نشریه علمی-پژوهشی مهندسی و مدیریت آبخیز
جلد ۱۲، شماره ۳، ۱۳۹۹، صفحات ۶۲۱-۶۳۴

Journal of Watershed Engineering and Management
Volume 12, Issue 3, 2020, Pages 621-634
DOI: 10.22092/ijwmse.2019.127500.1708

طبقه‌بندی حوضه آبی دریاچه پریشان به روش طبقه‌بندی جنگل تصادفی با استفاده از تصاویر ماهواره لندست

محمد صدیقی[۱] و امیرشاهرخ امینی[۲*]

[۱] دانشجوی کارشناسی‌ارشد سنجش از دور، ، واحد تهران جنوب، دانشگاه ازاد اسلامی، تهران، ایران و [۲] استادیار، گروه مهندسی نقشه‌برداری، واحد تهران جنوب، دانشگاه ازاد اسلامی، تهران، ایران

تاریخ دریافت: ۹۸/۰۶/۰۲ تاریخ پذیرش: ۹۸/۰۷/۱۷

چکیده

بحران کمبود آب موجود در جهان، مدیریت آب در بخش‌های مختلف از جمله حوضه‌های آبی و کشاورزی را تحت تاثیر قرار داده است. حوضه‌های آبی و زمین‌های اطراف آن‌ها درطول سال‌های اخیر با چالش‌های جدی از قبیل خشک شدن دریاچه‌ها و رودها، بیلان منفی آبخوان‌ها، تغییرات کاربری زمین‌های اطراف، افزایش سطح زیرکشت اراضی زراعی آبی و باغی و تغییر الگوی کشت از سمت محصولات کم‌مصرف به محصولاتی با نیاز آبی بالا مواجه شده‌اند. استفاده از تصاویر ماهواره‌ای به‌دلیل پوشش وسیع مکانی، قدرت تفکیک بالا، هزینه کم، آرشیو زمانی غنی تصاویر ماهواره‌ای و سهولت روش‌های تعیین کاربری، ابزاری مناسب و کارآمد برای کمک به مدیریت منابع آب و خاک است. در این پژوهش، چهار طبقه خاک، آب و مناطق مرطوب، شهری و کشاورزی انتخاب شده‌اند. سپس، از دو روش طبقه‌بندی جنگل تصادفی و ماشین‌های بردار پشتیبان، برای طبقه‌بندی تصاویر استفاده شده است. روش‌های طبقه‌بندی با محاسبه دو شاخص دقت کلی و ضریب کاپا با استفاده از داده‌های تست بررسی شدند. طبقه‌بندی جنگل تصادفی در چهار سال ۲۰۱۲، ۲۰۱۴، ۲۰۱۶ و ۲۰۱۸ و طبقه‌بندی ماشین‌های بردار پشتیبان در دو سال ۲۰۰۸ و ۲۰۱۰ بالاترین دقت را دارند. بنابراین، الگوریتم جنگل تصادفی در تفکیک طبقه‌ها به‌خصوص حوضه آبی به‌خوبی عمل کرده است و می‌توان به‌عنوان روشی قابل اعتماد از آن در این حوضه بهره برد.

واژه‌های کلیدی: بردار پشتیبان، طبقه‌بندی پوشش زمین، طبقه‌بندی عوارض، ضریب کاپا، مدیریت منابع

مقدمه

کشور ایران دارای حوضه‌های آبی بسیاری است. بحران کمبود آب باعث از بین رفتن این حوضه‌های آبی و زمین‌های اطراف شده است و مدیریت آب در بخش‌های مختلف از جمله کشاورزی را تحت تاثیر قرار داده است. از بین رفتن حوضه‌های آبی و تغییرات کاربری زمین‌های اطراف آن‌ها باعث نگرانی شده است. تهیه نقشه به‌روز به زمان زیادی احتیاج دارد. استفاده از تصاویر ماهواره‌ای و روش‌های طبقه‌بندی باعث تسریع در روند تهیه نقشه می‌شود. اگرچه دقیق‌ترین روش برای ارزیابی پوشش گیاهی، سطح پوشش منابع آب و به‌دست آوردن میزان تبخیر، اندازه‌گیری‌های مستقیم است، اما این روش‌ها بسیار وقت‌گیر و پرهزینه هستند و از طرفی، اگر در گذشته اندازه‌گیری‌هایی صورت نگرفته باشد و داده‌هایی در دسترس نباشند،

* مسئول مکاتبات: sh_amini@azad.ac.ir

به‌دلیل از دست دادن زمان گذشته، نمی‌توان ارزیابی در دوره دلخواه را به‌دست آورد.

یکی از مهمترین هدف‌های علم سنجش از دور، شناسایی و طبقه‌بندی عوارض از تصاویر هوایی و ماهواره‌ای است. شناسایی و طبقه‌بندی عوارض اهمیت زیادی در به‌روزرسانی نقشه‌ها و پایگاه داده‌ها، برنامه‌ریزی و مدیریت بحران‌ها دارد. تحقیقات زیادی در زمینه طبقه‌بندی و قطعه‌بندی برای شناسایی عوارض و مقایسه آن‌ها با هم صورت‌گرفته است. تصاویر ماهواره‌ای اطلاعاتی در زمینه طیف و بازتاب عوارض در محدوده مرئی و مادون‌قرمز در اختیار محققین قرار می‌دهند که از این اطلاعات می‌توان برای تولید ویژگی‌های طیفی و بافت استفاده کرد.

با توجه به‌دلایل ذکر شده، روش‌های زمینی در طول تاریخ به‌طور کافی مفید نبوده‌اند، از طرفی، با پیشرفت علم و حضور ماهواره‌ها و تصاویر ماهواره‌ای و ابزار قدرتمند GIS، استفاده از سنجش از دور و GIS به‌عنوان ابزاری مناسب و کارآمد برای کمک به مدیریت منابع آب و خاک شناخته شده است. فناوری جدید سنجش از دور یک روش مؤثر برای محاسبه پوشش گیاهی، سطح پوشش آب، میزان تبخیر، رطوبت خاک و غیره است که با وجود ویژگی‌های منحصر به فرد آن از جمله دید وسیع و پوشش‌های تکراری، سرعت بالای انتقال اطلاعات باعث شده است که در سال‌های اخیر از این علم استفاده بیشتری شود. استفاده از تصاویر ماهواره‌ای در این رابطه بسیار مقرون به‌صرفه و مفید است.

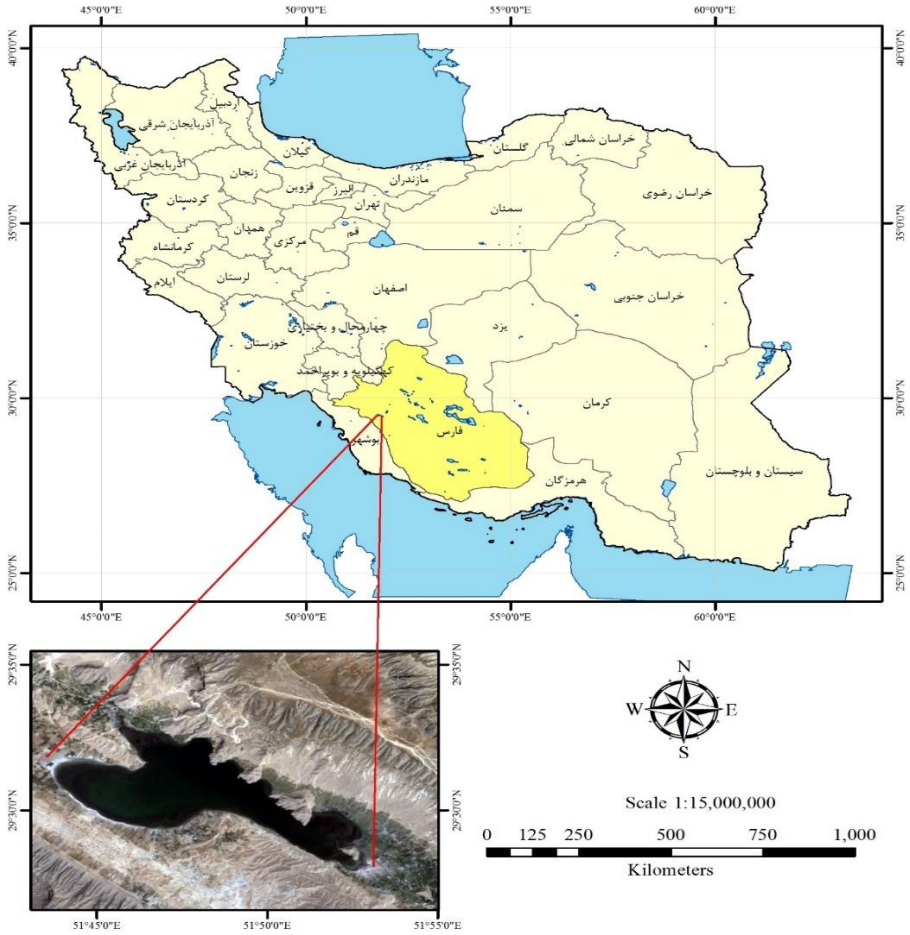
به‌طور کلی، طبقه‌بندی شامل سه روش کلی طبقه‌بندی نظارت نشده و طبقه‌بندی نظارت شده و شی‌گرا است. در روش طبقه‌بندی با نظارت، نوع و مکان برخی از انواع پوشش‌های زمین مثل منطقه شهری، کشاورزی، آب و از قبیل آن شناخته شده هستند. لازم است، کاربر مکان‌های مخصوصی را در داده‌های تصویری که نماینده مناطق یک‌دستی از پوشش‌های زمین که شناخته شده هستند را معین کند. این مکان‌ها را به‌دلیل خصوصیات طیفی آن‌ها در آموزش الگوریتم طبقه‌بندی برای تهیه نقشه نهایی، معمولا مناطق آموزشی می‌نامند. در طبقه‌بندی بدون نظارت، نوع و مکان انواع پوشش‌های زمین که قرار است به‌عنوان طبقه تعیین شوند، معمولا از قبل شناخته‌شده نیستند. بنابراین، از الگوریتم خواسته می‌شود که پیکسل‌های دارای خصوصیات طیفی مشابه از لحاظ برخی معیارهای آماری معین را در یک خوشه واحد گروه‌بندی کند. سپس، کاربر خوشه‌های طیفی را برچسب‌گذاری کرده و آن‌ها را ترکیب‌کرده و در طبقه‌های اطلاعات قرار می‌دهد.

Fathiyan و همکاران (۲۰۱۳) روند نزولی تراز آبی دریاچه ارومیه را مورد بررسی قرار دادند. ایشان در بررسی‌های خود از تصاویر سال‌های ۱۹۷۶ تا ۲۰۱۱ ماهواره لندست استفاده کردند. برای به‌دست آوردن نقشه کاربری منطقه از روش‌های طبقه‌بندی بیشینه شباهت، کمینه فاصله از میانگین و فاصله ماهالانوبیز استفاده کردند. در بین روش‌های طبقه‌بندی، روش بیشینه شباهت، به‌دلیل دقت بالاتر برای طبقه‌بندی مورد استفاده قرار گرفت. چهار روش آماری غیرپارامتری تحلیل روند شامل: من‌کندال، تایل‌سن، اسپیرمن و سنس‌تی و اطلاعات ۱۸ ایستگاه هواشناسی و آب‌سنجی واقع در منطقه را به‌کار گرفتند. نتایج بخش اول، روند معنی‌دار افزایشی دما را در کل منطقه نشان داد که در مورد بارندگی این چنین نبود. Ahmadpur و همکاران (۲۰۱۱) در پژوهشی، سه روش طبقه‌بندی نظارت شده کمینه فاصله از میانگین، شبکه‌های موازی و بیشینه احتمال را با استفاده از تصاویر لندست و IRS برای تشخیص گروه‌های گیاهی استان خراسان مورد مقایسه و ارزیابی قرار دادند. در ارزیابی با شاخص صحت کلی و ضریب کاپا، روش بیشینه احتمال بیشترین دقت و روش شبکه‌های موازی کمترین دقت را داشتند. Alimohammadi و همکاران (۲۰۰۹) در پژوهشی، روش‌های طبقه‌بندی پیکسل پایه، شی پایه و درخت تصمیم‌گیری را در تهیه نقشه تیپ‌های جنگل‌های آستارا با استفاده از تصاویر اسپات پنج مورد ارزیابی قرار دادند. اطلاعات زمینی را با استفاده از ۱۵۳ پلات به‌صورت تصادفی در مساحت های یک هکتاری تهیه کردند. نوع تیپ در هر پلات بر اساس درصد فراوانی کل نوع گونه‌ها تعیین شده، شامل تیپ‌های انجیلی آمیخته، پهن برگ مخلوط، ممرز آمیخته، راش آمیخته و حفاظتی بوده است. برای تهیه نقشه این تیپ‌ها، در

روش پیکسل پایه با استفاده از داده‌های طیفی و طبقه‌بندی کننده شبکه عصبی و در روش طبقه‌بندی شی پایه از روش نزدیک‌ترین همسایه و توابع فازی استفاده کردند. استفاده از داده‌های کمکی به‌خصوص مدل رقومی ارتفاع باعث افزایش دقت طبقه‌بندی شده است. روش درخت تصمیم‌گیری بیشترین دقت را در طبقه‌بندی داشته است. Genuer و همکاران (۲۰۱۰) از روش جنگل تصادفی برای طبقه‌بندی دو طبقه استفاده کردند و در تحقیق‌شان برای برطرف کردن خطای رگرسیون از روش Leo Breiman (۲۰۰۱) استفاده کردند. اولین اولویت ایشان پیدا کردن متغیرهای مهم و اولیت بعدی‌شان ارائه مدلی بهینه برای طبقه‌بندی بود. Smith (۲۰۱۰) از روش جنگل تصادفی به‌منظور طبقه‌بندی تصویر اسپات شمال داکوتا استفاده کرد. ایشان در روش خود برای انتخاب متغیرها از سنجش کمی مقیاس تصویر استفاده کردند و در این طبقه‌بندی به دقت ۸۵/۲ درصد رسیدند. Ghasemi Esfahan (۲۰۱۳) در پایان‌نامه کارشناسی ارشد خود برای بهبود طبقه‌بندی صحنه شهری از ترکیب روشی جدید مبتنی بر تلفیق روش شی‌گرا و طبقه‌بندی جنگل تصادفی استفاده کردند. برای مقایسه روش پیشنهادی، روش‌های طبقه‌بندی ماشین‌های بردار پشتیبان، شبکه‌های عصبی، جنگل تصادفی و بیشینه احتمال را نیز بر روی تصویر اجرا کردند و نتایج نشان‌دهنده این است که روش پیشنهادی می‌تواند عملکرد طبقه‌بندی را از نظر دقت و سرعت بهبود بخشد. Samani و همکاران (۲۰۰۹) به‌منظور ارزیابی روند تغییرات کاربری اراضی حوزه آبخیز طالقان در دوره ۱۳۶۶ تا ۱۳۸۰ از تصاویر لندست و روش طبقه‌بندی بیشینه احتمال استفاده کردند. Zebardast و همکاران (۲۰۱۰) روند تغییرات تالاب انزلی را در دوره ۱۰ ساله مورد ارزیابی قرار دادند. در این ارزیابی، از تصاویر لندست و طبقه‌بندی بیشینه احتمال استفاده شده است. نتایج این پژوهش بیان‌کننده تخریب تالاب در دوره ۱۰ ساله، تغییرات پوشش گیاهی و افزایش تغذیه‌گرایی مساحت بخش آبگیر تالاب است. علاوه‌بر آن، مساحت تالاب تحت تاثیر نوسانات دریای خزر و پیشروی و پسروی آب این دریا قرار دارد. Najafi و همکاران (۲۰۱۶) در پژوهشنامه مدیریت حوزه آبخیز، برای طبقه‌بندی کاربری اراضی حوضه چشمه کیله چالکرود از تصاویر ماهواره‌های لندست ۳ و روش ماشین‌های بردار پشتیبان استفاده کرده‌اند. چهار کرنل برای این روش شامل کرنل خطی، چند جمله‌ای، شعاعی و حلقوی برای فرایند طبقه‌بندی استفاده و کارآمدی طبقه‌بندی هر یک از آن‌ها را ارزیابی کرده‌اند. Shenani Hoveizeh و همکاران (۲۰۱۶) تغییرات کاربری اراضی حوزه آبخیز ابوالعباس را در دو دهه زمانی مورد بررسی قرار دادند. در این پژوهش، از تصاویر لندست و طبقه‌بندی شبکه عصبی مصنوعی استفاده کردند. تجزیه و تحلیل تغییرات نشان‌دهنده تغییر جنگل‌های انبوه به جنگل‌های تنک و مراتع و اراضی کشاورزی به مناطق مسکونی است. هدف این پژوهش، ارزیابی روش طبقه‌بندی جنگل تصادفی در بررسی تغییرات دریاچه‌ها و همچنین، مقایسه آن با روش ماشین‌های بردارپشتیبان است.

مواد و روش‌ها

منطقه مورد مطالعه: دریاچه پریشان یکی از بزرگ‌ترین دریاچه‌های آب شیرین ایران و خاورمیانه است. این دریاچه به‌نام‌های مور، شور، کازرون، یون، موز، توز و پریشم نیز شناخته شده است و در مختصات ″۴۴ ′۳۰ °۲۹ شمالی و ″۴۲ ′۴۶ °۵۱ شرقی واقع شده است. مساحت آن ۴۳۰۰ هکتار و ارتفاع آن از سطح آب‌های آزاد ۸۲۰ متر است. حوزه آبخیز آن ۲۶۶/۵ کیلومتر مربع است و بیشترین مساحت را در اردیبهشت ماه دارد. این دریاچه در کنوانسیون رامسر به‌عنوان تالاب بین‌المللی ثبت شده است و در تقسیم‌بندی مناطق جزء منطقه حفاظت‌شده محسوب می‌شود. این دریاچه در فاصله ۱۵ کیلومتری شهرستان کازرون واقع شده است و دور تا دور آن را کوه فراگرفته است. آب این دریاچه از صدها چشمه کوچک و بزرگ مناطق دشت ارژن و دشت فامور و چشمه‌های زیادی که در کف دریاچه وجود دارند، تأمین می‌شود. موقعیت منطقه مورد مطالعه در شکل زیر نشان داده شده است.



شکل ۱– موقعیت منطقه مورد مطالعه و تصویر لندست سال ۲۰۰۰ دریاچه پریشان

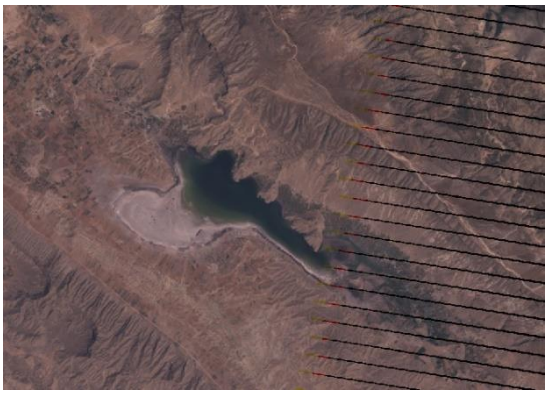
داده: در این تحقیق، از تصاویر ماهواره لندست استفاده شده است. در نتیجه، با توجه به این‌که به تصاویر دهه اخیر نیاز بوده از تصاویر سنجنده ETM^{+} و OLI استفاده شده است. همچنین، چون دریاچه پریشان در ماه اردیبهشت پر آب است، در این پژوهش سعی شده است، تصاویر مناسب از نظر میزان پوشش ابر و کیفیت تصویر در ماه اردیبهشت انتخاب شوند.

آماده‌سازی داده‌ها: محصولات ارائه شده به‌وسیله لندست دارای سطوح مختلف تصحیحات است. برای این پژوهش، از تصاویر تصحیح شده زمینی سطح یک[1] استفاده شده است. در تصاویر تصحیح شده سطح یک تصحیحات هندسی دقیق با استفاده از نقاط کنترل زمینی و مدل رقومی زمین انجام شده است.

خطای راه‌راه‌شدگی یا striping: بر روی تصاویر این سنجنده لندست ۷ از سال ۲۰۰۳ خطوطی نمایش داده شده، که فاقد داده هستند. در واقع از سال ۲۰۰۳ تصحیح‌کننده خط اسکن این سنجنده از کار افتاد. به همین دلیل مشاهده می‌شود که در حین تصویربرداری بسیاری از خطوط اسکن از سطح زمین تصاویر اخذ نمی‌شود و تصاویر ناقص است که به‌صورت گپ‌هایی مشاهده می‌شود. کمترین میزان خطا در مرکز تصاویر و بیشترین میزان خطا در حاشیه‌ها مشاهده می‌شود. هر چند روش‌های مختلفی برای اصلاح این تصاویر ارائه شده است، اما با این حال به هیچ عنوان نمی‌توان اطلاعاتی که در هر تصویر اسکن نشده و از بین رفته را بازگرداند. در این روش‌های اصلاح، انرژی واقعی از دست رفته الکترومغناطیسی که ثبت نشده را نمی‌توان برگرداند، فقط سعی می‌شود که نزدیک‌ترین انرژی معادل را بازسازی کرد. در این مقاله، برای اصلاح این خطا در نرم‌افزار ENVI از اکستنشن Gap/Fill استفاده شده که از میان روش‌های موجود از روش درون‌یابی از خطوط مجاور استفاده شده است. اگر ابتدا خطای بد

[1] Level 1 terrain corrected

اسکنر اصلاح شود و سپس، تصحیحات رادیومتریکی و اتمسفری اجرا شود، به‌صورت اتوماتیک نمی‌توان مقادیر تابش Radiance و بازتاب Reflectance را محاسبه کرد. چون مقادیر Gain و Offset که از طریق فراداده به‌دست آمده، از بین می‌رود. بنابراین، باید ابتدا تصحیحات رادیومتریکی انجام شود و سپس خطای راه‌راه‌شدگی برطرف شود.



۱۳۸۷/۰۲/۰۹ – ۲۰۰۸/۰۴/۲۸
تصویر لندست ۷

۱۳۸۹/۰۲/۳۰ – ۲۰۱۰/۰۵/۲۰
تصویر لندست ۷

۱۳۹۱/۰۲/۰۴ – ۲۰۱۲/۰۴/۲۳
تصویر لندست ۷

۱۳۹۳/۰۲/۱۷ – ۲۰۱۴/۰۵/۰۷
تصویر لندست ۸

۱۳۹۵/۰۲/۰۷ – ۲۰۱۶/۰۴/۲۶
تصویر لندست ۸

۱۳۹۷/۰۲/۲۸ – ۲۰۱۸/۰۵/۱۸
تصویر لندست ۸

شکل ۲- تصاویر لندست ۷ و ۸ از دریاچه پریشان

طبقه‌بندی جنگل تصادفی (Random Forest): الگوریتم جنگل تصادفی یک الگوریتم ترکیبی است که از درخت‌های تصمیم، برای الگوریتم‌های ساده و ضعیف خود استفاده می‌کند. یک الگوریتم درخت تصمیم، می‌تواند به‌راحتی عملیات طبقه‌بندی را بر روی داده‌ها انجام دهد. حال در الگوریتم جنگل تصادفی از چندین درخت تصمیم (برای مثال ۱۰۰ درخت تصمیم) استفاده می‌شود. در واقع مجموعه‌ای از درخت‌های تصمیم، با هم یک جنگل را تولید می‌کنند و این جنگل می‌تواند تصمیم‌های بهتری را اتخاذ کند. در الگوریتم جنگل تصادفی به هر کدام از درخت‌ها، یک زیرمجموعه‌ای از دادگان داده می‌شود. این درخت‌ها با همین زیرمجموعه، می‌توانند تصمیم بگیرند و مدل طبقه‌بندی خود را بسازند (Ghasemi Esfahan، ۲۰۱۳). روش RF برای طبقه‌بندی به‌طور خلاصه این‌گونه بیان می‌شود که در ابتدا، T نمونه خودراه‌انداز[1] از داده آموزشی بیرون کشیده می‌شود و سپس، از هر نمونه خودراه‌انداز β یک درخت طبقه‌بندی و رگرسیون (CART[2]) هرس نشده ایجاد می‌شود که برای انشعاب در هر گره CART، تنها یکی از M ویژگی انتخاب‌شده به‌صورت تصادفی، استفاده می‌شود. در نهایت، خروجی طبقه‌بندی بر اساس یک رای بیشترین از پیش‌بینی‌های تمام تک‌درخت‌های آموزش دیده، ایجاد می‌شود. روند کلی الگوریتم در شکل ۳ نشان داده شده است.

در یک طبقه‌بندی مبتنی بر RF، لازم است، دو پارامتر به‌وسیله کاربر مشخص شود. اندازه M یک زیرمجموعه تصادفی از ویژگی‌ها و تعداد درخت‌ها T. انتخاب پارامتر M بر روی نرخ خطای نهایی موثر است. اگر M افزایش یابد، هم وابستگی بین درخت‌ها و هم قدرت (صحت طبقه‌بندی) تک‌درخت در جنگل افزایش می‌یابد. نرخ خطا با وابستگی، رابطه مستقیم دارد (متناسب است)، اما نسبت عکس با قدرت دارد. معمولا، M برابر جذر تعداد ویژگی‌ها در نظر گرفته می‌شود. چون RF سریع است و overfit نمی‌شود، تعداد درخت‌ها T می‌تواند تا حد امکان زیاد باشد. با این حال، به‌علت محدودیت حافظه ماشین، معمولا T از چندصد تا چندهزار تا انتخاب می‌شود (Jin، ۲۰۱۲).

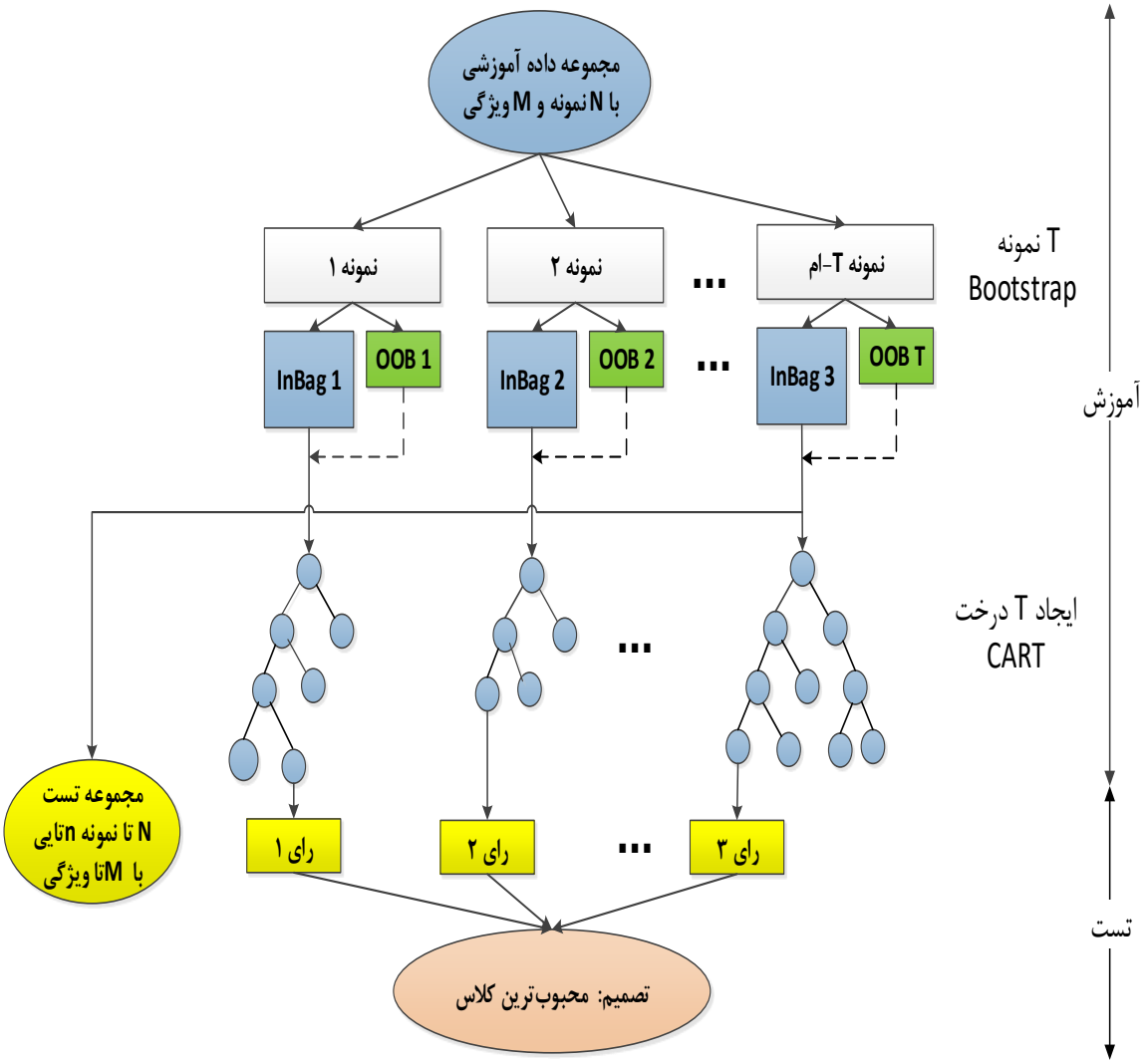
روش پژوهش: در این پژوهش، پس از دریافت تصاویر و تصحیح رادیومتریکی تصویر با الگوریتم FLAASH، تصاویر سال‌های ۲۰۰۸، ۲۰۱۰ و ۲۰۱۲ که دارای خطای راه‌راه‌شدگی هستند، برطرف شده و منطقه مورد نظر برش داده می‌شود. با توجه به دقت مکانی ۳۰ متری تصاویر لندست در نوارهای Optic، برای افزایش دقت طبقه‌بندی به کمک نوار پانکروماتیک که دارای دقت مکانی ۱۵ متری است، تصاویر بهبود کیفیت مکانی داده می‌شود. برای طبقه بندی از چهار طبقه آب (آب و مناطق مرطوب)، شهری (تمامی مناطق مسکونی)، کشاورزی (باغ و مناطق سبز و کشاورزی) و خاک (مناطق کوهستانی و صخره‌ای و فاقد پوشش گیاهی) استفاده شده است. بنابراین، در هر یک از تصاویر ماهواره‌ای به کمک نقشه کاربری اراضی، منطقه مورد مطالعه، برای هر طبقه نقاط آموزشی و آزمون انتخاب می‌شود. این نقاط به تعداد کافی و با پراکندگی مناسب به‌گونه‌ای انتخاب می‌شوند که حدود ۳۰ درصد از نقاط آموزشی به‌عنوان نقاط آزمون در نظر گرفته می‌شوند. در مرحله بعد به کمک داده‌های آموزشی هر یک از الگوریتم‌های طبقه بندی جنگل تصادفی و ماشین‌های بردار پشتیبان هدایت می‌شود تا مدل‌سازی انجام شود. پس از مدل-سازی و انجام طبقه‌بندی، به کمک داده‌های آزمون و ابزارهای ارزیابی دقت کلی و ضریب کاپا مدل‌ها مقایسه و ارزیابی می‌شود. روند کلی کار در شکل ۴ نمایش داده شده است.

نتایج و بحث

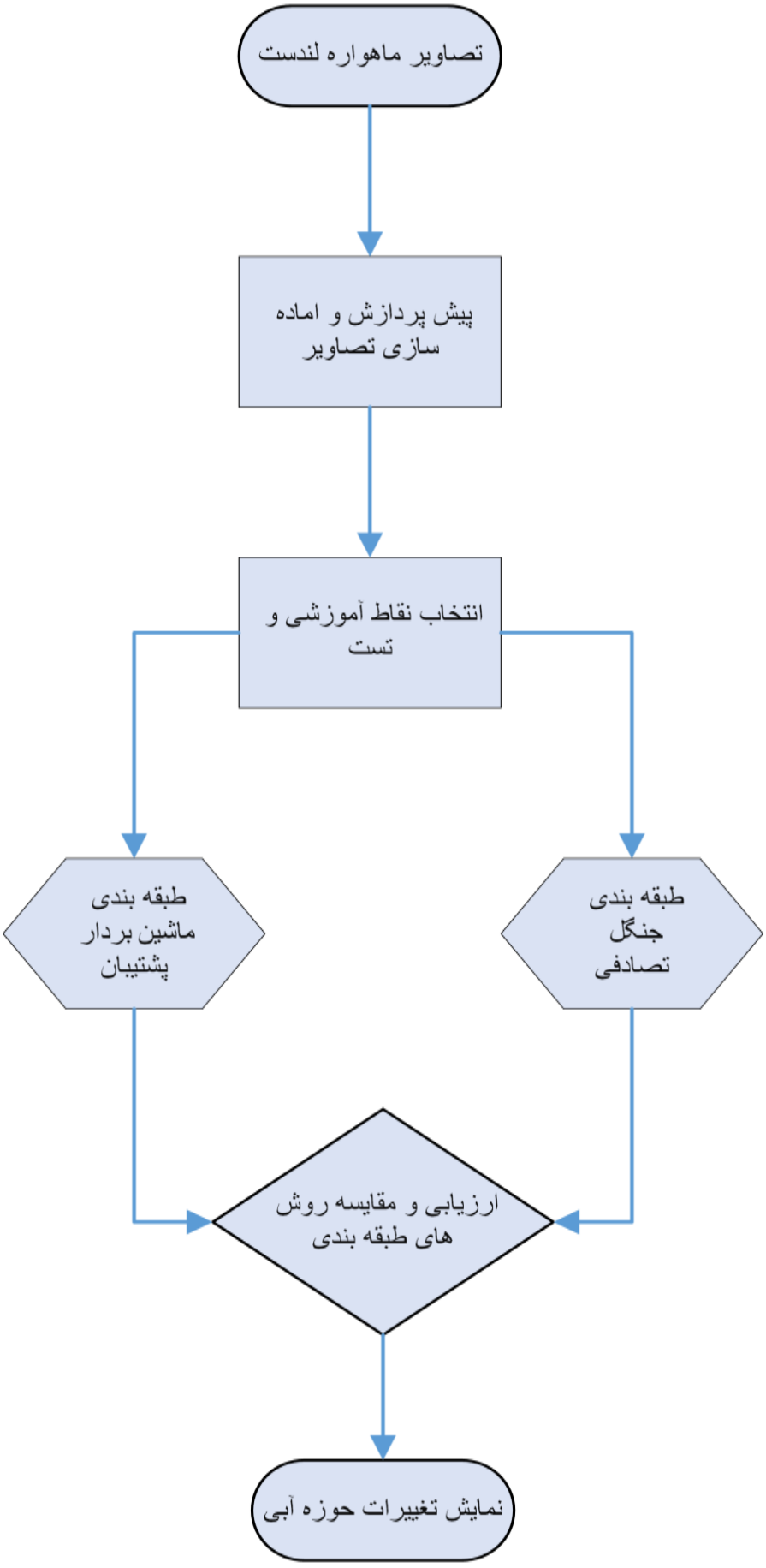
ارزیابی نتایج طبقه‌بندی یکی از مراحل مهم پس از طبقه‌بندی است. ارائه نتایج طبقه‌بندی بدون هیچ-گونه پارامتری که کیفیت یا صحت این نتایج را بیان کند، از ارزش آن‌ها می‌کاهد و در بعضی مواقع آن‌ها را بدون استفاده می‌کند. بنابراین، این نکته را باید در نظر داشت که همیشه در کنار انجام طبقه‌بندی، نتایج آن نیز ارزیابی شوند و عرضه شوند. متداول‌ترین روش برای ارزیابی کمی صحت طبقه‌بندی، انتخاب تعدادی پیکسل از پیکسل‌های نمونه معلوم و مقایسه طبقه آن‌ها با نتایج طبقه‌بندی است. این داده‌های معلوم را

[1] Bootstrap
[2] Classification And Regression Tree

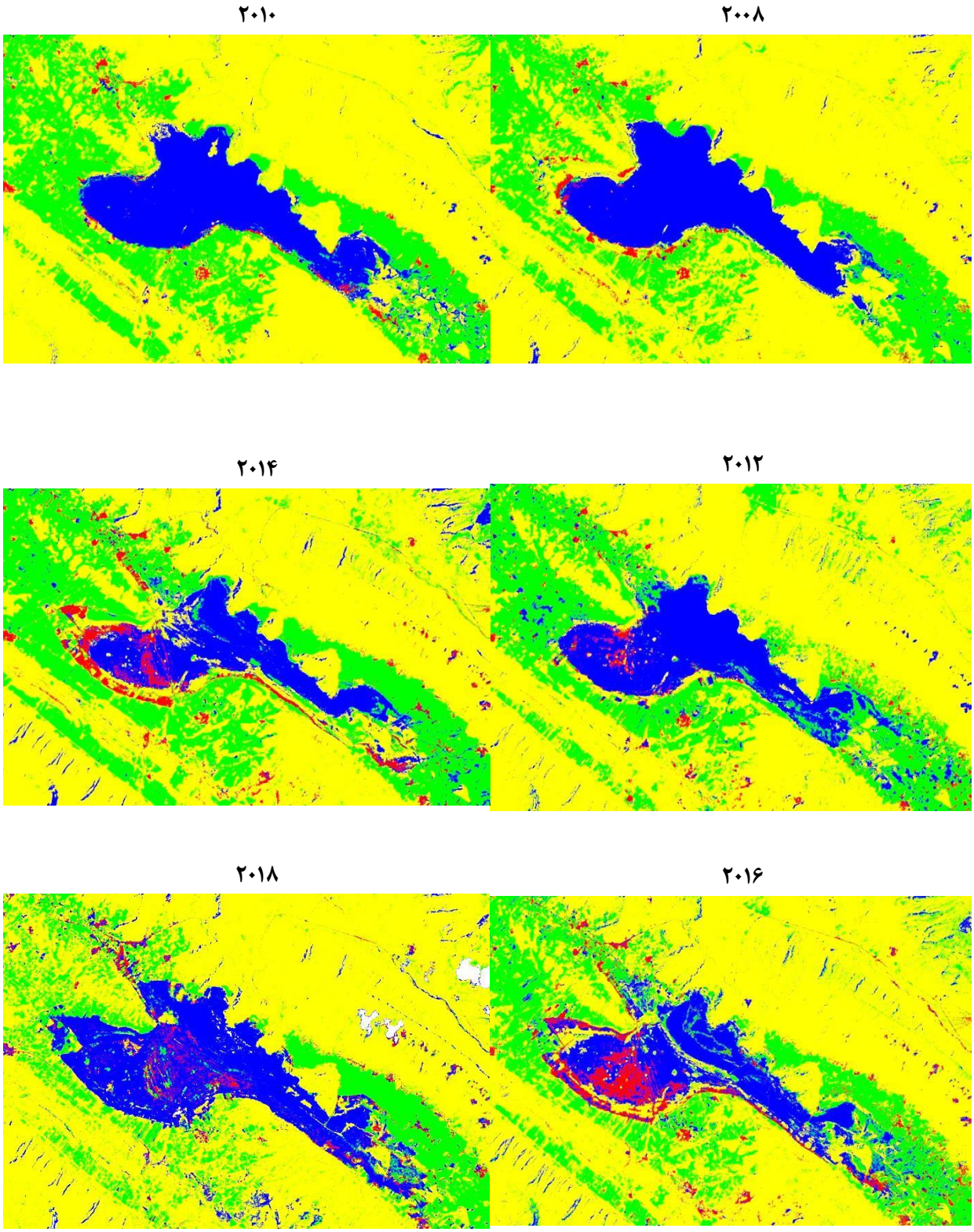
واقعیت زمینی یا داده‌های مرجع می‌نامند. انجام نمونه برداری از طریق بازدید زمینی یا با استفاده از داده‌های قبلی نظیر نقشه‌های موجود یا تصاویر هوایی-فضایی موجود، صورت می‌پذیرد. ارزیابی خطا و برآورد صحت طبقه‌بندی معمولا بر اساس پارامترهای آماری است که از ماتریس خطا استخراج می‌شوند. ماتریس خطا که ماتریس ابهام نیز نامیده می‌شود، حاصل مقایسه پیکسل به پیکسل، پیکسل‌های معلوم (در واقعیت زمینی) با پیکسل‌های متناظر در نتایج طبقه‌بندی است. دو پارامتر اصلی که برای ارزیابی صحت طبقه بندی استفاده می‌شوند، صحت کلی و ضریب کاپا هستند. صحت کلی میانگینی از طبقه‌بندی است که نسبت پیکسل‌های صحیح طبقه‌بندی شده به جمع کل پیکسل‌های معلوم را نشان می‌دهد. یکی دیگر از پارامترهای صحت که از ماتریس خطا استخراج می-شود، ضریب‌کاپا است. ضریب‌کاپا صحت طبقه‌بندی را نسبت به یک طبقه‌بندی کاملا تصادفی محاسبه می-کند. به این معنی که مقدار کاپا صحت طبقه‌بندی را نسبت به حالتی که یک تصویر کاملا به‌صورت تصادفی طبقه‌بندی شود، به‌دست می‌دهد. این کار را می‌توان به این صورت معنی‌کرد که پس از حذف تاثیر شانس در طبقه‌بندی مقدار تطابق با واقعیت‌زمینی محاسبه خواهد شد.



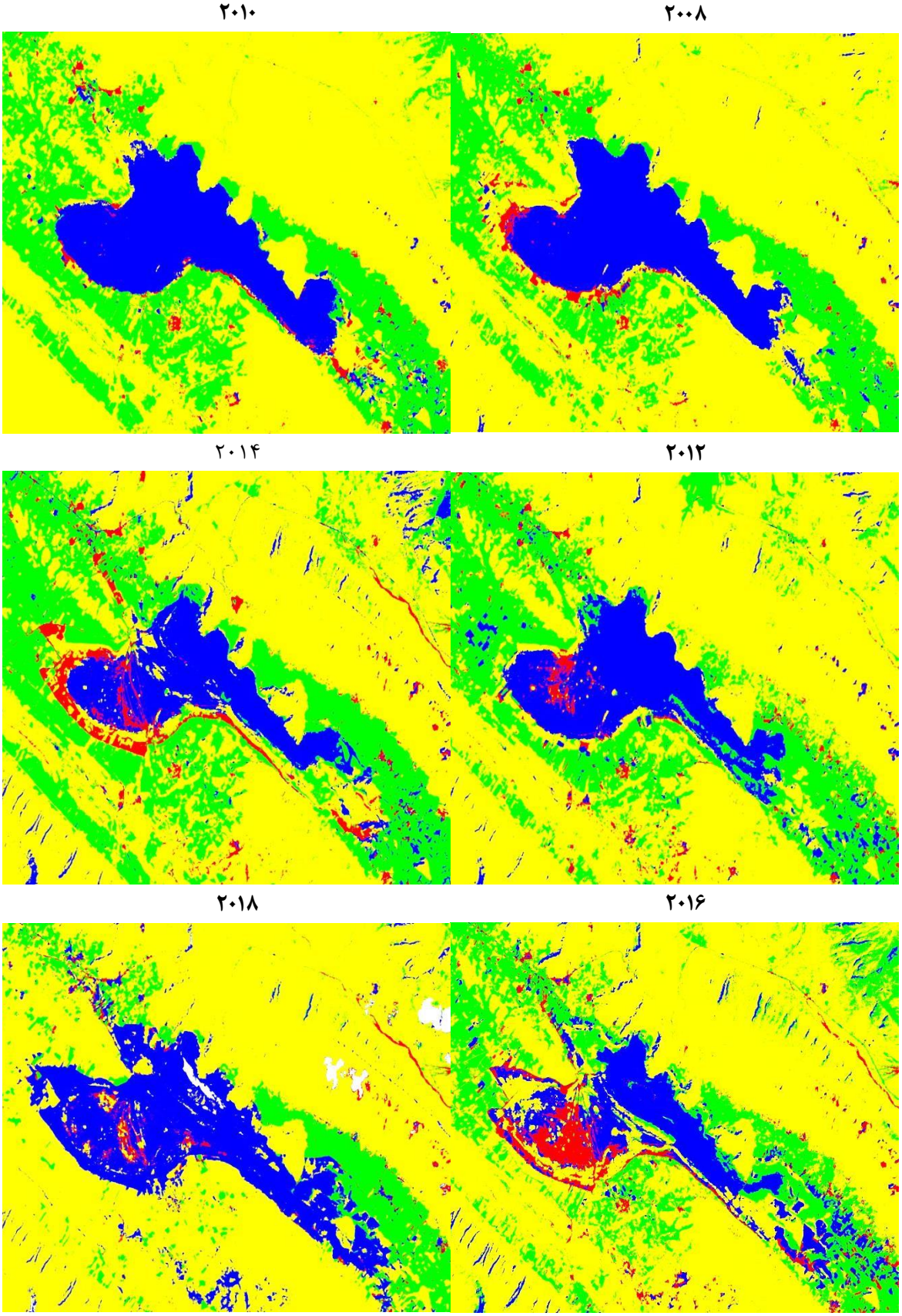
شکل **Error! No text of specified style in document.**- روند کلی الگوریتم جنگل تصادفی



شکل ۴- روند کلی پژوهش



شکل ۵- نتایج طبقه‌بندی به روش جنگل تصادفی در سال‌های مختلف



شکل ۶- نتایج طبقه‌بندی به روش ماشین بردار پشتیبان در سال‌های مختلف

جدول ۱- نتایج دقت کلی طبقه‌بندی (درصد)

Overall Accuracy	SVM	RF
۲۰۰۸	۸۵/۴۵۴۵	۸۴/۷۵۵۲
۲۰۱۰	۸۸/۲۰۹	۸۶/۲۶۸۷
۲۰۱۲	۷۵/۶۶۶۷	۷۹/۳۳۳۳
۲۰۱۴	۸۰/۱۱۱۱	۸۱
۲۰۱۶	۶۸/۷۳۴	۷۴/۴۳۰۴
۲۰۱۸	۷۴/۶۸۰۹	۷۹/۷۸۷۲

با توجه به جدول ۱، روش جنگل تصادفی در سال‌های ۲۰۱۲، ۲۰۱۴، ۲۰۱۶ و ۲۰۱۸ دقت طبقه‌بندی بهتری نسبت به روش ماشین بردار پشتیبان داشته است و در سال‌های ۲۰۱۰ و ۲۰۰۸ با درصد کمی نسبت به طبقه‌بندی ماشین بردارپشتیبان ضعیف عمل کرده است.

با توجه به جدول ۲، بازه ضرایب کاپا در بازه قابل قبول قرار داشته که بیان‌کننده دقت نسبتا خوب طبقه‌بندی است. در ادامه، مساحت هر طبقه برای هر یک از طبقه‌بندی‌های انجام شده مورد بررسی قرار می‌گیرد.

جدول ۲- نتایج ضرایب کاپا

Kappa Coefficient	SVM	RF
۲۰۰۸	۰/۷۵۹۳	۰/۷۵۱۷
۲۰۱۰	۰/۷۸۳۷	۰/۷۵۰۹
۲۰۱۲	۰/۶۷۳۸	۰/۶۸۶۳
۲۰۱۴	۰/۶۹۱۹	۰/۷۰۹۲
۲۰۱۶	۰/۵۱۱	۰/۶۰۴۷
۲۰۱۸	۰/۶۱۴۲	۰/۶۹۹۴

جدول ۳- مساحت‌های روش طبقه‌بندی جنگل تصادفی

RF	آب	شهری	کشاورزی	خاک
۲۰۰۸	۴۰/۵۶۵۰	۵/۵۵۶۶	۶۹/۴۰۶۲	۱۸۹/۰۵۳۸
۲۰۱۰	۴۱/۳۳۵۹	۴/۵۷۴۵	۷۷/۲۴۹۷	۱۸۱/۴۲۱۶
۲۰۱۲	۴۱/۸۰۵۵	۶/۹۸۲۴	۷۷/۰۰۹۴	۱۸۰/۷۸۴۴
۲۰۱۴	۳۵/۲۵۴۱	۱۱/۶۹۵۱	۸۷/۶۰۶۵	۱۷۰/۰۲۶۰
۲۰۱۶	۳۷/۳۸۶۰	۱۴/۲۷۷۶	۶۹/۷۲۲۸	۱۸۳/۱۹۵۲
۲۰۱۸	۴۹/۶۶۹۰	۸/۶۲۵۲	۵۹/۰۴۳۶	۱۸۴/۸۱۲۵

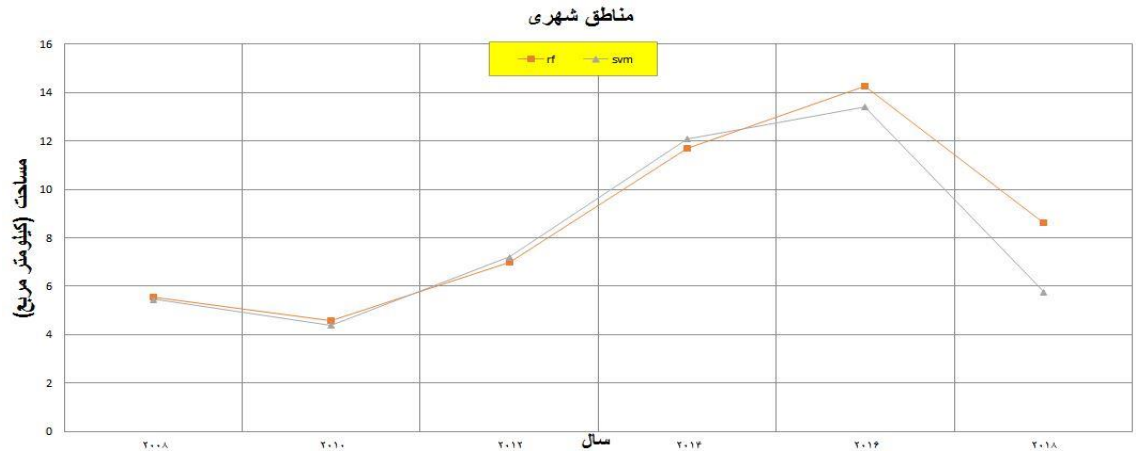
جدول ۴- مساحت‌های روش طبقه‌بندی ماشین بردارپشتیبان

SVM	آب	شهری	کشاورزی	خاک
۲۰۰۸	۳۹/۱۰۵۰	۵/۴۶۴۱	۶۷/۴۱۹۵	۱۹۲/۵۹۳۰
۲۰۱۰	۳۹/۷۰۸۵	۴/۳۹۲۰	۷۳/۳۸۳۵	۱۸۷/۰۹۷۶
۲۰۱۲	۳۶/۸۰۰۱	۷/۲۰۶۸	۷۵/۳۵۶۶	۱۸۵/۲۱۸۲
۲۰۱۴	۳۴/۲۷۶۵	۱۲/۰۹۳۱	۸۷/۵۹۶۱	۱۷۰/۶۱۵۹
۲۰۱۶	۳۱/۸۸۵۰	۱۳/۴۱۷۹	۶۵/۶۷۷۵	۱۹۳/۶۰۱۳
۲۰۱۸	۵۱/۳۶۳۹	۵/۷۶۷۷	۳۲/۱۴۶۲	۲۱۲/۵۱۹۹

با توجه به جدول‌های ۳ و ۴، مساحت‌های طبقه‌های مختلف در هر دو طبقه‌بندی به هم نزدیک بوده و از طرفی، مجموع مساحت‌های هر سال در هر طبقه‌بندی نشان از عملکرد خوب طبقه‌بندی است. در این قسمت، نمودار مساحت هر یک از چهار طبقه در سال‌های مختلف مقایسه می‌شود.

شکل ۷- مقایسه مساحت مناطق مرطوب به روش‌های طبقه‌بندی

شکل ۸- مقایسه مساحت مناطق کشاورزی به روش‌های طبقه‌بندی



شکل ۹- مقایسه مساحت مناطق شهری به روش‌های طبقه‌بندی

شکل ۱۰- مقایسه مساحت مناطق خاکی به روش‌های طبقه‌بندی

با توجه به نمودارها، مساحت طبقه‌های مختلف در هر سال به یکدیگر نزدیک بوده، تقریبا نمودارهای هر دو طبقه‌بندی در طبقه‌های گوناگون برهم منطبق و به واقعیت نزدیک است.

اگرچه دقیق‌ترین روش برای ارزیابی پوشش گیاهی، سطح پوشش منابع آب و به‌دست آوردن میزان تبخیر، اندازه‌گیری‌های مستقیم است، اما این روش‌ها بسیار وقت‌گیر و پرهزینه هستند. از طرفی، با پیشرفت علم و حضور ماهواره‌ها و تصاویر ماهواره‌ای و ابزار قدرتمند GIS، استفاده از سنجش از دور و GIS به‌عنوان ابزاری مناسب و کارآمد برای کمک به مدیریت منابع آب و خاک شناخته شده است. در این پژوهش، به کمک یکی از جدیدترین الگوریتم‌های هوش مصنوعی حوضه آبی طبقه‌بندی شده، با روش مرسوم و قدرتمند ماشین‌های بردارپشتیبان مقایسه و ارزیابی شد. نتایج دقت کلی و ضریب کاپا نشان‌دهنده دقت خوب این روش طبقه‌بندی بوده، مقایسه مساحت‌های طبقه‌های گوناگون نشان‌دهنده تقریب خوب این روش طبقه‌بندی در حوضه‌های آبی است.

منابع مورد استفاده

1. Ahmadpour, A., K. Soleimani, M. Shokri and J. Ghorbani. 2011. Comparison of the effectiveness of three common methods of supervised classification of satellite data in vegetation study. Journal of Remote Sensing and GIS in Natural Resources Science, 2(2): 69-80 (in Persian).
2. Alimohammadi, A., A.A. Metkan, P. Ziaeian and H. Tabatabaei. 2009. Comparison of base pixel, base object and decision tree classification methods in mapping forest types using remote sensing data , case study: Astara. Forest-Geographical Research Journal, 10(13): 1-20 (in Persian).
3. Fathiyan, F., S. Morid and S. Arshad. 2013. Evaluation of land use change trends using remote sensing technology and its relation to river flow process, case study: east of Urmia Lake basins. Journal of Water and Soil, 3: 642-655 (in Persian).
4. Genuer, R., J.M. Jean and C. Tuleau-Malot. 2010. Variable selection using random forests. Pattern Recognition Letters, 31(14): 2225-2236.
5. Ghasemi Esfahan, A. 2013. Investigation of stochastic forest method to improve urban land cover classification using satellite images. MSc Thesis, School of Surveying Engineering, Khaje Nasireddin Toosi University, 121 pages (in Persian).
6. Jin, J. 2012. A random forest based method for urban land cover classification using LiDAR data and aerial imagery. UWSpace, 142 page.
7. Najafi, A., S. Azizi Ghalati and M.H. Mokhtari. 2016. Application of support vector machine in land use classification of Cheshmeh Kileh Basin, Chalkroud. Journal of Watershed Management, Eighth Year, 15: 92-101 (in Persian).
8. Samani, A.A., M. Ghorbani and H.R. Kouhbani. 2010. Evaluation of land use changes in Taleghan Watershed in the period of 1366 to 2001. Scientific Research Journal, 3: 442-451 (in Persian).
9. Shenani Hoveizeh, S.M. and H. Zarei. 2016. Investigating land use changes over two decades of time, case study: Abol Abbas Watershed. Journal of Watershed Management, 14(7): 237-244 (in Persian).

10. Smith, A. 2010. Image segmentation scale parameter optimization and land cover classification using the random forest algorithm. Journal of Spatial Science, 55(1): 69-79 (in Persian).
11. Zebardast, L. and H.R. Jafari. 2010. Evaluation of trends in Anzali Wetland using remote sensing and management solution. Environmental Studies, 37(57): 57-64 (in Persian).