارای بهره اطلاعاتی (رابطه ۴) بالایی باشد. از معیار کاهش آنتروپی (رابطه ۵) برای بررسی سودمندی متغیر انتخاب شده استفاده می‌کند. (Adu-Poku, 2012).

$$E(t) = \sum_j p(j|t) \text{Log}p(j|t) \quad (\text{رابطه ۴})$$

$p(j|t)$ فراوانی نسبی کلاس j در گره t می‌باشد.

بنابراین بهره اطلاعات برای هر یک از تقسیم‌ها به صورت

زیر به دست می‌آید:

$$\text{GAIN}_{\text{split}} = E(p) - \left(\sum_{i=1}^k \frac{n_i}{n} E(i) \right) \quad (\text{رابطه ۵})$$

محاسبه متغیرهای محیطی

استخراج متغیرهای محیطی با هدف تعیین متغیرهای پیش-بینی‌کننده در مدل اسکورپن انجام می‌شود. در این مطالعه از متغیرهای محیطی مشتق شده از مدل رقومی ارتفاع (DEM) و تصاویر ماهواره Landsat 8 OLI/TIRS استفاده شد. مدل رقومی ارتفاع (DEM) نمایی از زمین‌چهر و ویژگی‌های آن و یک روش کلی برای پشتیبانی توپوگرافی برای اهداف مدل‌سازی زیست محیطی است (Hutchinson and Gallant, 2000). در این مطالعه از مدل رقومی ارتفاع (DEM) ماهواره ASTER با وضوح ۳۰ متر برای محاسبه تعداد ۱۹ متغیر از ویژگی‌های زمین و از تصویر ماهواره Landsat 8 (path 166, row 36) محاسبه ۱۱ متغیر شامل شاخص روشنایی (Metternicht and Zink, 2003)، نسبت کانی رسی (Boettinger *et al.*, 2008)، شاخص گچ (Nield *et al.*, 2007)، نسبت اکسید آهن (Segal, 1982)، شاخص کربناته (Boettinger *et al.*, 2008)، شوری (and Zink, 2003) و شاخص پوشش گیاهی تفاضل (Tucker, 1979)، شاخص پوشش گیاهی تفاضل نرمال (Gitelson, 1996)، شاخص گیاهی تنظیم‌شده خاک (Huete, 1988)، شاخص گیاهی نسبی (Pearson and Miller, 1972) و دمای سطح زمین (Due *et al.*, 2015) استفاده شد. مدل رقومی ارتفاع و تصویر ماهواره‌ای ابتدا در محیط نرم‌افزار Arc GIS 10.2.2 پردازش شد و برای استخراج متغیرهای محیطی

که در آن گره والد P به اجزاء k تقسیم می‌شود و n_i هم تعداد رکوردهای موجود در جزء i می‌باشد. به این ترتیب شباهت میان داده‌های موجود در یک گره و تفاوت میان مجموعه داده‌های موجود در گره‌ها بیشینه می‌شود. الگوریتم C5.0 از روش هرس بدبینانه برای حذف طبقه‌بندی اشتباه استفاده می‌کند (Shirali, 2016).

الگوریتم بوستینگ

بوستینگ یک متالگوریتم ترکیبی در حوزه رایان‌آموزی است که برای کاهش عدم توازن و همچنین واریانس به کار می‌رود. این روش، برای تبدیل سیستم‌های یادگیری ضعیف به قوی بر اساس ترکیب نتایج طبقه‌بندی‌های مختلف است. در واقع با استفاده از روش بوستینگ یک توالی از درخت‌های تصمیم‌گیری توسعه می‌یابد و هر درخت تلاش می‌کند که نرخ خطای طبقه‌بندی اشتباه را کاهش دهد. سپس هر درخت یک پیش‌بینی ایجاد می‌کند و در نهایت از بین این پیش‌بینی‌ها رأی‌گیری می‌شود. نهایتاً یک پیش‌بینی با بیشترین تعداد رأی به عنوان پیش‌بینی نهائی انتخاب می‌شود. در واقع بوستینگ مجموعه آموزشی را وزن‌دار می‌کند. به این صورت که ابتدا اولین درخت را از روی مجموعه آموزشی با وزن برابر با یک ($W_j=1$) می‌سازد، سپس درخت تولیدشده بررسی می‌شود و وزن مجموعه آموزشی بر اساس پیش‌بینی‌های درخت تغییر می‌کند. به این صورت که وزن آن بخش از داده‌ها (مجموعه آموزش) که به درستی توسط درخت پیش‌بینی شده باشند کاهش یافته و وزن آن بخش از داده‌ها که توسط درخت درست پیش‌بینی نشده‌اند افزایش می‌یابد. به این ترتیب داده‌هایی که وزن بیشتری دارند (طبقه‌بندی‌های اشتباه) در مرحله بعد به منظور تصحیح اهمیت آنها در ساخت درخت دوم بیشتر می‌شود و این فرایند به همین شکل ادامه می‌یابد تا اینکه در درخت آخر پیش‌بینی نهایی با بیشترین دقت توسط درخت انجام می‌شود.

مدل‌سازی

متغیرهای محیطی استخراج‌شده از تصاویر لندست ۸ و مدل رقومی ارتفاع (DEM) به نرم‌افزار R 3.1.3 منتقل شد و برای مدل‌سازی کلاس خاک مبتنی بر متغیرهای محیطی به کار رفت. مراحل طی شده در ساخت مدل شامل دو بخش آموزش مدل و ارزیابی صحت مدل است که در ادامه به توضیح آن می‌پردازیم.

به منظور آموزش مدل مجموعه داده‌ها (کلاس خاک به همراه متغیرهای محیطی استخراج‌شده مربوط به هر کلاس) به صورت تصادفی به دو مجموعه با نسبت ۷۵:۲۵ تقسیم شد. ۷۵ درصد داده‌ها برای آموزش یا به عبارتی واسنجی مدل و ۲۵

درصد دیگر به عنوان داده‌های اعتبار سنجی برای ارزیابی مورد استفاده قرار گرفت و برای جلوگیری از بیش‌برازش در روند مدل‌سازی از روش هرس پیش از رشد درخت (که موجب کاهش پیچیدگی می‌شود) استفاده شد. به این منظور از آماره فاکتور اطمینان^۱ که بین ۰ تا ۱ متغیر است استفاده شد. در این مطالعه از آماره فاکتور اطمینان برابر با ۰/۹۵ برای هرس درخت تصمیم‌گیری استفاده شد. علاوه بر این در این مطالعه از روش بوستینگ برای بهینه‌سازی الگوریتم C5.0 با تعداد ۱۰۰ درخت استفاده شد.

ارزیابی مدل

یک نکته بسیار مهم در تولید نقشه خاک، کیفیت و دقت آن می‌باشد که نقشه‌برداری رقومی خاک امکان محاسبه کمی دقت نقشه خاک را از طریق ارزیابی مدل‌های به کار برده شده در تهیه نقشه ممکن ساخته است. روش‌های مختلفی برای ارزیابی دقت مدل وجود دارد از جمله:

آمار توصیفی چند متغیره

آمار توصیفی چند متغیره شرح داده‌شده توسط (Jensen, 1996) برای ارزیابی سازگاری بین کلاس‌های خاک یا ویژگی‌های موجود و پیش‌بینی مدل استفاده می‌شود. آمار توصیفی چند متغیره شامل محاسبه دقت کاربر، دقت تولیدکننده و صحت عمومی نقشه در ماتریس خطای داده‌ها می‌باشد. ماتریس خطا (ماتریس درهمی) یک ماتریس $n \times n$ می‌باشد که در آن تعداد کلاس‌ها = n می‌باشد (Malone, 2013).

آمار گسسته چند متغیره

با استفاده از نمایه سازگاری کاپا (Cohen, 1960) می‌توان روش‌های مختلف طبقه‌بندی را از طریق مقایسه ماتریس‌های به دست آمده ارزیابی کرد. نمایه سازگاری کاپا برای اندازه‌گیری صحت یا توافق بین دسته‌های پیش‌بینی شده در داده‌های مرجع به کار می‌رود (Congalton, 1991). نمایه سازگاری کاپا به صورت بسیار متداولی برای ارزیابی دقت پیش‌بینی نقشه‌های طبقه‌بندی به کار رفته است.

$$K = \frac{p_o - p_e}{1 - p_e} \quad (۶)$$

که در آن K نمایه سازگاری کاپا، p_o دقت کلی و p_e دقت مورد انتظار است.

نمایه سازگاری کاپا بیانگر هماهنگی بین مشاهدات و پیش‌بینی یک طبقه‌بندی است. مقادیر ضریب کاپا از -۱ تا ۱

عریضی هستند که موجب تشکیل خاک‌های ورتی‌سول شده است.

- خاک‌های موجود در مناطق پست: این خاک‌ها دارای شیب خیلی کم و در برخی اوقات از سال تحت تأثیر آب‌های تحت‌الارضی بوده که موجب تجمع مواد آلی در سطح و شرایط احیاء در افق‌های خاک گردیده است.

سنگ‌های منطقه تحت تأثیر آب‌وهوا هوادیده شده و توسط آب و نیروی ثقل منتقل شده و تحت تأثیر عوامل مکانیکی، شیمیایی و بیولوژیکی، خاک‌های دشت میان‌دریاب را تشکیل داده‌اند. خاک‌های تولیدشده بسته به سرعت انتقال و شیب به تدریج بر جای گذاشته‌شده که نتیجه آن تشکیل خاک‌هایی با بافت سنگین و سنگریزه‌دار در مجاورت ارتفاعات و مخروط افکنه‌ها و خاک‌هایی با بافت خیلی سنگین در دشت‌های آبرفتی رودخانه‌ای و دامنه‌ای و زمین‌های پست می‌باشد.

بر اساس نتایج نشان داده‌شده در شکل (۵) مشاهده می‌شود که خاک‌های دشت دریاب شامل ۶ گروه بزرگ و ۱۴ زیرگروه است.

تغییر می‌کند و هرچه به ۱ نزدیک‌تر باشد دقت بالاتر است (Malone, 2013).

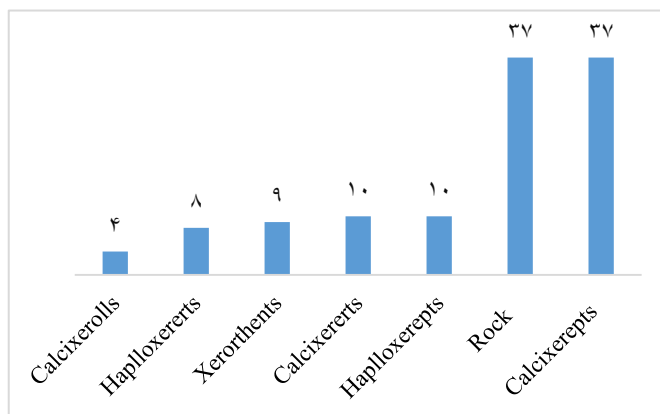
نتایج و بحث

رده‌بندی خاک‌ها

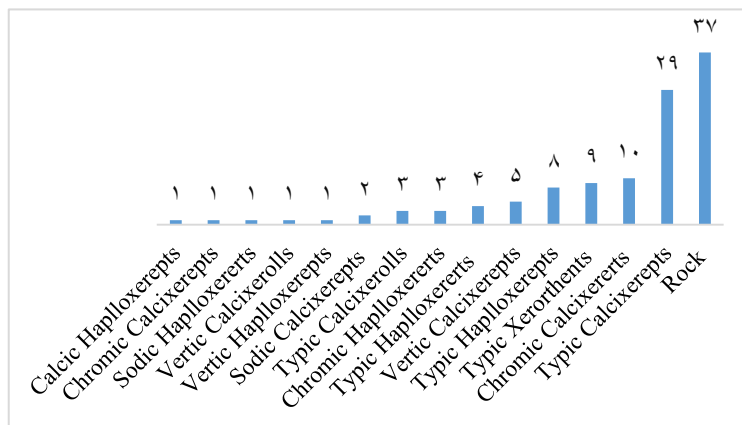
خاک‌ها از نظر موقعیت قرارگیری شامل موارد زیر می‌شوند.

- خاک‌های منطقه مورد مطالعه روی مخروط افکنه‌ها: این خاک‌ها تحت تأثیر آب زیرزمینی نبوده و شامل خاک‌های با بافت سنگین تا خیلی سنگین هستند که زمان کافی برای تکامل، انتقال و تجمع کربنات‌ها در اختیار داشته و در نتیجه خاک‌هایی با افق‌های آهکی را تشکیل داده‌اند.

- خاک‌های موجود در دشت‌های آبرفتی دامنه‌ای و رودخانه‌ای: این خاک‌ها دارای شیب کم و در بعضی قسمت‌ها همراه با کمی شیب و پستی بلندی بوده و تحت تأثیر آب‌های جاری شده از ارتفاعات هستند. این خاک‌ها دارای تکامل پروفیلی و افق‌های Calcic و Cambic و در بخش وسیعی به علت حضور رس‌های با قابلیت انقباض و انبساط، دارای درز و شکاف‌های



الف



ب

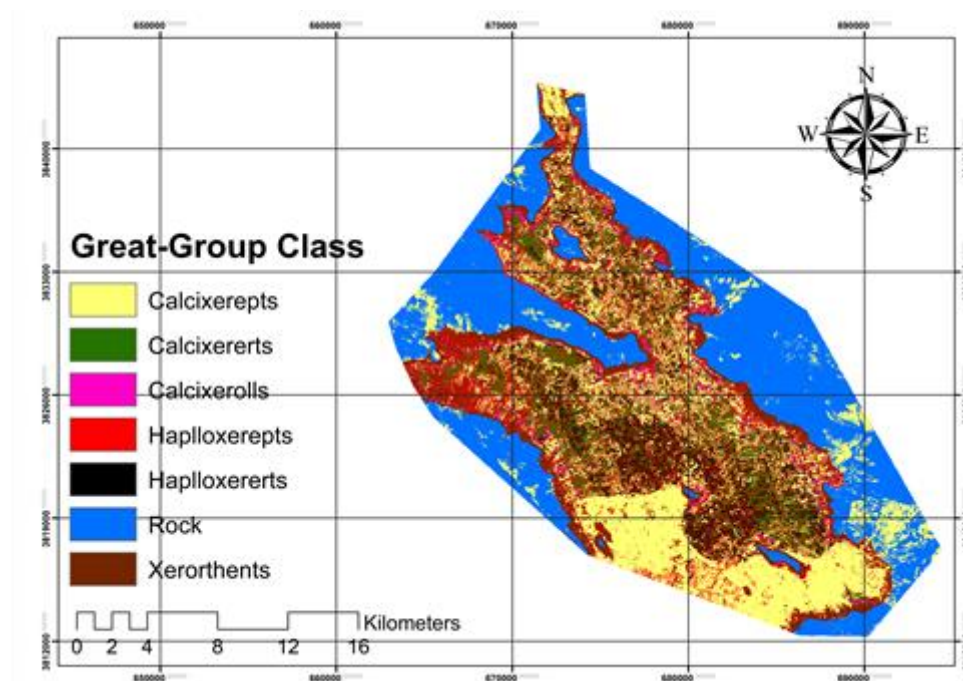
شکل ۵- تعداد پدان مطالعه‌شده در سطح گروه بزرگ (الف) و زیرگروه (ب)

پیش‌بینی الگوی پراکنش خاک بر اساس سطوح گروه بزرگ و زیرگروه با و بدون کاربرد روش بوستینگ

سطح گروه بزرگ

گروه بزرگ‌های شناسایی‌شده در دشت میان‌دربند شامل Haploxerepts، Calcixererts، Calcixerols، Calcixerepts و Haploxererts است. نقشه خاک تولیدشده با الگوریتم C5.0 بدون روش بوستینگ در شکل (۶) موجود است. نتایج ارزیابی نشان می‌دهد که صحت عمومی (جدول ۱) برابر با ۷۳ درصد و نمایه سازگاری کاپا (جدول ۳) برابر با ۰/۶۱ بود. مدل ۴ گروه بزرگ را با موفقیت پیش‌بینی کرد. اما در پیش-

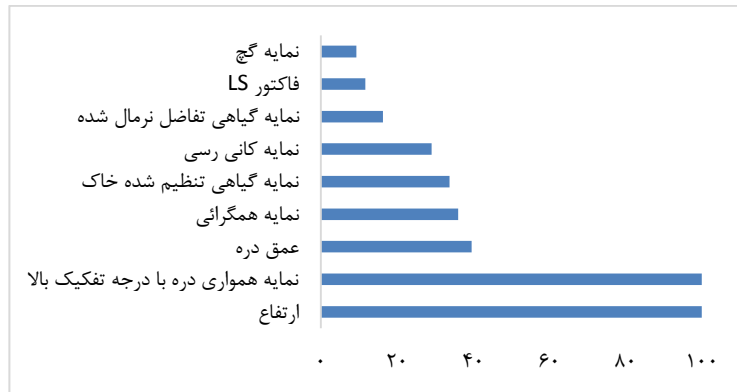
بینی گروه بزرگ Calcixerols و Haploxererts موفق نبود. بیشینه دقت تولیدکننده در گروه بزرگ Haploxerepts برابر با ۱۰۰ درصد و کمینه آن در گروه بزرگ Xerorthents برابر با ۵۰ درصد بود. همچنین بیشینه دقت کاربر در گروه بزرگ Calcixererts و Xerorthents برابر با ۱۰۰ درصد و کمینه دقت کاربر در Haploxerepts برابر با ۳۴ درصد بود. همچنین متغیرهای محیطی مؤثر در تمایز خاکرخ‌ها شامل نمایه‌های توپوگرافی و به‌ندرت نمایه‌های مواد مادری و پوشش گیاهی بود (شکل ۷).



شکل ۶. نقشه پیش‌بینی الگوریتم C5.0 در سطح گروه بزرگ

جدول ۱. ماتریس درهمی پیش‌بینی الگوریتم C5.0 در سطح گروه بزرگ

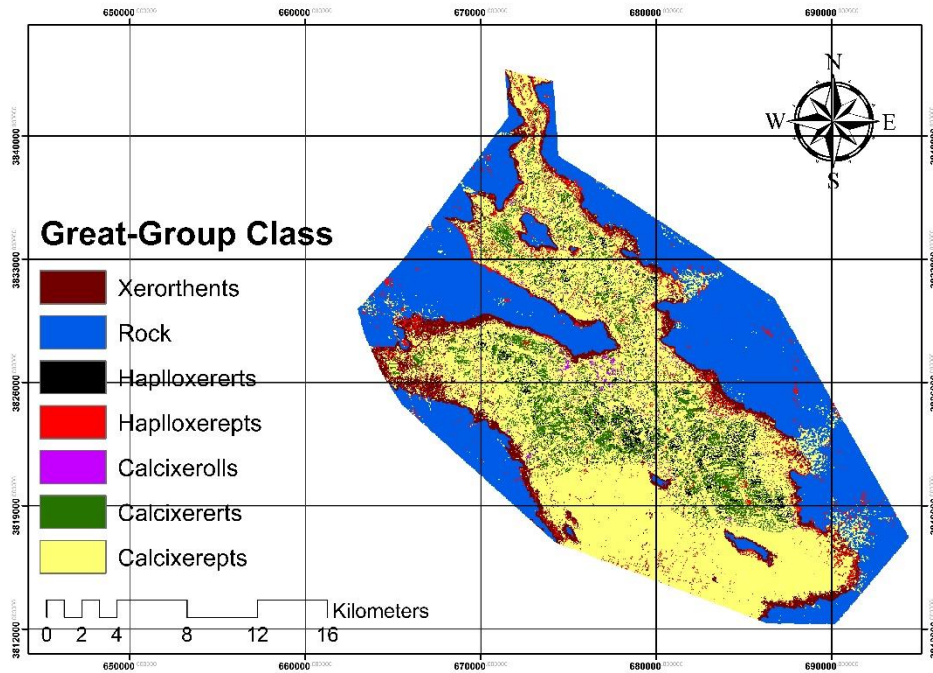
گروه بزرگ	دقت کاربر							
	a	b	c	d	e	f	g	
a: Calcixerepts	۷	۱	۱	۰	۰	۱	۰	۷۰
b: Calcixererts	۰	۲	۰	۰	۰	۰	۰	۱۰۰
c: Calcixerolls	۰	۰	۰	۰	۰	۰	۰	۰
d: Haploxerepts	۱	۰	۰	۱	۰	۰	۱	۳۴
e: Haploxererts	۲	۰	۰	۰	۰	۰	۰	۰
f: Rock	۱	۰	۰	۰	۰	۱۰	۰	۹۱
g: Xerorthents	۰	۰	۰	۰	۰	۰	۱	۱۰۰
دقت تولیدکننده	۶۴	۶۷	۰	۱۰۰	۰	۹۱	۵۰	۷۳



شکل ۷. اهمیت متغیرهای محیطی در پیش‌بینی الگوریتم C5.0

مکان ۱۸۷ نقطه مشاهداتی را تعیین کردند. آنها از انواع متغیرهای توپوگرافی، تصویر ماهواره‌ای، نقشه زمین‌شناسی و کاربری اراضی بهره گرفتند. سپس با استفاده از الگوریتم‌های ژنتیک، شبکه عصبی، رگرسیون لجستیک چند متغیره، آنالیز تشخیصی و الگوریتم C5.0 گروه بزرگ خاک را پیش‌بینی کردند. نتایج آنها نشان داد که الگوریتم C5.0 به‌عنوان الگوریتم برتر با صحت عمومی ۶۷/۵ درصد به‌طور موفقیت‌آمیز گروه‌های بزرگ خاک را پیش‌بینی می‌کند. همچنین تقی‌زاده مهرجردی و همکاران (۱۳۹۲) در مطالعه‌ای دیگر در همین منطقه از ترکیب فاصله تاکسونومیکی و درخت تصمیم‌گیری برای پیش‌بینی گروه بزرگ خاک بهره گرفتند. آنها توانستند گروه بزرگ خاک را با صحت عمومی ۷۰ درصد پیش‌بینی کنند. آنها توصیه کردند که از الگوریتم درخت تصمیم در سایر مناطق در ایران به کار گرفته شود. (Jafari et al., 2012) در منطقه زرنند کرمان با مساحت ۹۰۰۰۰ هکتار ۱۲۶ مکان مشاهده را تعیین کردند. برای تعیین متغیرهای محیطی از پارامترهای زمین، تصاویر ماهواره‌ای و نقشه زمین‌شناسی استفاده کردند. آنها توانستند گروه بزرگ خاک را با صحت عمومی ۶۷/۵ درصد پیش‌بینی کنند. همچنین Moran and Bui (2002) در حوزه موری-دارلینگ در شرق استرالیا با کاربرد الگوریتم بوستینگ درخت تصمیم‌گیری گروه بزرگ خاک را با دقت ۷۰ درصد پیش‌بینی کردند. این محققین بیان کردند که الگوریتم درخت تصمیم‌گیری توان بالایی در آشکارسازی پراکنش کلاس خاک دارد.

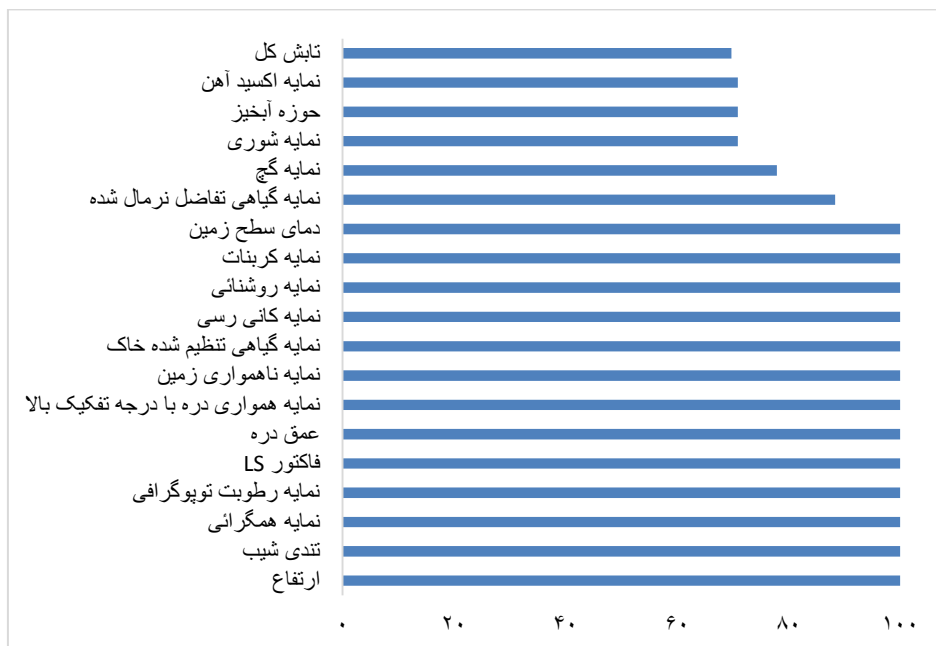
نقشه گروه بزرگ تولیدشده با الگوریتم بوستینگ C5.0 در شکل (۸) موجود است. نتایج ارزیابی مدل نشان می‌دهد که صحت عمومی (جدول ۲) برابر با ۸۰ درصد و نمایه سازگاری کاپا (جدول ۳) برابر با ۰/۷۲ بود. مدل توانست ۵ گروه بزرگ را با موفقیت پیش‌بینی کند. اما در پیش‌بینی گروه بزرگ Haploxerepts موفق نبود. بیشینه دقت تولیدکننده در گروه بزرگ Calcixerolls و Xerorthents برابر با ۱۰۰ درصد و کمینه آن در گروه بزرگ Haploxererts برابر با ۳۴ درصد بود. همچنین بیشینه دقت کاربر در گروه بزرگ Haploxererts و Calcixerolls برابر با ۱۰۰ درصد و کمینه آن در گروه بزرگ Calcixererts برابر با ۳۴ درصد بود. با توجه به متغیرهایی که توسط درخت تصمیم‌گیری در پیش‌بینی به کار رفت مشخص شد که متغیرهای محیطی مسئول گوناگونی خاک‌ها در سطح گروه بزرگ شامل متغیرهای توپوگرافی و پس از آن مادری و به میزان کمتری نمایه‌های گیاهی بود (شکل ۹). به‌این ترتیب می‌توان این‌گونه بیان کرد که عامل توپوگرافی و پس از آن مواد مادری بیشترین تأثیر را در تشکیل گروه‌های بزرگ خاک در دشت میان‌دربند دارد. به کارگیری الگوریتم بوستینگ C5.0 منجر به افزایش صحت عمومی پیش‌بینی گروه‌های بزرگ به میزان ۷ درصد شد. درواقع الگوریتم بوستینگ با جلوگیری از ایجاد بیش‌برازش و کاهش نرخ طبقه‌بندی اشتباه موجب بهبود پیش‌بینی می‌شود. نتایج به‌دست‌آمده در این پژوهش از قبیل صحت عمومی و اهمیت توپوگرافی در تمایز خاک‌ها مشابه با مطالعات (Taghizadeh-Mehrjardi et al., Jafari et al., 2012) 2014; در مناطق خشک و نیمه‌خشک ایران است. Taghizadeh-Mehrjardi et al. (2014) در منطقه اردکان استان یزد با مساحت ۷۲۰۰۰ هکتار با استفاده از ابر مکعب لاتین



شکل ۸. نقشه پیش‌بینی الگوریتم بوستینگ C5.0 در سطح گروه بزرگ

جدول ۲. ماتریس درهمی پیش‌بینی الگوریتم بوستینگ C5.0 در سطح گروه بزرگ

گروه بزرگ	a	b	c	D	e	f	g	دقت کاربر
a: Calcixerepts	۷	۱	۰	۱	۰	۰	۰	۷۸
b: Calcixererts	۰	۱	۰	۰	۲	۰	۰	۳۴
c: Calcixerolls	۰	۰	۱	۰	۰	۰	۰	۱۰۰
d: Haploxerepts	۱	۰	۰	۰	۰	۰	۰	۰
e: Haploxererts	۰	۰	۰	۰	۱	۰	۰	۱۰۰
f: Rock	۱	۰	۰	۰	۰	۱۰	۰	۹۱
g: Xerorthents	۰	۰	۰	۰	۰	۰	۳	۱۰۰
دقت تولیدکننده	۷۸	۵۰	۱۰۰	۰	۳۴	۱۰۰	۱۰۰	۸۰



شکل ۹. اهمیت متغیرهای محیطی در پیش‌بینی الگوریتم بوستینگ C5.0 در سطح گروه بزرگ

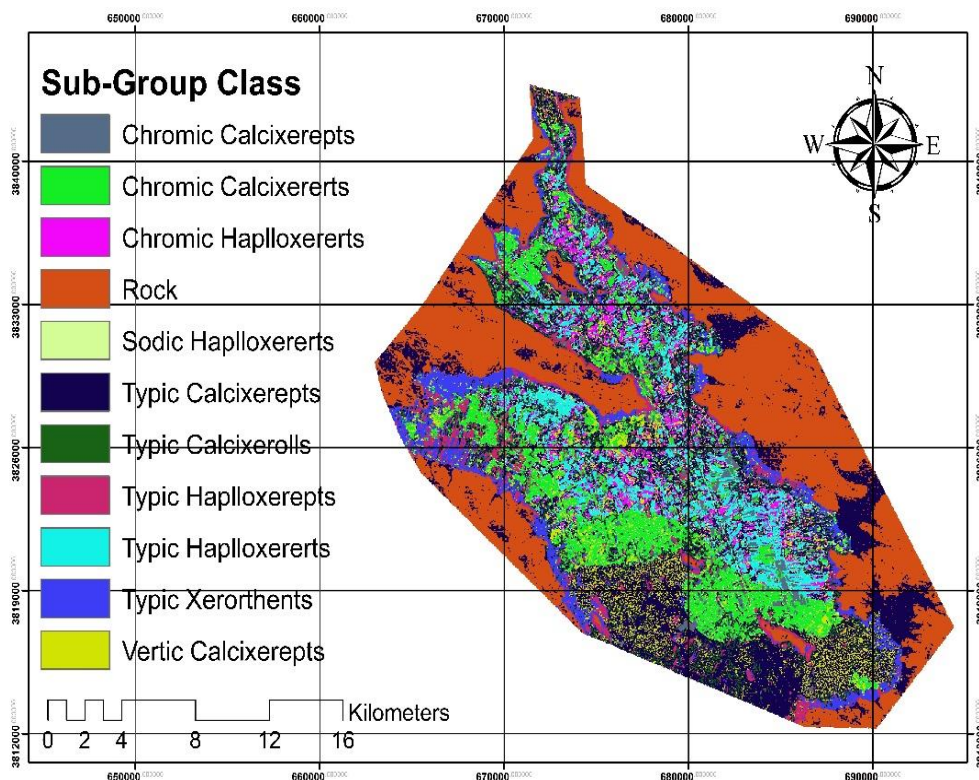
بينى آنها صفر مى‌شود. به‌اين ترتيب اگر در مجموعه اعتبار سنجى قرار بگيرند در مرحله آموزش حضور نداشته و توسط مدل آموخته و طبيعياً پيش‌بينى نمى‌شوند. در مرحله ارزىابى، هنگام تطابق كلاس‌هاى پيش‌بينى‌شده و كلاس‌هاى مشاهده‌شده محروم مى‌شوند. درنتيجه هميشه دقت پيش‌بينى براى اين كلاس‌ها صفر مى‌شود. با توجه به جدول (۴) مدل توانست ۵ كلاس را به‌طور موفقيت‌آمى پيش‌بينى كند. به‌طورى كه بيشينه دقت توليدكننده در زيرگروه Chromic Haploxererts و Typic Xerorthents برابر با ۱۰۰ درصد و كمينه آن در زيرگروه Chromic Calcixererts و Typic Haploxererts برابر با ۵۰ درصد بود. همچنين بيشينه دقت كاربر در زيرگروه Chromic Haploxererts برابر با ۵۰ درصد و بيشينه آن در زيرگروه Typic Calcixererts برابر با ۷۵ درصد بود و كمينه آن در زيرگروه Typic Haploxererts برابر ۲۵ درصد بود. اما در پيش‌بينى ساير زيرگروه‌ها موفقيتى حاصل نشد. با توجه به شكل (۱۱) مشخص شد كه متغيرهاى محيطى اثرگذار بر تشكيل خاك‌ها عمدتاً متغيرهاى توپوگرافى و به ميزان كمتر نمايه‌هاى مواد مادرى و پوشش گياهمى بودند.

جدول ۳- نمايه سازگارى كاپا در سطح گروه بزرگ

مدل (گروه بزرگ)	نمايه كاپا
الگوريتم C5.0	۰/۶۱
الگوريتم بوستينگ C5.0	۰/۷۲

سطح زيرگروه

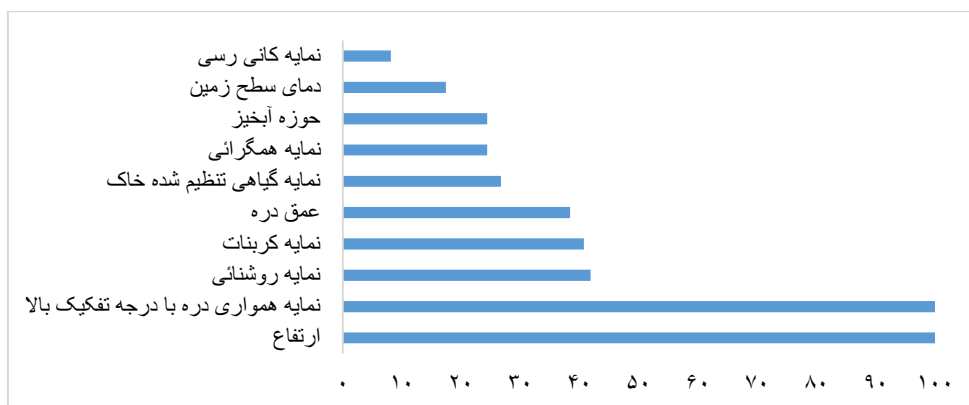
تعداد ۱۴ زيرگروه در منطقه مطالعاتى شناسايى شد. نقشه زيرگروه با استفاده از الگوريتم C5.0 در شكل (۱۰) موجود است. نتايج ارزىابى نشان مى‌دهد كه صحت عمومى (جدول ۴) برابر با ۷۳ درصد و نمايه سازگارى كاپا (جدول ۶) برابر با ۰/۶۳ بود. نقشه حاصل داراى ۱۱ كلاس بود. اين بدین معنى است كه الگوريتم قادر به يادگيرى ۴ زيرگروه Calcic Haploxerepts, Sodic Calcixererts, Vertic Calcixerolls و Vertic Haploxerepts نيست. ۴ زيرگروه مذكور به همراه زيرگروه Chromic Calcixererts داراى فراوانى يك بودند. هنگامى كه پاينگاه داده به دو دسته آموزش و اعتبارسنجى تقسيم مى‌شود اين ۵ زيرگروه يا در مجموعه آموزشى، يا در مجموعه اعتبار سنجى قرار مى‌گيرند. به‌اين ترتيب زمانى كه در مجموعه آموزشى قرار بگيرند، در مرحله اعتبارسنجى داراى عضويت نيستند و طبيعياً در راستى آزمائى ماتريس درهمى دقت پيش-



شكل ۱۰. نقشه پيش‌بينى الكوربتم C5.0 در سطح زيرگروه

جدول ۴. ماتریس درهمی پیش‌بینی الگوریتم C5.0 در سطح زیرگروه

زیرگروه	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	دقت کاربر
a: Calcic Haploxerepts	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
b: Chromic Calcixererts	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
c: Chromic Calcixererts	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۵۰
d: Chromic Haploxererts	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۵۰
e: Rock	۰	۰	۰	۰	۱۳	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۱۰۰
f: Sodic Calcixererts	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
g: Sodic Haploxererts	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
h: Typic Calcixererts	۰	۰	۰	۰	۰	۰	۰	۳	۰	۰	۰	۰	۱	۰	۰	۷۵
i: Typic Calcixerolls	۰	۰	۰	۰	۰	۰	۰	۱	۰	۰	۰	۰	۰	۰	۰	۰
j: Typic Haploxerepts	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
k: Typic Haploxererts	۰	۰	۱	۰	۰	۰	۰	۱	۱	۱	۰	۰	۰	۰	۰	۲۵
l: Typic Xerorthents	۰	۰	۰	۰	۰	۰	۰	۰	۱	۰	۲	۰	۰	۰	۰	۶۷
m: Vertic Calcixererts	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
n: Vertic Calcixerolls	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
o: Vertic Haploxerepts	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰
دقت تولیدکننده	۰	۰	۵۰	۱۰۰	۱۰۰	۰	۰	۷۵	۰	۰	۵۰	۱۰۰	۰	۰	۰	۷۳



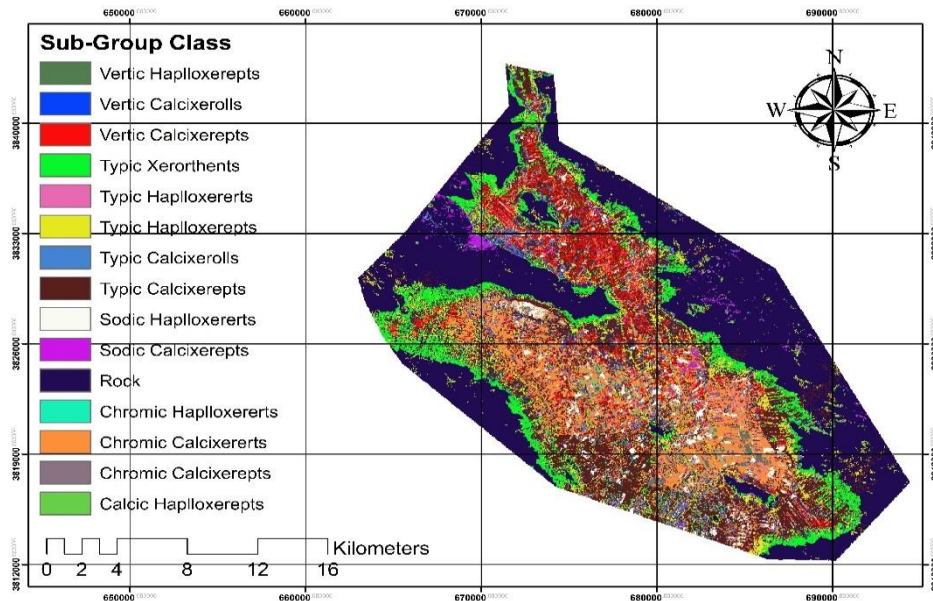
شکل ۱۱. اهمیت متغیرهای محیطی در پیش‌بینی الگوریتم C5.0 در سطح زیرگروه

درصد بود. با توجه به شکل (۵-ب) مربوط به فراوانی خاکرخ‌ها در سطح زیرگروه و جدول‌های (۴ و ۵) مربوط به ماتریس درهمی دو الگوریتم مشخص شد که فراوانی مشاهدات خاکرخ به شدت دقت نقشه زیرگروه را تحت تأثیر قرار داده است. این یافته با نتایج ارائه شده توسط (Barthold *et al.*, 2013; Marchetti *et al.*, 2011; Kim *et al.*, 2012; Hengl *et al.*, 2007; pahlavan rad *et al.*, 2015b; Jafari *et al.*, 2013) هماهنگ است. آنها بیان کردند که دقت مدل‌های رایان‌آموزی در پیش‌بینی کلاس‌های با مشاهدات کم کاهش می‌یابد. به‌طور کلی افزایش تعداد کلاس‌ها و کاهش مشاهدات، عملکرد مدل‌های رایان‌آموزی را تحت تأثیر قرار می‌دهد. این موضوع نشان می‌دهد که علاوه بر تأثیر قابل توجه انتخاب متغیر محیطی، فراوانی خاکرخ‌ها نیز دقت پیش‌بینی را متأثر می‌سازد.

نقشه زیرگروه تولید شده با استفاده از الگوریتم C5.0 با روش بوستینگ در شکل (۱۲) موجود است. با کاربست الگوریتم بوستینگ C5.0 نتایج ارزیابی نشان می‌دهد که صحت عمومی (جدول ۵) برابر با ۷۶ درصد و نمایه سازگاری کاپا (جدول ۶) برابر با ۰/۶۶ به دست آمد. نقشه زیرگروه حاصل دارای ۱۵ کلاس بود. این نشان‌دهنده این واقعیت است که روش بوستینگ موجب توانایی بالا در یادگیری فضای آموزش در الگوریتم C5.0 شده است. بیشینه دقت تولیدکننده برای زیرگروه Typic Xerorthents برابر با ۱۰۰ درصد و کمینه آن برای زیرگروه Chromic Calcixererts برابر با ۲۵ درصد بود. همچنین بیشینه دقت کاربر در زیرگروه Chromic Calcixererts برابر با ۱۰۰ درصد و کمینه آن در زیرگروه Typic Calcixererts برابر با ۶۷ درصد

را با صحت عمومی ۶۲ درصد با موفقیت پیش‌بینی کند. مساحت منطقه مطالعاتی در این پژوهش ۵۰۰۰۰ هکتار و تعداد مشاهدات ۱۱۵ خاکرخ بود که تعداد مشاهدات تقریباً یک‌سوم مشاهدات آن‌ها بود. این مطلب نشان‌دهنده این است که الگوریتم درخت تصمیم‌گیری در دشت دربند کارایی بالایی داشته است.

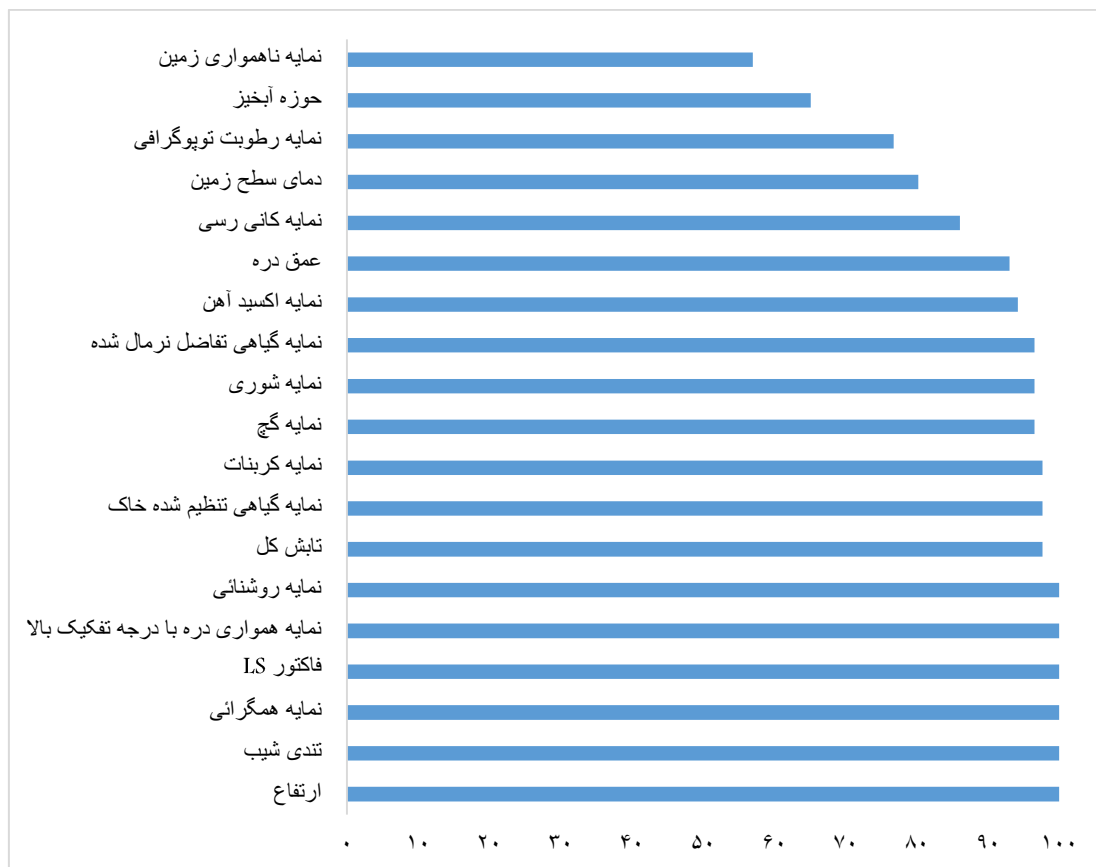
با توجه به شکل (۱۳) مشاهده می‌شود که متغیرهای محیطی اثرگذار در مدل‌سازی بیشتر شامل توپوگرافی و پس از آن نمایه‌های مواد مادری و پوشش گیاهی هستند. (Brungard *et al.*, 2015) در منطقه مطالعاتی خود در جنوب غربی میلفورد واقع در یوتا با مساحت ۳۰۰۰۰ هکتار با ۳۰۰ نمونه مشاهداتی اقدام به پیش‌بینی زیرگروه خاک کرد. او توانست زیرگروه خاک



شکل ۱۲. نقشه پیش‌بینی الگوریتم بوستینگ C5.0 در سطح زیرگروه

جدول ۵. ماتریس درهمی پیش‌بینی الگوریتم بوستینگ C5.0 در سطح زیرگروه

زیرگروه	a	b	c	d	f	g	h	i	j	k	l	m	n	o	دقت کاربر
a: Calcic Haploxerepts
b: Chromic Calcixererts
c: Chromic Calcixererts	.	.	۱	۱۰۰
d: Chromic Haploxerepts
e: Rock	۱۲	۱۰۰
f: Sodic Calcixererts
g: Sodic Haploxerepts
h: Typic Calcixererts	.	.	۲	۱	.	.	۶	۶۷
i: Typic Calcixerolls
j: Typic Haploxerepts	۱
k: Typic Haploxerepts
l: Typic Xerorthents	۱	.	۳	.	.	.	۷۵
m: Vertic Calcixererts	.	.	۱	.	.	.	۱
n: Vertic Calcixerolls
o: Vertic Haploxerepts
دقت تولیدکننده	.	.	۲۵	.	۱۰۰	.	۷۵	.	.	.	۱۰۰	.	.	.	۷۶



شکل ۱۳. اهمیت متغیرهای محیطی در پیش‌بینی الگوریتم بوستینگ C5.0 در سطح زیر گروه

هموار که متغیرهای زمین‌نما تغییرات چندانی ندارند نیز دارای قابلیت پیش‌بینی خوبی است. اما به خاطر هموار بودن زمین متغیرها دارای تغییرپذیری کم بوده و دقت الگوریتم را تحت‌الشعاع قرار می‌دهد. همچنین چون منطقه مسطح می‌باشد، زمانی که پارامترهای مورفومتری استخراج می‌شوند دارای نویز زیاد هستند به‌عنوان مثال برای شیب قابل محاسبه نیست یا دارای نویز می‌باشد و می‌تواند پیش‌بینی را با دشواری روبه‌رو کند. علاوه بر این تعداد کلاس زیاد و درعین حال تعداد مشاهده کم در کلاس‌ها نقشه پیش‌بینانه تولید شده را تحت تأثیر قرار می‌دهند. این الگوریتم می‌تواند با مدل‌سازی روابط خاک زمین-سرزمین برای کمک به مدیران و برنامه‌ریزان منابع طبیعی به کار گرفته شود. درخت تصمیم‌گیری با انتخاب متغیر بر اساس افزایش بهره‌های اطلاعاتی موجب کاهش پیچیدگی شده است. علاوه بر آن ساختار درختی اطلاعات قابل‌فهم و با قابلیت تفسیر آسان در مورد توانایی پیش‌بینی کلاس‌ها ایجاد کرد. به‌طور کلی نتایج نشان داد که روش‌های رایان‌آموزی به خصوص درخت تصمیم-گیری می‌توانند نقشه‌برداری خاک‌ها را با صرفه‌جویی در زمان و هزینه با دقت قابل‌قبول برای مناطق وسیع انجام دهد. همچنین باعث افزایش سرعت و دقت در استخراج و انتقال اطلاعات شود.

جدول ۶- نمایه‌سازی کاپا در سطح گروه بزرگ

مدل	نمایه کاپا
الگوریتم C5.0	۰/۶۳
الگوریتم بوستینگ C5.0	۰/۶۶

نتیجه‌گیری کلی

نقشه پیش‌بینانه تولیدشده در این مطالعه با برآورد واقعی از عدم قطعیت و به‌کارگیری هوش مصنوعی در انتخاب متغیر و پردازش رابطه متغیرهای محیطی با کلاس خاک می‌تواند اطلاعات قابل‌اعتمادی از پراکنش مکانی خاک در دشت میان-دریوند در اختیار کاربران منابع قرار دهد. علاوه بر این استخراج اطلاعات از این نقشه‌ها آسان‌تر بوده و با نمایش آن در سیستم-های اطلاعات جغرافیایی می‌توان با حداکثر بزرگ‌نمایی به آسانی از این نقشه‌ها استفاده کرد. نقشه‌های تولیدشده در سطح گروه بزرگ دارای صحت قابل‌قبولی بودند به‌جز در برخی کلاس‌ها که الگوریتم از تعداد مشاهدات تأثیر پذیرفته بود. همچنین نتایج نشان از برتری الگوریتم C5.0 تقویت‌شده با روش بوستینگ و بهبود دقت در جداسازی خاک‌ها بود. با توجه به عملکرد الگوریتم درخت تصمیم‌گیری به‌کاررفته در این مطالعه می‌توان نتیجه گرفت که درخت تصمیم‌گیری در مناطق

REFERENCES

- Adu-Poku, S. (2012). Comparing classification algorithms in data mining. MSc. dissertation. Central Connecticut State University.
- Banaii, M. H. 1977. Soil moisture and temperature regimes map of Iran. Soil and Water research Institute of Iran. Ministry of Agriculture, Tehran, Iran. (In Farsi).
- Barthold, F. K., Wiesmeier, M., Breuer, L., Frede, H. G., Wu, J., and Blank, F. B. (2013). Land use and climate control the spatial distribution of soil types in the grasslands of Inner Mongolia. *Journal of arid environments*, 88, 194-205.
- Boettinger, J. L., Ramsey, R. D., Bodily, J. M., Cole, N. J., Kienast-Brown, S., Nieli, S. J., and Stum, A. K. (2008) Digital soil mapping with limited data. In A. E. Hartemink., A. B. McBratney., and de Lourdes Mendonça-Santos, M. (Ed.). *Landsat spectral data for digital soil mapping*. (Vol. 16). (pp. 193-202). Springer Science and Business Media.
- Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., Edwards Jr., T. C., (2015). Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, 239-240, 68-83.
- Bui, E. N., Henderson, B. L. and Viergever, K. (2006). Knowledge discovery from models of soil properties developed through data mining. *Ecological Modelling*, 191(3-4), 431-446.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Cole, N. J. and Boettinger, J. L. (2006). . Developments in Soil Science. In P. Lagacherie, A.B. McBratney, M. Voltz (Ed.), *Pedogenic Understanding Raster Classification Method for Mapping Soils, Powder River Basin, Wyoming, USA*. Developments in Soil Science, (Vol. 31). (pp. 377-388).
- Congalton, R.G. (1991). A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of Environment*, 37, 35-46.
- Cook, S. E., Jarvis, A., and Gonzalez, J. P. (2008) Digital Soil Mapping with Limited Data. In A. E. Hartemink., A. B. McBratney., de Lourdes Mendonça-Santos. M., (Ed.), *A new global demand for digital soil information*. (Vol. 3). (pp. 31-41).
- Daigle, J. J., Hudnall, W. H., Gabriel, W. J., Mersiovsky, E., and Nielson, R. D. (2005). The National Soil Information System (NASIS): Designing soil interpretation classes for military land-use predictions. *Journal of terramechanics*, 42(3), 305-320.
- Du, C., Ren, H., Qin, Q., Meng, J., and Zhao, S. (2015). A practical split-window algorithm for estimating land surface temperature from Landsat 8 data. *Remote Sensing*, 7(1), 647-665.
- Elnaggar, A. A. (2007). Development of predictive mapping techniques for soil survey and salinity mapping. Ph. D. dissertation, Oregon State University.
- Franklin, J. (1995). Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in physical geography*, 19(4), 474-499.
- Grunwald, S., Thompson, J. A., and Boettinger, J. L. (2011). Digital soil mapping and modeling at continental scales: Finding solutions for global issues. *Soil Science Society of America Journal*, 75 (4): 1201-1213.
- Hash, S. J. (2008). Use of decision tree analysis for predictive soils mapping and implementation on the Malheur County. Ph. D. dissertation, Oregon State University.
- Henderson, B. L., Bui, E. N., Moran, C. J. and Simon, D. A. P. 2005. Australia-wide predictions of soil properties using decision trees. *Geoderma*, 124(3-4), 383-398.
- Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., and Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* 265, 62-77.
- Hengl, T., Toomanian, N., Reuter, H. I., and Malakouti, M. J. (2007). Methods to interpolate soil categorical variables from profile observations: lessons from Iran. *Geoderma*, 140(4), 417-427.
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote sensing of environment*, 25(3), 295-309
- Hutchinson, F. M., and Gallant, J. C. (2000) Terrain analysis: principles and applications. In Gallant J. C., Willson. J. P., (Ed.), *Digital elevation models and representation of terrain shape*. (Vol. 2). (pp. 29-50). New York: Wiley.
- Jafari, A., Finke, P. A., Vande Wauw, J., Ayoubi, S., and Khademi, H. (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. *European Journal of Soil Science*, 63(2), 284-298.
- Jafari, H. Khademi, SH. Ayoubi. (2013). Digital Mapping of Soil Diagnostic Horizons and Great Groups in Zarand Region of Kerman. *Journal of Water and Soil Science*, 16 (62), 177-193. (In Farsi)
- Jensen, J. R. (1996). *Introductory digital image processing: A remote sensing perspective*, (4th ed.). Prentice Hall, Inc., Upper Saddle River, New Jersey.
- Kim, J., Grunwald, S., Rivero, R. G., & Robbins, R. (2012). Multi-scale modeling of soil series using remote sensing in a wetland ecosystem. *Soil Science Society of America Journal*, 76(6), 2327-2341.
- Klute, A. (1986). *Methods of soil analysis. Part 1. Physical and mineralogical properties*. (2th ed.). Wisconsin USA: Madison

- Lagacherie, P., and McBratney, A. B. (2006). Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. *Developments in soil science*, 31, 3-22.
- Malone, B. (2013). Use of R for digital soil mapping. Soil Security Laboratory, University of Sydney, Australia. (pp. 1-209).
- Mark, D. M., and Csillag, F. (1989). The nature of boundaries on 'area-class' maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 26(1), 65-78.
- Marchetti, A., Piccini, C., Santucci, S., Chiuchiarelli, I., and Francaviglia, R. (2011). Simulation of soil types in Teramo province (Central Italy) with terrain parameters and remote sensing data. *Catena*, 85(3), 267-273.
- McBratney, A. B., Odeh, I. O., Bishop, T. F., Dunbar, M. S., and Shatar, T. M. (2000). An overview of pedometric techniques for use in soil survey. *Geoderma*, 97(3), 293-327.
- McBratney, A. B., Santos, M. M., and Minasny, B. 2003. On digital soil mapping. *Geoderma*, 117(1), 3-52.
- Metternicht, G. I., and Zinck, J. A. (2003). Remote sensing of soil salinity: potentials and constraints. *Remote sensing of Environment*, 85(1), 1-20.
- Msanya, B. M., Magoggo, J. P., and Otsuka, H. (2002). Development of soil surveys in Tanzania. *Japanese Society of Pedology*, 46 (2), 79-88.
- Moran, C. J., and Bui, E. N. (2002). Spatial data mining for enhanced soil map modelling. *International Journal of Geographical Information Science*, 16(6): 533-549.
- Nabiollahi, K., Haidari, A., and Taghizade-mehrjardi, R. (2014). Digital Mapping of Soil Texture Using Regression Tree and Artificial Neural Network in Bijar, Kurdistan. *Journal of Water And Soil*, 28(5), 1025-1036. (In Farsi)
- Nauman, T. (2009). Digital soil-landscape classification for soil survey using ASTER satellite and digital elevation data in Organ Pipe Cactus National Monument, Arizona. M. Sc. dissertation, The University of Arizona.
- Nelson, M. A., and Odeh, I. O. A. (2009). Digital soil class mapping using legacy soil profile data: a comparison of a genetic algorithm and classification tree approach. *Soil Research*, 47(6), 632-649.
- Nield, S. J., Boettinger, J. L., and Ramsey, R. D. (2007). Digitally mapping gypsic and natric soil areas using Landsat ETM data. *Soil Science Society of America Journal*, 71(1), 245-252.
- pahlavan rad, M., Khormali, F., Toomanian, N., Kiani, F., Komaki, B. (2015a). Digital soil mapping using Random Forest model in Golestan province. *Journal of Soil and Water Conservation*, 21 (6), 73-93. (In Farsi)
- pahlavan rad, M., Khormali, F., Toomanian, N., Kiani, F., Komaki, B. (2015b). Forecasting soil classes with random decision making and logistic regression methods in Golestan province. In: 14th Iranian Soil Science Congress - Genesis, Classification, Soil and Land Landscape Evaluation, 7-9 September., Iran, Rafsanjan , University of Rafsanjan, pp. 101-115.
- Sparks, D. L., Page, A. L., Helmke, P. A., Leopert, R. H., Soltanpour, P. N., Tabatabai, M. A., Johnston, C. T., Sumner, M. E. (Ed.). (1996) Methods of soil analysis. Part 3. Chemical Method. Wisconsin USA: Madison.
- Pearson, R. L., Miller, L. D., (1972). Remote mapping of standing crop biomass for estimation of the productivity of the short-grass prairie, Pawnee National Grassland, Colorado. 8th In: Proceedings of 8th International Symposium on Remote Sensing of Environment, pp. 1357-1381.
- Schoeneberger, P. J., D. A. Wysocki, E.C. Benham, and Soil Survey Staff. (2012). Field book for describing and sampling soils, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.
- Segal, D. B. (1982). Theoretical basis for differentiation of ferric-iron bearing minerals using Landsat MSS data. In: Proceedings. International Symposium on Remote Sensing of Environment, 2nd Thematic Conference, *Remote Sensing for Exploration Geology*, pp. 949-951.
- Scull, P., Franklin, J., and Chadwick, O. A. 2005. The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological modelling*, 181(1): 1-15.
- Shirali, R. (2016). Classification Trees and Rule-Based Modeling Using the C5. 0 Algorithm for Self-Image across Sex and Race in St. MSc. dissertation, Washington University St. Louis.
- Subburayalu, S. K., Jenhani, I., and Slater, B. K. (2014). Disaggregation of component soil series on an Ohio County soil survey map using possibilistic decision trees. *Geoderma* 213, 334-345.
- Taghizadeh-Mehrjardi, R., Sarmadian, F., and Omid, M., Toomanian, F., RoUsta, M. J., and Rahimian, M. H. (2015). Digital mapping of soil classes using different data mining techniques in Ardakan region, Yazd province. *Journal of Agricultural Engineering*, 37 (2), 101-115. (In Farsi)
- Taghizadeh-Mehrjardi, R., Sarmadian, F., Minasny, B., Triantafilis, J., and Omid, M. (2014). Digital mapping of soil classes using decision tree and auxiliary data in the Ardakan region, Iran. *Arid Land Research and Management*, 28(2), 147-168.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2), 127-150.