

مقاله علمی-پژوهشی

## کاربرد دو تکنیک داده‌کاوی برای تهیه نقشه پراکنش مکانی کربن آلی خاک (مطالعه موردی: کرانه شرقی دریاچه ارومیه)

امین موسوی<sup>۱</sup> - فرزین شهبازی<sup>۲\*</sup> - شاهین اوستان<sup>۳</sup> - علی اصغر جعفرزاده<sup>۴</sup> - بودیمین میناسنی<sup>۵</sup>

تاریخ دریافت: ۱۳۹۸/۰۸/۲۰

تاریخ پذیرش: ۱۳۹۹/۰۲/۰۸

### چکیده

در این پژوهش، از دو مدل جنگل تصادفی و کیوبیست به عنوان یکی از پرکاربردترین تکنیک‌های نوین داده‌کاوی برای تهیه نقشه رقومی کربن آلی خاک در ساحل شرقی دریاچه ارومیه استفاده شد. بدین منظور با استفاده از روش نمونه‌برداری تصادفی مرتب شده در منطقه‌ای به وسعت ۵۰۰ کیلومتر مربع تعداد ۱۳۱ نمونه خاک سطحی (عمق ۱۰-۰ سانتی متری) و از دوسایت جداگانه برداشت شد. متغیرهای کمکی مورد استفاده در این تحقیق شامل شش باند مستقل برگرفته از تصویر OLI ماهواره لندست ۸ (باندهای ۲ تا ۷)، تجزیه به مؤلفه‌های اصلی (PCA) باندها و همچنین تعداد ۱۴ شاخص ترکیبی مربوط به تیرماه سال ۱۳۹۶ می‌باشد. نتایج پیش‌بینی مدل در مرحله آزمون (۲۵ درصد داده‌ها) نشان داد که مدل جنگل تصادفی با مقادیر  $(ME=0/01$  و  $RMSE=0/16$ ،  $R^2=0/89$ ) صحت و کارایی بالاتری نسبت به مدل کیوبیست ( $ME=0/03$  و  $RMSE=0/21$ ،  $R^2=0/85$ ) دارد. همچنین نتایج رتبه‌بندی اهمیت متغیرهای کمکی برای پیش‌بینی کربن آلی خاک نشان داد که پارامترهای شاخص مرئی مقاومت اتمسفریک (VARI)، شاخص گیاهی نرمال شده (NDVI)، شاخص سنگی شده آهک دو (CRI2) و شاخص سنگی شده آهک یک (CRI1) دارای بیشترین تأثیر و شاخص گج (GI) و برخی باندهای مستقل از جمله باند ۵ (B5) و باند ۳ (B3) اهمیت کمتری نسبت به سایر شاخص‌ها دارند. به‌طور کلی نتایج نشان داد که مدل جنگل تصادفی در مقایسه با مدل کیوبیست به نحو مطلوبی قادر به مدل‌سازی و پیش‌بینی پراکنش مکانی کربن آلی خاک در منطقه مورد مطالعه بوده است.

**واژه‌های کلیدی:** جنگل تصادفی، کیوبیست، متغیرهای کمکی، مدل‌سازی، نقشه‌برداری رقومی

### مقدمه

لذا پژوهش‌های انجام گرفته در خاک‌های مناطق مختلف کشور می‌تواند منجر به ایجاد یک پایگاه داده از خاک‌ها شده و نیازهای کاربران مختلف را تأمین نماید.

در حال حاضر بیش از هر زمان دیگری به اطلاعات و نقشه‌های خاک با دقت و صحت بالاتری نیاز می‌باشد و برای تهیه نقشه‌هایی با قدرت تفکیک مکانی بالا و یا حتی کمی نمودن مدیریت اراضی بایستی از تحول رقومی شدن بهره گرفت (۴۱). اصلاح و مدیریت اصولی منابع خاک و همچنین تهیه نقشه‌های پوشش گیاهی و خصوصیات مختلف خاک نیاز به اطلاعات کمی، به روز و دقیق دارد و این امر با کاربرد تصاویر ماهواره‌ای که قابلیت زیادی دارند بیش از پیش آسان شده است. تهیه چنین نقشه‌هایی با استفاده از داده‌های سهل‌الوصول برگرفته از تصاویر ماهواره‌ای در مقایسه با روش‌های سنتی و معمول به‌عنوان لایه اطلاعاتی در محیط سامانه اطلاعات

خاک‌ها به عنوان یکی از عوامل اصلی و ارکان مهم در توسعه پایدار و تولید محسوب می‌شوند به‌طوری که آگاهی از ویژگی‌ها و چگونگی توزیع آن‌ها از جمله پایش و ارزیابی تغییرات کربن آلی خاک<sup>۱</sup> به عنوان یک شاخص مهم در ارزیابی حاصلخیزی، تخریب اراضی، مطالعات زیست‌محیطی و کاهش اثرات تغییرات اقلیمی می‌تواند در مدیریت مناسب آن‌ها مؤثر واقع شود (۳۵). به‌علاوه بیشتر برنامه‌های عمرانی کشور نیازمند شناسایی دقیق منابع خاک می‌باشد،

۱، ۲، ۳ و ۴- به ترتیب دانشجوی دکتری، دانشیار و استادان، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران  
(\*- نویسنده مسئول: Email: shahbazi@tabrizu.ac.ir)

۵- استاد، دانشکده علوم محیطی و کشاورزی، دانشگاه سیدنی، سیدنی، استرالیا  
DOI: 10.22067/jsw.v34i3.84154

کاربرد تکنیک‌ها و مدل‌های مبتنی بر داده‌کاوی<sup>۱۱</sup> و روش‌های ماشین یادگیری با توجه به دارا بودن نتایج دقیق‌تر در فرایند نقشه‌برداری رقومی خاک رواج گسترده‌ای یافته و مورد توجه محققان مختلفی قرار گرفته است. در مطالعه‌ای ور و همکاران (۶۲) روش‌های مختلف ماشین یادگیری شامل رگرسیون بردار پشتیبان<sup>۱۲</sup>، شبکه عصبی مصنوعی<sup>۱۳</sup> و جنگل تصادفی را در پیش‌بینی مقدار کربن آلی خاک مورد مقایسه و بررسی قرار دادند نتایج نشان داد که روش رگرسیون بردار پشتیبان با توجه به مقدار ریشه مربعات میانگین خطا<sup>۱۴</sup> بهترین روش در بین روش‌های بکار رفته بود. گریم و همکاران (۱۳) مدل جنگل تصادفی را برای پیش‌بینی کربن آلی خاک در نواحی گرمسیری استفاده نموده و دریافتند که جنگل تصادفی یک ابزار نیرومند و قوی برای تخمین مکانی و درک ارتباطات خاک-زمین نما است. اسرینواس و همکاران (۵۴) کربن آلی و غیرآلی خاک را تا عمق یک متری در سراسر هندوستان با استفاده از ۱۱۹۸ نمونه خاک و بهره‌گیری از روش جنگل تصادفی مدل‌سازی و نشان دادند که ضریب تعیین<sup>۱۵</sup> برای کربن آلی و غیرآلی به ترتیب ۰/۸۲ و ۰/۸۶ می‌باشد. وایسه و لاگاچری (۶۱) نقشه خصوصیات خاک مانند رس، شن، سیلت، کربن آلی خاک، اسیدیته و ظرفیت تبادل کاتیونی خاک را در فرانسه در سطح ۲۷۲۳۶ کیلومتر مربع با استفاده از ۲۰۲۴ نمونه خاک با چهار روش مختلف مدل‌سازی نموده و مشاهده کردند که روش جنگل تصادفی با داشتن ضریب  $R^2$  برابر با ۰/۹۹ دقت بیشتری نسبت به سایر روش‌ها داشته است. همچنین در این مطالعه، مشتقات DEM با قدرت تفکیک مکانی ۹۰ متر به عنوان عامل مهمی در فرایند مدل‌سازی معرفی شدند. برونگارد (۱۱) از روش جنگل تصادفی برای پیش‌بینی مکانی کلاس‌ها و خصوصیات خاک استفاده کرده و آن را به عنوان روشی مفید برای پیش‌بینی کلاس‌های خاک، خصوصیات خاک و توزیع پوشش گیاهی معرفی نمود. لایب و همکاران (۳۱) از روش جنگل تصادفی با موفقیت برای پیش‌بینی توزیع بافت خاک استفاده کرده و دقت بالای آن را در مقایسه با روش درخت رگرسیون گزارش کردند. پهلوان‌راد و همکاران (۴۰) از روش نمونه‌برداری مکعب لاتین شرطی<sup>۱۶</sup> و مدل جنگل تصادفی جهت به‌روزرسانی نقشه‌های خاک در استان گلستان استفاده کردند و

جغرافیایی<sup>۱</sup> با صرف هزینه و وقت کمتر و با دقت بیشتر انجام می‌شود (۱۵). برای غلبه بر محدودیت‌های روش‌های نقشه‌برداری سنتی و رایج که عمدتاً بر قابلیت‌های فردی و میزان فهم و درک کارشناس از روابط خاک و محیط استوار می‌باشند و در اغلب موارد نیز با خطای زیادی همراه می‌باشند (۶۴)، مک براتنی و همکاران (۳۴) با مطالعه مدل‌های مختلف برآورد مکانی خصوصیات مختلف خاک و با توجه به اطلاعات کلیدی آن‌ها، یک چارچوب تعمیم یافته (مدل اسکورپن<sup>۲</sup>) برای تخمین و برآورد مکانی خصوصیات خاک از طریق کمی‌سازی و بهره‌گیری از مدل رقومی ارتفاع<sup>۳</sup> و سنجش از دور<sup>۴</sup> و سایر تکنیک‌های پهنه‌بندی ارائه نمودند. مدل اسکورپن در حال حاضر پایه و اساس نقشه‌برداری نوین خاک را تشکیل می‌دهد و به طور گسترده‌ای در تحقیقات نقشه‌برداری رقومی خاک مورد استفاده قرار می‌گیرد.

تقی‌زاده مهرجردی و همکاران (۵۶) در منطقه بانه برای به‌دست آوردن اطلاعات و داده‌های دقیق و پیوسته از کربن آلی خاک با استفاده از نقشه‌برداری رقومی خاک و شاخص‌های دورسنجی و مشتقات DEM نقشه‌های پراکنش کربن آلی خاک در عمق‌های مختلف خاک را تهیه کردند. شهبازی و همکاران (۴۸) نیز نقشه توزیع مکانی نیتروژن، فسفر و بور در عمق‌های مختلف خاک را با استفاده از مشتقات DEM و شاخص‌های دورسنجی و همچنین کاربرد مدل‌های جنگل تصادفی و کیوبیست در محیط R را تهیه و ضمن گزارش ارجحیت مدل کیوبیست، مشاهده کردند که قسمت‌های غربی محدوده مطالعاتی مستعد به آلودگی بور می‌باشد. برای پیش‌بینی و برآورد ویژگی‌ها یا کلاس‌های خاک نیز مدل‌های مختلف دیگری مورد بررسی قرار گرفته‌اند که از آن جمله می‌توان به سیستم‌های تخصصی<sup>۵</sup> (۵۳ و ۵۸) طبقه‌بندی نظارت‌نشده<sup>۶</sup> (۵۷) و ماشین یادگیری<sup>۷</sup> (۱۲) اشاره نمود.

در سال‌های اخیر انواع مختلفی از روش ماشین یادگیری برای نقشه‌برداری رقومی خاک استفاده شده است که از آن جمله می‌توان به روش‌های درختی شامل مدل درخت تصمیم‌گیری<sup>۸</sup> و رگرسیون و همچنین مدل‌های جنگل تصادفی<sup>۹</sup> و کیوبیست<sup>۱۰</sup> اشاره کرد. امروزه

- 1- Soil Organic Carbon (SOC)
- 1- Geographic Information System (GIS)
- 2- SCORPAN
- 3- Digital Elevation Model (DEM)
- 4- Remote Sensing (RS)
- 5- Expert systems
- 6- Unsupervised classification
- 7- Machine learning
- 8- Decision Tree (DT)
- 9- Random Forests (RF)

- 10- Cubist (Cu)
- 11- Data mining
- 12- Support Vector Regression (SVR)
- 13- Artificial Neural Network (ANN)
- 14- Root Mean Square Error (RMSE)
- 15- Determination of coefficient ( $R^2$ )
- 16- Conditioned Latin Hypercube Sampling (cLHS)

است. بر این اساس ابتدا بر روی نقشه‌های پایه محدوده دو سایت مورد نظر (سایت‌های مورد مطالعه برای نقشه‌برداری و مدل‌سازی) به ترتیب هر یک به مساحت ۲۲۳ و ۲۷۵ کیلومترمربع واقع در ساحل شرقی منطقه مورد مطالعه بر اساس تکنیک نمونه‌برداری تصادفی مرتب شده<sup>۱</sup> موقعیت نقاط نمونه‌برداری مشخص گردید و سپس در مجموع تعداد ۱۳۱ نمونه خاک سطحی از عمق ۱۰-۰ سانتی‌متری برداشت گردید. سپس تمامی نمونه‌های مورد نظر پس از هواخشک کردن، از الک ۲ میلیمتری عبور داده شده و مقدار کربن آلی خاک به روش والکی- بلک اصلاح شده (۳۸) اندازه‌گیری شد.

### انتخاب متغیرهای کمکی دورسنجی

در این مرحله برای تهیه نقشه رقومی کربن آلی خاک بر اساس مدل اسکورپن، علاوه بر داده‌های به‌دست آمده از اندازه‌گیری کربن آلی نمونه‌های خاک مورد بررسی به عنوان داده‌های زمینی نیاز به دو سری داده کمکی از جمله داده‌های کمکی مستخرج از تصاویر ماهواره‌ای سنجنده OLI<sup>۲</sup> لندست ۸ با قدرت تفکیک مکانی ۳۰ متر و مدل رقومی ارتفاع می‌باشد. لازم به ذکر است که با توجه به اینکه منطقه مورد مطالعه دارای شرایط توپوگرافی یکنواختی می‌باشد لذا از مشتقات DEM به عنوان داده‌های کمکی در این تحقیق استفاده نشد. در ادامه پس از انجام پیش‌پردازش‌های اولیه از داده‌های کمکی مربوط به ماه جولای ۲۰۱۷ (تیرماه ۱۳۹۶) این ماهواره برای انجام مدل‌سازی و نقشه‌برداری رقومی کربن آلی خاک استفاده شد. در این بخش برخی از مهم‌ترین شاخص‌های دورسنجی و متغیرهای کمکی مربوط به منطقه مورد مطالعه از جمله شاخص نسبت گیاهی نرمال شده<sup>۳</sup> (۷)، شاخص مرئی مقاومت اتمسفریک<sup>۴</sup> (۲۱)، شاخص نسبت گیاهی<sup>۵</sup> (۷)، شاخص گیاهی تعدیل کننده اثر خاک<sup>۶</sup> (۲۰)، شاخص اصلاح شده گیاهی تعدیل کننده اثر خاک<sup>۷</sup> (۴۵)، شاخص نسبت رطوبتی نرمال شده<sup>۸</sup> (۵۲)، شاخص اصلاح شده نسبت رطوبتی نرمال شده<sup>۹</sup> (۶۳)، شاخص کمی شدت سوختن اراضی یک<sup>۱۰</sup> (۴۲)، شاخص کمی شدت سوختن اراضی دو<sup>۱۱</sup> (۴۲).

- 1- Stratified random sampling
- 2- Operational Land Imagery (OLI)
- 3- Normalized Difference Vegetation Index (NDVI)
- 4- Visible Atmospherically Resistant Index (VARI)
- 5- Ratio Vegetation Index (RVI)
- 6- Soil-Adjusted Vegetation Index (SAVI)
- 7- Modified Soil Adjusted Vegetation Index (MSAVI)
- 8- Normalized Difference Moisture Index (NDMI)
- 9- Modified Normalized Difference Water Index (MNDWI)
- 10- Normalized Burn Ratio (NBR)
- 11- Normalized Burn Ratio2 (NBR2)

مشاهده کردند که روش مذکور از قابلیت بالایی برای پیش‌بینی کلاس‌های مختلف خاک برخوردار است. در سال‌های اخیر نیز شهبازی و همکاران (۴۹) با استفاده از برخی تکنیک‌های داده‌کاوی و به کمک شاخص‌های دورسنجی نقشه رقومی آهن متبلور را در ساحل شرقی دریاچه ارومیه تهیه و گزارش نمودند که مدل رگرسیون خطی چندگانه برای پیش‌بینی مقدار رس خاک و مدل کیوبیست برای پیش‌بینی تراکم آهن کل خاک و اکسید آهن متبلور مناسب می‌باشد. به‌طور کلی از آنجایی که داده‌کاوی به‌عنوان مفهومی از شکل پیشرفته پشتیبانی از فرایند تصمیم‌به‌شمار می‌رود که با پایگاه‌های داده‌های بزرگی آمیخته شده است و در صورتی که به‌طور دقیق استفاده شود قادر خواهد بود به ابزار تحلیل پیچیده‌ای تبدیل شود و به دنبال استخراج اطلاعات مفید و ناشناخته از داده‌های خام و اولیه و کشف الگوهای مختلف بپردازد. علاوه بر این با توجه به اهمیت روش‌های جنگل تصادفی و کیوبیست به عنوان یکی از تکنیک‌ها و روش‌های نوین و کارآمد و نسبتاً جدید داده‌کاوی در نقشه‌برداری رقومی خاک که در سال‌های اخیر کارایی و انعطاف‌پذیری مطلوب و قابل قبولی در تفسیر نتایج نسبت به روش‌های معمول و قدیمی‌تر پیدا کرده‌اند لذا این پژوهش با هدف مقایسه و ارزیابی قابلیت هر یک از این دو مدل برای پیش‌بینی مقدار و تغییرات کربن آلی خاک در ساحل شرقی دریاچه ارومیه انجام گرفت.

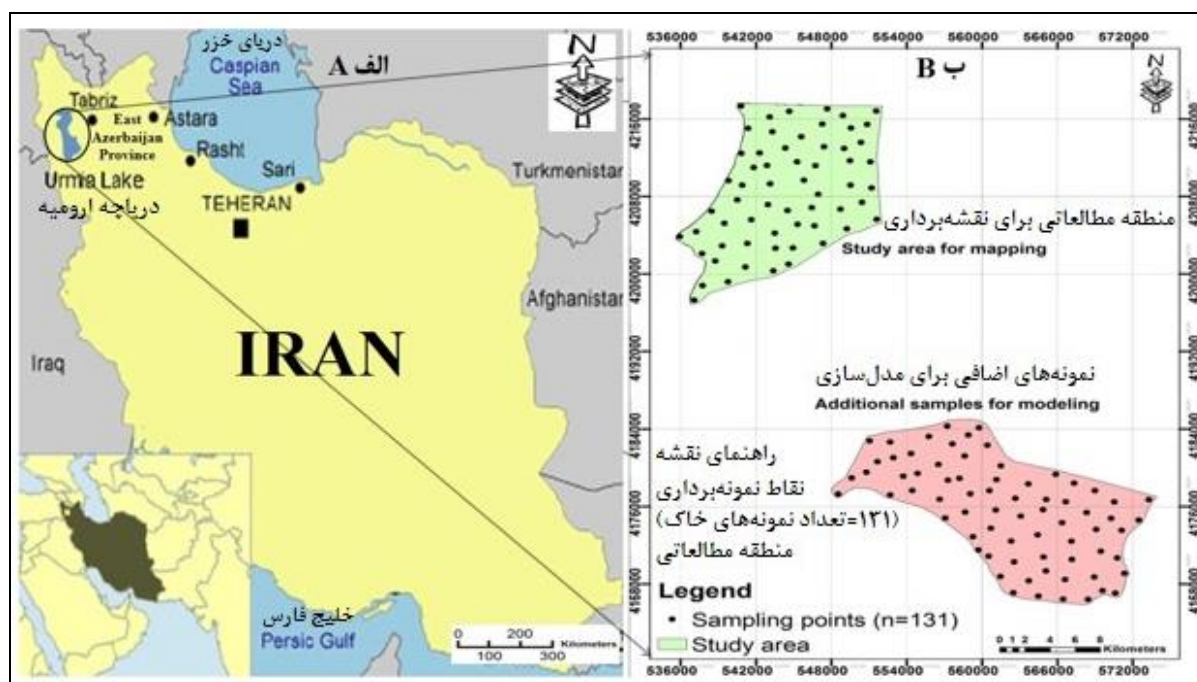
### مواد و روش‌ها

#### مشخصات منطقه مورد مطالعه

منطقه مورد مطالعه بخشی از اراضی ساحل شرقی دریاچه ارومیه با وسعتی حدود ۵۰۰ کیلومتر مربع است که بین مختصات جغرافیایی  $37^{\circ} 14' 58''$  تا  $38^{\circ} 07' 12''$  عرض شمالی و  $45^{\circ}$  تا  $46^{\circ} 07' 17''$  طول شرقی واقع شده است (شکل ۱). بر اساس داده‌های هواشناسی در یک دوره ۲۰ ساله (۱۳۹۶-۱۳۷۶) میانگین سالانه دمای منطقه  $10/83$  درجه سلسیوس و میانگین بارندگی این منطقه نیز  $345/37$  میلی‌متر در سال است (۲۸). رژیم‌های رطوبتی و حرارتی خاک منطقه به‌ترتیب زیریک و مزیک می‌باشد (۴). اراضی این منطقه از نظر فیزیوگرافی بر روی دشت رسوبی با شیب ملایم واقع شده‌اند. مواد مادری خاک‌های این منطقه بر اساس نقشه زمین‌شناسی ارومیه (۵۰) از نوع رسوبات کواترنر متشکل از پهنه‌های رسی نمکی می‌باشند که به دلیل تجمع بسیار زیاد نمک ناشی از دریاچه ارومیه به صورت اراضی غیرقابل کشت درآمده و به عنوان مرتع فصلی مورد استفاده قرار می‌گیرند.

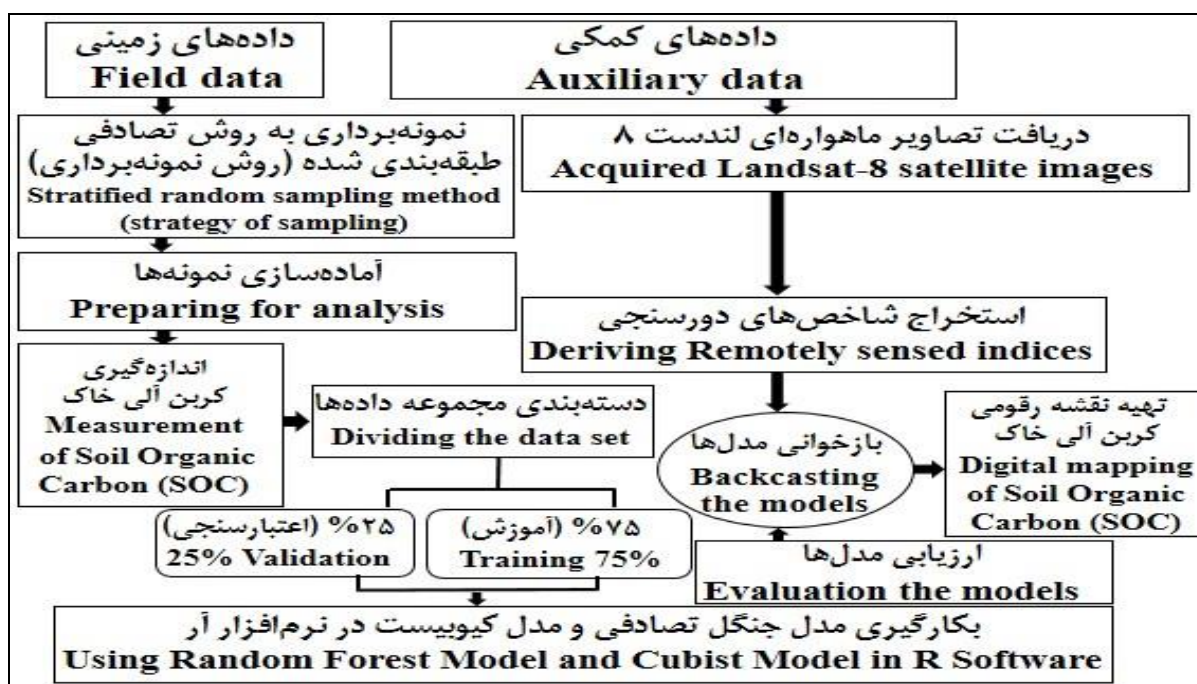
#### نمونه‌برداری و تجزیه‌های آزمایشگاهی

روند کلی چگونگی انجام تحقیق در شکل ۲ نشان داده شده



شکل ۱- الف- موقعیت دریاچه ارومیه و محدوده مطالعاتی در ایران، ب- مناطق مطالعاتی انتخاب شده برای نقشه برداری و مدل سازی در ساحل شرقی دریاچه ارومیه همراه با نقاط نمونه برداری (۱۳۱=تعداد نمونه های خاک)

Figure 1-A: The location of Urmia Lake and study area in Iran; B: The selected locations of the study area for mapping and modelling in the east shore of Urmia Lake with sampling points (Number of soil samples= 131)



شکل ۲- روندنمای روش مورد استفاده برای نقشه برداری رقومی خاک در این مطالعه  
Figure 2- The flowchart of method used for digital soil mapping in this study



مقادیر "ntree" از ۱۰۰ تا ۱۰۰۰۰ با افزایش ۱۰۰ مرتبه‌ای بهینه‌سازی شدند (۲۴). در این روش تکرارپذیری، میزان دقت پیش‌بینی‌ها براساس داده‌های خارج از بسته<sup>۶</sup> در هر مدل جنگل تصادفی به حداکثر می‌رسد. بنابراین مقدار ۱۰۰۰ برای پارامتر ntree استفاده شد.

### مدل کیوبیست

مدل کیوبیست یک الگوریتم پیشرفته از روش درخت رگرسیون است که مدل‌های مختلفی را از داده‌های آموزشی ایجاد می‌کند (۲۶). هر مدل متشکل از چندین قاعده است و هر قاعده در یک یا چندین شرط خلاصه می‌شود. وقتی همه شرایط یک قاعده تأمین می‌گردد، رابطه خطی مربوطه برای پیش‌بینی متغیر هدف بکار گرفته می‌شود. قواعد مدل بر اساس روند کاهشی اهمیت به وسیله الگوریتم کیوبیست رتبه‌بندی می‌شود. این بدان معنی است که اولین قاعده بیشترین و آخرین آن کمترین مشارکت را در دقت مدل دارا می‌باشد. بسته کیوبیست به صورت خودکار اهمیت متغیرهای کمکی را به شکل درصد مشارکت بیان و بر اساس متغیرهای مؤثر مقدار متغیر هدف را پیش‌بینی می‌کند و تعداد قواعد بر اساس بهترین مدل رگرسیونی برازش داده شده تعیین می‌شود.

در ادامه بر اساس مدل‌های ذکر شده و با استفاده از نتایج واسنجی هر یک از این مدل‌ها مقدار کربن آلی خاک در ۱۳۱ نقطه مشاهداتی با داده‌ها و اطلاعات حاصل از مهمترین متغیرهای کمکی در منطقه نقشه پیوسته کربن آلی خاک در نرم‌افزار R تهیه شد (۴۴).

### انتخاب مؤثرترین داده‌ها و متغیرهای کمکی

برای انتخاب مؤثرترین متغیرهای کمکی در پیش‌بینی کربن آلی خاک در مدل جنگل تصادفی از دستور varImp و در مدل کیوبیست از بسته Cubist در نرم‌افزار R استفاده شد (۴۶). مدل جنگل تصادفی در یک آنالیز درخت رگرسیون از معیار افزایش در میانگین مربع خطا<sup>۷</sup> (۱۰) و مدل کیوبیست بر اساس درصد مشارکت هر متغیر در قواعد مدل (۱) برای نمایش اهمیت و انتخاب مهمترین متغیرهای کمکی استفاده می‌کند.

### واسنجی و اعتبارسنجی مدل

برای ارزیابی میزان دقت هر مدل داده‌های خاک به دو دسته و مجموعه داده تصادفی از کل داده‌های مشاهده‌ای شامل آموزش ( ۷۵ درصد) و آزمون ( ۲۵ درصد) تقسیم گردیدند. به‌منظور واسنجی و

شاخص نسبت شوری<sup>۱</sup> (۳۶)، شاخص گج<sup>۲</sup> (۷)، شاخص سنگی شده آهک یک<sup>۳</sup> (۷)، شاخص سنگی شده آهک دو<sup>۴</sup> (۷) و شاخص کانی‌های رس<sup>۵</sup> (۷) با استفاده از نرم‌افزار ArcGIS و از ترکیب باندها محاسبه و تهیه گردید.

به‌طور کلی در این مطالعه از تعداد ۲۱ متغیر کمکی به عنوان لایه‌های اطلاعاتی و ورودی‌های مدل برای پیش‌بینی کربن آلی خاک بهره گرفته شد. در نهایت نیز از مدل جنگل تصادفی و کیوبیست در محیط نرم‌افزار R با استفاده از دستورات و کدهای مربوط به اجرای هر یک از این مدل‌ها برای مدل‌سازی و پیش‌بینی پراکنش مکانی کربن آلی خاک و همچنین با بهره‌گیری از اطلاعات این داده‌های کمکی و همچنین داده‌های زمینی حاصل از ۱۳۱ نمونه خاک به عنوان مشاهدات نقطه‌ای استفاده گردید.

### مدل‌های پیش‌بینی کننده با استفاده از روش‌های

#### داده‌کاوی

در این پژوهش دو مدل کاربردی و متداول داده‌کاوی برای نقشه‌برداری رقومی خاک شامل جنگل تصادفی و کیوبیست به منظور تهیه نقشه کربن آلی خاک سطحی (عمق ۱۰-۰ سانتی‌متری) مورد بررسی و مقایسه قرار گرفت.

### مدل جنگل تصادفی

مدل جنگل تصادفی به عنوان روشی توسعه یافته برای بهبود و بهسازی روش‌های درخت طبقه‌بندی و درخت رگرسیون به‌کار گرفته می‌شود. این مدل دارای دقت و قابلیت انطباق و سازگاری تقویت شده است و در انجام محاسبات از سرعت قابل قبولی برخوردار است. همچنین در انجام فرایند مدل‌سازی با جنگل تصادفی نیازی به نرمال‌سازی داده‌ها وجود ندارد (۱۰) لذا در این تحقیق نیز هیچ گونه عملیاتی برای نرمال‌سازی توزیع داده‌های اصلی صورت نگرفت. به منظور اجرای مدل جنگل تصادفی، بسته "randomForest" در نرم‌افزار R بازخوانی و نصب گردید. پارامترهای اصلی که می‌توان تصحیح کرد شامل تعداد درختان برای ساخت (ntree) و تعداد متغیرهای کمکی (mtry) است که به طور تصادفی به عنوان نمونه در هر تقسیم درخت تصمیم‌گیری جداشده (تفکیک‌شده) نمونه‌برداری می‌شوند. در این تحقیق، دو پارامتر فوق‌الذکر با تکرار کردن مقادیر "mtry" از ۱ تا ۱۵ برابر با تعداد کل متغیرهای کمکی و همچنین

- 1- Salinity Ratio (SR)
- 2- Gypsum Index (GI)
- 3- Carbonate Rock Index1 (CRI1)
- 4- Carbonate Rock Index2 (CRI2)
- 5- Clay Minerals Ratio (CMR)

6- Out of Bag (OOB)

7- Mean Square Error (MSE)

نیمه خشک) و وضعیت خاک منطقه مورد مطالعه از نظر قرار گرفتن در معرض فرسایش و تخریب خاک سبب کاهش میزان کربن آلی خاک می‌گردد. به‌طور کلی وجود این شرایط منجر به تغییرات شدید کربن آلی در منطقه شده است. همچنین نتایج به‌ترتیب نشان دهنده مقادیر چولگی و کشیدگی برابر با ۰/۴۳ و ۱/۳۴- برای کربن آلی می‌باشد که بیانگر چولگی به سمت چپ کربن آلی در مجموعه داده‌ها است. در ادامه نیز در شکل ۳ توزیع پراکنش داده‌ها به صورت نمودار جبه‌ای نشان داده شده است.

#### ارزیابی و مقایسه کارایی هر یک از مدل‌ها

نتایج ارزیابی و مقایسه کارایی هر یک از مدل‌های جنگل تصادفی و کیوبیست در مدل‌سازی و پیش‌بینی پراکنش مکانی کربن آلی خاک بر اساس داده‌های آزمون (۲۵ درصد از داده‌ها) و داده‌های آموزش (۷۵ درصد از داده‌ها) بعد از به دست آوردن قوانین مرتبط با هر یک از این مدل‌ها در جدول ۲ خلاصه و نشان داده شده است.

به‌طور کلی مدل‌های مطلوب و دارای دقت بالاتر مقادیر  $R^2$  برابر یا نزدیک به یک و RMSE و ME کوچکتر و نزدیک به صفر دارند. نتایج نشان داد که در برآورد و پیش‌بینی کربن آلی خاک در داده‌های واسنجی، مدل جنگل تصادفی دارای بیشترین مقدار  $R^2$  و کمترین مقدار RMSE به ترتیب برابر با ۰/۸۹ و ۰/۱۶ نسبت به مدل کیوبیست است که نشان دهنده دقت بیشتر این مدل است. با توجه به اندازه‌گیری عملکرد و مقادیر میانگین خطا (اریب) و نزدیکی این معیار به صفر هیچگونه خطای ارزیابی در این مطالعه مشاهده نشد که این خود نیز بیانگر کارایی و دقت بالای این مدل‌ها به ویژه مدل جنگل تصادفی می‌باشد. بنابراین می‌توان نتیجه گرفت که مدل جنگل تصادفی با بهترین عملکرد و با استفاده از شاخص‌های طیفی سنسور از دور به عنوان متغیرهای کمکی می‌تواند به‌خوبی و به‌طور موفقیت‌آمیزی کربن آلی خاک را در سراسر منطقه مورد مطالعه پیش‌بینی و شبیه‌سازی کنند. به‌طور کلی نتایج این بررسی با مقایسه نتایج واسکیوس و همکاران (۶۰) که پنج روش داده‌کاوی شامل درخت تصمیم، شبکه عصبی مصنوعی، ترکیب شبکه عصبی مصنوعی - الگوریتم ژنتیک، رگرسیون لاجیستیک چندجمله‌ای و جنگل تصادفی را برای پیش‌بینی ماده آلی خاک و مونجون و همکاران (۳۷) که برای پیش‌بینی واحد مختلف خاک از دو روش داده‌کاوی شبکه عصبی و درخت تصمیم و همچنین لوتو و جرت (۳۳) که روش‌های مختلف داده‌کاوی از جمله رگرسیون، تصمیم‌گیری درختی و شبکه عصبی را برای نقشه‌برداری سطوح ژئومورفیک در فنلاند به‌کار بردند مشابهت دارد که همه این موارد نشان از کارآمدی بهتر روش‌های مختلف داده‌کاوی در مقایسه با روش‌های معمول دارد.

اعتبارسنجی مدل (مدل‌های جنگل تصادفی و کیوبیست برای سال ۲۰۱۷) از معیارهای آماری و دقت‌سنجی مختلف مانند ضریب تبیین، میانگین ریشه مربعات خطا (رابطه ۱) و همچنین میانگین خطا<sup>۱</sup> یا همان معیار اریب یا بیاس (رابطه ۲) (به عنوان یک عامل متعادل و تنظیم‌کننده) به منظور ارزیابی و مقایسه کارایی، بیش‌برآوردی، کم‌برآوردی و اریبی پیش‌بینی هر یک از مدل‌ها جهت پیش‌بینی مقادیر کربن آلی خاک استفاده شد.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (obs_i - pred_i)^2}{n}} \quad (1)$$

$$ME = \frac{\sum_{i=1}^n (obs_i - pred_i)}{n} \quad (2)$$

در این رابطه‌ها  $obs_i$  و  $pred_i$  به ترتیب مقادیر مشاهده شده (واقعی) و پیش‌بینی شده متغیر هدف توسط مدل و یا برآوردگر آماری و  $n$  نیز تعداد مشاهدات است.

#### نتایج و بحث

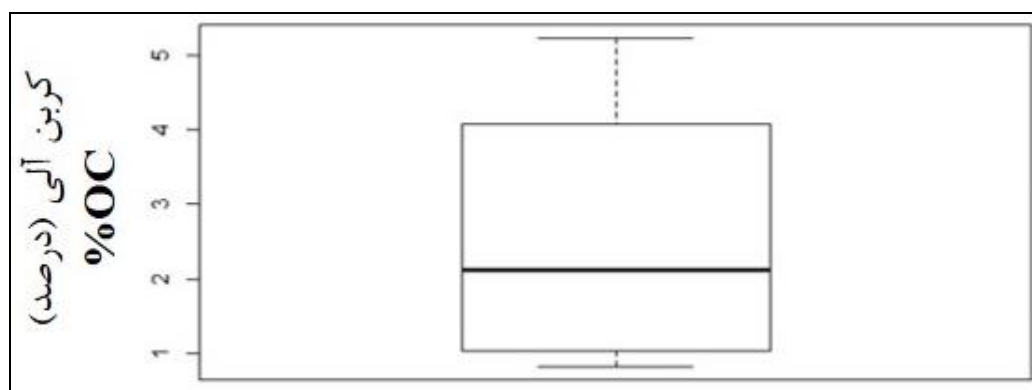
##### توصیف آماری داده‌ها

در جدول ۱ خلاصه توصیف آماری کربن آلی خاک در عمق ۱۰- سانتی‌متری آمده است. لازم به ذکر است که بر اساس آزمون کولموگروف-اسمیرنوف در نرم‌افزار SPSS داده‌های حاصل از کربن آلی دارای توزیع نرمال می‌باشند. مقادیر میانگین و میانه برای کربن آلی خاک به ترتیب برابر ۲/۵۲ و ۲/۱۱ محاسبه گردید. همچنین مقادیر حداقل و حداکثر آن نیز به ترتیب برابر با ۰/۸۳ تا ۵/۲۲ درصد تعیین گردید. علاوه بر این ضریب تغییرات کربن آلی برابر ۵۷/۹۴- به دست آمد که نشان دهنده تغییرات زیاد و پویا بودن آن در منطقه مورد مطالعه می‌باشد که بر اساس دسته‌بندی فانگ و همکاران (۱۸) این مقدار را می‌توان در دسته متوسط (بین ۱۰٪ تا ۹۰٪) طبقه‌بندی کرد. بر این اساس در منطقه مورد مطالعه ما به‌نظر می‌رسد که جایگاه و نقش عامل زمین‌نما در توزیع مکانی رطوبت خاک و پوشش گیاهی مؤثر باشد. به‌علاوه بیشترین مقدار کربن آلی خاک به میزان ۵/۲۲ درصد در مناطق باتلاقی (لجن‌زار) به‌علت داشتن دارای مقادیر بالای بقایای آلی تازه و میزان نسبتاً کم تجزیه این بقایا در طی زمان که به خاک وارد می‌شوند و کمترین آن در مناطق غیرباتلاقی (غیرلجن‌زار) و دارای پوشش گیاهی ضعیف‌تر به‌دلیل تجمع زیاد نمک حاصل از دریاچه ارومیه که در حدود ۳۵ سال اخیر به صورت اراضی غیر قابل کشت درآمده‌اند مشاهده شد و این عوامل همراه با دخیل بودن سایر عوامل از جمله شرایط اقلیمی (قرار گرفتن در مناطق خشک و

1- Mean Error (ME)

جدول ۱- توصیف آماری مقادیر کربن آلی خاک برای نمونه‌ها مورد استفاده در این مطالعه (n=131)  
Table 1- Descriptive statistics of soil organic carbon for samples used in this study (n=131)

تعداد Number	حداقل Minimum	حداکثر Maximum	متوسط Mean	چارک اول First quintile	میانه Median	چارک سوم Third quintile	انحراف معیار Standard deviation	ضریب تغییرات Coefficient of variation	چولگی Skewness	کشیدگی Kurtosis
کربن آلی خاک (%) Soil organic carbon (%)	131	0.83	2.52	1.03	2.11	4.07	1.46	57.94	0.43	-1.34



شکل ۳- نمودار جعبه‌ای مقادیر کربن آلی خاک برای نمونه‌های مورد استفاده در این مطالعه (تعداد نمونه‌های خاک = 131)  
Figure 3- Box plot of soil organic carbon values for samples used in this study (Number of soil samples= 131)

جدول ۲- کارایی مدل جنگل تصادفی و کیوبیست در مدل‌سازی و پیش‌بینی پراکنش مکانی کربن آلی خاک در این مطالعه (n=131)  
Table 2- Performance of RF and Cubist models in modelling and predicting spatial distribution of soil organic carbon in this study (n=131)

	واسنجی (تعداد ۹۸ نمونه داخل بسته)			اعتبارسنجی (تعداد ۳۳ نمونه خارج از بسته)		
	R <sup>2</sup>	RMSE	Bias	R <sup>2</sup>	RMSE	Bias
مدل جنگل تصادفی Random Forest model	0.89	0.16	0.01	0.41	0.18	0.03
مدل کیوبیست Cubist model	0.85	0.21	0.03	0.38	0.24	0.05

داشته است و در نتیجه روش رگرسیون بردار پشتیبان بهترین روش در بین روش‌های بکار رفته شناخته شد. این درحالی است که ایوبی و همکاران (۲) با استفاده از داده‌های تصاویر ماهواره‌ای لندست و دو مدل شبکه عصبی مصنوعی و رگرسیون خطی چندگانه، کربن آلی خاک را در دو منطقه سمیرم در استان اصفهان و لردگان در استان چهارمحال و بختیاری پیش‌بینی کردند. نتایج آن‌ها نشان داد که برای دو منطقه سمیرم و لردگان مدل شبکه عصبی مصنوعی پرسپترون چند لایه (به ترتیب با ضریب  $R^2$  برابر با ۰/۹۹ و ۰/۸۴) دارای کارایی

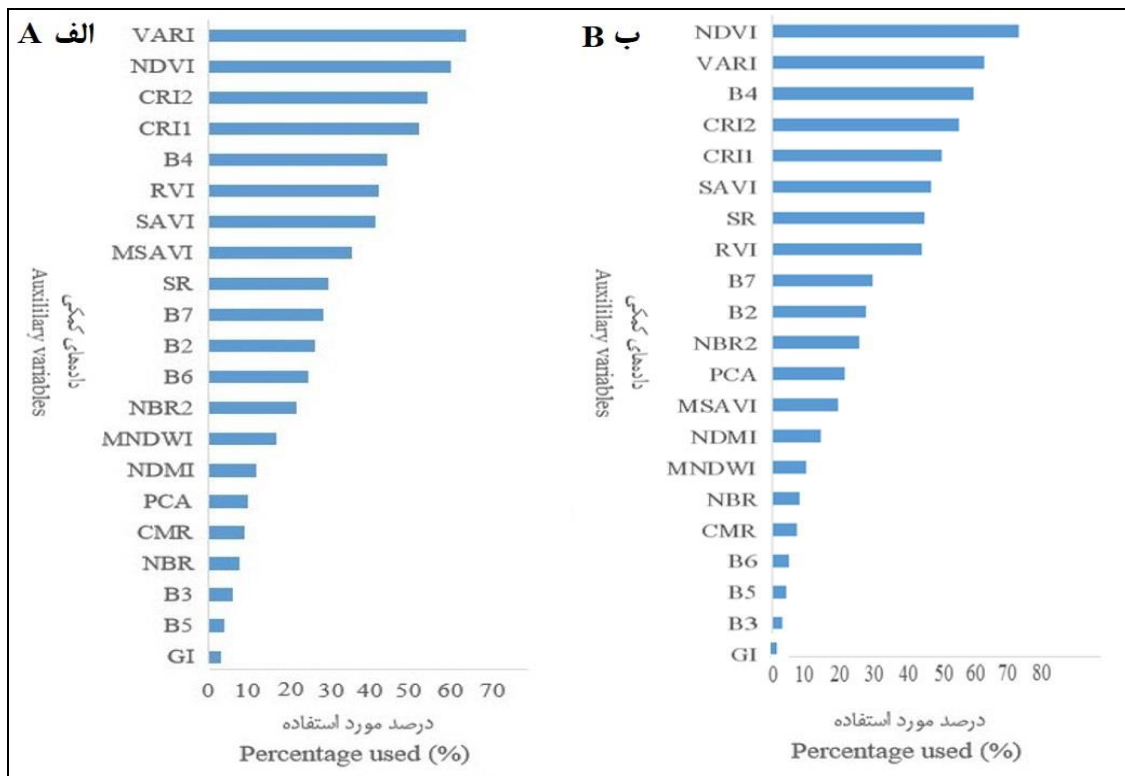
در سایر مطالعات انجام شده نیز محققین دیگری از مدل‌های مبتنی بر تکنیک‌های داده‌کاوی در نقشه‌برداری رقومی استفاده کردند. از جمله ویر و همکاران (۶۲) با استفاده از روش‌های جنگل تصادفی، رگرسیون بردار پشتیبان و شبکه عصبی مصنوعی اقدام به نقشه‌برداری رقومی و پیش‌بینی ذخایر کربن آلی خاک در کنیا کردند و به این نتیجه رسیدند که روش رگرسیون بردار پشتیبانی و شبکه عصبی مصنوعی کمترین میزان RMSE (به ترتیب ۱۵/۵ و ۱۴/۹ مگاگرم کربن بر هکتار) و بیشترین مقادیر ضریب  $R^2$  و برابر با ۰/۶ را

عنوان چهار شاخص مهم (دارای بیشترین اهمیت) برای پیش‌بینی کربن آلی خاک هستند (شکل ۴-الف) به طوری که شاخص VARI به عنوان مؤثرترین متغیر کمکی در این مدل دارای تأثیری حدود ۶۳ درصد می‌باشد. همچنین حضور متغیرهای کمکی مهم دیگری مانند RVI، SAVI و MSAVI در مدل جنگل تصادفی نشان‌دهنده این امر است که پوشش گیاهی و خاک یکی از مهم‌ترین فاکتورهای خاکسازي در منطقه مورد نظر می‌باشد. چرا که پوشش گیاهی و خاک هر منطقه یکی از ویژگی‌های مهم و تأثیرگذار بر ویژگی‌های خاک آن منطقه از جمله کربن آلی خاک می‌باشد. نتایج پژوهش‌های زیادی بیانگر رابطه میان پارامترهای زمین‌نما و تصاویر ماهواره‌ای با ویژگی‌های خاک بوده است. از جمله دای و همکاران (۱۷) توزیع مکانی کربن آلی خاک را با استفاده از پارامترهای ورودی مدل شامل ارتفاع، درجه حرارت، بارندگی (رطوبت) و شاخص NDVI و مدل شبکه عصبی مصنوعی- کریجینگ پیش‌بینی کردند. نتایج نشان داد که ضرایب همبستگی بین ماده آلی خاک با پارامترهای ارتفاع، بارندگی، درجه حرارت و شاخص NDVI به ترتیب ۰/۲۳، ۰/۱۶، ۰/۲۱ و ۰/۲ به دست آمد.

بالاتری نسبت به مدل رگرسیون خطی چندگانه (به ترتیب با ضریب  $R^2$  برابر با ۰/۷۷ و ۰/۵۴) بود.

### اهمیت متغیرهای کمکی

از آنجایی که در اغلب موارد بسیاری از ویژگی‌های غیرمرتبط در انتخاب متغیرهای کمکی کارایی الگوریتم یادگیری را کاهش می‌دهند و در برخی موارد دقت و سرعت مدل‌سازی را پایین می‌آورند بنابراین اطلاع از میزان اهمیت متغیرهای کمکی مؤثر در فرایند پیش‌بینی حائز اهمیت است. در این بین مدل جنگل تصادفی میزان اهمیت متغیرهای کمکی را نیز بر اساس اینکه اگر تعداد یک یا چند متغیر حذف شود چگونه پیش‌بینی انجام می‌شود را برآورد کرده و نمایش می‌دهد (۴۴). همچنین مدل کبویست نیز میزان اهمیت هر یک از متغیرها و داده‌های کمکی را به صورت درصد مشارکت نشان می‌دهد. میزان اهمیت و رتبه‌بندی هر یک از متغیرهای بکار رفته در مدل‌سازی به صورت درصد برای پیش‌بینی پراکنش مکانی کربن آلی خاک با استفاده از هر دو مدل در شکل ۴ نشان داده شده است. نتایج رتبه‌بندی اهمیت متغیرهای کمکی در مدل جنگل تصادفی نشان داد که شاخص‌های VARI، NDVI، CRI2 و CRI1 به



شکل ۴- رتبه‌بندی اهمیت متغیرهای محیطی برای پیش‌بینی کربن آلی خاک در این مطالعه با استفاده از الف- مدل جنگل تصادفی و ب- مدل کبویست

Figure 4- Covariate importance rankings for predicting soil organic carbon in this study using A: RF model and B: Cubist model



کنترل توان ذخیره رسوبات و حفاظت کربن آلی خاک مطالعاتی صورت گرفته است. نتایج تحقیقات مختلف نشان داده است که تغییر در کربن آلی خاک با تغییرات ساختار و میزان ثبات خاک در ارتباط است (۶). به طوری که طبیعت پیوندهای آلی و مینرالوژیکی و همچنین نحوه توزیع آن در ارتباط با خاکدانه‌های خاک میزان تثبیت شیمیایی آن را تعیین می‌کنند (۴۷). طی تحقیقی در مراتع نیمه‌خشک مدیترانه در کشور تونس به بررسی ارتباط ویژگی‌های خاک با میزان کربن آلی خاک دو نوع خاک رسی و خاک شنی پرداخته شد. یافته‌های این مطالعه بیانگر آن است که میزان کربن آلی ذخیره‌ای خاک با ویژگی‌های فیزیکی و شیمیایی مانند وزن مخصوص ظاهری، میزان شن، سیلت، رس و اسیدیته در ارتباط است و تفاوت در نوع خاک می‌تواند تفاوت در این ویژگی‌ها را توجیه کند (۹). بین شاخص‌های پوشش گیاهی، میزان رطوبت خاک و دمای سطح اراضی به مثکله جهانی<sup>۳</sup> معروف است. پوشش گیاهی و دمای سطح اراضی وابستگی پیچیده‌ای با رطوبت خاک دارند. شاخص‌های پوشش گیاهی اطلاعاتی در مورد وضعیت رشد و شرایط پوشش گیاهی سطح اراضی ارائه می‌دهد و از طرفی دمای سطح اراضی نیز وضعیت رطوبت خاک را منعکس می‌کند، بنابراین ترکیب این دو می‌تواند برای آشکار نمودن وضعیت رطوبتی و میزان کربن آلی خاک رابطه‌ای مطلوب ارائه کند. علاوه بر این کارلسون و همکاران (۱۳) دریافتند که رابطه فراگیری بین مقدار رطوبت خاک با شاخص NDVI در مناطق دارای تنوع پوشش گیاهی (پوشش گیاهی متراکم، تنک و خاک عاری از پوشش گیاهی) و کربن آلی خاک و دمای سطح اراضی وجود دارد. همچنان که در رتبه‌بندی میزان اهمیت متغیرهای کمکی مشاهده گردید، شاخص GI و برخی باندهای مستقل از جمله B3 و B5 اهمیت کمتری نسبت به سایر شاخص‌های ترکیبی دارا بودند (شکل ۴-الف). بنابراین می‌توان نتیجه گرفت که اهمیت شاخص‌های طیفی مستقل و ترکیبی برای پیش‌بینی پراکنش مکانی کربن آلی خاک متفاوت است. همچنین طبق شکل ۴-ب در انتخاب متغیرهای کمکی مؤثر در مدل کیویست و پیش‌بینی کربن آلی خاک به ترتیب و تقریباً مشابه با مدل جنگل تصادفی شاخص‌های مربوط به پوشش گیاهی و خاک شامل NDVI، VARI، SAVI، RVI، MSAVI و پس از آن شاخص رطوبتی NDMI از اهمیت بالاتر و شاخص GI و برخی باندهای مستقل از جمله B3، B5 و B6 از اهمیت کمتری نسبت به سایر شاخص‌ها ترکیبی برخوردار هستند.

سایندای هیورا (۵۱) در مقایسه تکنیک‌های مختلف نقشه‌برداری رقومی خاک برای مقادیر کربن آلی و رس در فلات مرکزی بوروندی از ۱۶ متغیر کمکی از جمله متغیرهای مشتق شده از نقشه‌های خاک موجود، مشتقات مدل رقومی ارتفاع و تصاویر ماهواره‌ای لندست استفاده کردند و نشان دادند که بهر ترتیب پارامترها در مدل‌سازی شامل واحدهای سنگ شناسی، مشتقات مدل رقومی ارتفاع و شاخص NDVI بود. تقی‌زاده و همکاران (۵۶) برای تهیه نقشه رقومی کربن آلی خاک در منطقه بانه کردستان، از متغیرهای کمکی داده‌های سنجش از دور، پارامترهای اراضی و زمین‌نما و مدل‌های شبکه عصبی مصنوعی، جنگل تصادفی، رگرسیون لجستیک، درخت تصمیم و الگوریتم ژنتیک استفاده کردند و به این نتیجه رسیدند که مهم‌ترین متغیرهای کمکی جهت پهنه‌بندی کلاس‌های خاک و کربن آلی خاک شاخص NDVI، شاخص رس، شیب جهت شیب، انحنای سطح، شاخص همواری دره با درجه تفکیک بالا<sup>۱</sup> و شاخص خیسی توپوگرافی<sup>۲</sup> بودند. یکی از اجزای با اهمیت دیگر در این مدل، شاخص‌های مربوط به آب و یا رطوبت از جمله NDMI و MNDWI است که این شاخص‌ها نشانگر تأثیر و میزان تجمع رطوبت در خاک می‌باشد. به طور کلی این شاخص‌های رطوبتی از تلفیق با شاخص NDVI برای بررسی وضعیت خشکسالی و همچنین برآورد مقدار رطوبت خاک مورد استفاده قرار می‌گیرند. علاوه بر این بر اساس نتایج رتبه‌بندی اهمیت متغیرهای کمکی در شکل‌های ۴-الف و ۴-ب هر دو شاخص مربوط به آب یا رطوبت (NDMI و MNDWI) به عنوان متغیرهای کمکی مهم و با اهمیت در پیش‌بینی کربن آلی در هر دو مدل جنگل تصادفی و کیویست شناخته شدند. البته در مدل جنگل تصادفی شاخص MNDWI با تأثیر حدود ۲۰ درصد اهمیت بیشتری نسبت به دیگر شاخص مرتبط با رطوبت یعنی NDMI نشان داد. از طرفی دیگر مطالعه شاخص‌های رطوبتی و متغیرهای اقلیمی، ویژگی‌های محیطی، توپوگرافیکی و مواد مادری با میزان ذخیره کربن آلی در مناطق مختلف می‌تواند برای پیش‌بینی میزان کربن آلی خاک و شبیه‌سازی اثرگذاری تغییرات متغیرهای کمکی بر میزان کربن آلی خاک مفید باشد. بر این اساس لیو و همکاران (۳۲) نیز در مطالعه‌ای برای پیش‌بینی کربن آلی خاک شاخص‌های مربوط به آب یا رطوبت (NDMI و MNDWI) را به عنوان متغیرهای کمکی مهم و با اهمیت در پیش‌بینی کربن آلی خاک یافتند و طی این مطالعه خاک‌های دارای مقادیر کربن آلی بیشتر را جزو مناطق مستعد برای دریافت رسوبات نشان دادند. در زمینه نقش خصوصیات خاک در تعیین و

1- Multi-resolution Valley Bottom Flatness (MrVBF)

2- Topographic Wetness Index (TWI)

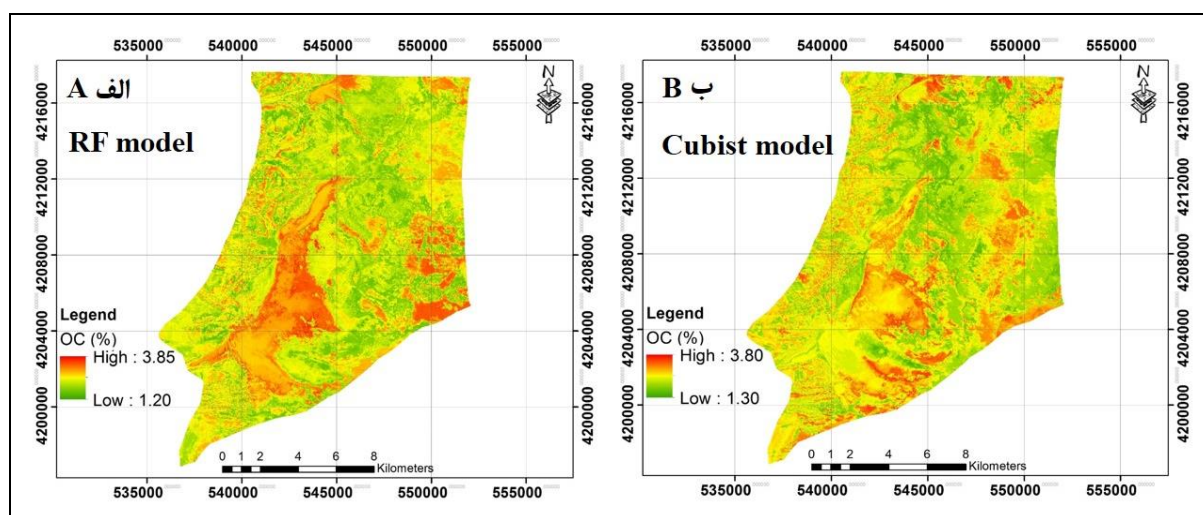
3- Universal Triangle

۵-الف و شکل ۵-ب) بیشترین مقادیر کربن آلی خاک در قسمت‌های مرکزی، شرقی و جنوبی قرار دارد که این بخش‌ها عمدتاً مناطق تحت پوشش باتلاق هستند که در اطراف دریاچه قرار گرفته‌اند و دارای کربن آلی خاک بالایی هستند که علت این امر به دلیل عدم وجود شرایط هوازی برای تجزیه می‌باشد. همچنین میزان کربن آلی خاک در قسمت‌های شمالی و غربی کاهش می‌یابد که دلیل آن دارابودن پوشش گیاهی کمتر در مناطقی که به‌عنوان مراتع فصلی برای چرای دام مورد استفاده قرار می‌گیرند می‌باشد. همچنین دیدم‌زار بودن اراضی منطقه مورد مطالعه می‌تواند مزید بر علت باشد و به‌عنوان عاملی منجر به کاهش کربن آلی خاک شود. طبق این نتایج و همچنین به‌علت انتخاب نمونه‌های خاک از دو منطقه مجزا تغییرات کربن آلی خاک در سراسر منطقه مورد مطالعه یکسان و یکنواخت نبوده و این بدان معنا است که منطقه مورد مطالعه بسیار پویا بوده و از بعد مکانی دچار تغییر و تحولات ناشی از خشک و مرطوب شدن متوالی خاک شده است.

### نقشه‌برداری رقومی و پهنه‌بندی کربن آلی خاک منطقه مورد مطالعه

پس از مدل‌سازی نقشه پراکنش مکانی کربن آلی خاک برای تیرماه سال ۱۳۹۶، با استفاده از متغیرهای کمکی مربوطه و با قدرت تفکیک مکانی ۳۰ متر با استفاده از دو روش جنگل تصادفی (شکل ۵-الف) و کیوبیست (شکل ۵-ب) در محیط R تهیه شد. پس از تهیه نقشه‌های رقومی، میانگین مقادیر عددی پیکسل‌ها که نشان دهنده میانگین مقادیر پیش‌بینی شده کربن آلی خاک با استفاده از هر دو مدل در منطقه مورد مطالعه می‌باشد تعیین گردید (جدول ۳).

بررسی‌های بیشتر به کمک نقشه‌های تهیه شده حاکی از آن است که مقدار متوسط کربن آلی با استفاده از مدل کیوبیست با مقدار متوسط واقعی نمونه‌ها (جدول ۱) همخوانی بیشتری داشته است ولی با استناد به معیارهای آماری مدل‌سازی به‌دلیل پایین بودن مقدار RMSE در این تحقیق مدل جنگل تصادفی به مراتب مدل بهتری معرفی گردید. همچنین بر اساس نقشه رقومی کربن آلی خاک (شکل



شکل ۵- پیش‌بینی پراکنش مکانی کربن آلی خاک در منطقه مورد مطالعه با استفاده از الف- مدل جنگل تصادفی و ب- مدل کیوبیست  
Figure 5- Predicted spatial distribution of soil organic carbon in study area using A: RF model and B: Cubist model

جدول ۳- میانگین مقادیر پیش‌بینی شده همراه با انحراف معیار کربن آلی خاک با استفاده از مدل جنگل تصادفی و کیوبیست در سال ۲۰۱۷  
Table 3- The mean predicted values associated with standard deviations of soil organic carbon using RF and Cubist models in 2017

	مدل جنگل تصادفی RF model	مدل کیوبیست Cubist model
کربن آلی خاک (%) Soil organic carbon (%)	2.53	2.58
انحراف معیار *Standard deviation	1.27	1.33

\*: انحراف معیار برای هر میانگین مقادیر پیش‌بینی شده کربن آلی خاک در این مطالعه برای سال ۲۰۱۷

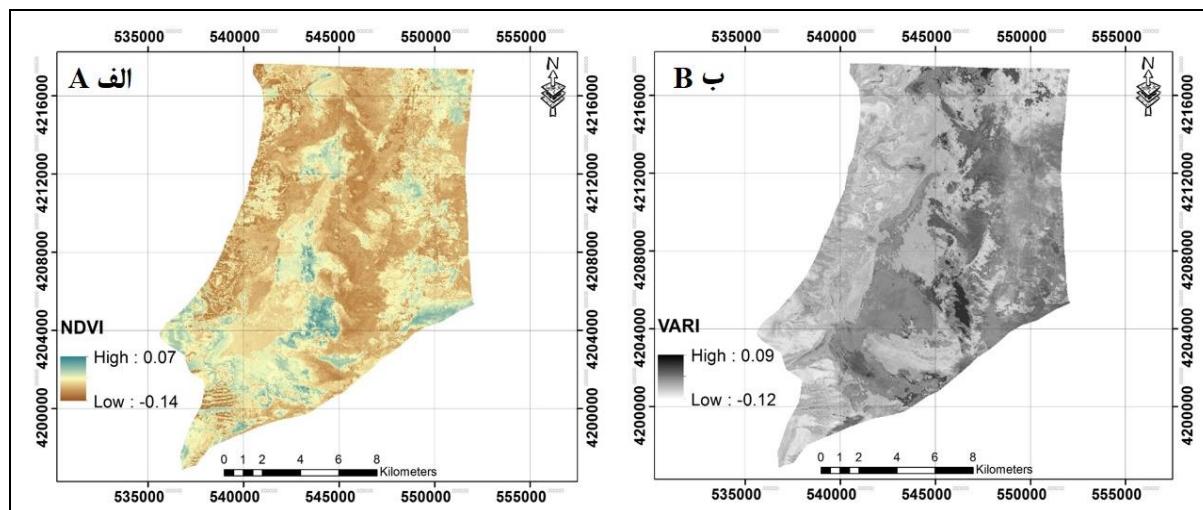
\*: Standard deviation for each predicted mean value of soil organic carbon at the time scenario 2017.

مثبت بین کربن آلی خاک و مقدار پوشش گیاهی، انتظار بر این است که مقادیر بیشتری از کربن آلی در این نواحی مشاهده شود که نقشه‌های رقمی تهیه شده و پهنه‌بندی کربن آلی خاک توسط مدل جنگل تصادفی و کیویست نیز مؤید این مطلب می‌باشد. همچنین یک دلیل احتمالی برای افزایش مقادیر پیش‌بینی شده کربن آلی در این زمان و در این مناطق را می‌توان به همبستگی مثبت بین کربن آلی با رطوبت، کاربری اراضی و نوع بافت خاک از جمله رسی بودن خاک (۱) نسبت داد. از آنجایی که سه مؤلفه مهم و کنترل‌کننده میزان کربن آلی خاک تحت تأثیر عوامل اقلیمی (میزان دما و رطوبت)، عوامل بیولوژیکی (ترکیب گیاهی و اضافه شدن بقایای گیاهی) و عوامل و ویژگی‌های فیزیکی و شیمیایی خاک مانند بافت خاک، میزان رس، کانی‌شناسی و اسیدیته می‌باشد (۴۳). عموماً میزان کربن آلی توسط این عوامل کنترل می‌شود. اما در مناطق نیمه‌خشک میزان ماده آلی توسط ویژگی‌های خاک با توجه به کمبود برگشت بقایای گیاهی و دمای بالای این مناطق اتفاق می‌افتد (۹). به عبارت دیگر در مناطق استوایی و گرم و مرطوب پوشش گیاهی و کانی‌شناسی مهم‌ترین عوامل کنترل‌کننده میزان ماده آلی خاک هستند در حالی که در مناطق نیمه‌خشک ماده آلی با توجه به کمبود برگشت بقایای گیاهی و لاشبرگ و بالا بودن دما بیشتر توسط عوامل فیزیکی و شیمیایی خاک کنترل می‌شود (۵). بر این اساس مقدار ذخیره کربن آلی خاک با مقدار و درصد پوشش گیاهی، لاشبرگ، رطوبت اشباع خاک رابطه مثبت و قوی دارد. همچنین مقدار کربن ذخیره شده در اکوسیستم‌های مختلف بیشتر تحت تأثیر زیتوده ریشه، رس و نیتروژن خاک، عامل ارتفاع از سطح دریا و میانگین بارش سالیانه قرار دارد (۳). علاوه بر این مشخصه درصد رس و نیتروژن خاک (۵۹) و بافت خاک، نیتروژن و نسبت کربن به نیتروژن (۳۹) به‌عنوان مهم‌ترین اجزای تأثیرگذار خاک بر مقدار کربن آلی گزارش شده‌اند. علاوه بر این ماده آلی به‌عنوان عاملی در ثبات و پایداری خاکدانه‌ها نقش دارد (۱۶). به‌طوری که به نظر می‌رسد تجزیه ماده آلی خاک در خاک‌های سطحی با بافت ریز کمتر از خاک‌هایی با بافت درشت است و ذرات رس و سیلت ریز میزان بالایی از کربن آلی را در خود حفظ می‌کنند (۲۹). علاوه بر این عواملی مانند بارندگی، دما و رس و بافت خاک در ارتباط با میزان کربن آلی خاک مهم ارزیابی شده‌اند (۲۷). به‌طور کلی تغییرات کربن آلی خاک در ارتباط با عواملی از جمله نوع کاربری اراضی، شکل اراضی، پایداری زمین‌نما<sup>۲</sup> و نوع و دامنه رفتارهای مدیریتی قرار دارد (۳۴). همچنین نوع کاربری اراضی، عوامل ژئومورفولوژی، توپوگرافی، پوشش گیاهی نیز به‌عنوان عوامل

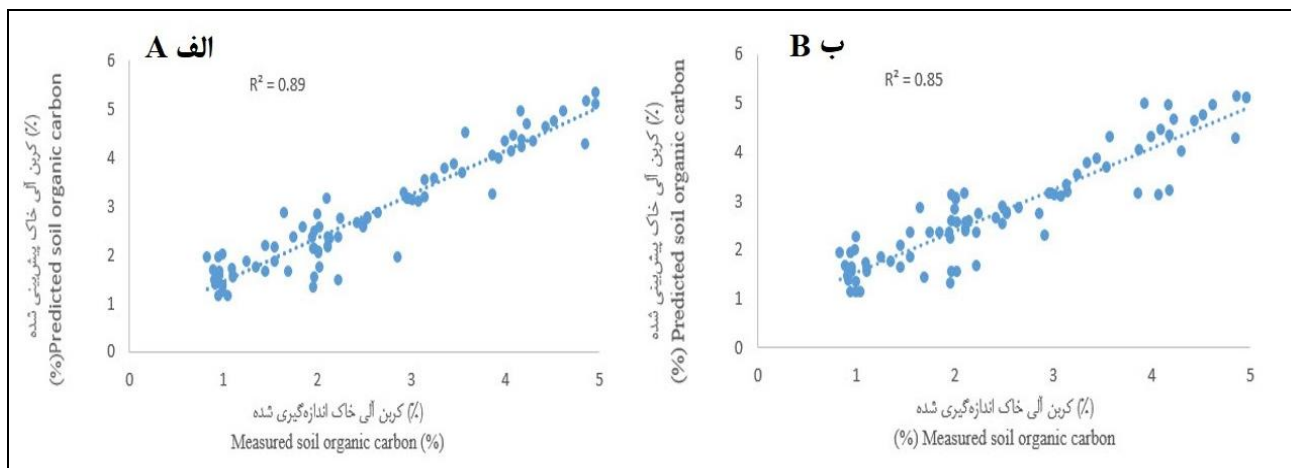
در این راستا هائسه و همکاران (۲۳) به اهمیت نقشه‌های پراکنش مکانی برای شناسایی و تشخیص فرایندهای مختلف رخ داده در خاک اشاره نموده‌اند. همچنین بونفاتی و همکاران (۸) در نقشه‌برداری رقمی خاک در جنوب برزیل بیان کردند که خاک‌های تحت پوشش مراتع دارای مقادیر بالاتر کربن آلی (۲۰۵ تن کربن بر هکتار)، ریشه‌های عمیق سبب تجمع کربن آلی تا عمق خاک می‌شود. خاک‌های تحت کشت محصولات زراعی و تاکستان کمترین مقدار کربن آلی (۱۶۳ تن کربن بر هکتار) را به دلیل کاهش ورود مواد آلی و افزایش تجزیه دارند. کومار و همکاران (۳۰) نیز در نقشه‌برداری رقمی کربن آلی خاک با استفاده از روش رگرسیون کریجینگ در مقایسه میزان کربن آلی ذخیره در خاک کاربری‌های مختلف در پنیسلوانیا نشان داد که اراضی زراعی عموماً مقادیر پایین‌تری ذخیره کربن آلی نسبت به ظرفیت بالقوه خود دارند. باتلاق‌ها در پنیسلوانیا حدود ۲/۳ درصد (۰/۰۲۷ پتاگرم) از مجموع (۱/۱۲ پتاگرم) ذخایر کربن آلی خاک را به خود اختصاص می‌دهند. شرایط بی‌هوازی طولانی مدت خاک‌ها در باتلاق‌ها، میزان تجزیه بقایای آلی را در مقایسه با سایر کاربری‌ها کاهش داده و باعث کاهش ذخیره کربن آلی خاک در آن‌ها می‌شود. کلایک (۱۴) هم در بررسی تغییر کاربری بر میزان ماده آلی خاک در سه کاربری جنگل، مرتع و اراضی زراعی نشان داد که درصد ماده آلی خاک اندازه‌گیری شده در اراضی زراعی در لایه صفر تا ده سانتی‌متر ۴۴ و ۴۸ درصد و در لایه صفر تا ۲۰ سانتی‌متری ۴۸ و ۵۰ درصد در مقایسه با کاربری جنگل و مرتع کاهش پیدا کرده است. نتایج برگرفته از این پژوهش‌ها توجیه‌کننده این نتیجه است که در قسمت‌های باتلاقی منطقه مورد مطالعه هم به علت عدم وجود شرایط بی‌هوازی جهت تجزیه، مقدار کربن آلی خاک بیش از قسمت‌های دارای کاربری مرتعی می‌باشد. شهبازی و همکاران (۴۷) نیز فرایند رسیدن خاک<sup>۱</sup> در ساحل شرقی دریاچه ارومیه را با استفاده از نقشه‌های رقمی آهن بلورین به‌خوبی نشان داده‌اند. همچنین اسرینیواس و همکاران (۵۴) در تحقیقی گزارش کردند که به علت وابستگی شدید کربن آلی به پوشش گیاهی، نقشه‌های رقمی پراکنش کربن آلی خاک می‌توانند در زمینه چگونگی اجرای عملیات و شیوه‌های مدیریت اراضی به‌کار گرفته شوند. بر اساس مشاهدات میدانی و شاخص‌های NDVI و VARI مربوط به سال ۲۰۱۷ (شکل ۶-الف و شکل ۶-ب) بیشترین مقدار این شاخص‌ها به ترتیب به میزان ۰/۰۷ و ۰/۰۹ در قسمت‌های مرکزی، شرقی و جنوبی منطقه مطالعاتی که دارای پوشش گیاهی بیشتری می‌باشند ملاحظه گردید و به همین دلیل با توجه به وجود همبستگی

حاصل از مقایسه مدل‌های جنگل تصادفی و کیوبیست بر اساس داده‌های آزمون و همچنین نمودار پراکنش مقادیر کربن آلی اندازه‌گیری شده (واقعی) در مقابل مقادیر پیش‌بینی شده نشان داد که مدل جنگل تصادفی با داشتن ضریب تعیین ۰/۸۹ به عنوان بهترین مدل از قابلیت و دقت بالاتری در پیش‌بینی و تخمین تغییرات مقدار کربن آلی خاک نسبت به مدل کیوبیست در منطقه مورد مطالعه برخوردار است (شکل ۷).

مؤثر بر تغییر مکانی کربن آلی خاک نقش دارند (۵۵). با این توصیفات آشکار می‌شود میزان ذخیره کربن آلی خاک عمدتاً به میزان تولید اولیه و نحوه تجزیه مواد آلی بستگی دارد و هر دو مؤلفه تولید و تجزیه مواد آلی شدیداً توسط آب و هوا و ویژگی‌های خاک مانند بافت، میزان مواد غذایی و آب قابل دسترس کنترل می‌گردد که در نهایت میزان جریان ماده آلی به داخل خاک و کیفیت و سرعت تجزیه آن را کنترل می‌کند (۱۹).  
در ادامه بررسی نقشه‌های پیش‌بینی کربن آلی خاک و نتایج



شکل ۶- الف - شاخص طیفی نسبت گیاهی نرمال شده (NDVI)، ب - شاخص مرئی مقاومت اتمسفریک (VARI) در این مطالعه  
Figure 6- A: Spectral indices normalized difference vegetation index (NDVI); B: Spectral indices visible atmospherically resistant index (VARI) in this study



شکل ۷- نمودار پراکنش مقادیر اندازه‌گیری شده (واقعی) و مقادیر پیش‌بینی شده کربن آلی خاک در منطقه مورد مطالعه با استفاده از الف - مدل جنگل تصادفی و ب - مدل کیوبیست

Figure 7- Scatter gram of measured and predicted values of soil organic carbon in study area using A: RF model and B: Cubist model



در ساحل شرقی دریاچه ارومیه نشان دهد. تهیه نقشه‌های رقومی پراکنش مکانی کربن آلی خاک توسط مدل مذکور می‌تواند بهره‌وری استفاده از اراضی منطقه را بهبود بخشد که نتایج مشابهی توسط برخی پژوهشگران نیز گزارش شده است (۴۰ و ۲۵). به‌طور کلی می‌توان نتیجه گرفت که استفاده از نقشه‌برداری رقومی خاک و داده‌های سهل‌الوصول در ایران که با محدودیت داده‌های خاک نیز روبرو می‌باشد نه تنها موجب صرفه‌جویی در وقت و هزینه می‌شود بلکه راهگشایی برای انجام سایر مطالعات علمی مرتبط با خاک و محیط زیست نیز می‌باشد.

## نتیجه‌گیری

نتایج این بررسی‌ها نشان داد که هرچند متغیرها و داده‌های کمکی اهمیت متفاوتی در پیش‌بینی پراکنش کربن آلی دارا بودند اما به‌طور کلی می‌توان با یافتن ارتباط بین آن‌ها و کربن آلی خاک از طریق مدل‌هایی مانند مدل جنگل تصادفی و مدل کیوبیست نقشه رقومی پیوسته کربن آلی خاک را با دقتی قابل قبول تهیه نمود. بنابراین می‌توان بیان نمود که الگوریتم و مدل جنگل تصادفی به عنوان راه‌حلی ساده، کم‌هزینه و با صرف زمانی اندک و با دقت پیش‌بینی بالا به‌خوبی توانسته تغییرات و پراکنش کربن آلی خاک را

## منابع

- 1- Adhikari K., Hartemink A.E., Minasny B., Bou Kheir R., and Greve M.B. 2014. Digital mapping of soil organic carbon contents and stocks in Denmark. PLoS One 9(8):1-13.
- 2- Ayoubi S., Shahri A.P., Karchegani P.M., and Sahrawat K.L. 2011. Application of artificial neural network (ANN) to predict soil organic matter using remote sensing data in two ecosystems. In Tech Publication 181-196.
- 3- Azarnivand H., Joneidi H., Zare Chahooki M.A., and Maddah Arefi H. 2011. Investigation of the effects of some ecological factors on carbon sequestration in *Artemisia sieberi* rangelands of Semnan province. Journal of Range and Watershed Management 64(1): 107-127. (In Persian with English abstract)
- 4- Banaei M.H. 1998. Soil moisture and temperature regime map of Iran. Soil and Water Research Institute. Ministry of Agriculture, Tehran, Iran.
- 5- Bernoux M., Carvalho M.C.S., Volkoff B., and Cerri C.C. 2002. Brazil's soil carbon stocks. Soil Sciences Society American Journal 66(3): 888-896.
- 6- Bicheldey T.K., and Latushkina E. 2010. Biogas emission prognosis at the landfills. International Journal of Environmental Science and Technology 7(4): 623-628.
- 7- Boettinger J.L., Ramsey R.D., Bodily J.M., Cole N.J., Kienast-Brown S., Nield S.J., Saunders A.M., and Stum A.K. 2008. Landsat spectral data for digital soil mapping, in: Hartemink A.E., McBratney A.B., and Mendonca-Santos M. (eds.) Digital Soil Mapping with Limited Data. Springer, Dordrecht. p. 193-203.
- 8- Bonfatti B., Hartemink A.E., Giasson E., Tornquist C.G., and Adhikari K. 2016. Digital mapping of soil carbon in a viticultural region of Southern Brazil. Catena 261: 204-221.
- 9- Brahim N., Blavet D., Gallali T., and Bernoux M. 2011. Application of structural equation modeling for assessing relationships between organic carbon and soil properties in semiarid Mediterranean region. International Journal of Environmental Science and Technology 8(2): 305-320.
- 10- Breiman L. 2001. Random forests. Machine Learning 45: 5-32.
- 11- Brungard C.W. 2009. Alternative Sampling and Analysis Methods for Digital Soil Mapping in Southwestern Utah. Thesis for Master of Science, Utah State University, USA.
- 12- Brungard C.W., Boettinger J.L., Duniway M.C., Wills S.A., and Edwards T.C. 2015. Jr. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239-240: 68-83.
- 13- Carlson T., Gillies R., and Perry E. 1994. A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover. Remote Sensing Reviews 9: 1-2.161-173.
- 14- Celik I. 2005. Land-use Effects on Organic Matter and Physical Properties of Soil in a Southern Mediterranean Highland of Turkey. Soil and Tillage Research 83: 270-277.
- 15- Chitsaz V. 1999. Investigation of Soil Salinity and Alkalinity Map in Isfahan East Area Using Digital Data. M.Sc. Degree in Desertification, Faculty of Natural Resources, Isfahan University of Technology.
- 16- Christensen B.T. 1996. Carbon in primary and secondary organomineral complexes. In structure and organic matter storage in agricultural soils in: Cater, M.R.; Stewart, B.A., (eds.) 97-165. CRC Press, Boca Raton.
- 17- Dai P.F., Qigang Z., Zhiqiang L.V., Xuemei W., and Gangcai L. 2014. Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau. Ecological Indicators 45: 184-194.
- 18- Fang X., Xue Z.J., Li B.C., and An S.S. 2012. Soil organic carbon distribution in relation to land use and its storage



- in a small watershed of the Loess Plateau, China. *Catena* 88: 6-13.
- 19- Garcia-Pausas J., Casals P., Camarero L., Hugué C., Sebastia M.T., Thompson R., and Romanya J. 2007. Soil organic carbon storage in mountain grasslands of the Pyrenees: effects of climate and topography. *Biogeochemistry* 82: 279-289.
  - 20- Gilabert M.A., Gonzalez-Piqueras J., Garcia-Haro F.J., and Melia J. 2002. A generalized soil adjusted vegetation index. *Remote Sensing of Environment* 82: 303-310.
  - 21- Gitelson A.A., Kaufman Y.J., Stark R., and Rundquist D. 2002. Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment* 80: 76-87.
  - 22- Grimm R., Behrens T., Märker M., and Elsenbeer H. 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island—digital soil mapping using random forests analysis. *Geoderma* 146: 102-113.
  - 23- Haese R.R., Wallmann K., Dahmke A., Kretzmann U., Muller P.J., and Schulz H.D. 1997. Iron species determination to investigate early diagenetic reactivity in marine sediments. *Geochimica et Cosmochimica Acta* 61: 63-72.
  - 24- Hengl T., Heuvelink G.B., Kempen B., Leenaars J.G., Walsh M.G., Shepherd K.D., Sila A., MacMillan R.A., de Jesus J.M., Tamene L., and Tondoh, J.E. 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLoS One* 10: 1-26.
  - 25- 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.
  - 26- Holmes G., Hall M., and Frank E. 1999. Generating rule sets from model trees. In: Foo N. (ed.) *Advanced Topics in Artificial Intelligence. Lecture Notes in Artificial Intelligence*, 1-12.
  - 27- Hontoria C., Rodríguez-Murillo J.C., and Saa A. 1999. Relationships between soil organic carbon and site characteristics in Peninsular Spain. *Soil Science Society of America Journal* 63(3): 614-621.
  - 28- IRIMO, 2012. Islamic Republic of Iran Meteorological Organization.
  - 29- Koutika L.S., Choné T., Andreux F., Burtin G., and Cerri C.C. 1999. Factors influencing carbon decomposition of topsoils from the Brazilian Amazon Basin. *Biology and Fertility of Soils* 28(4): 436-438.
  - 30- Kumar S., Lal R., and Liu D. 2012. A geographically weighted regression kriging approach for mapping soil organic carbon stock. *Geoderma* 189-190: 627-634.
  - 31- LieB M., Glaser B., and Huwe B. 2012. Uncertainty in the spatial prediction of soil texture comparison of regression tree and random forest models. *Geoderma* 170: 70-79.
  - 32- Liu F., Zhang G., Sun Y., Zhao Y., and Li D. 2013. Mapping the Three-Dimensional Distribution of Soil Organic Matter across a Subtropical Hilly Landscape. *Soil Science Society America Journal* 77: 1241-1253.
  - 33- Luoto M., and Hjort J. 2005. Evaluation of current statistical approaches for predictive geomorphological mapping. *Geomorphology* 67: 299-315.
  - 34- McBratney A.B., Mendonça Santos M.L., and Minasny B. 2003. On digital soil mapping. *Geoderma* 117: 3-52.
  - 35- McBratney A.B., Stockmann U., Angers D., Minasny B., and Field D. 2014. Challenges for Soil Organic Carbon Research. In Alfred E. Hartemink, Kevin McSweeney (eds.) *Soil Carbon*, p. 3-16. Cham: Springer.
  - 36- Metternicht G.I., and Zinck J.A. 2003. Remote sensing of soil salinity: potentials and constraints. *Remote Sensing of Environment* 85: 1-20.
  - 37- Moonjun R., Farshad A., Shrestha D.P., and Vaiphasa C. 2010. Artificial Neural Network and Decision Tree in Predictive Soil Mapping of Hoi Num Rin Sub-Watershed, Thailand. In: Boettinger J.L. Howell D.W. Moore A.C. Hartemink A.E and Kienast-Brown S. (eds.) *Digital Soil Mapping, Progress in Soil Science* 2, p. 151- 164.
  - 38- Nelson D.W., and Sommers L.E. 1996. Total Carbon, Organic Carbon, and Organic Matter. p. 961-1010. In: Sparks D.L. (ed.), *Methods of Soil Analysis. Chemical Methods. Part 3. The American Society of Agronomy and Soil Science Society of America, Madison, Wisconsin*.
  - 39- Nobakht A., Pourmajidian M., Hojjati S.M., and Fallah A. 2011. A comparison of soil carbon sequestration in hardwood and softwood monocultures (Case study: Dehmian forest management plan, Mazandaran). *Iranian Journal of Forest* 3(1): 13-23. (In Persian with English abstract)
  - 40- Pahlavan-Rad M.R., Toomanian N., Khormali F., Brungard C.W., Komaki C.B., and Bogaert P. 2014. Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. *Geoderma* 232-234: 97-106.
  - 41- Palm C., Sanchez P., Ahamed S., and Awiti A. 2007. Soils: A contemporary perspective. *Annual Review of Environment and Resources* 32: 99-129.
  - 42- Parks S.A., Dillon G.K., and Miller C. 2014. A new metric for quantifying burn severity: the relativized burn ratio. *Remote Sensing* 6: 1827-1844.
  - 43- Paustian K., Levine E., Post W.M., and Ryzhova I.M. 1997. The use of models to integrate information and understanding of soil C at the regional scale. *Geoderma* 79(1-4): 227-260.
  - 44- Prasad A.M., Iverson L.R., and Liaw A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9(2): 181-19.
  - 45- Qi J., Chehbouni A.R., Kerr Y.H., and Sorooshian S. 1994. A Modified Soil Adjusted Vegetation Index. *Remote*

- Sensing of Environment 48: 119-126.
- 46- R Development Core Team. 2015. R: a language and environment for statistical computing. R. Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.informaworld.com/contentUa/V.24-2-2004/article9.htm> (visited 5 September 2010).
- 47- Randall E.W., Mahieu D.S., Powlson D.S., and Christensen B.T. 1995. Fertilization effects on organic matter in physically fractionated soils as studied by <sup>13</sup>C NMR: results from two long-term field experiments. *European Journal of Soil Science* 46(4): 449-459.
- 48- Shahbazi F., Hughes P., McBratney A., Minasny B., and Malone B. 2019a. Evaluating the spatial and vertical distribution of agriculturally important nutrients-nitrogen, phosphorous and boron-in North West Iran. *Catena* 173: 71-82.
- 49- Shahbazi F., McBratney A., Malone B., Oustan S., and Minasny B. 2019b. Retrospective monitoring of the spatial variability of crystalline iron in soils of the east shore of Urmia Lake, Iran using remotely sensed data and digital maps. *Geoderma* 337: 1196-1207.
- 50- Shahrabi M. 1994. The report of 1:250000 scale geological map of Urmia. Publication of Geological survey and Mineral Exploration Organization of Iran.
- 51- Sindayihebura A., Ottoy S., Dondeyne S., Meirvenne M.V., and Orshoven J.V. 2017. Comparing digital soil mapping techniques for organic carbon and clay content: Case study in Burundi's central plateaus. *Catena* 156: 161-175.
- 52- Skakun R.S., Wulder M.A., and Franklin S.E. 2003. Sensitivity of the thematic mapper enhanced wetness difference index to detect mountain pine beetle red-attack damage. *Remote Sensing of Environment* 86: 433-443.
- 53- Smith C.A.S., Daneshfar B., and Frank G. 2012. Use of weights of evidence statistics to define inference rules to disaggregate soil survey maps. p. 215-220. In: Minasny B., Malone B.P., and McBratney A. (eds.) *Digital Soil Assessments and Beyond: Proceedings of the 5<sup>th</sup> Global Workshop on Digital Soil Mapping*. CRC Press, Sydney.
- 54- Sreenivas K., Dadhwal V.K., Kumar S., Sri Harsha G., Mitran T., Sujatha G., Janaki Rama Suresh G., Fyzee M.A., and Ravisankar T. 2016. Digital organic and inorganic carbon mapping of India. *Geoderma* 269: 160-173.
- 55- Stoorvogel J.J., Kempen, B., Heuvelink, G.B.M., and de Bruin S. 2009. Implementation and evaluation of existing knowledge for digital soil mapping in Senegal. *Geoderma* 149: 169-170.
- 56- Taghizadeh-Mehrjardi R., Nabiollahi K., and Kerry R. 2016. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* 266: 98-110.
- 57- Triantifilis J., Earl N.Y., and Gibbs I.D. 2012. Digital soil-classmapping across the Edgeroi district using numerical clustering and gamma-ray spectrometry data. p. 187-191. In: Minasny B., Malone B.P., and McBratney A. (eds.) *Digital Soil Assessments and Beyond: Proceedings of the 5<sup>th</sup> Global Workshop on Digital Soil Mapping*, CRC Press, Sydney.
- 58- Van Zijl G.M., le Roux P.A.L., and Smith A.B. 2012. Rapid soil mapping under restrictive conditions in Tete, Mozambique. p. 335-339. In: Minasny B. Malone B.P. and McBratney A. (eds.) *Digital Soil Assessments and Beyond: Proceedings of the 5<sup>th</sup> Global Workshop on Digital Soil Mapping*. CRC Press, Sydney.
- 59- Varamesh S., Hosseini S.M., Abdi N., and Akbarinia M. 2010. Increment of soil carbon sequestration due to forestation and its relation with some physical and chemical factors of soil. *Iranian Journal of Forest* 2(1): 25-35. (In Persian with English abstract)
- 60- Vasques G.M., Grunwald S., and Sickman J.O. 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma* 146: 14-25.
- 61- Vaysse K., and Lagacherie K. 2015. Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc- Roussillon (France). *Geoderma Regional* 4: 20-30.
- 62- Were K., Bui D.T., Dick Q.B., and Singh B.R. 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape. *Ecological Indicators* 52: 394-403.
- 63- Xu, H. 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing* 27: 3025-3033.
- 64- Zhang S., Huang Y., Shen C., Ye H., and Du Y. 2012. Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. *Geoderma* 171: 35-43.

## Application of Two Data Mining Techniques for Mapping the Spatial Distribution of Soil Organic Carbon (Case Study: East Shore of Urmia Lake)

A. Mousavi<sup>1</sup>- F. Shahbazi<sup>2\*</sup>- S. Oustan<sup>3</sup>- A.A. Jafarzadeh<sup>4</sup>- B. Minasny<sup>5</sup>

Received: 11-11-2019

Accepted: 27-04-2020

**Introduction:** Soils are considered as one of the most important parameters to be achieved the sustainable agriculture at any place in the world. Additionally, the digital environment needs to have a soil continuous maps at local and regional scales. However, such information are always not available at the required scale and mapping with high accuracy. Digital soil mapping (DSM) is a key for quantifying and assessing the variation of soil properties such as organic carbon (OC) especially in un-sampled and scarcely sampled areas. Using remotely sensed indices as an important auxiliary information relevant to the study area and data mining techniques were the pathway to create digital maps. Previous studies showed that digital elevation model (DEM) and remotely sensed data are the most commonly useful ancillary data for soil organic carbon prediction. the importance of DEM and derivative data in soil spatial modelling, it was not carried out in our research because there were no sharp differences in relief, and climate for that matter, across the study area. This research aims to investigate the spatial distribution of soil organic carbon (SOC) in a study area in north-western Iran using 21 remotely sensed indices as well as two data mining techniques namely Random Forests (RF) and Cubist.

**Materials and Methods:** This study was performed on the east shore of Urmia Lake located in the east Azerbaijan province, Iran. The area extension is about 500 km<sup>2</sup>. Based on the synoptic meteorological station report, the average annual precipitation and temperature of the study area is 345.37 mm and 10.83°C, respectively. Soil moisture and temperature regimes are Xeric and Mesic, respectively. Using stratified random soil sampling method, 131 soil samples (for the depth of 0-10 cm) were collected. Soil organic carbon (SOC) were then measured. The next step was to gather a suite of auxiliary data or environmental covariates thought to be useful (and available) for predicting SOC within a DSM framework for the region studied. Then, a number of remotely sensed imagery scenes from the Landsat 8-OLI acquired were collected in July 2017. The RF and Cubist models were applied to establish a relationship between soil organic carbon and auxiliary data. Both reflectance of the individual bands and indices derived from combinations of the individual bands were used. Fourteen spectral indices relevant to four types of data including: i) vegetation and soil; ii) water; iii) landscape; and iv) geology were gathered. Three different statistics was used for evaluating the performance of model in predicting SOC, namely the coefficient of determination ( $R^2$ ), mean error or bias (ME) and root mean square error (RMSE).

**Results and Discussion:** The results of the descriptive statistics of determined and calculated SOC for 131 soil samples showed that the mean and median values for SOC were 2.52% and 2.11%, respectively. Also, the CVs was recorded 57.94%. Minimum and maximum recorded values for SOC were 0.83% and 5.22%, respectively. The contents of SOC was left-skewed in the data set. The RF model prediction was quite good with calibration ( $R^2= 0.89$ , MSE = 0.16 and ME = 0.01). While, in the Cubist calibration data set, the Value of RMSE and ME were increased ( $R^2= 0.85$ , MSE = 0.21 and ME = 0.03). In terms of  $R^2$ , The RF model showed the higher value (0.89) compared with the Cubist model (0.85) for the validation dataset. Generally, the remote sensing (RS) spectral indices can successfully predict various SOC across the study area. The covariate importance rankings showed that VARI, NDVI, CRI2 and CRI1 were the four important covariates to predict SOC in the study area. Accordingly, the changes in SOC over space were not uniform across the study area and also it means that the study area is very dynamic and evolved over time.

**Conclusion:** The results of this study showed that although variables and auxiliary data had different importance in predicting the distribution of SOC, in general it can be found by modelling the relationship

1, 2, 3 and 4- Ph.D. Student, Associate Professor and Professors, Department of Science and Soil Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran, respectively.

(\*- Corresponding Author Email: shahbazi@tabrizu.ac.ir)

5- Professor, Sydney Institute of Agriculture and School of Life and Environmental Sciences, The University of Sydney, Sydney, Australia

DOI: 10.22067/jsw.v34i3.84154

between them and SOC through the model. The results revealed that the RF model was suitable for the target variable. Accordingly, the auxiliary variables had different importance in predicting the spatial distribution of SOC. Remote sensing imagery, particularly those encompassing the combined indices played an important role in the prediction of SOC. The obtained results also indicated that the Visible Atmospherically Resistant Index (VARI) and Normalized Difference Vegetation Index (NDVI) were important to predict SOC. The current study revealed that DSM using important environmental covariates can be successfully used in Iran which there is no sufficient soil databases. This research also provided a pathway to start further works in the future such as DSM relevant to the soil erosion, soil ripening, trace elements and so on.

**Keywords:** Cubist, Digital soil mapping, Environmental covariates, Modelling, Random Forest