

ارزیابی عملکرد سه روش طبقه‌بندی تصویر (جنگل تصادفی، ماشین بردار پشتیبان و بیشترین شباهت) در تهیه نقشه کاربری اراضی

فرشید جهانبخشی و محمدرضا اختصاصی^{*۱}

(تاریخ دریافت: ۱۳۹۶/۸/۱۵؛ تاریخ پذیرش: ۱۳۹۶/۱۲/۵)

چکیده

نقشه‌های کاربری/ پوشش اراضی ورودی پایه برای بسیاری از مدل‌های شبیه‌سازی محیط طبیعی است؛ بنابراین، صحت نقشه‌های حاصل از طبقه‌بندی تصاویر ماهواره‌ای، عدم قطعیت را در مدل‌سازی کاهش می‌دهد. این مطالعه با هدف ارزیابی صحت نقشه‌های کاربری اراضی تولید شده توسط روش‌های طبقه‌بندی مبتنی بر یادگیری ماشین (الگوریتم جنگل تصادفی و ماشین بردار پشتیبان) و مقایسه آن با روش متداول بیشترین شباهت انجام شد. بدین منظور تصویر سنجنده OLI ماهواره لندست ۸، مربوط به منطقه مورد مطالعه (حوضه سد ستارخان در آذربایجان شرقی)، پس از انجام تصحیحات اولیه، مورد استفاده قرار گرفت. پنج کاربری شهر، کشاورزی آبی، کشاورزی دیم، مرتع و پهنه آبی مورد توجه قرار گرفت. داده‌های واقعیت زمینی در قالب دو مجموعه داده‌های تعلیمی (۷۰ درصد از نمونه‌ها) و داده‌های آزمون (۳۰ درصد) برای انجام طبقه‌بندی نظارت شده استفاده شد. صحت نقشه‌های حاصل از سه الگوریتم، با استفاده از شاخص‌های ارزیابی صحت مورد مقایسه قرار گرفت. همچنین به منظور بررسی اختلاف معنادار آماری میان نتایج طبقه‌بندی از آزمون مک-نمار استفاده شد. نتایج نشان داد، صحت کل برای روش ماشین بردار پشتیبان، جنگل تصادفی و بیشترین شباهت به ترتیب برابر با ۹۶/۶، ۹۰/۸ و ۹۰/۸ درصد و ضریب کاپا به ترتیب ۰/۹۳۴، ۰/۸۱۳ و ۰/۸۳۴ بوده است. نتایج آزمون مک-نمار نیز معناداری اختلاف عملکرد در سطح پنج درصد آماری روش ماشین بردار پشتیبان با دو روش دیگر را تأیید کرد.

واژه‌های کلیدی: یادگیری ماشین، طبقه‌بند ناپارامتری، آزمون مک-نمار، الگوریتم جنگل تصادفی

۱. گروه مرتع و آبخیزداری، دانشکده منابع طبیعی، دانشگاه یزد، یزد

*: مسئول مکاتبات: پست الکترونیکی: mr_ekhtesasi@yazd.ac.ir

مقدمه

عملکرد روش جنگل تصادفی شی‌گرا را نسبت به روش‌های ماشین بردار پشتیبان و بیشترین شباهت با استفاده بهتر دانستند و این روش را برای استخراج نقشه اراضی کشت آبی پیشنهاد دادند.

مروری بر مطالعات گذشته نشان می‌دهد، الگوریتم ناپارامتری جنگل تصادفی در تهیه نقشه کاربری/ پوشش اراضی در کشور کمتر مورد توجه قرار گرفته است. از طرفی در ارزیابی کارایی روش‌های مختلف طبقه‌بندی تصاویر ماهواره‌ای اغلب تنها به گزارش صحت کل و شاخص کاپای روش‌ها اکتفا شده است و معناداری آماری اختلاف عملکرد این روش‌ها مورد توجه قرار نگرفته است. استفاده از آزمون آماری مناسب می‌تواند بهتر بودن عملکرد یک روش طبقه‌بندی نسبت به روش دیگر را از لحاظ معناداری آماری اثبات کند (۲۲). این نکته در انتخاب یک الگوریتم نسبت به الگوریتم دیگر اهمیت دارد. چنانچه دو روش یا الگوریتم طبقه‌بندی از نظر معناداری آماری عملکرد، مورد آزمون قرار گیرند و چنانچه علی‌رغم تفاوت مقادیر شاخص‌های صحت، اختلاف آنها معنادار نباشد، بهتر خواهد بود که از روش ارزان‌تر، سریع‌تر و کاراتر در استخراج نقشه استفاده شود (۱۳). مطالعه حاضر با هدف ارزیابی صحت نقشه‌های کاربری/ پوشش اراضی استخراج شده از تصویر ماهواره لندست ۸ توسط روش‌های طبقه‌بندی مبتنی بر یادگیری ماشین (الگوریتم جنگل تصادفی RF و ماشین بردار پشتیبان SVM) و مقایسه آن با روش متداول بیشترین شباهت (Maximum Likelihood Classification, MLC) انجام شد.

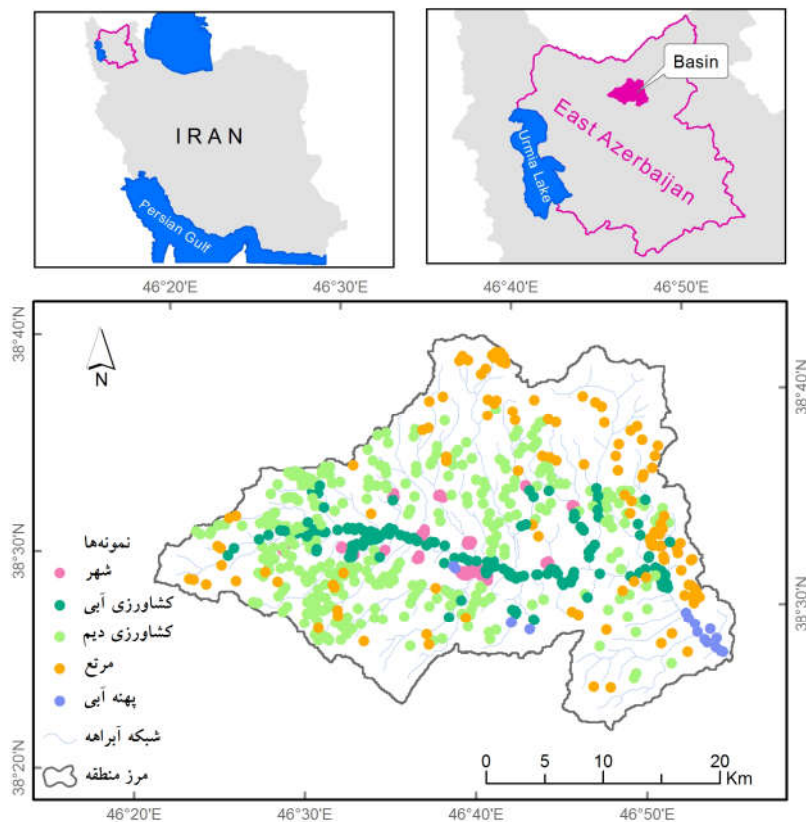
مواد و روش‌ها

منطقه مورد مطالعه

منطقه مورد مطالعه با وسعت ۹۳۰ کیلومتر مربع بین عرض جغرافیایی ۲۰° ۳۸' تا ۴۵° ۳۸' شمالی و ۴۶° ۲۰' تا ۴۷° ۰۰' شرقی، در بالادست سد ستارخان حوضه آبخیز اهرچای در شمال غرب ایران است (شکل ۱). متوسط دمای سالانه منطقه ۱۱ درجه سانتی‌گراد و میانگین میزان بارش ۲۸۵ میلی‌متر است.

نقشه‌های کاربری/ پوشش اراضی مستخرج از تصاویر ماهواره‌ای، اطلاعات پایه را برای مدیریت و پایش سیستم‌های محیطی فراهم می‌کنند. ارزش و قابلیت استفاده از هر نقشه تولید شده از تصاویر ماهواره‌ای به درجه صحت آن بستگی دارد (۲). صحت نقشه‌ها وابسته به فاکتورهایی نظیر صحت داده‌های اولیه و روش طبقه‌بندی تصویر ماهواره‌ای است (۲۵). طبق تعریف طبقه‌بندی اطلاعات ماهواره‌ای، شامل جداسازی مجموعه‌های طیفی مشابه و تقسیم‌بندی طبقاتی آنها از لحاظ رفتار طیفی یکسان است. به عبارت بهتر، در طبقه‌بندی تصویر ماهواره‌ای، الگوریتم طبقه‌بندی با توجه به ارزش عددی یک پیکسل، در مورد تعلق آن به کلاس یا پدیده‌ای خاص تصمیم می‌گیرد (۱).

در ارزیابی عملکرد الگوریتم‌های طبقه‌بندی جدید در تهیه نقشه‌های صحیح کاربری/ پوشش اراضی، به‌طور معمول کارایی این الگوریتم‌ها با روش‌های متداول نظیر حداکثر احتمال مورد مقایسه قرار می‌گیرد (۴، ۱۰ و ۲۳). آرخی (۲) در مقایسه کارایی چند روش طبقه‌بندی در تهیه نقشه کاربری اراضی، عملکرد ماشین بردار پشتیبان (Support Vector Machine, SVM) را با گزارش صحت کل (Overall Accuracy) و ضریب کاپا (Kappa's Coefficient) از روش حداکثر احتمال بهتر ارزیابی کرد. در مطالعه‌ای دیگر رینولدز و همکاران (۲۰) به‌منظور تهیه نقشه پوشش اراضی با هدف تفکیک زیستگاه‌های جانوری از روش طبقه‌بندی جنگل تصادفی (Random Forest, RF) استفاده کردند، ایشان دقت این روش را با ضریب کاپای ۶۶ درصد، شاخص صحت کل ۸۸ درصد و صحت کاربر ۷۲ درصد خوب و قابل قبول ارزیابی نمودند. اسحاقی و شتایی جویباری (۳) نیز عملکرد روش جنگل تصادفی با صحت کل ۷۵ درصد در تهیه نقشه خطر آتش سوزی را نسبت به دو روش شبکه عصبی مصنوعی (صحت کل ۷۲) و ماشین بردار پشتیبان (صحت کل ۶۸) بهتر ارزیابی کردند. باسوکلا و همکاران (۱۰) با هدف ارائه بهترین الگوریتم برای تهیه نقشه زمین‌های کشاورزی آبی،



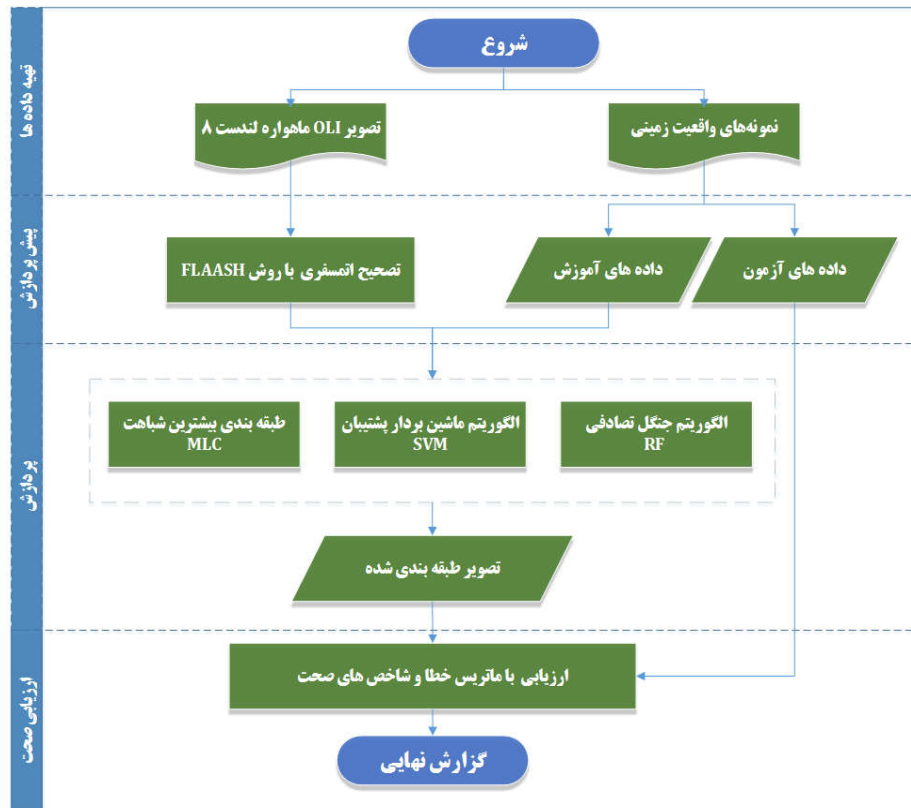
شکل ۱. موقعیت حوضه مورد مطالعه در ایران و استان آذربایجان شرقی و موقعیت نمونه‌های مورد استفاده

و از آنجا که خطای هندسی قابل توجهی تشخیص داده نشد از تصحیح هندسی تصویر صرف نظر شد. به منظور تصحیح اتمسفری تصویر، ابتدا تبدیل DN به رادیانس مطابق رویه (<http://landsat.usgs.gov/>) که در بسته Envi ارائه شده است، انجام شد. سپس تصحیح اتمسفری روی انعکاس سطحی با استفاده از الگوریتم FLAASH، انجام شد (۷، ۱۵ و ۲۱). در نهایت تصویر اصلاح شده برای استخراج نقشه کاربری اراضی استفاده شد.

تهیه نمونه‌های تعلیمی و واقعیت زمینی امری ضروری در طبقه‌بندی نظارت شده است. امروزه با گسترش داده‌های تصویری رایگان، با وضوح بالا و با دقت مکانی بالای گوگل ارث بخش قابل توجهی از مراجعات هزینه‌بر به عرصه کاهش یافته است، به طوری که متخصص آشنا به منطقه می‌تواند با مراجعه به این تصاویر نسبت به تهیه نمونه‌ها اقدام کند (۱۹).

ارتفاع حوضه بین ۱۴۰۰ تا ۲۹۵۰ متر از سطح دریا است. رودخانه اهرچای با امتداد غربی- شرقی از زیرحوضه‌های رودخانه ارس است. کشاورزی آبی، کشاورزی دیم، مرتع، سطوح توسعه یافته (شهری، روستایی و صنعتی) و پهنه‌های آبی کاربری/ پوشش‌های غالب در منطقه هستند.

شکل ۲ نمودار جریانی انجام تحقیق را نشان می‌دهد. در این مطالعه تصویر سنجنده OLI (Operational Land Imager) ماهواره Landsat8 (مسیر ۱۶۸، ردیف ۳۳، تاریخ ۱۴ ژوئن ۲۰۱۵) فاقد ابر و غبار از پایگاه داده USGS Earth Explorer (<http://earthexplorer.usgs.gov/>) تهیه شد. تصاویر لندست ۸ توسط سازمان زمین‌شناسی ایالات متحده آمریکا به صورت زمین مرجع شده ارائه می‌شوند و نیازی به تصحیح هندسی ندارند (۸). با این حال، راه‌های ارتباطی استخراج شده از نقشه توپوگرافی سازمان نقشه برداری کشور با تصویر انطباق داده شد



شکل ۲. نمودار جریان‌ی انجام پژوهش

به‌منظور تهیه پلی‌گون‌های آموزش الگوریتم طبقه‌بندی و آزمون صحت، از تصاویر با وضوح بالای گوگل ارث استفاده شد (۲۰). این پلی‌گون‌ها توسط کارشناس آشنا به منطقه، به‌صورت تصادفی، با پراکنش و تعداد مناسب بسته به وسعت کلاس‌های کاربری (شهر، کشاورزی آبی، کشاورزی دیم، مرتع و پهنه آبی) تهیه شد. به‌منظور اطمینان از صحت نمونه‌ها، تعدادی از آنها با مراجعه به عرصه مورد بازدید قرار گرفت و صحت محل و کاربری پلی‌گون‌ها تأیید شد. درنهایت ۲/۳ از کل ۷۳۰ پلی‌گون برای آموزش و استخراج خصوصیات طیفی جهت طبقه‌بندی نظارت شده تصویر و ۱/۳ باقیمانده برای آزمون استفاده شد (شکل ۱).

طبقه‌بندی تصویر

فرایند انجام طبقه‌بندی تصویر در این مطالعه با استفاده از بسته‌های نرم‌افزاری Envi/IDL 5.3، Google Earth Pro 7.1،

روش طبقه‌بندی بیشترین شباهت
روش بیشترین شباهت Maximum Likelihood Classification از متداول‌ترین روش‌های طبقه‌بندی نظارت شده در سنجش از دور است. اصل تصمیم در این روش بر اساس احتمال تعلق یک پیکسل به یک کلاس خاص است. معادله پایه این روش فرض می‌کند که احتمالات برای تمامی کلاس‌ها یکسان است و باندهای ورودی دارای توزیع نرمال هستند (۱۸). در این روش احتمال تعلق یک پیکسل به تمامی کلاس‌ها محاسبه شده و

الگوریتم ماشین بردار پشتیبان

ماشین بردار پشتیبان نیز از جمله الگوریتم‌های یادگیری ماشین ناپارامتری است. در تعریف ساده، SVM الگوریتمی است که به وسیله یک ابرصفحه جداکننده که روی داده‌های آموزشی تعریف می‌شود، دو کلاس خاص از هم جدا و مشخص می‌شوند (۶). از جمله قابلیت‌های ماشین بردار پشتیبان، توانایی غلبه بر مشکل توزیع غیر خطی داده‌های آموزشی است. در این حالت با استفاده از توابع کرنل، داده‌ها به فضایی به بعد بزرگ‌تر انتقال می‌یابند که در آن تفکیک‌پذیری بهتر انجام می‌شود و ابرصفحه جداکننده در آن فضا تعیین می‌شود. در سنجش از دور کاربردی، کرنل تابع پایه شعاعی Radial Base Function RBF گوسی از نظر کارایی و زمان پردازش مناسب با ارزش است (۳). تابع کرنل پایه شعاعی گوسی به صورت زیر تعریف شده است:

$$k(x_i, x) = \exp\left(-g|x_i - x|^2\right) \quad (2)$$

برای مدل‌سازی و برآورد مشخصه‌های کمی توسط مدل SVM و کرنل تابع پایه شعاعی انتخاب مقدار پارامتر g و c لازم است. هر دو این پارامترها، وابسته به بازه توزیع داده‌ها هستند و از یک مسأله طبقه‌بندی به طبقه‌بندی دیگر متفاوت هستند. روش معمول برای انتخاب مقادیر مناسب پارامتر c و g استفاده از یک شبکه جستجوی دو بعدی است که در این مطالعه بدین منظور از جستجوی شبکه در نرم‌افزار EnMap-Box استفاده شد و مقدار بهینه پارامترها در طبقه‌بندی لحاظ شد.

ارزیابی صحت نقشه‌ها

از آنجا که هر کدام از شاخص‌های مستخرج از ماتریس صحت، اطلاعات متفاوتی را ارائه می‌دهند و در مقایسه عملکرد الگوریتم‌ها، بسته به هدف استفاده کننده، می‌توان چندین شاخص را مد نظر قرار داد، بنابراین در مطالعه حاضر شاخص‌های صحت تولید کننده، صحت کاربر، واحد FI، صحت کلی و ضریب کاپا محاسبه و گزارش شد. همچنین از آماره S آزمون مک نمار برای بررسی معنادار بودن اختلاف

پیکسل به کلاس با بیشترین احتمال تعلق می‌گیرد (۴).

طبقه‌بندی بیشترین شباهت برای یک پیکسل X در تصویر با n باند به صورت زیر بیان می‌شود (۱۶):

$$p_x(X) = \text{Exp}\left(-\frac{1}{\gamma} \ln \left| \sum_k \left(-\frac{1}{\gamma} X - m_k\right)^{\sum_k (X - m_k)} \right.\right) \quad (1)$$

که در آن n تعداد باندها، X داده تصویر در n باند، $p_x(X)$ احتمال تعلق پیکسل X به کلاس k ، k واریانس - کواریانس ماتریس کلاس k و $| \bullet_k |$ دترمینان \bullet_k است.

الگوریتم جنگل تصادفی

الگوریتم جنگل تصادفی در حال حاضر یکی از بهترین الگوریتم‌های یادگیری است (۵). این الگوریتم یک الگوریتم یادگیری ماشین ناپارامتری مبتنی بر دسته‌ای از درخت‌های تصمیم است. تعداد بسیاری درخت تصمیم در طبقه‌بندی الگوریتم RF رشد می‌یابند. پیکسل یا پدیده طبقه‌بندی نشده با توجه به ویژگی‌های همراه خود در یک کلاس گنجانده می‌شود. این امر با عبور دادن پیکسل از میان X درخت تصمیم انجام می‌شود. به گونه‌ای که تعلق پیکسل به یکی از کلاس‌ها (Y تعداد کلاس) میان درختان تصمیم به رای گذاشته می‌شود و هر درخت تصمیم رای خود را در مورد گنجاندن پیکسل در یک کلاس اعلام می‌کند. جنگل پیکسل به رای گذاشته شده را به کلاسی که دارای بیشترین رای درختان جنگل باشد، اختصاص می‌دهد. درختان تصمیم‌گیری به صورت منفرد از مجموعه نمونه آموزشی (مجموعه آموزشی کیسه) رشد می‌کنند. با N بار نمونه‌گیری با جایگزینی، دو سوم از مجموعه داده‌های اصلی برای آموزش یک درخت استفاده می‌شود، N تعداد نمونه‌ها در مجموعه داده‌های اصلی است. بدین طریق با نمونه‌گیری با جایگزینی، یک سوم باقیمانده داده‌ها در آموزش درخت دخالت نخواهند داشت و از کیسه کنار گذاشته می‌شوند تا برای صحت‌سنجی درونی الگوریتم مورد استفاده قرار گیرند. کارایی RF در مورد مجموعه‌های بسیار بزرگ به اثبات رسیده است از این رو می‌تواند در تجزیه و تحلیل داده‌های ماهواره‌ای مورد استفاده قرار گیرد (۱۰ و ۱۱).

عملکرد بین دو الگوریتم استفاده شد.

$$\hat{K} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} * x_{+i})} \quad (4)$$

که در آن r تعداد سطر در ماتریس خطا، x_{ij} تعداد مشاهدات در سطر i و ستون j ، x_{i+} و x_{+i} به ترتیب مجموع سطر i ام و ستون i ام و N تعداد کل مشاهدات است.

آزمون مک نمار

آزمون مک نمار توسط دیتریچ (۱۴) برای بررسی اختلاف معنادار میان دو روش طبقه‌بندی پیشنهاد شد (۱۲). این آزمون می‌تواند اختلاف معنادار آماری میان دو ماتریس خطای حاصل از دو الگوریتم طبقه‌بندی را تعیین کند، به عبارت دیگر می‌تواند بهتر بودن روش الف نسبت به روش ب را از لحاظ معناداری آماری اثبات کند. برای اجرای این آزمون دو الگوریتم A و B با قسمتی (R) از کل داده‌ها (D) آموزش می‌بینند و مدل‌های \hat{f}_A و \hat{f}_B ایجاد می‌شود. سپس این دو مدل با قسمت دیگر داده‌ها (T) مورد آزمون قرار می‌گیرند. برای هر نمونه $x \in T$ ، چگونگی طبقه‌بندی آن ثبت می‌شود و جدول احتمالاتی زیر تشکیل می‌شود:

n_{01} : تعداد نمونه‌هایی که با مدل \hat{f}_A نادرست طبقه‌بندی شده‌اند اما با \hat{f}_B نه

n_{11} : تعداد نمونه‌هایی که نه با مدل \hat{f}_A و نه با مدل \hat{f}_B نادرست طبقه‌بندی شده‌اند

n_{00} : تعداد نمونه‌های نادرست طبقه‌بندی شده توسط هر دو مدل \hat{f}_A و \hat{f}_B

n_{10} : تعداد نمونه‌هایی که با مدل \hat{f}_B نادرست طبقه‌بندی شده‌اند اما با \hat{f}_A نه

که در آن $n = n_{00} + n_{01} + n_{10} + n_{11}$ تعداد کل نمونه‌های آزمون (T) است. تحت فرض صفر این آزمون، دو الگوریتم باید نرخ خطای یکسانی داشته باشند ($n_{01} = n_{10}$). آزمون مک نمار برای نکویی برآزش بر پایه آزمون کای اسکور χ^2 استوار است. بر این اساس آماره مک نمار S به صورت زیر تعریف می‌شود:

صحت تولید کننده (Producer's accuracy) و صحت کاربر (User's accuracy) دو شاخص معمول در ارزیابی دقت طبقه‌بندی تصویر هستند. بنا به تعریف صحت تولید کننده احتمال نسبت دادن کلاس خاص به یک پیکسل توسط الگوریتم طبقه‌بندی است، در شرایطی که کلاس واقعی آن پیکسل برای تولید کننده مشخص باشد. این شاخص از تقسیم عنصر قطری هر کلاس در ماتریس خطا به مجموع مقادیر ستون همان کلاس به دست می‌آید. صحت کاربر نیز بیانگر احتمال طبقه‌بندی یک کلاس خاص مطابق با همان کلاس در نقشه واقعیت زمینی را بیان می‌کند و در ماتریس خطا برای هر کلاس، از طریق تقسیم عنصر قطری به مجموع عناصر سطر همان کلاس حاصل می‌شود (۲۲).

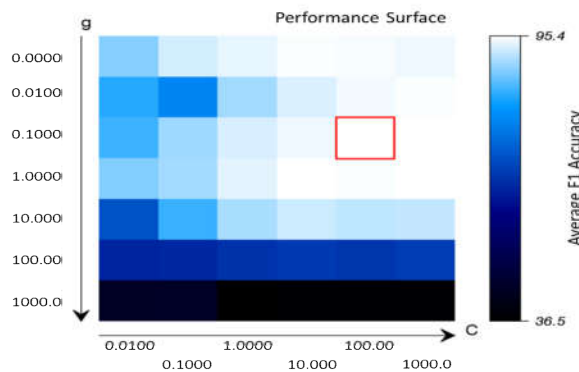
واحد F1 (F1 measure) میانگین همساز وزنی صحت تولید کننده (PA) و صحت کاربر (UA) است. برخی این شاخص را شاخص میانگین صحت هلدن (Hellden's mean accuracy index) نیز نامیده‌اند. این شاخص برای کلاس i به صورت رابطه (۳) محاسبه می‌شود (۱۷):

$$F1_i = \frac{2 \times UA_i \times PA_i}{(UA_i + PA_i)} \quad (3)$$

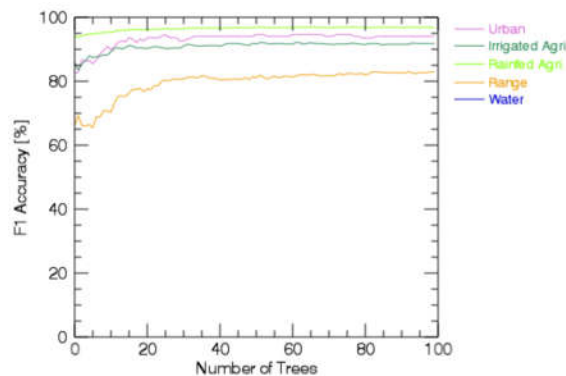
میانگین صحت F1 از میانگین حسابی F1 کلاس‌ها حاصل می‌شود.

صحت کلی (Overall accuracy) از جمع عناصر قطر اصلی ماتریس خطا تقسیم بر تعداد کل پیکسل‌ها به دست می‌آید. صحت کلی طبقه‌بندی که بیانگر میزان اعتبار طبقه‌بندی انجام شده است، در نقشه‌های کاربری استخراج شده از تصاویر ماهواره‌ای باید بیش از ۸۵ درصد باشد (۴).

تحلیل کاپا یک روش چند متغیره گسسته است که در ارزیابی صحت مورد استفاده قرار می‌گیرد. نتیجه انجام تحلیل کاپا تخمینی از کاپا است که آماره KHAT نامیده می‌شود. این آماره یکی از معیارهای متداول ارزیابی میزان توافق یا صحت است (۱۳). آماره KHAT برای یک ماتریس خطا به صورت زیر محاسبه می‌شود (۲۲):



شکل ۴. شبکه جستجوی دو بعدی برای تعیین بهینه مقادیر پارامترهای c و g در هنگام مدل‌سازی SVM



شکل ۳. منحنی یادگیری هنگام ساخت مدل RF

برای مدل‌سازی RF مناسب بوده است.

شکل ۴ شبکه جستجوی دو بعدی برای تعیین بهینه مقادیر پارامترهای c و g در هنگام مدل‌سازی SVM را نشان می‌دهد. بر این اساس هنگام مدل‌سازی با داده‌های آموزش، با انتخاب مقادیر ۱۰۰ و ۱۰٪ به ترتیب برای پارامتر c و g، کارایی مدل بر اساس میانگین صحت F1 (۹۵/۴) بیشینه خواهد بود.

شکل ۵ بخشی از تصویر طبقه‌بندی شده توسط الگوریتم‌های مورد استفاده را نشان می‌دهد. در مقایسه بصری در مورد دو کلاس شهر و پهنه آبی می‌توان گفت که روش MLC نتوانسته است دو پهنه آبی موجود در گوشه راست پایین (ربع جنوب شرقی) را تشخیص دهد، در حالی که دو روش در تشخیص این دو عارضه عملکرد بهتری داشته‌اند. در مورد کلاس شهر نیز باید گفت روش MLC پیکسل‌های بیشتر و پراکنده‌ای را برای کلاس شهر اختصاص داده است. این مورد را با دقت در ربع جنوب غربی می‌توان دریافت به گونه‌ای که در اطراف پهنه آبی (سد خاکی موجود در منطقه) پیکسل‌هایی را به کلاس شهر اختصاص داده است که خلاف واقع است.

جدول ۱ ماتریس خطای الگوریتم‌های مورد استفاده را نشان می‌دهد. با دقت در ستون‌های این جدول مشخص می‌شود که هر سه الگوریتم، پیکسل‌های آزمون متعلق به دو کاربری مرتع و کشاورزی دیم را در حداقل سه کلاس توزیع کرده‌اند. اگرچه

$$S = \frac{(n_{01} - n_{10})^2}{n_{01} + n_{10}} \quad (5)$$

از آنجا که توزیع کای اسکور پیوسته است، در حالی که آماره S گسسته است، رابطه فوق به صورت زیر اصلاح می‌شود:

$$S = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}} \quad (6)$$

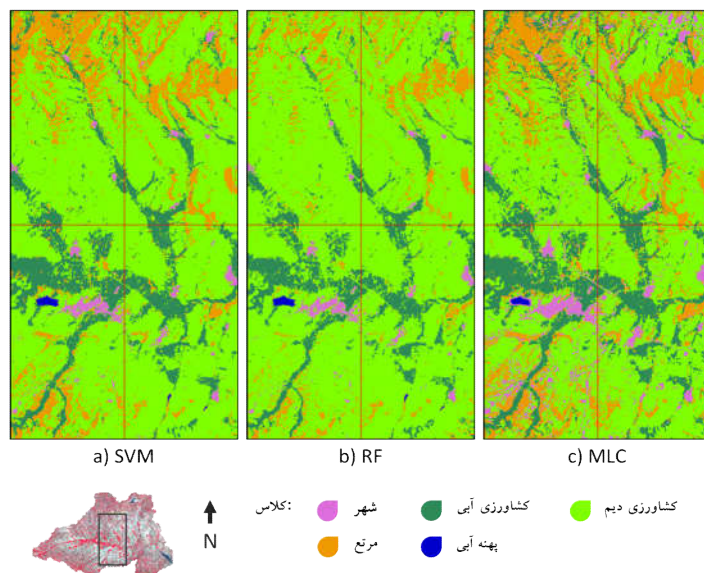
چنانچه فرض صفر (برابری کارایی دو الگوریتم) صحیح باشد، احتمال بزرگ‌تر بودن S از مقدار $\chi^2_{0.95} = 3/841$ کمتر از ۰/۰۵ (p-value) است. بنابراین چنانچه مقدار آماره S بزرگ‌تر از ۳/۸۴۱ باشد، می‌توان فرض صفر را به نفع فرض تفاوت معنادار کارایی دو الگوریتم که با مجموعه داده‌های R آموزش دیده‌اند، در سطح پنج درصد رد کرد (۱۲ و ۱۴).

در این مطالعه شاخص‌های صحت کل، کاپا، صحت تولید کننده و کاربر با استفاده از نرم‌افزار Envi 5.3 و آماره مک نماز از طریق کدنویسی در محیط IDL محاسبه شد.

نتایج و بحث

اجرای الگوریتم‌ها

شکل ۳ منحنی یادگیری هنگام ساخت مدل RF را نشان می‌دهد. در این منحنی با افزایش تعداد درخت تصمیم، واحد صحت F1 برای کلاس‌ها افزایش می‌یابد تا اینکه به مقدار ثابتی می‌رسد که نشان می‌دهد انتخاب تعداد ۱۰۰ درخت تصمیم




شکل ۵. بخشی از تصویر طبقه‌بندی شده با الگوریتم (a) بردار ماشین پشتیبان، (b) جنگل تصادفی و (c) بیشترین شباهت

جدول ۱. ماتریس‌های خطای الگوریتم‌های ماشین بردار پشتیبان SVM، جنگل تصادفی RF و بیشترین شباهت MLC

SVM						
شهر	کشاورزی آبی	کشاورزی دیم	مرتع	پهنه آبی	مجموع ردیف	
۴۱	۰	۰	۰	۰	۴۱	شهر
۰	۱۳۵	۲	۶	۰	۱۴۳	کشاورزی آبی
۰	۰	۶۸۷	۱۸	۰	۷۰۹	کشاورزی دیم
۰	۰	۵	۱۳۲	۰	۱۳۷	مرتع
۰	۰	۰	۰	۱۱	۱۱	پهنه آبی
۴۵	۱۳۵	۶۹۴	۱۵۶	۱۱	۱۰۴۱	مجموع ستون
RF						
شهر	کشاورزی آبی	کشاورزی دیم	مرتع	پهنه آبی	مجموع ردیف	
۳۹	۰	۰	۰	۰	۳۹	شهر
۰	۱۳۳	۱۲	۳	۰	۱۴۸	کشاورزی آبی
۰	۲	۶۶۸	۵۹	۰	۷۳۵	کشاورزی دیم
۰	۰	۱۴	۹۴	۰	۱۰۰	مرتع
۰	۰	۰	۰	۱۱	۱۱	پهنه آبی
۴۵	۱۳۵	۶۹۴	۱۵۶	۱۱	۱۰۴۱	مجموع ستون
MLC						
شهر	کشاورزی آبی	کشاورزی دیم	مرتع	پهنه آبی	مجموع ردیف	
۴۵	۰	۹	۰	۱	۵۵	شهر
۰	۱۳۴	۴۴	۱۱	۰	۱۸۹	کشاورزی آبی
۰	۰	۶۱۲	۱	۰	۶۱۳	کشاورزی دیم
۰	۱	۲۹	۱۴۴	۰	۱۷۴	مرتع
۰	۰	۰	۰	۱۰	۱۰	پهنه آبی
۴۵	۱۳۵	۶۹۴	۱۵۶	۱۱	۱۰۴۱	مجموع ستون

جدول ۲. (رنگی در نسخه الکترونیکی) شاخص‌های ارزیابی عملکرد به تفکیک کلاس کاربری برای الگوریتم‌های مورد بررسی

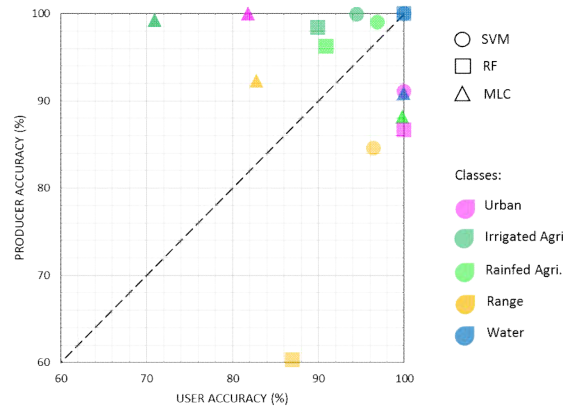
	FI (%)			UA (%)			PA (%)			
	MLC	RF	SVM	MLC	RF	SVM	MLC	RF	SVM	
شهر	۹۰	۹۲/۹	۹۵/۳	۸۱/۸	۱۰۰	۱۰۰	۱۰۰	۸۶/۷	۹۱/۱	
کشاورزی آبی	۸۲/۷	۹۴	۹۷/۱	۷۰/۹	۸۹/۹	۹۴/۴	۹۹/۳	۹۸/۵	۱۰۰	
کشاورزی دیم	۹۳/۶	۹۳/۵	۹۷/۹	۹۹/۸	۹۰/۹	۹۶/۹	۸۸/۲	۹۶/۳	۹۹	
مرتع	۸۷/۳	۷۱/۲	۹۰/۱	۸۲/۸	۸۷	۹۶/۴	۹۲/۳	۶۰/۳	۸۴/۶	
پهنه آبی	۹۵/۲	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۱۰۰	۹۰/۹	۱۰۰	۱۰۰	
										
							۹۰/۸	۹۰/۸	۹۶/۶	صحت کل (%)
							۸۹/۷	۹۰/۳	۹۶/۱	صحت FI (%)
							۰/۸۳۴	۰/۸۱۳	۰/۹۳۴	ضریب کاپا

الگوریتم‌ها به تفکیک کلاس‌های کاربری از شاخص‌هایی نظیر واحد FI که ترکیبی از PA و UA هستند، استفاده شود. بر این اساس، با توجه به واحد FI، در تمام کلاس‌ها الگوریتم SVM بهتر از دو الگوریتم دیگر بوده است. از طرفی در سه کلاس شهری، کشاورزی آبی و پهنه‌های آبی الگوریتم RF بهتر از MLC بوده است. در کلاس مرتع الگوریتم MLC نسبت به RF بهتر عمل کرده است، به ترتیب با FI حدود ۸۷ درصد و ۷۱ درصد. عملکرد این دو الگوریتم در مورد کلاس مرتع با واحد FI برابر ۹۳ درصد، یکسان بوده است.

با توجه به روش محاسبه دو شاخص PA و UA، چنانچه مقدار PA از مقدار UA در یک کلاس بیشتر باشد، می‌توان گفت که حداقل در نمونه‌های مورد بررسی، الگوریتم در آن کلاس دچار بیش تخمینی شده است. به عبارت بهتر، الگوریتم پیکسل‌های بیشتر از واقعیت را به آن کلاس اختصاص داده است. هنگام کمتر بودن مقدار PA از UA نیز عکس این نکته صادق است و الگوریتم در آن کلاس دچار کم تخمینی شده است. برای مثال در تفسیر بصری شکل ۵ مشخص شد که الگوریتم بیشترین شباهت از پیش‌بینی دو پهنه آبی موجود ناتوان بوده است، با دقت در جدول ۲ در کلاس پهنه آبی نیز مشخص می‌شود که مقدار PA کمتر از UA (به ترتیب ۹۰/۹ و

در این مورد الگوریتم SVM به نسبت دو الگوریتم دیگر بهتر عمل کرده است به گونه‌ای که در این الگوریتم و در کلاس‌های کشاورزی دیم و مرتع به ترتیب ۱ و ۱۵ درصد از پیکسل‌ها از بودن در کلاس اصلی خود محروم گشته‌اند (خطای محرومیت Omission)؛ این در حالی است که این مقادیر خطا در الگوریتم جنگل تصادفی حدود ۴ و ۴۰ درصد و در الگوریتم بیشترین شباهت در حدود ۱۲ و ۸ درصد به ترتیب برای کلاس‌های کشاورزی دیم و مرتع بوده است. این مورد را می‌توان به همپوشانی بالای طیفی میان کلاس‌های مرتع و کشاورزی دیم نسبت داد. یونس‌زاده جلیلی و همکاران (۹) نیز همپوشانی بالای این دو کلاس و در نتیجه افزایش خطای الگوریتم را گزارش کردند.

جدول ۲ شاخص‌های ارزیابی عملکرد الگوریتم‌های مورد مطالعه به تفکیک کلاس کاربری اراضی را نشان می‌دهد. در این جدول به منظور سهولت در مقایسه اعداد از طیف رنگی، رنگ قرمز برای کوچک‌ترین عدد موجود در جدول (۶۰/۳) و رنگ سبز برای بیشترین اعداد (۱۰۰)، نیز استفاده شده است. در این جدول مقادیر صحت تولید کننده PA، صحت کاربر UA و واحد FI برای پنج کلاس کاربری اراضی ارائه شده است. لاباتوت و چریفی (۱۷) پیشنهاد کردند، برای مقایسه عملکرد



شکل ۶. مقایسه مقادیر PA و UA مربوط به الگوریتم‌های مورد بررسی در کلاس‌های کاربری

جدول ۳. نتایج آزمون مک نمار برای معناداری اختلاف بین عملکرد الگوریتم‌های طبقه‌بندی مورد بررسی

p-value	S	n10	n01	
۰/۰۰۰	۵۹/۰۱۶	۶۱	۰	SVM
				RF
۰/۰۰۰	۳۸/۷۰۹	۷۷	۱۶	SVM
				MLC
۰/۹۲۵	۰/۰۰۸	۵۷	۵۷	RF
				MLC

با ۹۶/۶، ۹۶/۱ درصد و ۰/۹۳۴ نسبت به دو الگوریتم دیگر بهتر عمل کرده است. دو الگوریتم RF و MLC نیز از نظر شاخص صحت کل، عملکرد یکسان داشته‌اند. در شاخص صحت F1 الگوریتم RF نسبت به روش MLC با اختلاف اندک ۰/۶ درصد عملکرد بهتر اما در ضریب کاپا با اختلاف ۲/۱ درصد نسبت به MLC عملکرد ضعیف‌تری را داشته است.

جدول ۳ نتایج آزمون مک نمار در بررسی اختلاف معنادار میان سه الگوریتم ماشین بردار پشتیبان، جنگل تصادفی و بیشترین شباهت نشان می‌دهد. با توجه به مقدار آماره S آزمون مک نمار در این جدول، نتایج نشان داد بین عملکرد الگوریتم ماشین بردار پشتیبان و جنگل تصادفی ($S = ۵۹/۰۱۶ < ۳/۸۴۱$) و همچنین بین ماشین بردار پشتیبان و بیشترین شباهت ($S = ۳۸/۷۰۹ < ۳/۸۴۱$) اختلاف معنادار آماری در سطح اطمینان ۹۵ درصد وجود دارد. اما بین عملکرد الگوریتم جنگل

۱۰۰ درصد) است و الگوریتم MLC پیکسل‌های کمتر از واقعیت را به این کلاس اختصاص داده است. بر این اساس شکل ۶ مقایسه مقادیر PA و UA در کلاس‌های مختلف را نشان می‌دهد. در این شکل در مورد کلاس‌هایی که بالای محور قطری (خط ۱:۱) قرار گرفته‌اند، مقدار PU بیشتر از AU بوده است در نتیجه الگوریتم‌ها تعداد پیکسل بیشتر از واقعیت را به این کلاس‌ها اختصاص داده‌اند. عکس این موضوع در مورد کلاس‌های زیر خط چین ۱:۱ صادق است، برای مثال با دقت در شکل ۶ می‌توان دریافت که الگوریتم جنگل تصادفی در کاربری مرتع دچار کم‌تخمینی شدید بوده است (فاصله زیاد از خط قطری در راستای عمودی) و تنها ۶۱ درصد از تعداد کل پیکسل‌های مربوط به مرتع را درست پیش‌بینی کرده است.

با توجه به جدول ۲، نتایج نشان داد که الگوریتم SVM در هر سه شاخص صحت کل، صحت F1 و ضریب کاپا (به ترتیب

آزمون مورد استفاده در این تحقیق، در تمام شاخص‌های ارزیابی عملکرد نسبت به دو الگوریتم دیگر بوده است. مقادیر صحت کل، صحت F1 و ضریب کاپا برای این الگوریتم به ترتیب برابر ۹۶/۶، ۹۶/۱ درصد و ۰/۹۳۴ به دست آمد. دو الگوریتم RF و MLC نیز از نظر شاخص صحت کل عملکرد یکسان داشته‌اند. در شاخص صحت F1 الگوریتم RF نسبت به روش MLC با اختلاف اندک ۰/۶ درصد عملکرد بهتر اما در ضریب کاپا با اختلاف ۲/۱ درصد نسبت به MLC عملکرد ضعیف‌تری را نشان داد. در نهایت آزمون مک-نمار نیز ضمن تأیید بهتر بودن کارایی الگوریتم SVM، وجود اختلاف آماری معنادار در سطح اطمینان ۹۵ درصد میان عملکرد این الگوریتم و دو الگوریتم دیگر را تأیید کرد. مطابق نتایج این آزمون، اختلاف معناداری میان عملکرد دو الگوریتم RF و MLC در استخراج نقشه کاربری اراضی منطقه مورد مطالعه وجود ندارد. در نهایت، با توجه به نتایج، استفاده از نقشه کاربری اراضی حاصل از الگوریتم SVM برای مدل‌سازی، مدیریت و پایش سیستم‌های محیطی در منطقه بالادست سد ستارخان پیشنهاد می‌شود.

تصادفی و بیشترین شباهت اختلاف معنادار در سطح اطمینان ۹۵ درصد وجود ندارد. بر این اساس، نتایج، عملکرد بهتر الگوریتم ماشین بردار پشتیبان نسبت به دو الگوریتم دیگر را با توجه به نمونه‌های آموزشی و آزمون استفاده شده در این مطالعه و در منطقه مورد مطالعه، تأیید می‌کند.

نتیجه‌گیری

پژوهش حاضر به منظور مقایسه عملکرد دو الگوریتم ناپارامتری ماشین بردار پشتیبان و جنگل تصادفی و نیز روش پارامتری بیشترین شباهت در استخراج نقشه کاربری اراضی از تصویر سنجنده OLI ماهواره لندست ۸ و در نهایت انتخاب مناسب‌ترین الگوریتم طبقه‌بندی برای منطقه بالادست سد ستارخان در آذربایجان شرقی انجام شد. به منظور ارزیابی عملکرد سه الگوریتم مورد بررسی از شاخص‌های صحت تولید کننده، صحت کاربر، صحت کل، صحت FI و ضریب کاپا استفاده شد. آزمون مک نمار نیز به منظور بررسی اختلاف معنادار میان عملکرد الگوریتم‌ها مورد استفاده قرار گرفت. نتایج نشان داد الگوریتم ماشین بردار پشتیبان با توجه به نمونه‌های آموزش و

منابع مورد استفاده

۱. ابراهیمی، پ.، اصلاح، م. و ج. سلیمی کوچی. ۱۳۹۵. بررسی کارایی زنجیره مارکوف در برآورد تغییرات کاربری اراضی و پوشش زمین با استفاده از تصاویر ماهواره ای Landsat. *علوم و مهندسی آبخیزداری* ۱۰(۳۴): ۸۵-۹۲.
۲. آرخی، ص. ۱۳۹۳. تهیه نقشه کاربری اراضی دشت عباس ایلام با استفاده از روش های شبکه عصبی مصنوعی، ماشین بردار پشتیبان و حداکثر احتمال. *نشریه مرتعداری* ۱(۲): ۳۰-۴۳.
۳. اسحاقی، م. ا. و ش. شتایی جویباری. ۱۳۹۵. تهیه نقشه خطر وقوع آتش سوزی با استفاده از الگوریتم‌های ماشین بردار پشتیبان، جنگل تصادفی و شبکه عصبی مصنوعی (مطالعه موردی: پارک ملی گلستان، شمال شرقی ایران). *پژوهش‌های علوم و فناوری چوب و جنگل* ۲۳(۴): ۱۵۴-۱۳۳.
۴. اکبری، ا.، زنگنه اسدی، م. ع. و ا. تقوی مقدم. ۱۳۹۵. پایش تغییرات اراضی با استفاده از روش‌های مختلف تئوری آموزش آماری منطقه نیشابور. *آمایش جغرافیایی فضا* ۶(۲۰): ۳۵-۵۰.
۵. طالبی، ع.، س. گودرزی و ح. ر. پورقاسمی. ۱۳۹۶. بررسی امکان تهیه نقشه خطر زمین لغزش با استفاده از الگوریتم جنگل تصادفی (محدوده مورد مطالعه: حوضه آبخیز سردارآباد، استان لرستان). *مخاطرات محیط طبیعی* ۷(۱۶): ۴۵-۶۴.
۶. فتحی‌زاد، ح. ع. صفری، م. بازگیر و غ. خسروی. ۱۳۹۵. ارزیابی و مقایسه روش‌های ماشین بردار پشتیبان با کرنل‌های خطی، چند

- جمله‌ای و پایه شعاعی با شبکه عصبی مصنوعی برای طبقه‌بندی کاربری اراضی. *تحقیقات مرتع و بیابان ایران* ۲۳(۴): ۷۴۳-۷۲۹.
۷. مختاری، ح. و ا. نجفی. ۱۳۹۴. مقایسه روش‌های طبقه‌بندی ماشین بردار پشتیبان و شبکه عصبی مصنوعی در استخراج کاربری‌های اراضی از تصاویر ماهواره ای لندست TM. *علوم آب و خاک (علوم و فنون کشاورزی و منابع طبیعی)* ۱۹(۷۲): ۴۴-۳۵.
۸. مکاری م، ب. قهرمان و س. ح. ثنائی‌نژاد. ۱۳۹۴. بهینه کردن شاخص SAVI و برآورد تبخیر- تعرق واقعی با استفاده از تصاویر ماهواره ای لندست ۸. *آبیاری و زهکشی ایران* ۳(۹): ۴۶۹-۴۵۹.
۹. یونس‌زاده جلیلی، س. م. کمالی و پ. دانش کار آراسته. ۱۳۹۵. بررسی تحلیلی کاربری اراضی (اراضی کشت آبی و پهنه‌های آبی) در حوضه آبخیز دریاچه ارومیه با استفاده از تصاویر ماهواره لندست. *علوم آب و خاک (علوم و فنون کشاورزی و منابع طبیعی)* ۲۰(۷۸): ۲۸-۱۵.
10. Basukala, A. K., C. Oldenburg, J. Schellberg, M. Sultanov and O. Dubovyk (2017). "Towards improved land use mapping of irrigated croplands: performance assessment of different image classification algorithms and approaches. *European Journal of Remote Sensing* 50(1): 187-201.
11. Breiman, L. and A. Cutler. 2017. Avalaibe online at: Random Forests. from [http://www. stat. berkeley.edu/~breiman/RandomForests/cc_home. htm](http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm).
12. Canty, M. J. 2014. Image Analysis, Classification and Change Detection in Remote Sensing: with Algorithms for ENVI/IDL and Python, CRC Press. Boca Raton, Florida.
13. Congalton, R. G. and K. Green. 2008. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, CRC Press. Boca Raton, Florida.
14. Dietterich, T. G. 1998. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation* 10(7): 1895-1923.
15. Dube, T. and O. Mutanga. 2015. Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *Photogrammetry and Remote Sensing* 101:36-46.
16. Huang, C. and K. Song. 2012. Forest-Cover Change Detection Using Support Vector Machines. Remote Sensing of Land Use and Land Cover, CRC Press: Boca Raton, Florida.
17. Labatut, V. and H. Cherifi. 2012. Accuracy measures for the comparison of classifiers. In: Proceeding of the the 5th International Conference on Information Technology. University of Jordan, Amman, Jordan: PP. 1-5.
18. Moreno, A. and J. De Larriva. 2012. Comparison between new digital image classification methods and traditional methods for land-cover mapping. Remote sensing of land use and land cover. CRC Press, Boca Raton.
19. Özdoğan, M. 2015. Image Classification Methods in Land Cover and Land Use. Remotely Sensed Data Characterization, Classification and Accuracies, CRC Press, Boca Raton.
20. Reynolds, J., K. Wesson, A. L. Desbiez, J. M. Ochoa-Quintero and P. Leimgruber. 2016. Using remote sensing and Random Forest to assess the conservation status of critical Cerrado Habitats in Mato Grosso do Sul, Brazil. *Land* 5(2): 12.
21. Sonobe, R., Y. Yamaya, H. Tani, X. Wang, N. Kobayashi and K. Mochizuki. 2017. Mapping crop cover using multi-temporal Landsat 8 OLI imagery. *International Journal of Remote Sensing* 38(15): 4348-436.
22. Thenkabail, P. S. 2015. Remotely Sensed Data Characterization, Classification, and Accuracies, CRC Press, Boca Raton.
23. Ustuner, M., F. B. Sanli and B. Dixon. 2015. Application of support vector machines for landuse classification using high-resolution RapidEye images: a sensitivity analysis. *European Journal of Remote Sensing* 48(1): 403-422.
24. Van der Linden, S., A. Rabe, M. Held, B. Jakimow, P. Leitão, A. Okujeni, M. Schwieder, S. Suess and P. Hostert. 2015. The EnMAP-Box- A Toolbox and Application Programming Interface for EnMAP Data Processing. *Remote Sensing* 7(9): 11249.
25. Waske, B., S. Van der Linden, C. Oldenburg, B. Jakimow, A. Rabe and P. Hostert. 2012. imageRF – A user-oriented implementation for remote sensing image analysis with Random Forests. *Environmental Modelling and Software* 35: 192-193.

Performance Evaluation of Three Image Classification Methods (Random Forest, Support Vector Machine and the Maximum Likelihood) in Land Use Mapping

F. Jahanbakhshi and M. R. Ekhtesasi^{1*}

(Received: November 6-2017 ; Accepted: February 24-2018)

Abstract

Land use/cover maps are the basic inputs for most of the environmental simulation models; hence, the accuracy of the maps derived from the classification of the satellite images reduces the uncertainty in modeling. The aim of this study was to assess the accuracy of the maps produced by machine learning based on classification methods (Random Forest and Support Vector Machine) and to compare them with a common classification method (Maximum Likelihood). For this purpose, the image of the OLI sensor of Landsat 8 for the study area (Sattarkhan Dam's basin in the Eastern Azerbaijan) was used after the initial corrections. Five land uses including urban, irrigated and rain-fed agriculture, range and water body were considered. For conducting the supervised classification, ground truth data were used in two sets of educational (70% of the total) and test (30%) data. Accuracy indexes were used and the McNemar test was employed to show the significant statistical difference between the performances of the methods. The results indicates that the overall accuracy of Support Vector Machine, Random Forest, and Maximum Likelihood methods was 96.6, 90.8, and 90.8 %, respectively; also the Kappa coefficient for these methods was 0.93, 0.81 and 0.83, respectively. The existence of a significant statistical difference at the 95% confidence between the performances of the Support Vector Machine algorithm and the other two algorithms was confirmed by the McNemar test.

Keywords: Machine learning; Non-parametric classifier; McNemar test; Random forest algorithm

1. Department of Watershed Management, Faculty of Natural Resources, Yazd University, Yazd, Iran.

*: Corresponding Author, Email: mr_ekhtesasi@yazd.ac.ir