

تعیین قابلیت داده‌های سنجنده ASTER و روش‌های درخت طبقه‌بندی و رگرسیونی و جنگل تصادفی در تهیه نقشه تپ جنگل

- ❖ اصغر فلاح*؛ دانشیار گروه علوم جنگل‌داری، دانشکده منابع طبیعی، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران
- ❖ سیاوش کلبی؛ دانشجوی دکتری گروه علوم جنگل‌داری، دانشکده منابع طبیعی، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران
- ❖ شعبان شتایی جویباری؛ دانشیار گروه جنگل‌داری، دانشکده منابع طبیعی، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، گرگان، ایران
- ❖ امید کرمی؛ دانشجوی دکتری گروه علوم جنگل‌داری، دانشکده منابع طبیعی، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران

چکیده

شناسایی واحدهای همگن و تفکیک آن‌ها و نهایتاً برنامه‌ریزی برای هر واحد، اصولی‌ترین راه برای مدیریت واحدهای جنگلی محسوب می‌شود. هدف از این مطالعه، تهیه نقشه تپ جنگل با استفاده از داده‌های سنجنده ASTER و نیز استفاده از دو الگوریتم درخت طبقه‌بندی و رگرسیونی و جنگل تصادفی در سری یک جنگل آموزشی و پژوهشی دارابکلا است. برای این هدف، ابتدا با استفاده از شبکه آماربرداری 50×350 متر تعداد ۱۵۰ قطعه نمونه در کل سری برداشت شد. پس از پیش‌پردازش‌های لازم مانند تصحیح هندسی، اتمسفریک، پردازش‌های نسبت‌گیری باندها، ایجاد شاخص‌های متداول گیاهی، تجزیه مؤلفه اصلی، و تولید شاخص تسلدگپ (روشنایی، سبزی‌نگی، نمناکی) انجام شد. سپس مقادیر طیفی متناظر با قطعات نمونه بر روی باندهای اصلی و پردازش‌شده استخراج شد و ارزش‌های سایر پیکسل‌ها با استفاده از الگوریتم‌های مورد بررسی طبقه‌بندی شد. ارزیابی صحت نتایج طبقه‌بندی با تعدادی قطعه نمونه که در فرایند طبقه‌بندی شرکت نداشتند صورت گرفت. نتایج نشان داد که نقشه تهیه‌شده با استفاده از روش جنگل تصادفی با صحت کلی ۶۶ درصد و ضریب کاپای ۰/۵۸ در مقایسه با روش الگوریتم درخت طبقه‌بندی و رگرسیونی به ترتیب با صحت کلی ۵۸ درصد و ضریب کاپای ۰/۴۹ دارای صحت بالاتری است. در مجموع نتایج نشان داد که داده‌های سنجنده ASTER و الگوریتم‌های درخت طبقه‌بندی و رگرسیونی و جنگل تصادفی نتایج مقبولی در تهیه نقشه تپ جنگل در جنگل دارابکلا دارند.

واژگان کلیدی: تپ جنگل، درخت طبقه‌بندی و رگرسیونی، جنگل تصادفی، سنجنده ASTER.

مقدمه

جنگل‌های شمال ایران به‌عنوان تنها جنگل‌های تجاری کشور خدمات اکوسیستمی زیادی دارند. این خدمات شامل تهیه مواد غذایی، حفاظت از تخریب خاک، تنوع زیستی بسیار بالا، ذخیره کربن، و غیره است. خدمات اکوسیستم‌ها با تغییر تیپ‌های جنگلی تغییر می‌کند. در نتیجه اطلاع از وسعت و پراکنش تیپ‌های گوناگون جنگلی برای مدیریت پایدار جنگل‌ها بسیار مفید است. به هر حال داده‌های کمی موجود از وسعت و پراکنش تیپ‌های جنگلی کامل نیست و گاهی نیز صحیح نیست. نقشه تیپ در جنگل‌های شمال، به‌علت دسترسی نداشتن به تمام قسمت‌های جنگل و هزینه بالا و وسعت زیاد تیپ‌ها اغلب دقیق نیست. داده‌های سنجش از دور در صورت دقت زیاد می‌تواند یکی از به‌صرفه‌ترین راه‌ها برای تهیه نقشه تیپ جنگل باشد. تا به حال مطالعات زیادی برای تعیین قابلیت داده‌های سنجش از دور به‌منظور تهیه نقشه تیپ جنگل در داخل و خارج از کشور انجام شده است [۱-۳]. یکی از راه‌ها برای استخراج اطلاعات از تصاویر، طبقه‌بندی است. روش‌های گوناگونی برای طبقه‌بندی داده‌های سنجش از دور وجود دارد. روش‌های قدیمی، مانند جعبه‌ای، حداکثر احتمال و حداقل فاصله از میانگین، که برای طبقه‌بندی استفاده می‌شود، مشکلات اساسی دارند، زیرا این روش‌ها نمی‌توانند با مشخصات واقعی داده‌های سنجش از دور تطبیق شوند [۴]. یا به‌عبارتی در این روش‌ها معمولاً شرایط اولیه (همانند توزیع گوسن برای انعکاس طیفی یا همگنی پراکنش) با داده‌های سنجش از دور مطابقت ندارد [۴]. علاوه بر این، در این

روش‌ها داده‌ها نیاز به پیش‌فرض‌هایی همانند توزیع نرمال دارند. در اکثر مطالعات، خصوصاً در ایران، از روش‌های پارامتریک همانند روش حداکثر احتمال (ML)^۱ و روش بیزین و غیره برای طبقه‌بندی داده‌ها برای تهیه نقشه تیپ استفاده شده [۱، ۲، ۴، ۵] و اکثر نتایج فاقد صحت بالا برای مدیریت توده بوده است [۱، ۴، ۵]. این بدان معنی است که قدرت این داده‌ها به‌طور جدی در طبقه‌بندی تصاویر دارای مشکل است. روی‌هم‌رفته فرضیه نرمال بودن داده‌های طیفی در مناظر طبیعی نقص دارد. علاوه بر این، کم‌بودن نمونه‌های آموزشی و وجود نمونه‌های آموزشی غیر مرتبط به عدم قطعیت طبقه‌بندی با روش‌های پارامتریک کمک می‌کند. تا کنون نتیجه بسیاری از مطالعات نشان داده است که روش‌های ناپارامتری نتایج بهتری در مقایسه با روش‌های پارامتری دارند [۶]. مطالعه ساروناس [۷] نشان داد که حتی با نمونه‌های آموزشی کم، روش‌های ناپارامتریک دارای نتایج بهتری در مقایسه با روش‌های پارامتریک‌اند. مطالعه کورتیجو و بلانکا [۸] نشان داد که حتی اگر نمونه‌های تعلیمی دارای هم‌پوشانی باشند، روش‌های ناپارامتری به‌عنوان روش طبقه‌بندی قابلیت بهتری دارند. در روش‌های ناپارامتری به پیش‌فرض نرمال بودن داده‌ها و هیچ پارامتر آماری برای جداکردن طبقات تصاویر نیاز نیست [۳]. امروزه روش‌های آماری ناپارامتری زیادی گسترش یافته است که از این روش‌ها می‌توان به الگوریتم درخت طبقه‌بندی و رگرسیونی^۲ و جنگل تصادفی^۳ اشاره کرد. هیم و همکاران [۹] از الگوریتم

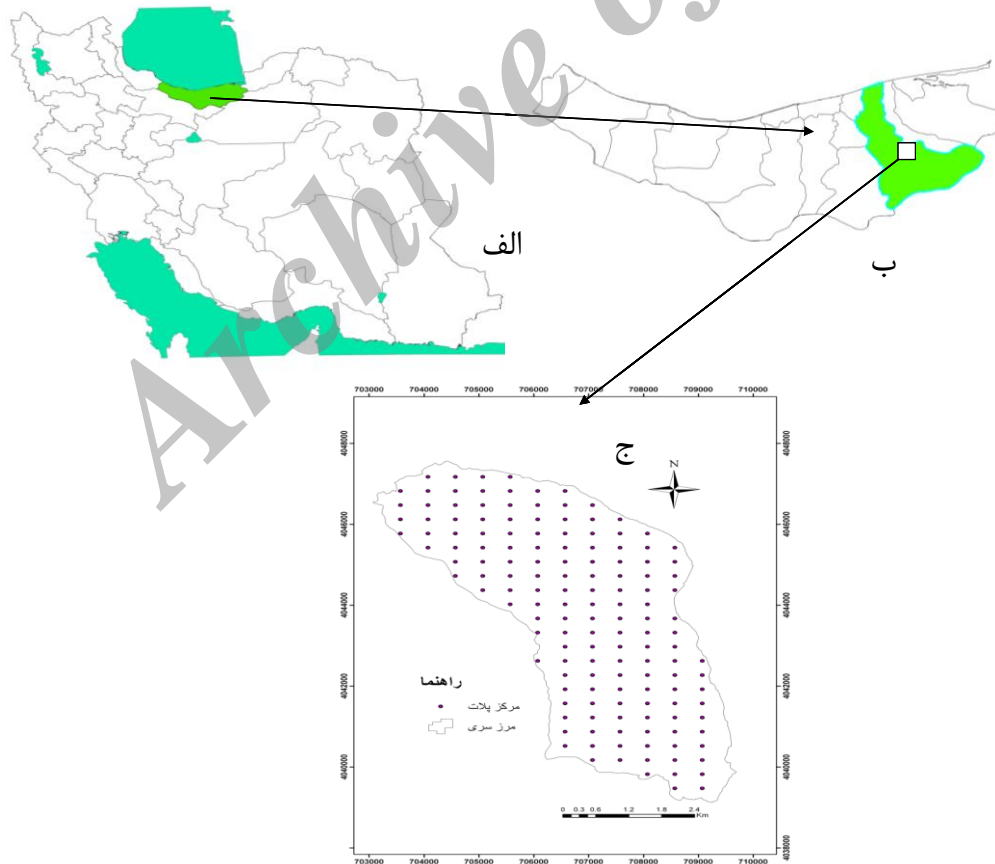
1. Maximum likelihood
2. Classification and Regression Tree
3. Random Forest

روش شناسی

منطقه مورد مطالعه

این تحقیق در جنگل آموزشی و پژوهشی دارابکلای ساری و با مساحتی حدود ۲۶۱۲ هکتار انجام گرفت که در هشت سال اخیر تحت مدیریت دانشگاه علوم کشاورزی و منابع طبیعی ساری بوده است. شکل ۱ موقعیت منطقه مورد مطالعه را نشان می‌دهد. گونه‌های اصلی موجود در این سری شامل راش، مرز، انجیلی، افرا، توسکا، و بلوط است. این منطقه در محدوده ارتفاعی بین ۱۴۰ تا ۸۸۰ متر ارتفاع از سطح دریا واقع شده و متوسط بارش سالیانه آن ۹۳۸/۸ میلی‌متر است. بر اساس روش آمبرژه منطقه مورد مطالعه اقلیم سرد و مرطوب دارد.

جنگل تصادفی برای طبقه‌بندی داده‌های فراطیفی استفاده کردند. مطالعه لاورنس و رایت [۱۰] نشان داد که آنالیز طبقه‌بندی درختی روش بسیار مناسبی برای طبقه‌بندی داده‌های سنجنش از دور در مواقعی است که داده‌های اولیه پیوستگی زیاد دارند. هدف اصلی این مطالعه تعیین قابلیت تهیه داده‌های سنجنده ASTER برای تهیه نقشه تیپ جنگل آموزشی و پژوهشی دارابکلا با دقت بالاست. هدف دیگر این مطالعه، مقایسه روش‌های ناپارامتری الگوریتم درخت طبقه‌بندی و رگرسیونی و جنگل تصادفی در تهیه نقشه تیپ جنگل است. این نتایج در صورت دقت بالا می‌تواند برای توده‌های دیگر و با داده‌های سنجنده‌های دیگر استفاده شود.



شکل ۱. موقعیت منطقه مورد مطالعه در ایران (الف)، استان مازندران (ب)، و شبکه آماربرداری و پلات‌ها (ج)

برداشت اطلاعات زمینی

تعیین تعداد قطعات نمونه مورد نیاز برای برآورد صحت نقشه‌های حاصل از طبقه‌بندی تصاویر ماهواره‌ای با توجه به هدف و وضعیت منطقه، متفاوت است. برای تعیین قطعات نمونه مورد نیاز جهت تهیه نقشه موضوعی مطالعات متعددی صورت گرفته است. گزارش‌های گوناگونی نیز حداقل تعداد پیکسل نمونه مورد نیاز برای طبقه‌بندی از هر کلاس را ۵۰ نمونه بیان کردند [۱۱]. در این مطالعه ۱۵۰ قطعه نمونه به شکل مربع با مساحت ۳۶۰۰ متر مربع با استفاده از یک شبکه آماربرداری با ابعاد ۵۰×۳۵۰ متر از شهر یور تا مهر ۱۳۹۰ برداشت شد (شکل ۱ ج). پس از ثبت نسبتاً دقیق (به روش میانگین‌گیری زمانی) موقعیت مرکز قطعه نمونه با استفاده از دستگاه موقعیت‌یاب، اطلاعات مربوط به نوع گونه و قطر برابر سینه درختان با قطر بیشتر از ۱۲/۵ سانتی‌متر برداشت شد. برای تعیین نوع تیپ جنگل از اصول کلی روش کوچلر استفاده شد. تیپ‌های موجود در منطقه به ۵ طبقه شامل راش، راش آمیخته، ممرز آمیخته، پهن‌برگ آمیخته، و جنگل کاری (طبقات ۱، ۲، ۳، ۴، ۵) و یک طبقه هم از نوع جاده و فضای خالی (طبقه ۶) نام‌گذاری شد. در این مطالعه مطابق با پیشنهاد درویش‌صفت و همکاران [۵] تیپ‌هایی با سطح کوچک در نظر گرفته نشد.

داده‌های ماهواره‌ای مورد استفاده در این مطالعه

در این مطالعه از داده‌های سنجنده ASTER مربوط به ۲۰۰۷/۱۰/۱۰ برابر با ۲۰ مهرماه ۱۳۸۸ با سطوح تصحیحات A۱ استفاده شد. این سنجنده دارای سه نوع باند VNIR^۱، SWIR^۲، و TIR^۳ است. قدرت تفکیک مکانی این سه نوع داده به ترتیب برابر ۱۵، ۳۰، و ۹۰ متر است. در این مطالعه به علت قدرت تفکیک مکانی و طیفی نامناسب از داده‌های حرارتی استفاده نشد.

پیش‌پردازش و پردازش تصاویر ماهواره‌ای

تصحیح هندسی تصاویر همراه با حذف اثر توپوگرافی بر هندسه تصویر با استفاده از مدل رقومی زمین منطقه^۴ (که از نقشه توپوگرافی ۱:۲۵۰۰۰ با دقت ارتفاعی ۱۰ متر استخراج شده است) و تعداد ۲۱ نقطه کنترل زمینی^۵ صورت گرفت. تصاویر VNIR با میزان خطای RMSE^۶ برابر با ۰/۶۳ پیکسل و تصاویر SWIR با میزان ۰/۶۱ پیکسل مورد تصحیح هندسی قرار گرفت. با عمل نمونه‌گیری مجدد و برای هماهنگ‌سازی ابعاد پیکسل‌های تمامی تصاویر، مقدار ابعاد پیکسل‌های تصاویر SWIR به ۱۵ متر تبدیل شد. برای بررسی میزان دقت تصحیحات هندسی صورت گرفته از لایه‌های وکتوری جاده‌ها استفاده شد و از صحت این تصحیحات اطمینان حاصل شد. برای کاهش اثر اتمسفر بر روی تصاویر از روش کلی COST استفاده شد. همچنین برای یکسان‌سازی میزان نوردهی^۷ و کاهش اثر توپوگرافی بر رادیومتری تصویر با استفاده از مدل رقومی زمین با اندازه سلول ۱۰ متری و زاویه‌های زینت^۸ و آزیموت^۹ خورشید اثر پستی و بلندی در نرم‌افزار Erdas Imaging 9.2 نرمال شد. در این روش از نسبت‌گیری‌های گوناگون، تولید شاخص‌های گیاهی، تجزیه مؤلفه اصلی برای باندهای VNIR و SWIR به صورت مجزا، و برای هر ۹ باند به صورت مشترک استفاده شد. همچنین سه مؤلفه تسلدگپ (روشنایی، سبزیگی، نمناکی) برای باندهای VNIR و SWIR طبق ضرایب جدول ارائه شده توسط یاربورگ [۱۲] تولید شد. در جدول ۱ تعدادی از شاخص‌های مورد استفاده در این مطالعه آورده شده است.

4. Digital Elevation Model

5. Ground control points

6. Root Mean Square error

7. Illumination

8. Zenith

9. Azimuth

1. Visible and Near Infrared

2. Short Wave Infrared

3. Thermal Infrared

جدول ۱. تعدادی از شاخص‌های گیاهی استفاده شده در این مطالعه

شاخص گیاهی	رابطه
Stress Index (SI)	Red/NIR
Differential Vegetation Index (DVI)	NIR-RED
Normalized Differential Vegetation Index (NDVI)	NIR-Red/NIR+Red
Moisture Stress Index (MSI)	SWIR/NIR
Modified Soil Adjusted Vegetation Index (MSAVI2)	$(0.5) \times (2 \times (NIR + 1) - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - R)})$
Simple Ratio (SR)	NIR/Red
Normalized Difference Water Index (NDWI)	NIR-SWIR/NIR+SWIR

با زیرمجموعه‌های نمونه است که در این مطالعه از روش برآورد با نمونه آزمون استفاده شده است؛ یعنی داده‌ها به دو قسمت، قسمتی برای طبقه‌بندی و قسمت دیگر برای ارزیابی طبقه‌بندی تقسیم شده‌اند. از مهم‌ترین بخش‌هایی که در مورد این روش باید در نظر گرفته شود شاخه‌بندی درخت طبقه‌بندی است. انتخاب شاخه‌های ساختار درخت طبقه‌بندی باید به‌گونه‌ای باشد که یکنواختی نمونه‌ها در تمامی گروه‌های پایانی به میزان حداکثر میزان ممکن برسد. این یکنواختی با استفاده از معیار عدم خلوص اندازه‌گیری می‌شود. مهم‌ترین و پرکاربردترین اندازه عدم خلوص، شاخص جینی است [۱۴]؛ و زمانی که شاخه‌بندی در ساختار درختی نیاز به توقف دارد نیز باید مشخص شود. در عمل ساختار درختی شده باید تا زمانی ادامه یابد که همه گروه‌های پایانی خالص یا اینکه هر گروه پایانی شامل حداقل تعداد معینی نمونه باشد (روش حداقل تعداد) یا اینکه سهم هر یک از گروه‌های پیامد بیش از یک مقدار معین نباشد [۱۴]. پارامتر دیگری که در این روش باید تعیین شود هرس کردن ساختار درختی و انتخاب ساختار با اندازه مناسب است. انتخاب ساختار درختی با اندازه مناسب موضوعی مهم در طبقه‌بندی درختی است. ساختارهای بزرگ و پیچیده درختی معمولاً در تفسیر نتایج با مشکلات زیادی همراه است. هدف این

طبقه‌بندی

درخت طبقه‌بندی و رگرسیونی

مدل درخت طبقه‌بندی و رگرسیونی یک روش ناپارامتری است که برای اولین بار از سوی بریمن و همکاران [۱۳] عمومیت یافت. این روش قادر به پیش‌بینی متغیرهای کمی (درخت رگرسیونی) و متغیرهای طبقه‌بندی شده (درخت طبقه‌بندی) بر اساس مجموعه‌ای از متغیرهای کمی و کیفی است. یک مدل درخت طبقه‌بندی و رگرسیونی متشکل از چند شاخه^۱ و چند گره^۲ است. اولین گره که مشتمل بر کلیه نمونه‌ها یا مشاهدات است گره والد^۳ نامیده می‌شود. بقیه گره‌ها فرزند نامیده می‌شود. سپس بر اساس وضعیت یکی از متغیرهای پیش‌بینی‌کننده، دو شاخه شکل می‌گیرد و این وضعیت ادامه می‌یابد تا به گره انتهایی، که معمولاً شامل گروه‌های هموزن از طبقات گوناگون پیامد است، ختم شود. معمولاً سه اندازه برای تعیین دقت پیش‌بینی یا طبقه‌بندی مدل طبقه‌بندی درختی و رگرسیونی وجود دارد که شامل برآورد با جایگزینی مجدد، برآورد با نمونه آزمون، و برآورد اعتباربخشی

1. Split
2. Node
3. Parent node

گرفته شده) با جایگزینی مجموعه داده‌های آموزشی اصلی ساخته می‌شوند. از مزایای این روش می‌توان به عدم نیاز به هرس درختان در موقع مدل‌سازی و طبقه‌بندی اشاره کرد. برای انجام این روش نیز باید چند پارامتر تعیین شود. اولین پارامتر تعداد درختان پیش‌بینی است که در این مطالعه برای تعیین درختان بهینه، تعداد ۲۰۰ درخت اولیه تولید شد. دومین پارامتر تعداد متغیر پیش‌بینی است. ساده‌ترین روش برای تعیین تعداد متغیر پیش‌بینی، محاسبه آن با استفاده از مجذور متغیرهای مستقل است [۱۹].

انتخاب بهترین باندها

استفاده از تعداد باندهای زیاد برای طبقه‌بندی، علاوه بر اینکه هیچ کمکی به بهبود فرایند طبقه‌بندی نمی‌کند، سبب کاهش سرعت و دقت طبقه‌بندی می‌شود. در این روش برای انتخاب بهترین باندها برای انجام دادن فرایند طبقه‌بندی از معیار اهمیت متغیر استفاده شد. اگرچه برای روش‌هایی مانند روش جنگل تصادفی نیاز به انتخاب بهترین متغیرها (باندها) برای طبقه‌بندی نیست، اما برای بالابردن قدرت تفسیر و تعیین اثر متغیرهایی که بیشترین تأثیر را در فرایند مدل‌سازی داشتند بسیار استفاده می‌شود. انتخاب بهترین متغیرها در مدل الگوریتم‌های درختان تصمیم‌گیری مشابه سایر مدل‌های پیش‌بینی نیست. به عبارت دیگر در مدل‌های پارامتریک مثل رگرسیون لجستیک معمولاً ابتدا رابطه تک تک متغیرهای مستقل با پیامد مورد نظر ارزیابی می‌شود، سپس متغیرهایی که رابطه معنی‌داری با پیامد مورد نظر داشته باشند وارد مدل چندمتغیره می‌شوند [۱۳]. اما در این مدل می‌توان متغیرهایی را وارد مدل کرد که از نظر منطقی می‌توانند با پیامد رابطه داشته باشند. برای تعیین اینکه کدام یک از متغیرها در پیش‌بینی پیامد اهمیت بیشتری دارند از اندازه‌ای به نام معیار اندازه اهمیت متغیر^۲

مرحله از تحلیل این است که به ساختار مناسبی از پیش‌بینی برسیم که ضمن داشتن دقت مقبول، در تفسیر نتایج به محقق کمک کند. در عمل برای رسیدن به این هدف می‌توان از دو اندازه یا روش اعتباربخشی نمونه‌آزمون و اعتباربخشی با زیرمجموعه نمونه استفاده کرد. به عبارت دیگر این دو اندازه هم در تعیین دقت (اعتبار) مدل و هم در انتخاب ساختار درختی با اندازه مناسب کاربرد دارند. این روش محاسن و معایبی نیز دارد که از محاسن آن قابلیت تفسیر بسیار بالا و ناپارامتری بودن، و از مشکلات آن صحت پایین‌تر نسبت به سایر روش‌های طبقه‌بندی است [۱۵].

روش جنگل تصادفی

روش جنگل تصادفی یک روش غیرپارامتری و متعلق به خانواده روش‌های دسته‌جمعی^۱ است که در اواخر قرن نوزدهم از روش‌های ماشین یادگیری به دست آمد [۱۶]. این الگوریتم را که شامل مجموعه‌ای از درخت‌های طبقه‌بندی و رگرسیونی است اولین بار بریمن [۱۷] توسعه داد. روش جنگل تصادفی از مجموعه‌ای از درخت‌های رگرسیونی که با بازسازی داده‌های آموزشی به کار گرفته شده است، ساخته می‌شود. در حالت عادی، مجموعه‌ای از نمونه‌ها هستند که به شکل تصادفی با جایگزینی داده‌های آموزشی اصلی شکل می‌گیرند. ترکیب سه پارامتر در الگوریتم روش جنگل تصادفی، ضروری است. اول اینکه چه تعداد درخت باید ساخته شود؛ دیگر اینکه چه تعداد از متغیرها در ایجاد یک گره برای هر شبکه شرکت می‌کنند؛ و پارامتر سوم، به اندازه گره برمی‌گردد که عمق درخت رگرسیون ساخته شده را نشان می‌دهد [۱۸]. در زمان ساخته شدن یک درخت رگرسیون، مجموعه آموزشی جدید (نمونه‌های به کار

2. Measure of Variable Importance

1. Ensemble

نتایج و بحث

تپ‌های شناسایی شده

تپ‌های موجود در منطقه به ۵ طبقه شامل راش، راش آمیخته، ممرز آمیخته، پهن‌برگ آمیخته، و جنگل کاری (طبقات ۱، ۲، ۳، ۴، ۵) و یک طبقه هم از نوع جاده و فضای خالی (طبقه ۶) نام‌گذاری شد.

انتخاب بهترین باندها

در این مطالعه برای انتخاب بهترین متغیرها جهت طبقه‌بندی از معیار اهمیت متغیر استفاده شد. نتایج مربوط به بهترین باندها با $P < 0/01$ و میزان اهمیت آن‌ها در جدول ۲ آورده شده است. بهترین متغیرها بر اساس معیار میزان اهمیت متغیر به ترتیب اهمیت شامل تجزیه مؤلفه اصلی برای ۹ باند (VNIR و SWIR)، مؤلفه سبزی‌نگی، شاخص NDVI، باند مادون قرمز نزدیک، و مؤلفه‌های نمناکی و روشنایی است. این نتایج مطابق با نتایج مطالعه رشیدی و همکاران [۴] است که نشان دادند باند مادون قرمز و مؤلفه‌های شاخص تسلدک اهمیت زیادی در تفکیک تپ‌های گیاهی دارند.

تهیه نقشه پوشش گیاهی با استفاده از طبقه‌بندی تصویر در صورتی کامل می‌شود که صحت طبقه‌بندی معلوم شود [۲]. نتایج نشان داد روش جنگل تصادفی نتایج بهتری در مقایسه با روش درخت طبقه‌بندی و رگرسیونی دارد. نتایج صحت کلی و ضریب کاپای به دست آمده برای طبقه‌بندی با استفاده از روش‌های جنگل تصادفی و درخت طبقه‌بندی و رگرسیونی در جدول ۳ و ۴ آورده شده است. در این مطالعه روش جنگل تصادفی با صحت کلی ۶۶ درصد و ضریب کاپای ۰/۵۷ نتایج بهتری در مقایسه با الگوریتم درخت طبقه‌بندی و رگرسیونی با صحت کلی ۵۸ درصد و ضریب کاپای ۰/۴۹ دارد.

استفاده می‌شود. این شاخص از رابطه ۱ محاسبه می‌شود.

$$M(X_m) = \sum_{t \in T} \Delta i(s_m^*, t) \quad (1)$$

بدیهی است در هر شاخه t برای هر متغیر m بیشترین کاهش در عدم خلوص جینی محاسبه می‌شود و از مجموع این مقادیر برای هر متغیر اندازه اهمیت آن متغیر تعیین می‌شود. برای مقایسه میزان اهمیت هر متغیر در مدل نسبت به سایر متغیرها به یک اندازه نسبی نیاز است. شاخص نسبی اهمیت متغیر، اندازه‌ای بین صفر تا ۱۰۰ دارد که از رابطه ۲ محاسبه می‌شود:

$$M_R(X_m) = \frac{M(X_m)}{\max M(X_m)} \times 100 \quad (2)$$

به عبارت دیگر، متغیری که بیشترین اندازه اهمیت را داشته باشد شاخص نسبی برابر با ۱۰۰ را پیدا می‌کند.

ارزیابی صحت طبقه‌بندی

در این مطالعه برای ارزیابی صحت از روش اعتباربخشی با نمونه آزمون استفاده شد. تقریباً حدود دوسوم نمونه‌ها برای طبقه‌بندی (۱۰۰ قطعه برای طبقه‌بندی) و یک سوم قطعات نمونه برای ارزیابی صحت طبقه‌بندی (۵۰ قطعه نمونه) استفاده شد. از معیارهای صحت کلی طبقه‌بندی و ضریب کاپا برای ارزیابی صحت طبقه‌بندی استفاده شد.

مقایسه صحت نقشه‌های موضوعی

هنگام مقایسه نقشه‌های موضوعی و برای انتخاب نقشه دارای صحت بالاتر، یک محقق می‌تواند به سادگی نقشه‌ای را که صحت کلی یا ضریب کاپای بالاتری دارد انتخاب کند؛ اما معلوم نمی‌شود که از نظر آماری بین آن‌ها تفاوت معنی‌داری وجود دارد و بنابراین یک آزمون آماری نیاز است. در این مطالعه از آزمون مک‌نمار [۲۰] استفاده شد.

1. McNemar's test

جدول ۲. بهترین متغیرها و میزان اهمیت‌ها آن‌ها در طبقه‌بندی

متغیر	F	P	متغیر	F	p
PCI9	۵۷	۰/۰۰۰	BRITNESS	۳۱/۷۳	۰/۰۰۰
GREENESS	۵۴/۸۶	۰/۰۰۰	PCI6	۲۸/۵۲	۰/۰۰۰
NDVI	۴۹/۴۱	۰/۰۰۰	PCI3	۱۹/۴۷	۰/۰۰۱
NIR	۳۶/۳۳	۰/۰۰۰	SWIR1	۱۶/۴	۰/۰۰۰
WITNESS	۳۵/۸۱	۰/۰۰۰	SWIR2	۱۵/۶۴	۰/۰۰۰۲

جدول ۳. جدول ماتریس خطای حاصل از روش درخت طبقه‌بندی و رگرسیونی

طبقه	راش	راش آمیخته	ممرز	ممرز آمیخته	پهن برگ	پهن برگ آمیخته	جنگل کاری	جاده و فضای خالی
راش	۶	۰	۲	۰	۰	۰	۰	۰
راش آمیخته	۳	۲	۲	۰	۳	۰	۰	۰
ممرز آمیخته	۲	۲	۵	۰	۱	۰	۰	۱
پهن برگ آمیخته	۱	۲	۰	۰	۵	۰	۰	۰
جنگل کاری	۰	۱	۰	۰	۰	۶	۰	۰
جاده و فضای خالی	۰	۰	۰	۰	۰	۰	۰	۶

صحت کلی: ۵۸ درصد؛ ضریب کاپا: ۰/۴۹

جدول ۴. جدول ماتریس خطای حاصل از روش جنگل تصادفی

طبقه	راش	راش آمیخته	ممرز	ممرز آمیخته	پهن برگ	پهن برگ آمیخته	جنگل کاری	جاده و فضای خالی
راش	۶	۰	۲	۰	۰	۰	۰	۰
راش آمیخته	۰	۶	۲	۰	۲	۰	۰	۰
ممرز آمیخته	۰	۲	۸	۰	۰	۰	۱	۰
پهن برگ آمیخته	۱	۳	۰	۰	۴	۰	۰	۰
جنگل کاری	۰	۰	۱	۰	۰	۰	۶	۰
جاده و فضای خالی	۰	۰	۰	۰	۰	۰	۳	۳

صحت کلی: ۶۶ درصد؛ ضریب کاپا: ۰/۵۷

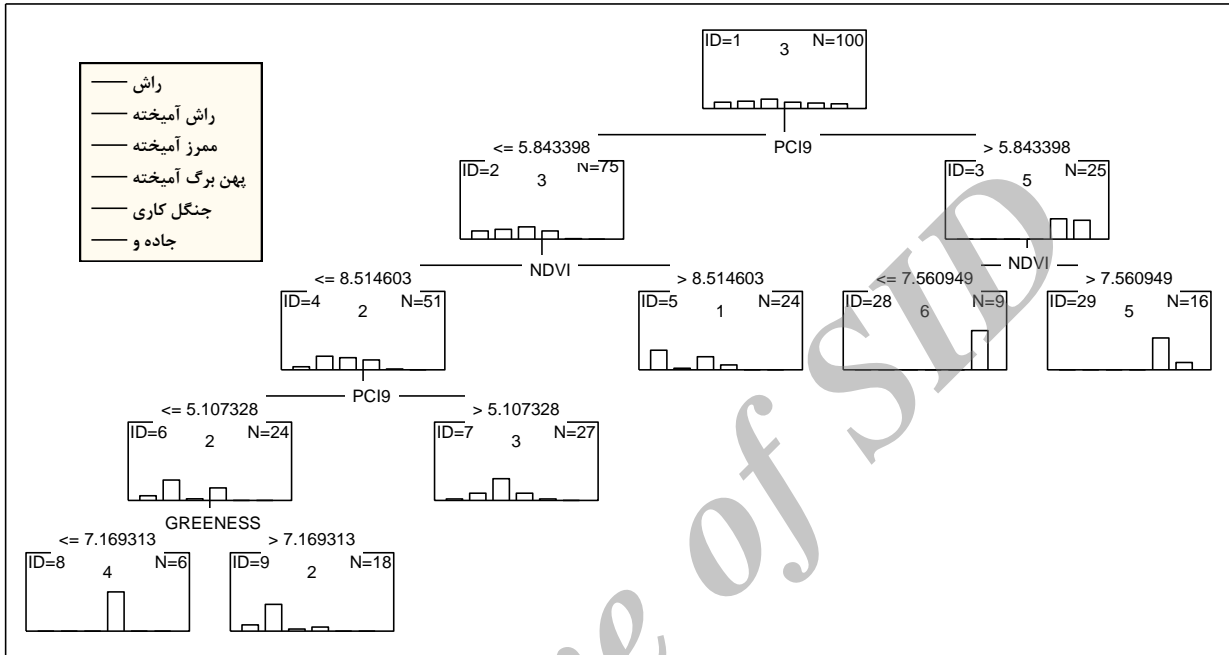
رگرسیونی آورده شده است. درخت مربوط به روش درخت طبقه‌بندی و رگرسیونی شامل ۵ گره فرزند غیر نهایی و ۶ گره نهایی است که باند مصنوعی PC9 گره والد است. در این درخت N تعداد نمونه موجود برای طبقه‌بندی و ID شماره گره است. سپس بر اساس وضعیت باند مصنوعی PC9، دو شاخه شکل گرفته است. این درخت شامل سه متغیر

نتایج آزمون مک‌نمار نشان داد که روش جنگل تصادفی به‌طور معنی‌داری در مقایسه با الگوریتم درخت طبقه‌بندی و رگرسیونی بهتر است ($X^2=20.23$, $P < 0.0001$). بریمن [۱۷] نشان داد استفاده از این روش برای طبقه‌بندی، باعث افزایش چشمگیری در نتایج می‌شود. در شکل ۲ درخت طبقه‌بندی حاصل از روش درخت طبقه‌بندی و

۶ گره نهایی است که باند NIR گره والد است. این درخت شامل چهار متغیر NIR، NDVI، WITNESS، و Greeness است.

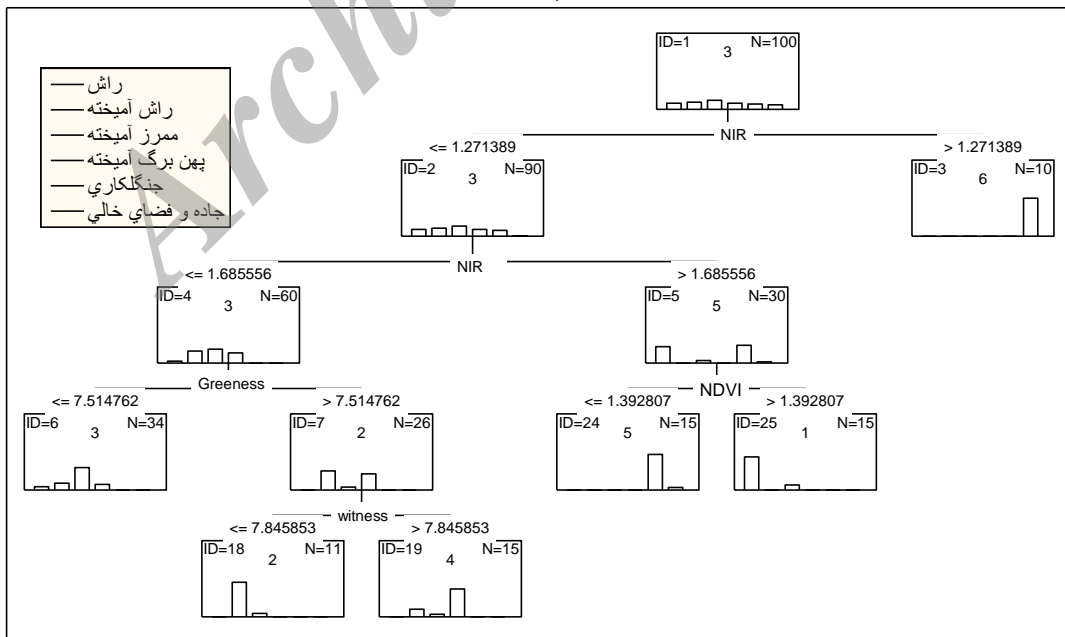
Greeness و NDVI، PC9 است. در شکل ۳ درخت حاصل از روش طبقه‌بندی جنگل تصادفی آورده شده است. این درخت نیز شامل ۵ گره فرزند غیر نهایی و

Tree 8 graph for az
Num. of non-terminal nodes: 5, Num. of terminal nodes: 6

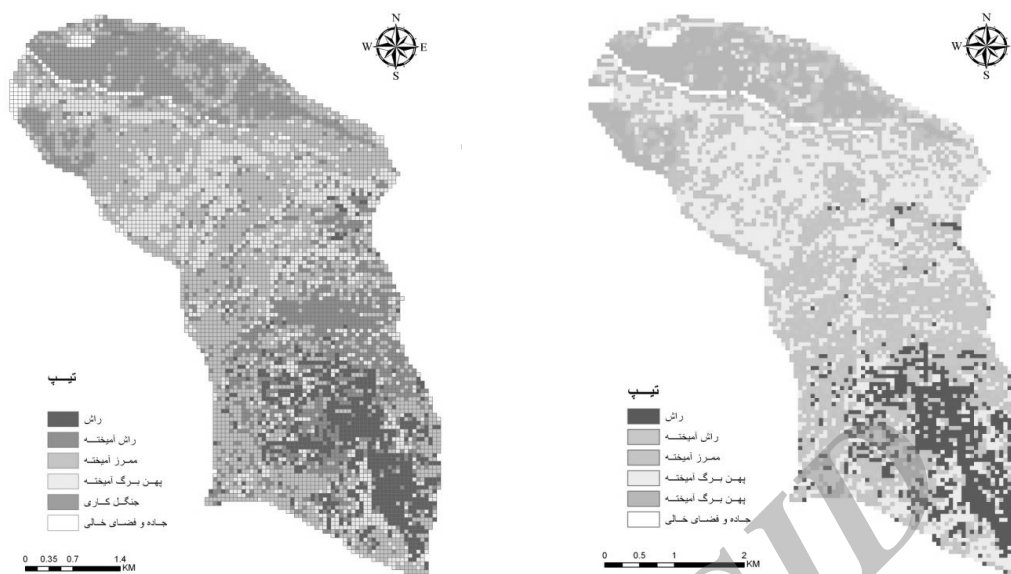


شکل ۲. درخت حاصل از روش درخت طبقه‌بندی و رگرسیونی

Tree 3 graph for az
Num. of non-terminal nodes: 5, Num. of terminal nodes: 6



شکل ۳. درخت حاصل از روش جنگل تصادفی



شکل ۴. نقشه تیب با استفاده از روش‌های درخت طبقه‌بندی و رگرسیونی و جنگل تصادفی

به‌طور کلی بدون در نظر گرفتن دیگر شرایط) در مقایسه با اکثر مطالعات انجام‌شده در ایران برای تیپ‌بندی جنگل، خصوصاً جنگل‌های شمال، وضعیت بهتری داشت [۱، ۴، ۵]. از تفاوت‌های این نتایج با نتایج دیگران می‌توان به نوع انتخاب داده‌های آموزشی و همچنین الگوریتم‌های به‌کاررفته در این مطالعه اشاره کرد. یکی دیگر از دلایل کسب چنین نتیجه‌ای را می‌توان در یکسان‌سازی میزان نوردهی تصویر و کاهش اثر توپوگرافی بر رادیومتری تصویر دانست که در بعضی از تحقیقات انجام‌شده از آن غافل بوده‌اند؛ اما در مقایسه با نتایج مطالعه باتووی و لی‌ودن [۲] صحت کمتری داشت. شاید یکی از دلایل اصلی زیاد بالانبودن صحت طبقه‌بندی این نتایج در مقایسه با سایر مطالعات انجام‌شده در خارج نوع تیپ و همچنین وجود پیکسل‌های مرزی باشد. مطابق با نتایج مطالعه رشیدی و همکاران [۴] با افزایش تعداد تیپ‌ها و جزئی‌تر شدن طبقه‌بندی میزان پیکسل‌های مرزی نیز افزایش می‌یابد و باعث بروز خطا می‌شود. برای افزایش صحت طبقه‌بندی استفاده از روش شی پایه با توجه به آمیختگی توده‌ها توصیه می‌شود.

تهیه نقشه تیب مربوط به هر تیپ

در شکل ۴ نقشه تیب با استفاده از روش‌های درخت طبقه‌بندی و رگرسیونی و جنگل تصادفی آورده شده است.

نتیجه‌گیری

تیپ‌های جنگلی به‌عنوان واحدهای نسبتاً همگن خدمات اکوسیستمی بسیار دارند. خدمات اکوسیستم‌ها با تغییر تیپ‌های جنگلی تغییر می‌کند، بنابراین تشخیص و شناسایی این واحدها برنامه‌ریزی را برای آن‌ها راحت‌تر می‌کند. در این مطالعه، برای اولین بار در کشور قابلیت دو روش ناپارامتریک درخت طبقه‌بندی و رگرسیونی و جنگل تصادفی برای تیپ‌بندی جنگل با استفاده از داده‌های سنجنده ASTER در جنگل آموزشی و پژوهشی دارابکلا مورد مقایسه و آزمون قرار گرفت. در این زمینه استفاده از داده‌های بارزسازی طیفی و داده‌های اصلی سبب بهبود طبقه‌بندی تیپ در هر دو روش طبقه‌بندی مورد استفاده در این مطالعه شد. صحت کلی به‌دست‌آمده در این مطالعه برای هر دو روش

References

- [1]. Shataee, Sh. (2003). Investigation on the possibility of beech forest type mapping using satellite data (Case study: Khyroud forest). Phd Thesis Tehran University, 155 pp.
- [2]. Baatuuwie, N.B., and Leeuwen I.L.V. (2011). Evaluation of three classifiers in mapping forest stand types using medium resolution imagery: a case study in the Offinso Forest District, Ghana. *African Journal of Environmental Science and Technology*, 5(1): 25-36.
- [3]. Quirós, E., Felicísimo Á.M., and Cuartero, A. (2009). Testing multivariate adaptive regression splines (MARS) as a method of land cover classification of TERRA-ASTER satellite images. *Sensors*, 9: 9011-9028.
- [4]. Rashidi, F., Babaie Kafaki, S., and Oladi, DJ. (2009). Investigation on the capability of digital data of ETM+ sensor in separating of forest types (Case study: Lafoor area of Savadkooh), *Iranian Journal of Forest and Poplar Research*, 17(13): 51-63.
- [5]. Darvishsefat, A.A., Abbasi, M., and Marvi Mohadjer, M.R. (2009). Investigation on the possibility of beech forest type mapping using Landsat ETM+ data (Case study: Khyroud forest), *Iranian Journal of Forest*, 2(9): 105-113.
- [6]. Kurt, I., Ture, M., and Kurum, A.T. (2008). Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert System Applied*, 34(1): 366-374.
- [7]. Sarunas, R. (1997). On dimensionality, sample size, and classification error of nonparametric linear classification algorithms *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19: 667-671.
- [8]. Cortijo, F.J., and de la Blanca, N.P. (1999). The performance of regularized discriminant analysis versus non-parametric classifiers applied to high-dimensional image classification. *International Journal Remote Sensing*, 20: 3345-336.
- [9]. Ham, J., Chen Y., Crawford, M.M., and Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):492-501.
- [10]. Lawrence, R., and Wright, A. (2001). Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric and Engineering Remote Sensing*, 67:1137-1142.
- [11]. Chen, D., and Stow, D. (2002). The effect of training strategies on supervised classification at different spatial resolutions. *Photogrammetric Engineering & Remote Sensing*, 68(11): 1155-1161.
- [12]. Yarbrough, L.D., Easson, G., and Kuszmaul, J.S. (2005). Using at-sensor radiance and reflectance tasseled cap transforms applied to change detection for the ASTER sensor. *IEEE Third International Workshop on the Analysis of Multi-temporal Remote Sensing Images*, 16-18.
- [13]. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone. C.J. (1984). *Classification and regression trees*. Wadsworth and Brooks/Cole, Monterey, California, USA.
- [14]. Yoneyama, Y., Suzuki, S., Sawa, R., Yoneyama, K., Power, G.G., and Araki, T. (2002). Increased plasma adenosine concentrations and the severity of preeclampsia. *Obstetrics & Gynecology*. 100(6): 66-70.
- [15]. Kalbi, S. (2010). Investigation estimation forest structure attributes using ASTER and SPOT-HRG satellite data (case study: Darabkola Forest). *MSC Thesis Sari University*. 107 pp.
- [16]. Dietterich, T. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization. *Machine Learning*, 40: 139-157.

- [17]. Breiman, L. (2001). Random forests. *Machine Learning*, 45:15–32.
- [18]. Yu. X., Hyyp, J., Vastaranta. M., Holopainen, M., Viitala., R. (2011). Predicting individual tree attributes from airborne laser point clouds based on the random forests technique. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66: 28–37.
- [19]. Shataee, S., Kalbi, S., Fallah, A., Pelz, D. (2012). Forest attribute imputation using machine-learning methods and ASTER data: comparison of k-NN, SVR and random forest regression algorithms. *International Journal of Remote Sensing*, 33(19): 6254-6280.
- [20]. Foody, G.M. (2004). Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 70: 627-634.

Archive of SID