

Comparison of Accuracy Between Support Vector Machine and Random Forest Classifiers for Land Use and Crop Mapping Using Multi-Temporal Sentinel-2 Images

Zeinab Ghodsi.¹, Mir Masoud Kheirkhah Zarkesh.^{2*} , Bagher Ghermezcheshmeh.³

1. M.Sc. Student, Department of RS-GIS, Faculty of Natural Resources and Environment, Science and Research Branch, Islamic Azad University, Tehran, Iran.
2. Associate Research Prof., Soil Conservation and Watershed Management Research Institute (SCWMRI) of the Agricultural Research and Education Organization (AREO), Tehran, Iran
3. Assistant Research Prof., Soil Conservation and Watershed Management Research Institute (SCWMRI) of the Agricultural Research and Education Organization (AREO), Tehran, Iran

Abstract

Land-cover/land-use maps are necessary for monitoring land changes and proper planning for managers in agriculture, natural resources and environment fields each year. The method of field data collection using GPS and land survey is time-consuming and costly. Therefore satellite images which have entire coverage and repetition of collection, low cost and real-time data, are usually used so that land-cover/land-use maps are produced. Accurate mapping using technique suitable for today is a key factor. Although in the past, conventional classification methods have been applied to images such as Landsat, using new satellite images and modern classifiers specially machine learning has been growing recently and their effectiveness in preparing land-cover/land-use maps has been very successful. Another advantage of satellite images is repetitious collection and according to that, vegetation changes through time can be used to differentiate land cover types. The Sentinel-2 satellite with the superiority of a pixel rating of 10 meters is one of the appropriate tools to discriminate land cover types. In the current study, Support Vector Machine and Random Forest classifiers on multi-temporal Sentinel-2 images were used to differentiate land use and crop types of Sanjabi plain in Ravansar and their accuracies were compared. To do so, after sampling, Principal Component Analysis was performed for four dates in crops' growing season and PC1,2,3 bands of the images were combined. The two techniques were implemented on the layerstacks of PC1,2,3 bands of the images and the training samples. Results of accuracy assessments showed that Support Vector Machine, with overall accuracy of 91.36% and Kappa coefficient of 0.8927, produces a more precise land use and crop map rather than Random Forest method.

Keywords: Support Vector Machine, Random Forest, Multi-temporal Sentinel-2 Images, Land Use, Crops.



سنجش از دور

GIS ایران



سنجش از دور و GIS ایران سال دوازدهم، شماره چهارم، زمستان ۱۳۹۹
Vol.12, No. 4, Winter 2021 Iranian Remote Sensing & GIS

۷۳-۹۲
مقاله پژوهشی

مقایسه دقت روش‌های ماشین بردار پشتیبان و جنگل تصادفی در تهیه نقشه کاربری اراضی

و محصولات زراعی، با استفاده از تصاویر چندزمانه سنتینل-۲

زینب قدسی^۱، میر مسعود خیرخواه زرکش^۲، باقر قرمزچشمه^۳

۱. دانشجوی کارشناسی ارشد گروه تخصصی سنجش از دور و سیستم اطلاعات جغرافیایی، دانشکده منابع طبیعی و محیط‌زیست، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران

۲. دانشیار پژوهشی، پژوهشکده حفاظت خاک و آبخیزداری سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران، ایران

۳. استادیار پژوهشی، پژوهشکده حفاظت خاک و آبخیزداری سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران، ایران

تاریخ پذیرش مقاله: ۱۳۹۹/۰۶/۰۲

تاریخ دریافت مقاله: ۱۳۹۸/۰۹/۰۴

چکیده

نقشه‌های پوشش/کاربری اراضی، برای پایش تغییرات عوارض و برنامه‌ریزی صحیح، هر ساله مورد نیاز مدیران حوزه کشاورزی، منابع طبیعی و زیست‌محیطی است. روش برداشت میدانی با جی‌پی‌اس (GPS) و نقشه‌برداری زمینی مستلزم صرف زمان و هزینه‌های بسیار است. بنابراین، اغلب از تصاویر ماهواره‌ای، که دارای پوشش سراسری و توالی برداشت، هزینه کم و اطلاعات به‌هنگام‌اند، استفاده می‌شود تا نقشه‌های پوشش/کاربری زمین به‌دست آید. تهیه نقشه دقیق با روش مناسب روز موضوعی کلیدی است. طی سال‌های اخیر، استفاده از تصاویر ماهواره‌ای جدید و روش‌های نوین طبقه‌بندی، به‌ویژه یادگیری ماشین، رشد فزاینده‌ای داشته و کارایی آنها در تهیه نقشه‌های پوشش/کاربری اراضی بسیار موفقیت‌آمیز بوده است. یکی دیگر از مزایای تصاویر ماهواره‌ای برداشت‌های متوالی است و براساس آن، می‌توان از تغییرات پوشش گیاهی در طول زمان، برای تفکیک نوع پوشش، استفاده کرد. ماهواره سنتینل-۲، با امتیاز پیکسل ۱۰ متری، یکی از ابزارهای مناسب برای تفکیک نوع پوشش محسوب می‌شود. در این تحقیق، برای تفکیک انواع کاربری اراضی و محصولات زراعی دشت سنجایی روانسر، از تصاویر چندزمانه سنتینل-۲ و روش‌های طبقه‌بندی ماشین بردار پشتیبان و جنگل تصادفی استفاده و دقت آنها با یکدیگر مقایسه شد. بدین‌منظور، پس از نمونه‌برداری، تحلیل مؤلفه‌های اصلی برای چهار تاریخ دوره رشد محصولات اجرا شد و باندهای PC₁، PC₂ و PC₃ تصاویر با هم ترکیب شدند. دو روش روی ترکیب باندهای PC₁، PC₂ و PC₃ تصاویر و نمونه‌های تعلیمی اعمال شدند. ارزیابی دقت‌ها نشان داد ماشین بردار پشتیبان، با صحت کلی ۹۱.۳۶٪ و ضریب کاپای ۰.۸۹۲۷، نقشه کاربری اراضی و محصولات دقیق‌تری، در قیاس با روش جنگل تصادفی، تولید می‌کند.

کلیدواژه‌ها: ماشین بردار پشتیبان، جنگل تصادفی، تصاویر چندزمانه سنتینل-۲، کاربری اراضی، محصولات زراعی.

* نویسنده مکاتبه‌کننده: تهران، کیلومتر ۱۰ جاده مخصوص کرج، خیابان شهید عاشری، خیابان شهید شفیعی، پژوهشکده حفاظت خاک و آبخیزداری سازمان تحقیقات، آموزش و ترویج کشاورزی. تلفن: ۰۹۱۲۵۰۰۵۵۷۳، ۰۲۱۴۴۹۰۱۲۱۵، ۰۲۱۴۴۹۰۱۲۱۵. دورنگار: ۰۲۱۴۴۹۰۵۷۰۹

۱- مقدمه

سطح زمین را دریافت و در قالب داده ذخیره می‌کنند و اغلب گیاهان زراعی دارای طیف بازتابی مشابهی‌اند، تفکیک آنها به راحتی امکان پذیر نیست. انواع گیاهان زراعی دوره رویشی متفاوتی دارند؛ در نتیجه، ممکن است رفتار طیفی متفاوتی داشته باشند اما تصاویر ماهواره‌ای نیز در زمان‌های متفاوتی برداشت می‌شوند؛ از این رو، می‌توان بر این نقیصه فائق آمد. در سال‌های اخیر و پس از شروع به کار ماهواره سنتینل-۲ (با دقت مکانی و زمانی بالا)، امکان استفاده از فناوری‌های طبقه‌بندی با تصاویر چندزمانه^۵ به وجود آمد. امروزه، به این روش‌ها توجه ویژه‌ای می‌شود و در این زمینه، تکنیک‌های گوناگونی به کار می‌رود. یکی از این روش‌ها استفاده از تجزیه مؤلفه‌های اصلی^۶ برای کاهش حجم داده‌هاست. در پژوهش‌های بررسی‌شده (Gómez et al., 2016; Guidici and Clark, 2017; Khatami et al., 2016; Rujoiu-Mare et al., 2017)، به منظور تفکیک عوارض زمین، به جای تصاویر تک‌زمانه، از تصاویر چندزمانه استفاده کردند و دقت طبقه‌بندی را به مراتب بهبود داده‌اند. قاسمی اسفهلان و همکاران (۱۳۹۲) و خاتمی^۷ و همکاران (۲۰۱۶) روش PCA را به کار برده‌اند.

طبقه‌بندی تصویر ماهواره‌ای رایج‌ترین روش در تهیه نقشه‌های مکانی پوشش/کاربری اراضی شمرده می‌شود (Mathur & Foody, 2008). در سال‌های اخیر، کاربرد روش‌های مدرن طبقه‌بندی روند روبه‌رشدی داشته است. در دسته‌بندی‌ای کلی، این روش‌ها را می‌توان براساس به‌کارگیری و یا به‌کارنبردن اطلاعاتی مانند داده‌های زمینی، عکس‌های هوایی و تصاویر ماهواره‌ای از سوی عامل انسانی، به دو دسته نظارت‌شده^۸ و

شناسایی پدیده‌های زمین در حوزه مدیریت توسعه شهری، منابع طبیعی و زیست‌محیطی از جایگاه ویژه‌ای برخوردار است. داشتن اطلاعات دقیق و مورد اعتماد شامل کشت محصولات و شرایط خاک برای کشاورزی پایدار و توسعه برنامه‌ها و سیاست‌های مؤثر در این زمینه ضروری است. بنابراین، تصمیم‌گیران و مقامات سازمان‌های گوناگون، به منظور برنامه‌ریزی استراتژیک و مدیریت توسعه پایدار منابع طبیعی و محیط‌زیست، به اطلاعات دقیقی در مورد کاربری و پوشش اراضی نیازمندند. تهیه نقشه‌ای صحیح و دقیق از پوشش/کاربری اراضی و استفاده از روش مناسب روز به موضوعی بسیار مهم تبدیل شده است. در گذشته، روش‌های معمول برداشت زمینی، به لحاظ زمانی و مالی و نیروی انسانی، هزینه بالایی در برداشتند. داده‌های رقومی سنجش از دور یکی از منابع اطلاعاتی مکانی است که در قیاس با سایر منابع متداول، مانند نقشه‌های کاغذی، دارای مزایای گوناگونی است؛ از جمله پوشش وسیع، کاهش هزینه و به‌هنگام و تکرارپذیر بودن اطلاعات (مهدوی و فلاح شمسی، ۱۳۹۱). اهمیت فراوان این تصاویر به لحاظ ارائه اطلاعات به‌هنگام، کم‌هزینه، رقومی، و امکان پردازش آنها در تهیه نقشه‌های پوشش/کاربری اراضی است. پیش‌تر، نقشه‌های متعددی با استفاده از تصاویر ماهواره‌ای متداول، مانند تصاویر لندست^۱ و مادیس^۲، ایجاد شده است. امروزه، تصاویر ماهواره‌ای جدید با قابلیت‌های بالا مورد توجه بسیاری از محققان گرفته است. برای طبقه‌بندی عوارض مناطق با وسعت کم، باید از تصاویر دارای قدرت تفکیک مکانی بالا استفاده شود. سنتینل-۲، که سازمان فضایی اروپا^۳ در سال ۲۰۱۵ آن را به فضا پرتاب کرد، دارای سیزده باند طیفی با اندازه پیکسل ۱۰، ۲۰ و ۶۰ متری است. رزولوشن مکانی و طیفی بالای این ماهواره از ویژگی‌هایی است که در تشخیص دقیق‌تر پدیده‌های روی زمین، کمک به‌سزایی می‌کند. از آنجاکه تصاویر ماهواره‌ای غیرفعال بازتاب

1. Landsat
2. MODIS
3. Sentinel-2
4. European Space Agency (ESA)
5. Multi-temporal images
6. Principal Component Analysis (PCA)
7. Khatami
8. Supervised classification

طبقه‌بندی جنگل تصادفی نیز از تکنیک‌های آموزش مجموعه‌ای از درختان تصمیم‌گیری با استفاده از داده‌های تعلیمی است. پس از ایجاد درختان تصمیم و اجرای الگوریتم، نتیجه پیش‌بینی هر درخت جداگانه مشخص می‌شود (Breiman, 2001). خروجی طبقه‌بندی، دسته‌هایی است که بیشترین نتیجه پیش‌بینی درختان را دربر دارد (Guidici & Clark, 2017). کارایی جنگل تصادفی در مورد مجموعه‌های بسیار بزرگ به اثبات رسیده است؛ از این رو، می‌توان آن را در تجزیه و تحلیل داده‌های ماهواره‌ای به کار برد (Basukala et al., 2017). این روش حجم داده‌ها را بسیار کاهش می‌دهد (Gómez et al., 2016) و سرعت پردازش بالایی در طبقه‌بندی داده‌های حجیم منطقه دارد. ساختار ساده و قابل درک الگوریتم جنگل تصادفی باعث انعطاف‌پذیری آن می‌شود و به آسانی می‌توان آن را با روش‌های دیگر ترکیب کرد. در مناطقی که بر اثر افزایش نمونه‌های آموزشی، داده‌های نویزدار پدیدار می‌شوند نیز، سرعت و قابلیت بالایی دارد (شفیعی و همکاران، ۱۳۹۶).

دشت سنجایی در شهرستان روانسر، با بارش چشمگیر و مزارع گوناگون از جمله گندم، جو، نخود و ذرت، ناحیه‌ای شاخص در سطح استان کرمانشاه محسوب می‌شود (صحبتلو، ۱۳۹۶). تهیه نقشه تفکیک اراضی با دقیق‌ترین روش، تصویر ماهواره‌ای مطلوب و تکرار تهیه آن در سری‌های زمانی معین برای پایش تغییرات و برنامه‌ریزی و مدیریت صحیح منطقه ضروری است.

نظارت‌نشده^۱ تقسیم کرد. روش‌های طبقه‌بندی نظارت‌شده به دو گروه پارامتریک و ناپارامتریک تقسیم می‌شوند. در روش‌های طبقه‌بندی پارامتریک، مانند حداکثر احتمال^۲ و حداقل فاصله تا میانگین^۳، از پارامترهای آماری نظیر میانگین و ماتریس واریانس - کوواریانس داده‌های آموزشی^۴ استفاده می‌شود. مشکل اصلی این روش‌ها وابستگی آنها به توزیع نرمال^۵ داده‌هاست (Mather & Tso, 2009). به همین دلیل، روش‌های ناپارامتریک از جمله شبکه‌های عصبی^۶، درخت تصمیم‌گیری^۷، جنگل تصادفی^۸ و ماشین بردار پشتیبان^۹، برای طبقه‌بندی تصاویر ماهواره‌ای به کار می‌روند. ماشین بردار پشتیبان و جنگل تصادفی، به‌منزله مدل‌های مدرن یادگیری ماشین^{۱۰}، در سنجش از دور استفاده گسترده‌ای دارند (Clark & Kilham, 2016). این روش‌ها در بسیاری از پلت‌فرم‌ها و پایگاه‌های داده، نتایج قوی‌تر از دیگر روش‌های طبقه‌بندی دارند (Fernandez-Delgado et al., 2014). الگوریتم ماشین بردار پشتیبان، با استفاده از نمونه‌های آموزشی و توانایی تعمیم آنها، ابرصفحه^{۱۱} بهینه‌ای را تعیین و با آن، دو طبقه را به بهترین نحو جدا می‌کند (میرزایی‌زاده و همکاران، ۱۳۹۴). این روش برای طبقه‌بندی مناطق پیچیده با تنوع پدیده‌ها مناسب است (Chuvieco, 2016) و فضای ویژگی^{۱۲} وسیع، مانند تصاویر چندزمانه و چندطیفی^{۱۳}، را می‌تواند به راحتی مدیریت کند اما، برای فضای ویژگی کوچک، کارایی اندکی دارد. به عبارتی، دسته‌ها باید، در باندهای طیفی تصویر، تنوع داشته باشند تا دقت بالا در طبقه‌بندی حاصل شود. این روش به نقاط آموزشی زیادی نیاز ندارد و به‌ازای نقاط آموزشی اندک نیز، منطقه را با دقت بالایی طبقه‌بندی می‌کند (Gómez et al., 2016). از قابلیت‌های ماشین بردار پشتیبان، توانایی غلبه بر مشکل توزیع غیرخطی نمونه‌هاست. در این حالت، با استفاده از توابع کرنل^{۱۴}، داده‌ها به فضایی با بعد بالاتر انتقال می‌یابند تا، در آن، تفکیک‌پذیری بهتر صورت پذیرد؛ ابرصفحه جداکننده نیز در آن فضا تعیین می‌شود (جهانبخشی و اختصاصی، ۱۳۹۷).

1. Unsupervised classification
2. Maximum Likelihood (ML)
3. Minimum Distance to Mean (MDM)
4. Training data
5. Normal distribution
6. Neural Networks (NN)
7. Decision Tree (DT)
8. Random Forest (RF)
9. Support Vector Machine (SVM)
10. Machine learning
11. Hyperplane
12. Feature space
13. Multispectral
14. Kernel

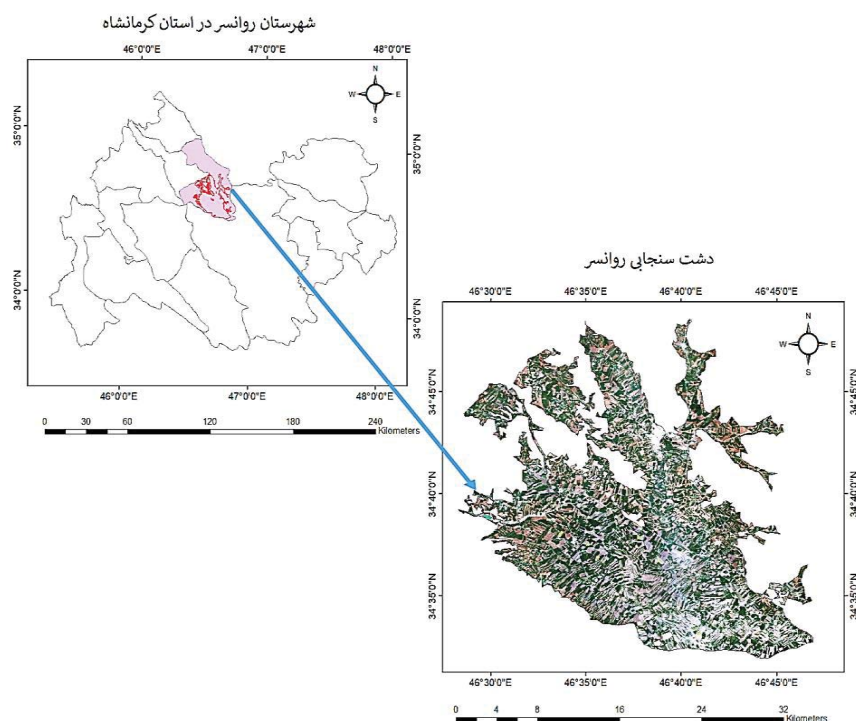
کرمانشاه، با مساحت ۳۷۹ کیلومترمربع، بین طول‌های جغرافیایی ۴۶ درجه و ۲۰ دقیقه تا ۴۶ درجه و ۵۰ دقیقه شرقی و عرض‌های جغرافیایی ۳۴ درجه و ۲۰ دقیقه تا ۳۴ درجه و ۵۳ دقیقه شمالی واقع شده است و از مرزهای شرقی به خط‌الرأس کوه خورین و کوه چال‌آباد، از جنوب به محدوده دشت ماهی‌دشت از طریق بلندی‌های نامحسوس، از شمال به ارتفاعات کوه بازن و از غرب به ارتفاعات کوه نواوله، کوه سنگ سیاه محدود می‌شود. آب‌وهوای آن سرد و نیمه‌مرطوب است و بارش سالیانه چشمگیری دارد. این منطقه، به‌دلیل داشتن زمین‌های کشاورزی مستعد و آب فراوان، در زمینه تولید محصولات زراعی مانند گندم، جو، نخود و به‌خصوص، ذرت دانه‌ای در سطح استان مطرح است. از دیرباز مهد کشاورزی بوده است و یکی از قطب‌های مهم زراعت و باغبانی در استان کرمانشاه محسوب می‌شود (صحبتلو، ۱۳۹۶). موقعیت جغرافیایی دشت سنجابی روانسر، در جایگاه منطقه مورد پژوهش، در شکل ۱ نشان داده شده است.

منابع مورد بررسی اغلب از تصاویر به‌صورت تک‌زمانه در طبقه‌بندی استفاده کرده‌اند. کمتر از چند تصویر با زمان‌های متفاوت و یا یک روش را برای تصاویر چندزمانه، به‌کار برده‌اند. در این تحقیق، به‌منظور تفکیک کاربری اراضی و مزارع گوناگون دشت سنجابی روانسر، از چهار تصویر سنتینل-۲ با تاریخ‌های متفاوت استفاده شد. این تصاویر مبتنی بر دوره ریشی محصولات بودند؛ به‌طوری‌که در آنها حداکثر اختلاف رفتار طیفی بین محصولات لحاظ شده بود. علاوه بر آن، برای کاهش تعداد باندها در هر تاریخ و افزایش میزان تفکیک‌پذیری طیفی، از روش PCA استفاده شد که در منابع مورد بررسی، کمتر به آن پرداخته شده است. همچنین، دو روش نوین طبقه‌بندی مقایسه شدند که در مجموع، از نوآوری‌های این پژوهش محسوب می‌شود.

۲- مواد و روش‌ها

۲-۱- منطقه پژوهش

دشت سنجابی در شهرستان روانسر، از توابع استان



شکل ۱. موقعیت جغرافیایی منطقه پژوهش

۲-۲- تصاویر ماهواره‌ای سنتینل-۲

تصاویر سنتینل-۲ دارای سیزده باند در محدوده طیف الکترومغناطیسی و قدرت تفکیک مکانی ۱۰، ۲۰ و ۶۰ متر هستند که از سال ۲۰۱۵، سازمان فضایی اروپا آنها را به‌منزله بخشی از برنامه کوپرنیکوس^۱ توسعه داده است. پیکسل‌های ۱۰ متری از امتیاز ویژه این ماهواره در شناسایی دقیق پدیده‌های زمین به‌شمار می‌رود. قدرت تفکیک رادیومتریک آن ۱۲ بیت و دوره بازگشت آن، به‌ازای سنتینل-2A، ده روز و به‌ازای سنتینل-2B، پنج روز است. در جدول ۱، مشخصات ماهواره سنتینل-۲ آمده است (ESA, 2015).

۲-۳- ماشین بردار پشتیبان

روش طبقه‌بندی ماشین بردار پشتیبان بر نظریه یادگیری آماری استوار است (Vapnik, 1995).

فاکتورهای ماشین بردار پشتیبان، عبارت‌اند از: ابرصفحه (یک زیرفضای 1-p بعدی در فضای p بعدی)، حاشیه (کمترین فاصله یکی از نقاط آموزشی تا ابرصفحه)، بردار پشتیبان (بردار شامل نزدیک‌ترین نقاط آموزشی به ابرصفحه‌های تفکیک‌کننده) و ابرصفحه با بیشترین حاشیه (ابرفصفحه با پهن‌ترین نوار که می‌تواند بین داده‌های دو کلاس قرار بگیرد). ایده اصلی آن یافتن ابرصفحه‌ای بهینه، به‌منزله سطح تصمیم‌گیری، به‌گونه‌ای است که حاشیه بین دو کلاس را به حداکثر برساند. در مسائلی که داده‌ها به‌صورت خطی جداپذیر نباشند، با یک کرنل غیرخطی، به فضای دارای ابعاد بیشتر نگاشت پیدا می‌کنند تا بتوان، در این فضای جدید، آنها را به‌صورت خطی جدا و ابرصفحه بهینه را تعیین کرد.

جدول ۱. مشخصات ماهواره سنتینل-۲

شماره باند	نام اصلی باند	نام فارسی باند	میانگین طول موج (میکرومتر)	پهنای باند (میکرومتر)	توان تفکیک مکانی (متر)
B 01	Coastal aerosol	گردوغبار ساحلی	۰.۴۴۳	۰.۰۲	۶۰
B 02	Blue	آبی	۰.۴۹۰	۰.۰۶۵	۱۰
B 03	Green	سبز	۰.۵۶۰	۰.۰۳۵	۱۰
B 04	Red	قرمز	۰.۶۶۵	۰.۰۳	۱۰
B 05	Vegetation Red Edge	لبه قرمز پوشش گیاهی	۰.۷۰۵	۰.۰۱۵	۲۰
B 06	Vegetation Red Edge	لبه قرمز پوشش گیاهی	۰.۷۴۰	۰.۰۱۵	۲۰
B 07	Vegetation Red Edge	لبه قرمز پوشش گیاهی	۰.۷۸۳	۰.۰۲	۲۰
B 08	NIR	فروسرخ نزدیک	۰.۸۴۲	۰.۱۱۵	۱۰
B 08A	Narrow NIR	فروسرخ نزدیک باریک	۰.۸۶۵	۰.۰۲	۲۰
B 09	Water Vapour	بخار آب	۰.۹۴۵	۰.۰۲	۶۰
B 10	SWIR – Cirrus	فروسرخ طول موج کوتاه-سیروس	۱.۳۷۵	۰.۰۳	۶۰
B 11	SWIR	فروسرخ طول موج کوتاه	۱.۶۱۰	۰.۰۹	۲۰
B 12	SWIR	فروسرخ طول موج کوتاه	۲.۱۹۰	۰.۱۸	۲۰

1. Copernicus

پارامتر C پارامتر تنظیم در ماشین‌های بردار پشتیبان است.

برای در نظر داشتن نویز موجود در داده و تداخل بین داده‌های آموزشی، از متغیر $\xi_j > 0$ استفاده می‌شود. وجود شرط رابطه (۲) ضمانت می‌دهد که داده‌ای در حاشیه قرار نمی‌گیرد. سطح تصمیم‌گیری بهینه با حل رابطه (۲)، بر مبنای روش لاگرانژ، طبق رابطه (۳) محاسبه می‌شود.

رابطه (۳) $f(x) = \sum_{x_i \in SV} y_i \alpha_i \varphi(x_i) \cdot \varphi(x) + b$
 α_i ضرایب لاگرانژ اند که در مرحله بهینه‌سازی^۱ محاسبه می‌شود. SV بردارهای پشتیبان‌اند و ضریب لاگرانژ متناظرشان بزرگ‌تر از صفر است.

این داده‌های آموزشی نزدیک‌ترین نمونه‌ها به ابرصفحه‌اند. همان‌طور که رابطه (۳) نشان می‌دهد، فقط بردارهای پشتیبان در مرحله تعلیمی شرکت می‌کنند؛ از این‌رو، به نمونه‌های آموزشی زیادی نیاز ندارند. در رابطه (۳)، ضرب داخلی بین دو کرنل نگاشت‌شده را می‌توان با کرنل آن دو نمونه محاسبه کرد.

روش کار ماشین بردار پشتیبان بدین ترتیب است: فرض می‌شود I داده‌های آموزشی موجود باشد که هر یک با (x_i, y_i) نشان داده می‌شود. x_i بردار ویژگی n بعدی و $y_i \in \{-1, 1\}$ برچسب آن است. هدف یافتن ابرصفحه‌ای است که دو کلاس با برچسب 1 و -1 را با بیشترین حاشیه، از هم جدا کند. فرمول این ابرصفحه در رابطه (۱) آمده است.

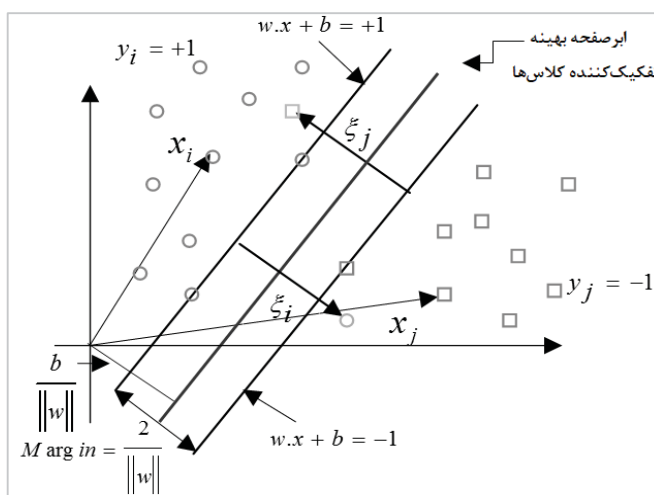
$$w^T \varphi(x) + b = 0 \quad (1)$$

w بردار وزن عمود بر ابرصفحه، b بردار بایاس برای اندازه‌گیری فاصله ابرصفحه تا مبدأ و $\varphi(x)$ کرنل برای انتقال داده به فضای با ابعاد بالاتر است. شکل ۲ طبقه‌بندی داده‌هایی را نشان می‌دهد که به‌طور خطی، با استفاده از ماشین‌های بردار پشتیبان تفکیک ناپذیرند (صمدزادگان و حسنی، ۱۳۹۱).

بیشینه‌کردن حاشیه بین دو کلاس معادل کمینه‌کردن اندازه w است که منجر به حل مسائل کمینه‌سازی رابطه (۲) می‌شود.

$$\text{رابطه (۲)} \quad \text{Min} \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^k \xi_i \right)$$

$$\text{Subject to: } y_i (w\varphi(x_i) + b) > 1 - \xi_i \quad i=1, \dots, k$$



شکل ۲. طبقه‌بندی داده‌هایی که، به‌طور خطی، با استفاده از ماشین‌های بردار پشتیبان تفکیک‌ناپذیرند

1. Optimization

ویژگی‌های استفاده‌شده در جداسازی‌ها برابر است با $\sqrt{N_f}$ (Guidici and Clark, 2017). این روش، برخلاف روش‌های معمول رگرسیون، شهودی و مبتنی بر درک مستقیم کاربر نیست (Jeong et al., 2016).

در این پژوهش، بر مبنای راهنمای OTB Team (۲۰۱۸)، در نرم‌افزار OTB روی ترکیب باندی PC_1 ، PC_2 و PC_3 تصاویر و داده‌های آموزشی، طبقه‌بندی جنگل تصادفی اجرا شد. پارامترهای حداکثر پنج سطح در عمق درختان، حداقل ده نمونه در سطح گره و حداکثر یکصد درخت در جنگل، به‌طور پیش‌فرض، تنظیم شدند.

۳- روش تحقیق

این تحقیق طی سه گام انجام شد. در گام نخست، جمع‌آوری داده و پیش‌پردازش^{۱۲} تصاویر ماهواره‌ای و نمونه‌برداری؛ در گام دوم، پردازش تصاویر و طبقه‌بندی آنها؛ و در گام سوم، پردازش نهایی^{۱۳} و تولید نقشه کاربری اراضی و محصولات زراعی منطقه صورت گرفت. داده‌ها در قالب داده‌های ماهواره‌ای و برداری، شامل پلی‌گون مرز منطقه، پلی‌گون مزارع گندم و نقاط برداشت زمینی، که مرجع نمونه‌برداری به‌شمار می‌روند، گرد آمدند. در شکل ۳، مراحل انجام‌شدن پژوهش آورده شده است.

۱-۳- پیش‌پردازش تصاویر

ابتدا دوازده تصویر سنتینل-2A از منطقه مورد مطالعه دریافت شد. این تصاویر در بازه‌های متفاوت، از آذر سال ۱۳۹۴ تا شهریور سال ۱۳۹۵، با پوشش ابر متوسط کمتر

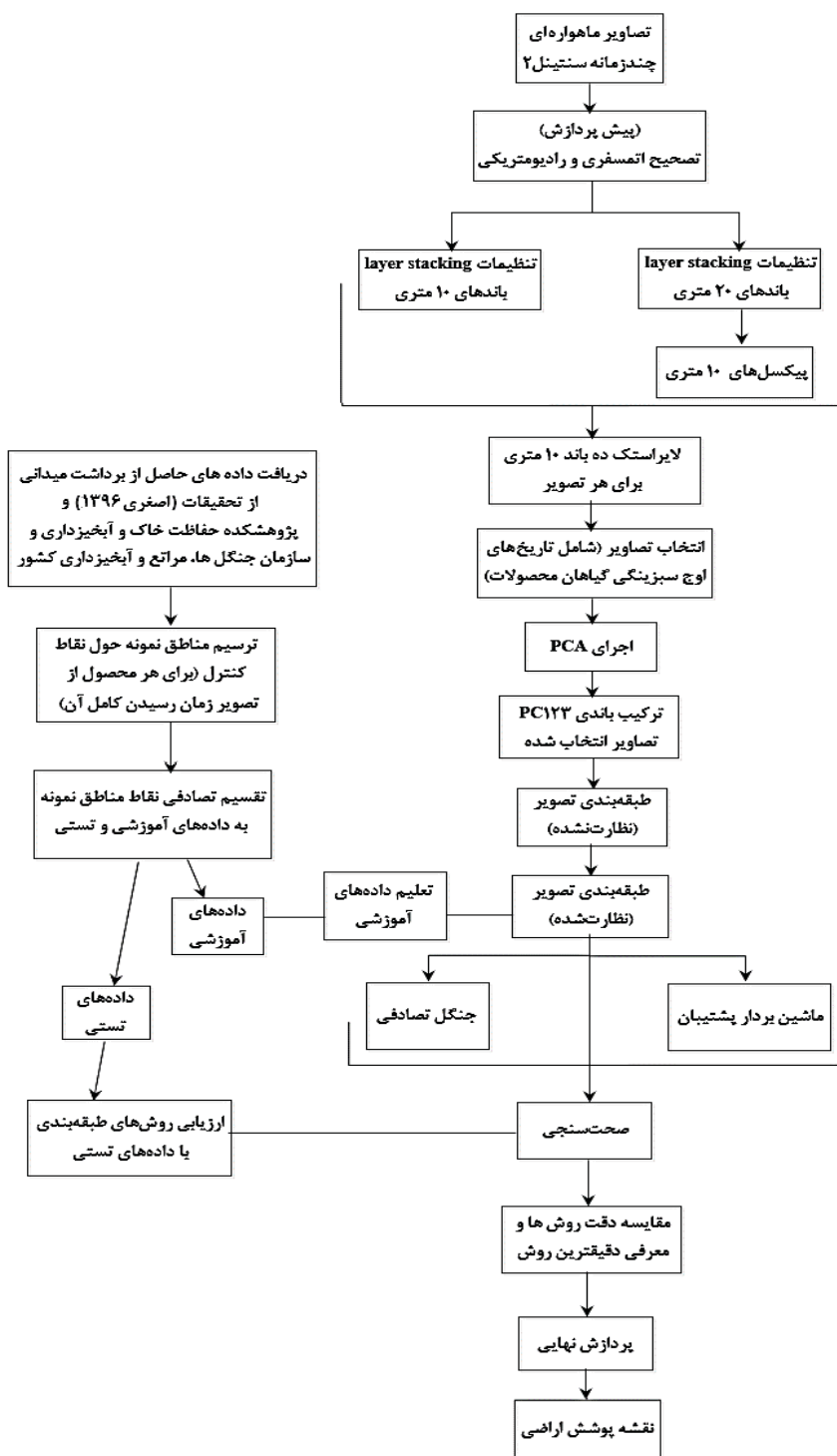
1. Linear
2. Gaussian Radial Basis Function
3. polynomial
4. Sigmoid
5. The ORFEO ToolBox
6. Ensemble
7. Aggregating
8. Bootstrap
9. Bagging
10. Majority vote
11. Out-of-Bag (OOB)
12. Pre-processing
13. Post-processing

کرنل‌های خطی^۱، تابع پایه شعاعی گوسین^۲، چندجمله‌ای^۳ و پیچشی^۴ از پرکاربردترین کرنل‌ها به حساب می‌آیند.

در این تحقیق، ابتدا، ترکیب باندی PC_1 ، PC_2 و PC_3 تصاویر و داده‌های آموزشی، که ۶۰٪ کل داده‌ها را پوشش می‌دهند، فراخوانی شدند. سپس، بر اساس راهنمای OTB Team (۲۰۱۸)، در نرم‌افزار متن‌باز^۵ OTB مدل SVM طراحی شد تا نمونه‌ها بر اساس آن تعلیم ببینند. در طبقه‌بندی ماشین بردار پشتیبان، کرنل سائز چندجمله‌ای و مدل C-SVM با پارامتر C برابر با ۱ و پارامتر Nu مساوی ۰.۵ به کار برده شد.

۴-۲- جنگل تصادفی

جنگل تصادفی مجموعه‌ای بزرگ^۶ از درختان تصمیم‌گیری است که طبق اصول یادگیری ماشین عمل می‌کند (Immitzer et al., 2016). جنگل تصادفی از الگوریتم‌های تجمیعی^۷ خودراه‌انداز^۸، مانند روش انباشتن^۹، استفاده می‌کند تا زیرمجموعه‌های متفاوتی از نمونه‌های آموزشی ایجاد کند و به مجموعه متنوعی از درختان برسد. درختان به دست آمده را می‌توان با فرمت مجموعه‌ای از دستورهای if-then نمایش داد تا بررسی آنها راحت‌تر شود (Mitchell, 2017). هر درخت، به‌ازای داده‌هایی که انتخاب نشده‌اند، یک خروجی طبقه‌بندی ارائه می‌دهد. خروجی نهایی جنگل تصادفی آن طبقه‌بندی‌ای است که رأی اکثریت^{۱۰} خروجی تعداد بی‌شماری از درختان مجزا را به خود اختصاص می‌دهد (Breiman, 2001). آن دسته از داده‌ها، که در مرحله تعلیمی روش خودراه‌انداز استفاده نمی‌شوند، «خارج از کیسه»^{۱۱} نام دارند. با استفاده از داده‌های خارج از کیسه، می‌توان برآوردی از نرخ خطای طبقه‌بندی داشت (Immitzer et al., 2016). پارامترهای جنگل تصادفی عبارت‌اند از: تعداد درختانی که در اجرا رشد می‌کنند، حداکثر تعداد ویژگی‌های مورد استفاده در هر گره تصمیم هر درخت، و حداقل تعداد نمونه‌ها در هر برگ. اگر N_f تعداد کل ویژگی‌ها باشد، تعداد



شکل ۳. مراحل پژوهش

برداشت زمینی صورت گرفت. تشخیص پهنه‌های آبی با استفاده از شاخص طیفی MNDWI^۵ انجام شد (Ali et al., 2019; Gulácsi & Kovács, 2020). برای نمونه‌برداری از دیگر کلاس‌ها مانند اراضی مسکونی، جاده‌ها و زمین‌های آیش و بایر، از تصاویر گوگل ارث^۶ کمک گرفته شد. جدول ۲ رفتار محصولات عمده را در بازه‌های گوناگون سال زراعی ۹۵-۱۳۹۴، نشان می‌دهد. برای نمونه‌برداری از هر محصول، تصویری ماهواره‌ای در نظر گرفته شد که تاریخ آن با اوج دوره رشد محصول در جدول ۲ مطابقت داشته باشد. در جدول ۳، تاریخ تصاویر ماهواره‌ای انتخاب‌شده برای نمونه‌برداری از محصولات آمده است.

جدول ۲. رفتار محصولات عمده منطقه، در بازه‌های گوناگون سال زراعی ۹۵-۱۳۹۴

زمان	محصول	وضعیت محصول
اوایل	گندم	رشد خوشه‌ها- اوج سبزیگی گیاه
اردیبهشت	گوجه‌فرنگی	شروع به کاشت بذر
	جو	گلدهی و آغاز تشکیل میوه
	گندم	شروع زردشدن
اواخر	ذرت	اواخر رشد و در حال زردشدن
خرداد	گوجه‌فرنگی	اوایل رشد و در حال سبزشدن
	جو	نمو میوه- رسیدن کامل محصول برداشت‌شده
	گندم	شروع به برداشت
اواسط تیر	ذرت	پیدایش گل‌آذین نر و کاکل
	گوجه‌فرنگی	تولید محصول و آغاز برداشت
	جو	کامل برداشت‌شده
	گندم	کامل برداشت‌شده
اواخر	ذرت	رسیدن کامل- اوج سبزیگی گیاه
مرداد	گوجه‌فرنگی	برداشت‌شده
	جو	کامل برداشت‌شده

1. United States Geological Survey (USGS)
2. Sentinel-2 Atmospheric and Topographic Correction
3. The Sentinel Application Platform
4. Resampling
5. Modified Normalized Difference Water Index
6. Google Earth

از ۲۰٪ و شماره کاشی T38SPD، از سایت سازمان فضایی اروپا و سازمان زمین‌شناسی آمریکا^۱ دریافت شدند. تاریخ تصاویر بر تاریخ‌های برداشت زمینی گندم و ذرت منطقه، که از تیم تحقیقاتی اصغری (۱۳۹۶) به‌دست آمد، انطباق داشت تا کلاس اراضی، با تکیه بر انواع محصولات، استخراج شود. تصاویر سنتینل ۲- دانلودشده، از ابتدا، زمین‌مرجع‌اند. پستی‌وبلندی زیادی در منطقه وجود نداشت؛ بنابراین، تصحیح هندسی لازم نبود. صرفاً برای رفع آثار جوّی، با استفاده از افزونه Sen2Cor^۲ نرم‌افزار SNAP^۳، تصحیح اتمسفری انجام شد. بدین‌صورت، محصول Level1C ماهواره و بازتاب بالای اتمسفر به محصول Level2A و به‌عبارتی، بازتاب پایین اتمسفر تبدیل شد. تصاویر از روی پلی‌گون مرز ناحیه، که از سازمان جنگل‌ها، مراتع و آبخیزداری کشور اخذ شد، برش پیدا کردند. اندازه پیکسل‌های ۲۰ متری، با عملیات نمونه‌برداری مجدد^۴، به ۱۰ متر تقلیل یافت و به‌ازای هر تصویر، ده باند ۱۰ متری با هم ادغام شدند. با اعمال رابطه (۴) با زبان برنامه‌نویسی IDL روی لایراستک‌های ایجادشده، مقیاس باندها تغییر یافت و ارزش رقومی هر باند نرمال و مقداری بین ۰ و ۱ شد. رابطه (۴)
$$b1 \leq 10000) * (\text{float}(b1) / 10000.0) + (b1 > 10000) * 1$$
 ارزش رقومی باندهای تصاویر است. بدین‌ترتیب، بازتاب واقعی زمین به‌دست آمد. در تصاویر دارای پوشش ابری، با ایجاد و اعمال ماسک ابر، اثر آن از پردازش‌ها حذف شد.

۳-۲- نمونه‌برداری

طبق آمار کارشناسان وزارت جهاد کشاورزی استان کرمانشاه، برای تفکیک زمین‌های زراعی منطقه، کلاس‌های گندم، ذرت، گوجه‌فرنگی و جو و برای طبقه‌بندی انواع دیگر کاربری اراضی منطقه، نواحی شهری شامل مسکونی و جاده، آب، سایر پوشش‌های گیاهی و زمین‌های آیش و بایر در نظر گرفته شدند. نمونه‌برداری از محصولات زراعی با مراجعه به منطقه و

جدول ۳. تاریخ تصاویر ماهواره‌ای انتخاب شده برای نمونه برداری

محصول	از محصولات منطقه	
	تاریخ تصویر جهت نمونه برداری از محصول	زمان رسیدن کامل محصول
گندم	۹۵/۰۲/۰۵	اوایل اردیبهشت
ذرت	۹۵/۰۵/۲۲	اواخر مرداد
گوجه‌فرنگی	۹۵/۰۳/۲۴	اواخر خرداد
جو	۹۵/۰۲/۰۵	اوایل اردیبهشت- مشابه گندم

جدول ۴. مشخصات مناطق نمونه

کد دسته	نام دسته	تعداد پیکسل‌ها	مساحت (مترمربع)
۱	گندم	۱۰۳۲	۱۰۳,۲۰۰
۲	ذرت	۳۶۰	۳۶,۰۰۰
۳	گوجه‌فرنگی	۹۰	۹,۰۰۰
۴	جو	۱۷۴	۱۷,۴۰۰
۵	مسکونی و جاده	۳۰۰	۳۰,۰۰۰
۶	آب	۲۲۳	۲۲,۳۰۰
۷	سایر- پوشش گیاهی	۵۳۵	۵۳,۵۰۰
۸	سایر- آیش و بایر	۵۰۱	۵۰,۱۰۰

ویژگی نبود اشتراک زمانی در تقویم زراعی محصولات ذکر شده در جدول ۲، به جز گندم و جو که رفتار طیفی مشابهی دارند، به کسب اطمینان از صحت انتخاب نمونه‌ها کمک کرد. وقتی محصول مورد نظر در دوره رسیدن کامل باشد، سایر مزارع پوشش سبز کمتری دارند، در حال شخم خوردن اند، کاشته نشده‌اند و یا در حال برداشت‌اند. هنگام تخصیص پیکسل تصویر به دسته یک محصول خاص، به تقویم زراعی آن در جدول ۲ مراجعه شد و دقت شد بازتاب پیکسل در هر تصویر، به صورت جداگانه، با رفتار طیفی استاندارد محصول در تاریخ مورد نظر، متناسب باشد. در مرحله بعد، مناطق نمونه با استفاده از پارامتر آماری انحراف معیار، هیستوگرام و اسکترپلات زوج باندها و نمودار میانگین انعکاس طیفی داده‌ها ارزیابی شد تا از همگن بودن و توزیع نرمال نمونه‌ها اطمینان به دست آید و میزان تفکیک پذیری آنها بررسی شود. هریک از مناطق نمونه محصولات گندم و جو نیز، از پیکسل‌های همسایه نقاط کنترل مخصوص به خود، انتخاب شدند. بدین ترتیب، پلی‌گون‌های نمونه، متناسب با وسعت کلاس‌ها در منطقه، ترسیم شدند که مشخصات آنها در جدول ۴ آمده است.

۳-۳- انتخاب ترکیب باندی مناسب طبقه‌بندی

از بین تصاویر ماهواره‌ای دانلود شده، تصاویری برای طبقه‌بندی انتخاب شدند که زمان رسیدن کامل محصولات را دربر داشته باشند؛ به گونه‌ای که با تفسیر چشمی هر تصویر، بیشترین سبزیگی فقط متعلق به گیاه مربوط به یک محصول خاص باشد. بنابراین، با مراجعه به رفتار فنولوژیکی محصولات در جدول ۲، تصاویر اوایل اردیبهشت، اواخر خرداد، اواسط تیر و اواخر مرداد ۱۳۹۵ در تاریخ‌های برداشت ۹۵/۰۲/۰۵، ۹۵/۰۳/۲۴، ۹۵/۰۴/۱۳ و ۹۵/۰۵/۲۲ در نظر گرفته شدند. سپس، به منظور کاهش ابعاد داده‌ها، حذف داده‌های هم‌بسته و افزایش میزان تفکیک پذیری باندها، تحلیل مؤلفه‌های اصلی روی تصاویر مورد نظر اجرا شد. این اقدام را قاسمی اسفهان و همکاران (۱۳۹۲) و خاتمی (۲۰۱۶) نیز انجام دادند. مؤلفه‌های اول، دوم و سوم PCA آنها با هم ادغام شد. بدین ترتیب، ترکیب باندی PC_1 ، PC_2 و PC_3 تصاویر اوایل اردیبهشت، اواخر خرداد، اواسط تیر، اواخر مرداد ۱۳۹۵، که مجموعه دوازده باند ۱۰ متری است، برای طبقه‌بندی انتخاب شد.

۴-۳- طبقه‌بندی

طبقه‌بندی هیبریدی^۱ در مواقعی که عوامل گوناگونی در یکنواخت‌نبودن انعکاس‌های طیفی سطح زمین تأثیر دارند، بسیار مفید است (زبیری و مجد، ۱۳۹۲). برای تفکیک اراضی متنوع این تحقیق، طبقه‌بندی هیبریدی انجام شد. ابتدا، طبقه‌بندی نظارت‌نشده IsoData^۲ و سپس، روش‌های طبقه‌بندی نظارت‌شده ماشین بردار پشتیبان و جنگل تصادفی اجرا شد. در طبقه‌بندی IsoData، پیکسل‌ها در تعداد مشخصی خوشه^۳ دسته‌بندی می‌شوند. سپس، پیکسل‌ها از خوشه‌ای به خوشه دیگر جابه‌جا می‌شوند تا زمانی که جابه‌جایی‌ها سبب ایجاد تغییر محسوس در خوشه‌های به‌دست‌آمده نشود (Mather and Tso, 2009). برای اجرای این روش، طبق آگاهی کارشناسان و با تفسیر چشمی تصاویر، حداقل هشت و حداکثر پانزده دسته در نرم‌افزار انوی^۴ لحاظ شد تا منطقه، در این بازه، طبقه‌بندی شود. به دلیل کاهش کنترل انسانی بر فرایند برچسب‌دهی، اغلب دقت نتایج این‌گونه روش‌ها پایین است (فاطمی و رضائی، ۱۳۹۳). حاصل چنین طبقه‌بندی‌ای پیکسل‌هایی است که در چند کلاس نامعلوم تقسیم‌بندی شده‌اند (خیرخواه زرکش و همکاران، ۱۳۹۶).

بدین ترتیب، تنها شمای کلی از وضعیت انواع اراضی منطقه به‌دست آمد. ماشین بردار پشتیبان و جنگل تصادفی از انواع مدل‌های هوش مصنوعی یادگیری ماشین به حساب می‌آیند. در این تحقیق، با رجوع به راهنمای OTB Team (۲۰۱۸) و با استفاده از نرم‌افزار متن‌باز OTB، پیکسل‌های داخل هر پلی‌گون با نقطه مرکزی آن متناظر شدند. به‌علاوه، مجموعه نمونه‌ها به‌طور تصادفی به داده‌های آموزشی و تستی^۵ تقسیم شد؛ به‌صورتی که اندازه داده آموزشی ۶۰٪ و اندازه داده تستی ۴۰٪ باشد. جدول ۵ تعداد نقاط آموزشی و تستی را نشان می‌دهد.

ارزش رقومی ترکیب باندی انتخاب‌شده به داده‌های آموزشی و تستی اختصاص پیدا کرد. سپس، الگوریتم ماشین بردار پشتیبان با کرنل چندجمله‌ای و مدل C-SVM (با پارامتر $C=1$ و $\gamma=0.5$) و الگوریتم جنگل تصادفی با پارامترهای پیش‌فرض همچون حداکثر عمق درختان (پنج سطح)، حداقل تعداد نمونه‌ها در سطح گره (ده نمونه) و حداکثر تعداد درختان در جنگل (یکصد درخت)، روی ترکیب باندی مورد نظر اعمال شد و داده‌های آموزشی تعلیم داده شدند.

جدول ۵. تعداد نقاط مربوط به داده‌های آموزشی و تستی

اندازه داده‌ها	جمع کل	گندم	ذرت	گوجه‌فرنگی	جو	مسکونی و جاده	آب	سایر- پوشش گیاهی	سایر- آیش و بایر
آموزشی	۱۹۲۹	۶۱۹	۲۱۶	۵۴	۱۰۴	۱۸۰	۱۳۴	۳۲۱	۳۰۱
تستی	۱۲۸۶	۴۱۳	۱۴۴	۳۶	۷۰	۱۲۰	۸۹	۲۱۴	۲۰۰
مجموع داده‌ها	۳۲۱۵	۱۰۳۲	۳۶۰	۹۰	۱۷۴	۳۰۰	۲۲۳	۵۳۵	۵۰۱

1. Hybrid Classification
2. Iterative Self-Organizing Data Analysis Technique
3. Cluster
4. Envi
5. Testing data

۵-۳- صحت‌سنجی روش‌های طبقه‌بندی

برای برآورد دقت نتایج حاصل از طبقه‌بندی کاربری اراضی و محصولات منطقه، ماتریس درهمی^۱ دو روش ماشین بردار پشتیبان و جنگل تصادفی محاسبه شد. داده‌های تستی، که در مرحلهٔ تعلیمی وارد نشده بودند، برای اعتبارسنجی^۲ نتایج طبقه‌بندی به کار می‌آیند (Raczko & Zagajewski, 2017)؛ بنابراین، ترکیب باندی حاوی ارزش‌های رقومی تمامی پیکسل‌های دسته‌های تعریف‌شده با نتیجهٔ طبقه‌بندی تلاقی داده می‌شود و از مقایسهٔ بین ارزش داده‌های تستی و ارزش پیکسل‌های متناظر در نقشهٔ خروجی، میزان خطای موجود در نقشه تعیین می‌شود. شاخص‌های اعتبارسنجی شامل صحت کلی^۳، ضریب کاپا^۴، صحت تولیدکننده^۵، صحت کاربر^۶، خطای امیشن^۷ و خطای کمیشن^۸ است. صحت کلی میانگینی از دقت طبقه‌بندی است که نسبت پیکسل‌های به‌درستی طبقه‌بندی‌شده را به جمع کل پیکسل‌های معلوم نشان می‌دهد و براساس رابطه (۵) به‌دست می‌آید.

$$O.A. = \frac{\sum_{i=1}^c E_{ii}}{N} \quad \text{رابطه (۵)}$$

O.A. صحت کلی طبقه‌بندی، c تعداد کلاس‌ها، N تعداد کل پیکسل‌های معلوم و E_{ii} اعضای قطری ماتریس درهمی است.

ضریب کاپا دقت را به‌نسبت یک طبقه‌بندی تصادفی، محاسبه می‌کند (Richards, 1995) و طبق رابطه (۶) محاسبه می‌شود.

$$K = \frac{N \sum_{i=1}^c X_{ii} - \sum_{i=1}^c X_{i+} X_{+i}}{N^2 - \sum_{i=1}^c X_{i+} X_{+i}} \quad \text{رابطه (۶)}$$

K ضریب کاپا، N تعداد کل پیکسل‌های واقعیت زمینی، X_{i+} مجموع عناصر سطر X_{+i} و X_{+i} مجموع عناصر ستون X_{+i} است.

$$N = \sum_{i=1}^r \sum_{j=1}^c X_{ij} \quad X_{+j} = \sum_{i=1}^r X_{ij} \\ X_{i+} = \sum_{j=1}^c X_{ij}$$

r و c، به‌ترتیب، تعداد سطرها و ستون‌های ماتریس درهمی‌اند. صحت کلی تخمینی خوشبینانه^۹ است و همیشه دقت طبقه‌بندی را بیشتر از مقدار واقعی بیان می‌کند. ضریب کاپا تخمینی بدبینانه^{۱۰} است و دقت را کمتر نشان می‌دهد (Liu, 2008).

۴- نتایج

۴-۱- نقشه‌های حاصل از روش‌های طبقه‌بندی

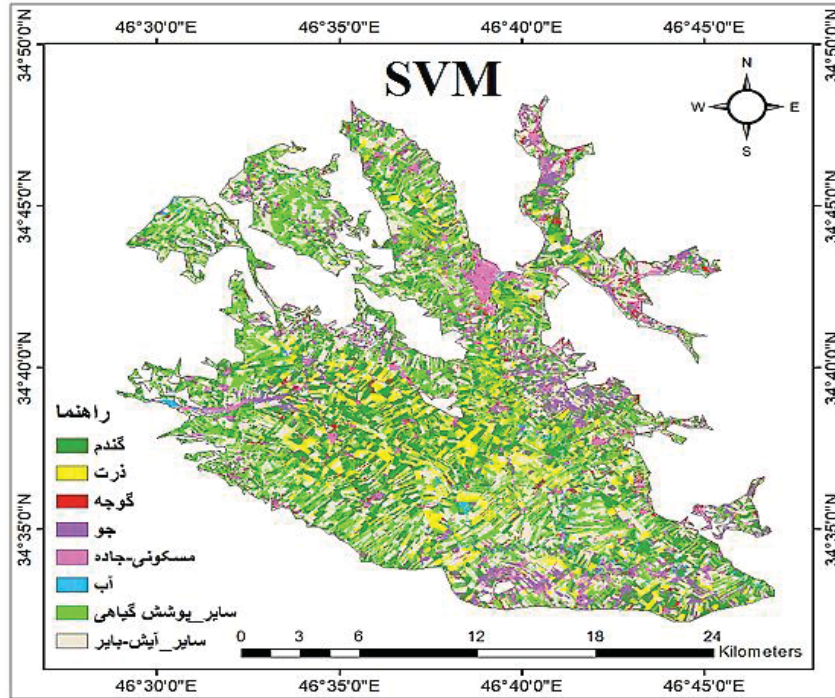
نقشهٔ کاربری اراضی و محصولات منطقه، با روش ماشین بردار پشتیبان و جنگل تصادفی، به‌ترتیب، در شکل‌های ۴ و ۵ آمده است.

۴-۲- مقایسهٔ دقت روش‌های طبقه‌بندی

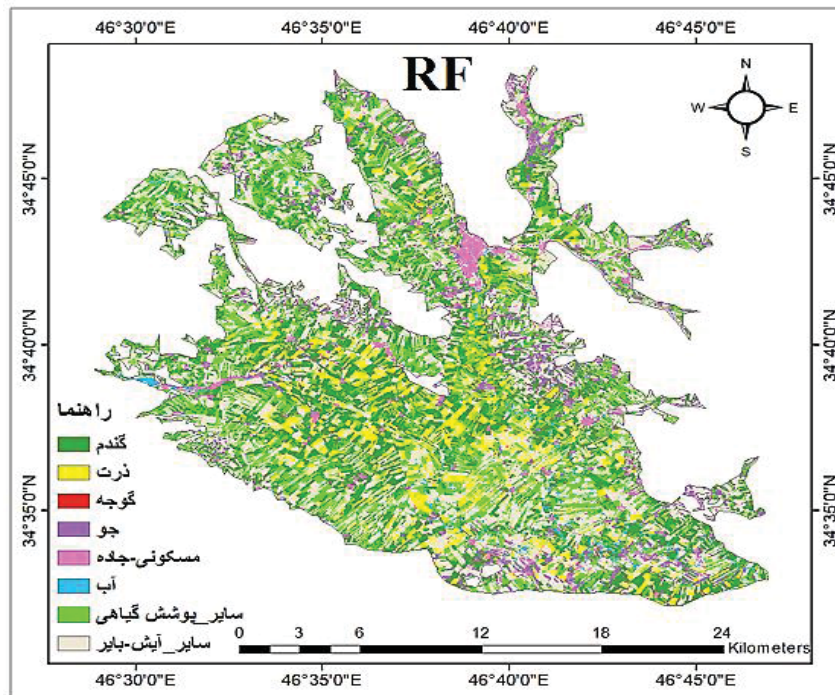
شاخص‌های اعتبارسنجی روش‌های طبقه‌بندی از طریق محاسبهٔ صحت کلی (رابطه (۵))، ضریب کاپا (رابطه (۶))، صحت تولیدکننده، صحت کاربر، خطای امیشن و خطای کمیشن به‌دست آمد. ماتریس درهمی نقشهٔ حاصل از طبقه‌بندی کاربری اراضی و محصولات منطقه به‌روش ماشین بردار پشتیبان و جنگل تصادفی، به‌ترتیب، در جدول‌های ۶ و ۷ آمده است. اعداد موجود در ستون‌ها نشان‌دهندهٔ داده‌های طبقه‌بندی‌شده و اعداد موجود در سطرها بیانگر داده‌های مرجع زمینی‌اند.

1. Confusion matrix
2. Validation
3. Overall Accuracy
4. Kappa Coefficient
5. Producer's Accuracy
6. User's Accuracy
7. Omission Error
8. Commission Error
9. Over estimation
10. Under estimation

مقایسه دقت روش‌های ماشین بردار پشتیبان و جنگل تصادفی در ...



شکل ۴. نقشه طبقه‌بندی کاربری اراضی و محصولات زراعی منطقه با روش ماشین بردار پشتیبان



شکل ۵. نقشه طبقه‌بندی کاربری اراضی و محصولات زراعی منطقه با روش جنگل تصادفی

زینب قدسی و همکاران

جدول ۶. ماتریس درهمی نقشه حاصل از طبقه‌بندی کاربری اراضی و محصولات زراعی منطقه به‌روش ماشین بردار پشتیبان

C _{e1}	U.A. _r	مجموع	سایر-		آب	مسکونی و جاده	جو	گوجه‌فرنگی	ذرت	گندم	کلاس
			آیش و بایر	پوشش گیاهی							
۱۳.۹۵	۸۶.۰۵	۴۱۳	۰	۱۴	۰	۰	۴	۰	۰	۳۹۵	گندم
۰	۱۰۰	۱۴۴	۰	۰	۰	۰	۰	۰	۱۴۴	۰	ذرت
۰	۱۰۰	۳۶	۰	۹	۰	۱	۰	۶	۰	۲۰	گوجه‌فرنگی
۲۴.۱۴	۷۵.۸۶	۷۰	۰	۴	۰	۰	۴۴	۰	۰	۲۲	جو
۰.۸۸	۹۹.۱۲	۱۲۰	۲	۳	۰	۱۱۳	۱	۰	۰	۱	مسکونی و جاده
۰	۱۰۰	۸۹	۰	۰	۸۹	۰	۰	۰	۰	۰	آب
۱۴.۰۲	۸۵.۹۸	۲۱۴	۰	۱۸۴	۰	۰	۹	۰	۰	۲۱	سایر- پوشش گیاهی
۰.۹۹	۹۹.۰۱	۲۰۰	۲۰۰	۰	۰	۰	۰	۰	۰	۰	سایر- آیش و بایر
		۱۲۸۶	۲۰۲	۲۱۴	۸۹	۱۱۴	۵۸	۶	۱۴۴	۴۵۹	مجموع
			۱۰۰	۸۵.۹۸	۱۰۰	۹۴.۱۶	۶۲.۸۵	۱۶.۶۶	۱۰۰	۹۵.۶۴	P.A. _r
			۰	۱۴.۰۲	۰	۵.۸۴	۳۷.۱۵	۸۳.۳۴	۰	۴.۳۶	O _{e۴}
											Kappa _Δ = ۰.۸۹۲۷ OA _۶ = ۹۱.۳۶
صحت کلی/ OA _۶ % ضریب کاپا Kappa _Δ خطای آمیشن/ O _{e۴} % صحت تولیدکننده/ P.A. _r % صحت کاربر/ U.A. _r % خطای کمیشن/ C _{e۱} %											

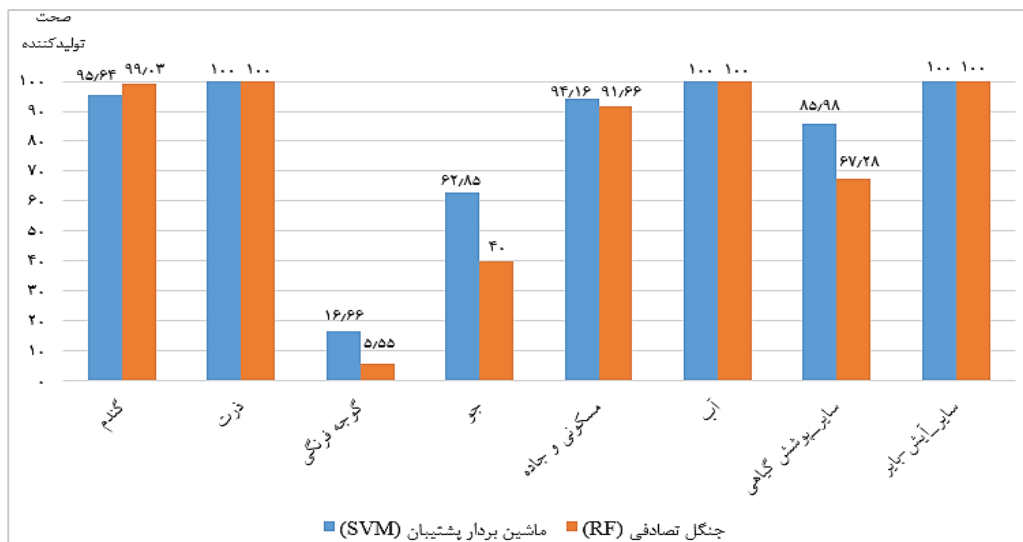
جدول ۷. ماتریس درهمی نقشه حاصل از طبقه‌بندی کاربری اراضی و محصولات زراعی منطقه به‌روش جنگل تصادفی

C _{e۱}	U.A. _r	مجموع	سایر-		آب	مسکونی و جاده	جو	گوجه‌فرنگی	ذرت	گندم	کلاس
			آیش و بایر	پوشش گیاهی							
۲۳.۹۸	۷۶.۰۲	۴۱۳	۰	۳	۰	۰	۱	۰	۰	۴۰۹	گندم
۰	۱۰۰	۱۴۴	۰	۰	۰	۰	۰	۰	۱۴۴	۰	ذرت
۰	۱۰۰	۳۶	۷	۷	۰	۰	۱	۲	۰	۱۹	گوجه‌فرنگی
۹.۶۸	۹۰.۳۲	۷۰	۰	۶	۰	۰	۲۸	۰	۰	۳۶	جو
۰	۱۰۰	۱۲۰	۳	۲	۰	۱۱۰	۰	۰	۰	۵	مسکونی و جاده
۰	۱۰۰	۸۹	۰	۰	۸۹	۰	۰	۰	۰	۰	آب
۱۱.۱۲	۸۸.۸۸	۲۱۴	۰	۱۴۴	۰	۰	۱	۰	۰	۶۹	سایر- پوشش گیاهی
۴.۷۷	۹۵.۲۳	۲۰۰	۲۰۰	۰	۰	۰	۰	۰	۰	۰	سایر- آیش و بایر
		۱۲۸۶	۲۱۰	۱۶۲	۸۹	۱۱۰	۳۱	۲	۱۴۴	۵۳۸	مجموع
			۱۰۰	۶۷.۲۸	۱۰۰	۹۱.۶۶	۴۰	۵.۵۵	۱۰۰	۹۹.۰۳	P.A. _r
			۰	۳۲.۷۲	۰	۸.۳۴	۶۰	۹۴.۴۵	۰	۰.۹۷	O _{e۴}
											Kappa _Δ = ۰.۸۴۳ OA _۶ = ۸۷.۵۵
صحت کلی/ OA _۶ % ضریب کاپا Kappa _Δ خطای آمیشن/ O _{e۴} % صحت تولیدکننده/ P.A. _r % صحت کاربر/ U.A. _r % خطای کمیشن/ C _{e۱} %											

بیشترین به کمترین، متعلق به دسته‌های ذرت، آب و سایر-آیش و بایر با مقدار ۱۰۰٪، گندم ۹۵.۶۴٪، مسکونی و جاده ۹۴.۱۶٪، سایر-پوشش گیاهی ۸۵.۹۸٪، جو ۶۲.۸۵٪ و گوجه‌فرنگی ۱۶.۶۶٪ است.

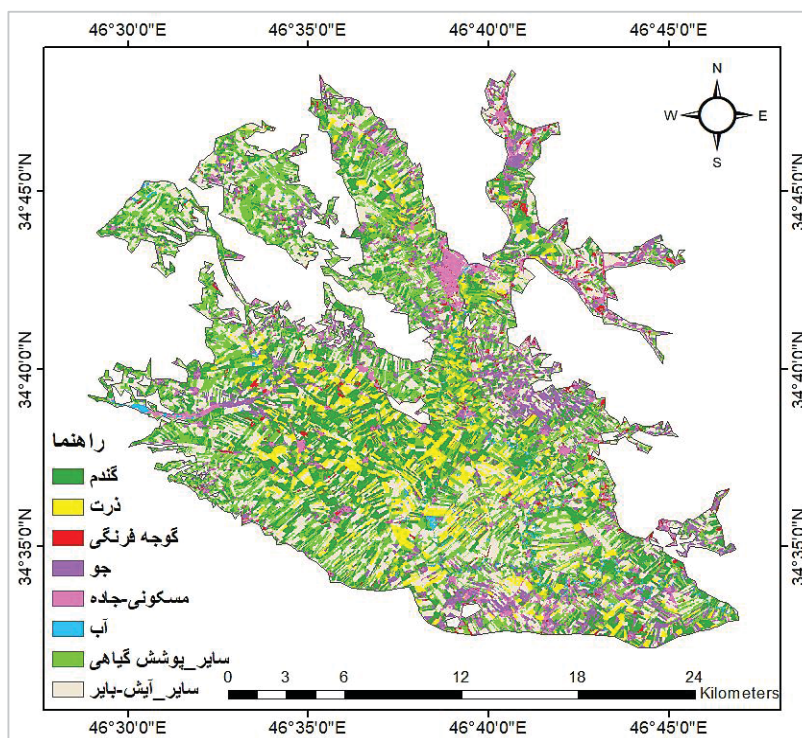
۳-۴- پردازش نهایی و استخراج نقشه کاربری اراضی و محصولات زراعی منطقه با دقیق‌ترین روش برای پردازش نهایی، فیلتر اکثریت^۱ با اندازه پنجره متحرک ۳×۳ روی دقیق‌ترین نقشه اعمال شد و پیکسل‌های پراکنده با پیکسل‌های همسایه مشابه ادغام شدند. بدین ترتیب، دقیق‌ترین نقشه طبقه‌بندی انواع کاربری اراضی و محصولات زراعی منطقه، به روش ماشین بردار پشتیبان و با استفاده از تصاویر چندزمانه سنتینل-۲، تهیه شد که در شکل ۷ نشان داده شده است.

در اندازه صحت کلی و ضریب کاپا، بین طبقه‌بندی با روش ماشین بردار پشتیبان و جنگل تصادفی مقایسه صورت گرفت. نتیجه مقایسه نشان داد نقشه حاصل از طبقه‌بندی منطقه به روش ماشین بردار پشتیبان با کرنل سائز چندجمله‌ای (با صحت کلی ۹۱.۳۶ و ضریب کاپا ۰.۸۹۲۷)، با اختلاف ۳.۸۱٪ در صحت کلی و ۰.۴۹۷ در ضریب کاپا، دقیق‌تر از نقشه حاصل از طبقه‌بندی به روش جنگل تصادفی است. اعداد نمودار میله‌ای شکل ۶ نشان‌دهنده صحت تولیدکننده و به عبارت دیگر، دقت کلاس‌های تولیدشده با استفاده از روش‌های طبقه‌بندی ماشین بردار پشتیبان و جنگل تصادفی‌اند. صحت تولیدکننده برای کلاس‌های متفاوت، به منزله میزان درستی اختصاص پیکسل‌های واقعیت زمینی یک دسته به آن دسته با روش طبقه‌بندی، در دو روش مقادیر متفاوتی دارد. دقت کلاس‌های به دست آمده با دقیق‌ترین روش، ماشین بردار پشتیبان، به ترتیب



شکل ۶. دقت کلاس‌های تولیدشده به روش‌های طبقه‌بندی ماشین بردار پشتیبان و جنگل تصادفی

1. Majority Filter
2. Moving Window



شکل ۷. دقیق ترین نقشه طبقه بندی کاربری اراضی و محصولات زراعی منطقه به روش ماشین بردار پشتیبان، با تصاویر چندزمانه سنتینل-۲

۴-۴- نتایج طبقه بندی به تفکیک کلاس ها در

دقیق ترین روش

آمار کلاس های تفکیک شده در دقیق ترین نقشه طبقه بندی کاربری اراضی و محصولات منطقه به روش ماشین بردار پشتیبان با استفاده از نرم افزار انوی محاسبه شد (جدول ۸). پیکسل های ماسک شده با ابر و پیکسل های خارج از مرز منطقه در این جدول قید نشده است.

۵- بحث و نتیجه گیری

تا کنون روش های متفاوتی برای طبقه بندی، از جمله پارامتریک و ناپارامتریک، روی تصاویر گوناگون اجرا شده اند. دقت نقشه های استخراج شده با توجه به نوع داده های ماهواره ای، روش طبقه بندی و شرایط منطقه تعیین می شود. در این تحقیق، به علت تنوع زیاد محصولات اراضی دشت سنجابی روانسر و جایگاه ویژه

جدول ۸. مشخصات کلاس های دقیق ترین نقشه طبقه بندی کاربری اراضی و محصولات زراعی منطقه

کد کلاس	نام کلاس	تعداد پیکسل ها	مساحت (هکتار)	مساحت در تصویر (%)
۱	گندم	۸۴۳.۷۰۴	۷۰۴۸.۴۳	۸.۴۷
۲	ذرت	۵۴۵.۳۲۶	۳۲۶۵.۴۵	۳.۹۲
۳	گوجه فرنگی	۱۹۸.۷۹	۷۹۱.۹۸	۰.۹۵
۴	جو	۵۰۲.۴۴۵	۴۴۵۵.۰۲	۵.۳۵
۵	مناطق مسکونی و جاده ها	۵۲۴.۲۲۲	۲۲۲۵.۲۴	۲.۶۷
۶	آب	۹۴۱.۴۱	۴۱۹.۴۱	۰.۵
۷	سایر- پوشش گیاهی	۶۵۹.۹۰۸	۹۰۸۶.۵۹	۱۰.۹۲
۸	سایر- آیش و بایر	۹۹۷.۹۵۱	۹۵۱۹.۹۷	۱۱.۴۴

(Gómez et al., 2016). همچنین، می‌تواند مشکلات موجود در اختلاف نامتعادل میان نمونه‌های آموزشی را حل کند (Jensen, 2005).

در هر دو روش طبقه‌بندی، ترکیب باندی PC_1 ، PC_2 و PC_3 تصاویر سنتینل-۲ مربوط به اوایل اردیبهشت، اواخر خرداد، اواسط تیر و اواخر مرداد سال ۱۳۹۵ به‌کار رفت. استفاده از تصاویر چندزمانه سنجش از دور و تعداد کافی داده زمینی، در جداسازی پدیده‌های منطقه با تنوع اراضی، نقش مهمی ایفا می‌کند و باعث بهبود دقت نقشه طبقه‌بندی می‌شود (Gómez et al., 2016)، (Guidici and Clark, 2017)، (Khatami et al., 2016) و (Rujoiu-Mare et al., 2017). در این تحقیق، مناطق شهری و سایر پوشش‌های گیاهی، با استفاده از ماشین بردار پشتیبان، به‌خوبی شناسایی شدند. ذرت، آب و اراضی آیش و بایر، با هر دو روش، با دقت بالا تفکیک شدند. ذرت به‌دلیل دوره رویشی متفاوت با دیگر محصولات، سایر اراضی به‌دلیل آیش و یا شخم‌بودن و پهنه‌های آبی نیز، به‌دلیل رفتار طیفی متفاوتی که دارند، به‌درستی تشخیص داده شدند. فقط در دسته گندم، دقت روش RF تا حدی بیشتر از دقت SVM شد. روش SVM در استخراج محصولات گوجه‌فرنگی و جو، با اختلاف تقریباً زیادی، موفق‌تر از روش RF بوده است. در مجموع، از بین محصولات کشاورزی، ذرت و گندم با دقت بالا و جو و گوجه‌فرنگی با دقت پایین، با دو روش، تفکیک شدند. گوجه‌فرنگی، به‌دلیل مساحت کم مزارع آن در منطقه و تعداد نمونه‌های کمتر و اشتراک دوره رویشی، و جو، به‌دلیل شباهت رفتار فنولوژیکی با گندم، به‌خوبی تفکیک نشدند. پیشنهاد می‌شود شبکه عصبی، الگوریتم ژنتیک^۱ و روش امضای طیفی^۲ با تصاویر چندزمانه، برای تفکیک اراضی و انواع محصولات منطقه، به‌کار رود. همچنین، روش فاکتور شاخص مطلوب^۳، که براساس واریانس و همبستگی باندها کار می‌کند، برای تعیین باندهای مناسب طبقه‌بندی به‌کار برود (علوی‌پناه، ۱۳۹۵).

منطقه در سطح استان کرمانشاه، به‌ویژه از لحاظ باغبانی و زراعت، شناسایی دقیق انواع کاربری زمین و محصولات کشاورزی برای پیش و مدیریت منطقه ضرورت داشت. برای این کار، چون محصولات متنوعی در منطقه کشت می‌شوند، از تصاویر چندزمانه سنتینل-۲ استفاده شد. برای کاهش ابعاد داده‌ها و افزایش میزان تفکیک‌پذیری طیفی، تحلیل مؤلفه‌های اصلی اعمال شد و بدین ترتیب، روش طبقه‌بندی ماشین بردار پشتیبان و جنگل تصادفی، با الگوریتم یادگیری ماشین، روی ترکیب باندی مؤلفه‌های اول، دوم و سوم PCA تصاویر چندزمانه سنتینل-۲ اجرا شد. از میان منابع مورد بررسی، در کمتر تحقیقی طبقه‌بندی روی ترکیب باندی PC_1 تصاویر چندزمانه سنتینل-۲ انجام شده است. دقت این دو روش از طریق مقادیر صحت کلی و ضریب کاپا در جدول ۶ و ۷ مورد ارزیابی قرار گرفت. مقایسه صحت کلی و ضریب کاپای بین دو روش نشان داد نقشه حاصل از طبقه‌بندی کاربری اراضی و محصولات زراعی دشت سنجایی روانسر به‌روشن ماشین بردار پشتیبان با کرنل سائز چندجمله‌ای، با صحت کلی ۹۱.۳۶ و ضریب کاپای ۰.۸۹۲۷ (با اختلاف ناچیز ۳.۸۱٪ در صحت کلی و ۰.۰۴۹۷ در ضریب کاپا)، در مقایسه با نقشه حاصل از طبقه‌بندی به‌روش جنگل تصادفی، از دقت بیشتری برخوردار بود و به این میزان به واقعیت زمینی شباهت بیشتری داشت. این نتایج با نتایج تحقیقات شنایی هویزه و زارعی (۱۳۹۵)، میرزایی‌زاده و همکاران (۱۳۹۴)، یوسفی و همکاران (۱۳۹۳)، خاتمی و همکاران (۲۰۱۶) و پژوهشی دیگر (Rujoiu-Mare et al., 2017)، که در آنها ثابت شده است روش طبقه‌بندی ماشین بردار پشتیبان دقتی بیشتر از دیگر روش‌ها دارد، دارای مطابقت است و از سوی دیگر، با یافته‌های پژوهش‌های قاسمی اسفهلان و همکاران (۱۳۹۲)، عینی زیناب و صاحبی (۱۳۹۴) و تحقیقی دیگر (Forkuor et al., 2017)، که حاکی از برتری جنگل تصادفی بودند، مطابقت ندارد. از مزایای روش ماشین بردار پشتیبان، می‌توان به این اشاره کرد که روش یادشده قابلیت بالایی در تفکیک تصاویر چندزمانه و چندطیفی در یک مجموعه ویژگی‌های وسیع دارد و به‌ازای نقاط آموزشی اندک، می‌تواند نتایج طبقه‌بندی بسیار قدرتمندی فراهم آورد

1. Genetic Algorithm
2. Spectral Signature
3. Optimum Index Factor (OIF)

۶- سیاست‌گذاری

در پایان، از پژوهشکده حفاظت خاک و آبخیزداری و سازمان جنگل‌ها، مراتع و آبخیزداری کشور، بابت همکاری در تحویل داده‌های مورد نیاز این تحقیق، قدردانی می‌شود.

۷- منابع

اصغری، ف.، ۱۳۹۶، تخمین میزان تولید محصول ذرت با استفاده از مدل سنجش از دوری GRAMI، پایان‌نامه کارشناسی ارشد رشته RS و GIS گرایش هواشناسی ماهواره‌ای، دانشکده محیط‌زیست و انرژی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات.

جهانبخشی، ف.، اختصاصی، م.ر.، ۱۳۹۷، ارزیابی عملکرد سه روش طبقه‌بندی تصویر (جنگل تصادفی، ماشین بردار پشتیبان و بیشترین شباهت) در تهیه نقشه کاربری اراضی، علوم آب و خاک (علوم و فنون کشاورزی و منابع طبیعی)، سال ۲۲، شماره ۴، صص. ۲۳۵-۲۴۷.

خیرخواه زرکش، م.م.، میرزاپور، س.، صفا، م.، ۱۳۹۶، آموزش نرم‌افزار ENVI 5.3، انتشارات مهرگان قلم، تهران.

زبیری، م.، مجد، ع.، ۱۳۹۲، آشنایی با فن سنجش از دور و کاربرد در منابع طبیعی، انتشارات دانشگاه تهران، تهران.

شفیعی، م.، سرکارگر اردکانی، ع.، وحیدنیا، م.ح.، ۱۳۹۶، مقایسه طبقه‌بندی با الگوریتم‌های جنگل تصادفی و شبکه‌های عصبی بر روی تصاویر شبیه‌سازی شده ابرطیفی، بیست و چهارمین همایش ملی ژئوماتیک، سازمان نقشه‌برداری کشور، تهران.

شنانی هویزه، س.م.، زارعی، ح.، ۱۳۹۵، مقایسه الگوریتم‌های طبقه‌بندی شبکه عصبی مصنوعی، ماشین بردار پشتیبان و حداکثر

احتمال در استخراج نقشه کاربری اراضی حوزه

آبخیز ابوالعباس، علوم و مهندسی آبخیزداری ایران، سال ۱۰، شماره ۳۳، صص. ۷۳-۸۵.

صحبتلو، ل.، ۱۳۹۶، تعیین نیاز آبی محصول گندم با استفاده از تکنیک‌های سنجش از دور (مطالعه موردی: شهرستان روانسر)، پایان‌نامه کارشناسی ارشد RS و GIS گرایش آب و خاک، دانشکده محیط‌زیست و انرژی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات.

صمدزادگان، ف.، حسنی، ح.س.، ۱۳۹۱، تعیین ماشین‌های بردار پشتیبان بهینه در طبقه‌بندی تصاویر فراترکیفی بر مبنای الگوریتم ژنتیک، فناوری اطلاعات و ارتباطات ایران، سال ۴، شماره‌های ۱۳ و ۱۴، صص. ۲۴-۹.

علوی پناه، س.ک.، ۱۳۹۵، کاربرد سنجش از دور در علوم زمین (علوم خاک)، انتشارات دانشگاه تهران، تهران.

عینی زیناب، ح.، صاحبی، م.ر.، ۱۳۹۴، طبقه‌بندی تصاویر ابرطیفی در مناطق شهری با روش جنگل تصادفی، بیست و دومین همایش ملی ژئوماتیک، سازمان نقشه‌برداری کشور، تهران.

فاطمی، س.ب.، رضائی، ی.، ۱۳۹۳، مبانی سنجش از دور، انتشارات آزاده، تهران.

قاسمی اسفهلان، ا.، ولدان‌زوج، م.ج.، صاحبی، م.ر.، ۱۳۹۲، طبقه‌بندی تصویر ابرطیفی و لیدار با استفاده از الگوریتم جنگل تصادفی، بیست و دومین همایش ملی ژئوماتیک، سازمان نقشه‌برداری کشور، تهران.

مهدوی ع.، فلاح شمسی، س.ر.، ۱۳۹۱، تهیه نقشه تغییرات سطح جنگل با استفاده از عکس‌های هوایی و تصاویر LISS-III ماهواره IRS (مطالعه موردی: شهرستان ایلام)، پژوهش‌های علوم و فناوری چوب و جنگل، سال ۱۹، شماره ۱، صص. ۹۱-۷۷.

- California with Simulated HypsIRI Imagery**, ISPRS Journal of Photogrammetry and Remote Sensing, 119, PP. 228-245.
- ESA (European Space Agency), 2015, **User Guide of Sentinel-2 MSI**.
<http://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi>.
- Fernandez-Delgado, M., Cernadas, E., Barro, S. & Amorim, D., 2014, **Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?**, Journal of Machine Learning Research, 15, PP. 3133-3181.
- Forkuor, G., Dimobe, K., Serme, I. & Ebagnerin Tondoh, J., 2017, **Landsat-8 vs. Sentinel-2 Examining the Added Value of Sentinel-2's Red-Edge Bands to Land-Use and Land-Cover Mapping in Burkina Faso**, GIScience & Remote Sensing, 55(3), PP. 331-354.
- Gómez, C., White, J.C. & Wulder, M.A., 2016, **Optical Remotely Sensed Time Series Data for Land Cover Classification: A Review**, ISPRS Journal of Photogrammetry and Remote Sensing, 116, PP. 55-72.
- Guidici, D. & Clark, M.L., 2017, **One-Dimensional Convolutional Neural Network Land-Cover Classification of Multi-Seasonal Hyperspectral Imagery in the San Francisco Bay Area, California**, Remote Sensing, 9(6), P. 629.
- Gulácsi, A. & Kovács, F., 2020, **Sentinel-1 Imagery-Based High-Resolution Water Cover Detection on Wetlands, Aided by Google Earth Engine**, Remote Sensing, 12(10), P. 1614.
- Immitzer, M., Vuolo, F. & Atzberger, C., 2016, **First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe**, Remote Sensing, 8(3), P. 166.
- میرزایی‌زاده و، نیک‌نژاد، م.، اولادی قصادیکلایی، ج.، ۱۳۹۴، **ارزیابی الگوریتم‌های طبقه‌بندی نظارت‌شده غیر پارامتریک در تهیه نقشه پوشش زمین با استفاده از تصاویر لندست ۸**، **سنجش از دور و سامانه اطلاعات جغرافیایی در منابع طبیعی**، سال ۶، شماره ۳، صص. ۲۹-۴۴.
- یوسفی، ص.، تازه، م.، میرزایی، س.، مرادی، ح.ر.، توانگر، ش.، ۱۳۹۳، **مقایسه الگوریتم‌های مختلف طبقه‌بندی تصاویر ماهواره‌ای در تهیه نقشه کاربری اراضی (مطالعه موردی: شهرستان نور)**، **سنجش از دور و سامانه اطلاعات جغرافیایی در منابع طبیعی**، سال ۵، شماره ۳، صص. ۶۷-۷۶.
- Ali, M.I., Dirawan, G.D., Hasim, A.H. & Abidin, M.R., 2019, **Detection of Changes in Surface Water Bodies Urban Area with NDWI and MNDWI Methods**, International Journal on Advanced Science, Engineering and Information Technology, 9(3), PP. 946-951.
- Basukala, A.K., Oldenburg, C., Schellberg, J., Sultanov, M. & Dubovyk, O., 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), PP. 187-201.
- Breiman, L., 2001, **Random Forests**, **Machine Learning**, 45(1), PP. 5-32.
- Chuvieco, E., 2016, **Fundamentals of Satellite Remote Sensing: An Environmental Approach** (2nd Ed.), Boca Raton/London/New York: CRC Press Taylor & Francis.
- Clark, M.L. & Kilham, N.E., 2016, **Mapping of Land Cover in Northern**

- Jensen, J., 2005, **Introductory Digital Image Processing: A Remote Sensing Perspective** (3rd Ed.), Prentice Hall, Upper Saddle River, NJ.
- Jeong, J., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.M., Gerber, J.S., Reddy, V.R. & Kim, S.H., 2016, **Random Forests for Global and Regional Crop Yield Predictions**, PLoS One, 11(6).
- Khatami, R., Mountrakis, G. & Stehman, S.V., 2016, **A Meta-Analysis of Remote Sensing Research on Supervised Pixel-Based Land-Cover Image Classification Processes- General Guidelines for Practitioners and Future Research**, Remote Sensing of Environment, 177, PP. 89-100.
- Liu, C.L., 2008, **Partial Discriminative Training for Classification of Overlapping Classes in Document Analysis**, International Journal on Document Analysis and Recognition, 11(2), PP. 53-65.
- Mather, P.M. & Tso, B., 2009, **Classification Methods for Remotely Sensed Data**, CRC Press, Boca Raton.
- Mathur, A. & Foody, G.M., 2008, **Crop Classification by Support Vector Machine with Intelligently Selected Training Data for an Operational Application**, International Journal of Remote Sensing, 29(8), PP. 2227-2240.
- Mitchell, T.M., 2017, **Machine Learning** (1st Ed.), Mc Graw Hill, India.
- OTB Team, 2018, **OTB CookBook Documentation, Release 6.6.1**. <http://www.orfeo-toolbox.org/packages/archives/Doc/CookBook-6.6.1.pdf>
- Raczko, E. & Zagajewski, B., 2017, **Comparison of Support Vector Machine, Random Forest and Neural Network Classifiers for Tree Species Classification on Airborne Hyperspectral APEX Images**, European Journal of Remote Sensing, 50(1), PP. 144-154.
- Richards, J.A., 1995. **Remote Sensing Digital Image Analysis: An Introduction** (2nd Ed.), Springer, Berlin.
- Rujoiu-Mare, M.R., Olariu, B., Mihai, B.A., Nistor, C. & Săvulescu, I., 2017, **Land Cover Classification in Romanian Carpathians and Sub-Carpathians Using Multi-Date Sentinel-2 Remote Sensing Imagery**, European Journal of Remote Sensing, 50(1), PP. 496-508.
- Vapnik, V.N., 1995, **The Nature of Statistical Learning Theory**, Springer-Verlag, New York.