

Comparative Analysis of LAI Retrieval from Hyperspectral Imagery Using Machine Learning Approaches

Behzad Mohammadi Sheikh Razi,¹ Mohammad Sharif Molla,¹ Ali Jafar Mousivand,^{2*} , Ali Shamsoddini.²

1. M.sc. in Remote Sensing and GIS, Tarbiat Modarres University, Tehran, Iran

2. Assistant prof. of Remote Sensing, Remote Sensing and GIS Dep., Tarbiat Modarres University, Tehran, Iran

Abstract

Vegetation biophysical and biochemical variables are key inputs to a wide range of modelling approaches for carbon, water, energy cycle, climate and agricultural applications. Leaf Area Index (LAI) is among the most important canopy variables, used by many different physiological and functional plant models. Several approaches have been developed for vegetation properties retrieval from remotely sensed hyperspectral data. Among them, nonparametric machine learning methods have increasingly gained attention in vegetation variable retrieval due to their flexibility and efficiency while working with data of high dimensionality over the last decades. Although these methods provide reasonable accuracy at relatively high speed, they are mainly restricted to estimate values within their training domain and often perform poorly on the marginal values (i.e. outside of the training domain). The performance of these methods has not been adequately studied in retrieving LAI on the marginal values. This study employs four well-known machine learning methods including SVR, GPR, ANN, and RF to retrieve LAI from a hyperspectral CHRIS-Proba image over Barax, Spain, in order to inspect their capability in retrieving marginal values. The results showed that although all the methods perform similarly well on retrieving LAI over the training domain values with RMSE values of less than 0.5 and relative error of less than 10%, GPR and SVR performed slightly better. However, ANN outperformed the other methods in estimating LAI on the marginal values, resulted in the generated LAI map more consistent with the NDVI map, as well as, the hyperspectral image of the region.

Keywords: Parameter retrieval, Leaf Area Index, Nonparameteric machine learning, CHRIS-Proba.

* Correspondence Address: Jalal AleAhmad, Nasr bridge, PO box: 14115-111, Tehran. Tel: + 98 (0)2182884698

Email: a.mousivand@modares.ac.ir

<https://doi.org/10.52547/gisj.12.3.47>



سبج از دور

GIS ایران

سنجش از دور و GIS ایران سال دوازدهم، شماره سوم، پاییز ۱۳۹۹
Vol.12, No. 3, Autumn 2020 Iranian Remote Sensing & GIS

۶۴-۴۷
مقاله پژوهشی



آنالیز مقایسه‌ای بازیابی شاخص سطح برگ

با استفاده از روش‌های یادگیری ماشین مبتنی بر داده‌های ابرطیفی

بهزاد محمدی شیخ‌رضی^۱، محمد شریف ملا^۱، علی جعفر موسیوند^{۲*}، علی شمس‌الدینی^۲

۱. کارشناسی ارشد سنجش از دور و GIS، دانشگاه تربیت مدرس، تهران، ایران

۲. استادیار گروه سنجش از دور و GIS، دانشگاه تربیت مدرس، تهران، ایران

تاریخ پذیرش مقاله: ۱۳۹۹/۰۳/۲۴

تاریخ دریافت مقاله: ۱۳۹۸/۰۶/۲۱

چکیده

متغیرهای بیوفیزیکی و بیوشیمیایی پوشش گیاهی، به‌منزله متغیرهای ورودی، برای مدل‌های متفاوت چرخه کربن، آب، انرژی و مدل‌های اقلیمی و کشاورزی دقیق نقش مهمی ایفا می‌کنند. یکی از مهم‌ترین متغیرهای مربوط به تاج‌پوشش گیاه، که کاربردهای فراوانی در مدل‌سازی‌های گوناگون خاک و گیاه و اتمسفر دارد، شاخص سطح برگ (LAI) است. روش‌های گوناگونی برای بازیابی LAI از تصاویر ابرطیفی به‌کار رفته‌اند که، از میان آنها، روش‌های ناپارامتریک غیرخطی یادگیری ماشین بسیار مورد توجه قرار گرفته‌اند زیرا، در مواجهه با داده‌های دارای ابعاد زیاد، انعطاف‌پذیرند. با این حال، در مطالعات پیشین، به بررسی عملکرد روش‌های یادگیری ماشین در بازیابی مقادیر LAI در مقادیر حاشیه‌ای (مقادیر خارج از دامنه نمونه‌گیری زمینی) و قابلیت این روش‌ها در تهیه نقشه متغیر توجه چندانی نشده است. در این تحقیق، عملکرد چهار روش پرکاربرد یادگیری ماشین شامل رگرسیون بردار پشتیبان، فرایند گاوسی، شبکه عصبی مصنوعی و جنگل تصادفی در بازیابی LAI از تصویر ابرطیفی ماهواره کریس- پروبا بررسی شده است. نتایج نشان داد که، به‌رغم کارایی هر چهار روش در بازیابی مقادیر LAI برای دامنه مقادیر اندازه‌گیری شده زمینی با RMSE بهتر از ۰.۵ و خطای نسبی کمتر از ۱۰٪، روش‌های فرایند گاوسی و رگرسیون بردار پشتیبان صحت بالاتری در مقایسه با سایر روش‌ها دارند. با وجود این، عملکرد روش شبکه عصبی مصنوعی، در تخمین LAI‌های دارای مقادیر حاشیه‌ای، بهتر از دیگر روش‌هاست و نقشه تهیه‌شده با این روش و تابع یادگیری GDA تطابق بیشتری با نقشه NDVI و تصویر ابرطیفی منطقه دارد.

کلیدواژه‌ها: بازیابی پارامتر، شاخص سطح برگ، داده ابرطیفی، روش‌های ناپارامتریک یادگیری ماشین، کریس- پروبا.

* نویسنده مکاتبه‌کننده: تهران، جلال آل احمد، پل نصر، صندوق پستی: ۱۱۱- تلفن: ۰۴۶-۸۲۸۸۵۰۲۱

۱- مقدمه

از بین روش‌های مطرح‌شده برای بازیابی پارامترهای پوشش گیاهی از داده‌های سنجنش از دوری، شامل رویکردهای تجربی - آماری و فیزیکی - مبنای رویکرد نخست، به دلیل سادگی و سرعت محاسباتی بالا، کاربرد فراوانی در مطالعات گوناگون داشته است (Baret and Fraaij, 2008; Kimes et al., 2000). این رویکرد دربرگیرنده فرایند آموزش یک مدل (مثلاً رگرسیون) با داده‌های واقعیت زمینی (یا شبیه‌سازی‌شده) و سپس، استفاده از مدل آموزش‌داده‌شده برای بازیابی پارامتر مورد نظر است. چالش اصلی در استفاده از این روش‌ها وابستگی فرایند آموزش به شرایط متفاوت همچون سنجنده، زمان و مکان خاص است که، با هر تغییری در این شرایط، مدل باید دوباره آموزش داده شود (Verstraete et al., 1996). رویکردهای تجربی- آماری نیز به روش‌های پارامتریک و ناپارامتریک تقسیم‌بندی می‌شوند (Verrelst et al., 2018). در روش‌های پارامتریک که کاربرد گسترده‌ای داشته‌اند، غالباً بین شاخص‌های پوشش گیاهی^۱ و متغیر مورد نظر روابط رگرسیونی برقرار می‌شود و این روابط برای تخمین میزان پارامتر مورد نظر در پیکسل‌های نامعلوم به کار می‌روند (Clevers, 2014). فراوانی پارامترهای پوشش گیاهی و تأثیرپذیری بازتابندگی از متغیرهایی مانند زاویه کجی برگ‌ها، خاک پس‌زمینه و سایر متغیرهای محیطی می‌تواند پیچیدگی زیادی پدید آورد که، با پیش‌فرض‌های ساده، این روش‌ها قادر به کارکردن با آنها نباشند (Asner, 1998). با وجود سادگی و سرعت محاسباتی بالای این روش‌ها، استفاده از شاخص‌های پوشش گیاهی، منتج از ترکیب دو یا چند باند طیفی، همراه با مشکل اشباع‌شدگی^۲ (مثلاً به نسبت مقادیر بالای شاخص سطح برگ) ممکن است مشکل مهمی در استفاده از این مدل‌ها ایجاد کند (Darvishzadeh, 2008). علاوه بر این، روش‌های پارامتریک اغلب تخمینی از میزان عدم اطمینان نتایج در اختیار قرار نمی‌دهند.

گیاهان موتور محرک حیات در کره زمین‌اند که، با جذب دی‌اکسیدکربن و آب از طریق پدیده فتوسنتز، اکسیژن و مواد عالی تولید می‌کنند. اهمیت و نقش ویژه گیاهان در تعدیل و تنظیم جریان‌های ماده و انرژی از طریق چرخه‌های کربن، آب و نیتروژن، در مقیاس‌های محلی تا جهانی در مطالعات گوناگون، بررسی شده است (Menenti et al., 2003; Meroni, et al., 2004; Sellers et al., 1997). در اغلب مدل‌های اقلیمی، مدل‌های هیدرولوژیکی، مدل‌های پیش‌بینی رشد گیاه و همچنین در مدل‌سازی چرخه‌های ماده و انرژی، پارامترها و متغیرهای بیوفیزیکی و بیوشیمیایی گیاهان به‌منزله ورودی‌های مهم استفاده می‌شوند (Baret et al., 2007; Mousivand, 2015). از این‌رو، تعیین میزان و تغییرات پارامترهای گوناگون گیاهی اهمیت به‌سزایی در مدل‌سازی‌های متفاوت از قبیل محیطی، هواشناسی و اقلیم، اکولوژی، و کشاورزی دارد (Asner, 1998). یکی از مهم‌ترین متغیرهای گیاهی، با کاربردی وسیع در مدل‌های گوناگون بیوشیمیایی، محیط‌زیست و بهره‌وری گیاهان زراعی در مقیاس‌های متفاوت، شاخص سطح برگ (LAI)^۱ است که به‌صورت مساحت کل یک سمت برگ‌ها در واحد سطح زمین تعریف می‌شود (Watson, 1947). شاخص سطح برگ یک کمیت ساختاری گیاه است که تعیین می‌کند چه مساحتی از پوشش گیاهی برای تبادل آب، حرارت و گاز مابین اتمسفر و خاک در دسترس قرار دارد. پارامترهای پوشش گیاهی را می‌توان، با اندازه‌گیری زمینی، به‌دقت تعیین کرد. با این حال، روشن است که اندازه‌گیری زمینی اغلب به‌صورت تخریبی، زمان‌بر و پرهزینه است و محدودیت‌های مکانی و زمانی خاص خود را دارد (Weiss et al., 2004). با توجه به پرهزینه‌بودن و کارآیی‌نداشتن اندازه‌گیری‌های زمینی در مقیاس‌های وسیع و برای زمان‌های متوالی، سنجنش از دور گزینه‌ای مناسب و به‌عبارتی، تنها گزینه ممکن را، برای تخمین این پارامتر/ متغیرها، در اختیار قرار می‌دهد (Baret and Buis, 2008).

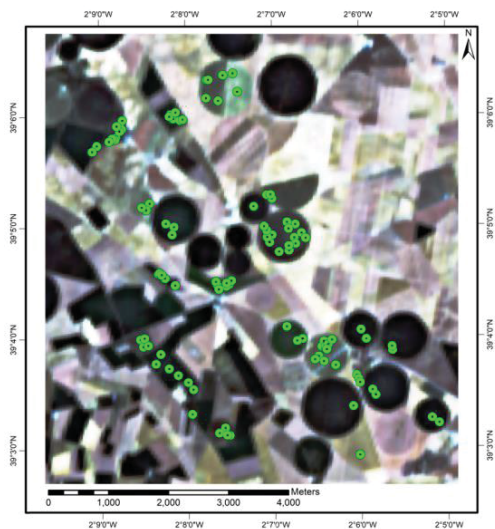
1. Leaf Area Index
2. Vegetation Indexes
3. Saturation

یادگیری ماشین برای بازیابی بوده‌اند. مورد دیگر تحلیل‌نشدن توانایی این مدل‌ها در بازیابی مقادیر حاشیه‌ای متغیر هدف است. اندازه‌گیری‌های زمینی، در اغلب موارد، محدود به تعدادی نمونه خاص‌اند که برآورد جامعی از پراکنش متغیر هدف روی زمین نمی‌دهند. در این‌گونه موارد، معمولاً سعی می‌شود بیشترین تمرکز نمونه‌برداری زمینی بر مقادیری باشد که بیشترین فراوانی در منطقه را دارند. مثلاً، در نمونه‌برداری از مقادیر LAI در منطقه‌ای کشاورزی، توجه چندانی به اندازه‌گیری مقادیر خیلی پایین و یا خیلی بالا نمی‌شود؛ درحالی‌که گاه مساحت وسیعی از منطقه دارای چنین مقادیری است و بازیابی آنها اهمیت فراوانی دارد. در چنین حالتی روشن است که، به‌دلیل آموزش روش‌های یادگیری ماشین براساس مقادیر معلوم (اندازه‌گیری‌های زمینی)، این روش‌ها معمولاً حساسیت بالایی به مقادیر حاشیه‌ای نشان می‌دهند و در اغلب موارد، عملکرد مناسبی در بازیابی پیکسل‌های با مقادیر خارج از دامنه آموزش داده‌شده نخواهند داشت. نظر به محدودیت اندازه‌گیری‌های زمینی در اغلب موارد و محدودبودن دامنه مقادیر اندازه‌گیری‌شده، این مورد اهمیت ویژه‌ای در بازیابی پارامترهای پوشش گیاهی دارد. با این حال و به‌رغم اهمیت این موضوع، توجه چندانی به آن نشده و اعتبارسنجی مدل، در مطالعات گوناگون، فقط براساس نمونه‌های اندازه‌گیری‌شده زمینی صورت گرفته و به توانایی مدل در تهیه نقشه مناسب از تمامی منطقه پرداخته نشده است. این تحقیق به‌منظور ارزیابی کارایی الگوریتم‌های یادگیری ماشین در بازیابی پارامترهای گیاهی از تصاویر ابرطیفی و بررسی پتانسیل این

از سوی دیگر، در روش‌های ناپارامتریک فرایند، یادگیری مدل به‌صورت مرحله‌ای و بدون پیش‌فرض در زمینه توزیع آماری نمونه‌های آموزشی، با بهینه‌سازی وزن‌هایی (ضرایبی) که اختلاف بین متغیر هدف و ورودی‌های مدل را کمینه می‌کند، انجام می‌شود (Camps-Valls, 2009). در این حالت، نیازی به پارامتری کردن فضای داده‌های ورودی نیست و ایجاد و بسط مدل، کمابیش، مسیر ساده‌ای را می‌پیماید. الگوریتم‌های ناپارامتریک با تبدیلات خطی^۱ (مانند رگرسیون خطی چندمتغیره)، به‌دلیل سرعت محاسباتی بالا و سادگی، بیشتر مورد توجه قرار گرفته‌اند؛ هرچند کارایی این الگوریتم‌ها در استفاده از داده‌های ابرطیفی، به‌دلیل همبستگی زیاد بین باندهای طیفی، محدود است (Verrelst et al., 2018). از این‌رو، الگوریتم‌های ناپارامتریک غیرخطی یادگیری ماشین^۲ کاربرد بیشتری در بازیابی پارامترهای بیوفیزیکی و بیوشیمیایی پوشش گیاهی از داده‌های ابرطیفی دارند (Verrelst et al., 2015). انعطاف‌پذیری بالا در مدل کردن روابط غیرخطی بین ضرایب بازتابندگی و متغیر هدف و امکان به‌کارگیری تمامی باندهای طیفی در فرایند یادگیری، بدون نیاز به استفاده از سایر تبدیلات مرسوم مانند مؤلفه‌های اصلی، از مزایای این روش‌هاست. با این حال، ممکن است انعطاف‌پذیری زیاد مدل موجب بیش‌برازشی^۳ شود؛ در این حالت، باید از تکنیک‌هایی مانند اعتبارسنجی متقاطع^۴ و تنظیم پارامترهای مدل استفاده کرد تا از پیچیدگی و انعطاف زیاد مدل جلوگیری شود (Verrelst et al., 2018).

مطالعات گوناگونی درباره بازیابی LAI، با استفاده از روش‌های یادگیری ماشین مانند رگرسیون بردار پشتیبان^۵ (Durbha et al., 2007)، فرایندهای گاوسی^۶، رگرسیون لبه‌کرنی^۷ (Verrelst et al., 2015)، شبکه‌های عصبی مصنوعی^۸ (Duveiller et al., 2011; Yi et al., 2011) و جنگل تصادفی^۹ (Yuan et al., 2017)، انجام شده است. اما این مطالعات اغلب با استفاده از تصاویر چندطیفی و نیز، محدود به استفاده از یک یا چند روش

1. linear transformation
2. Machine Learning Regression Algorithms (MLRA)
3. Overfitting
4. Cross-Validation
5. Support Vector Machine (SVR)
6. Gaussian Process Regression (GPR)
7. Kernel Ridge Regression (KRR)
8. Artificial Neural Network (ANN)
9. Random Forest (RF)



شکل ۱. تصویر ماهواره‌ای کریس- پروبا (ترکیب باندی ۸-۱۲-۲۵ RGB) از منطقه باراکس اسپانیا و پراکنش نمونه‌های برداشت‌شده زمینی LAI (نقاط سبز)

از منطقه مطالعاتی در مجموع، تعداد ۱۱۰ نمونه زمینی LAI با مقدار حداقل ۰.۳۱ و حداکثر ۵.۳۶، با میانگین ۲.۵۸ و انحراف معیار ۱.۵۸، با استفاده از LAI 2200C Plant Canopy Analyzer جمع‌آوری و پردازش شده است. این نمونه‌ها همراه با تصویر ابرطیفی کریس- پروبا در حالت نادیر ماهواره، برای بازیابی در این تحقیق به کار رفته‌اند. ماهواره کریس - پروبا اولین تصویربردار ابرطیفی چندزاویه‌ای است که سازمان فضایی اروپا، در ۲۲ اکتبر ۲۰۰۱، آن را در قالب مأموریتی تحقیقاتی به فضا پرتاب کرد. برداشت تصاویر در پنج زاویهٔ زنیث^۶ (اوج) متفاوت به این ماهواره امکان فوق‌العاده‌ای برای BRDF^۷ سطح زمین می‌دهد. مشخصات این ماهواره در جدول ۱ آورده شده است.

1. SPECTRA Barrax Campaigns (SPARC)
2. European space agency
3. Barrax
4. fractional vegetation cover
5. Chris-Proba
6. Zenith

الگوریتم‌ها در بازیابی مقادیر حاشیه‌ای متغیر LAI انجام شده است. از این‌رو، چهار الگوریتم شناخته‌شده یادگیری ماشین رگرسیون بردار پشتیبان، فرایند گاوسی، شبکه عصبی مصنوعی و جنگل تصادفی در مدل‌سازی متغیر LAI، با استفاده از تمامی باندهای تصویر ابرطیفی سنجنده کریس- پروبا، در منطقه باراکس اسپانیا به کار رفته‌اند. علاوه بر ارزیابی جامع روش‌های ذکر شده، توانایی پیش‌بینی هر یک از مدل‌ها برای مقادیر حاشیه‌ای و تهیه نقشهٔ پیوسته LAI بررسی شده است.

۲- داده‌ها و روش

۲-۱- داده‌ها و منطقه مورد مطالعه

داده‌های استفاده‌شده در این تحقیق مربوط به کمپین اسپارک^۱ سال ۲۰۰۴ است که سازمان فضایی اروپا^۲ در منطقه باراکس^۳ اسپانیا انجام داده است. منطقه باراکس یک مرکز کشاورزی در جنوب کشور اسپانیا و نزدیک شهر والنسیا است که، در سال‌های اخیر، میزبان چندین کمپین جمع‌آوری داده‌های زمینی برای کالیبراسیون و اعتبارسنجی مأموریت‌های ماهواره‌ای سازمان فضایی اروپا بوده است (Moreno et al., 2003). میانگین ارتفاع منطقه از سطح دریا ۷۰۰ متر است و تغییرات ارتفاعی بسیار ناچیزی (اختلاف ارتفاع ۲ متر بین مرتفع‌ترین و پست‌ترین نقطه) دارد. مسطح بودن زمین، همراه با وجود مزارع وسیع و تنوع زراعی، این منطقه را تبدیل به گزینه مناسبی برای مطالعات کالیبراسیون و اعتبارسنجی ماهواره‌ای کرده است. گیاهان گوناگونی شامل غلات و حبوبات (گندم، جو، نخود)، پیاز و سیب‌زمینی، چغندر قند، یونجه، ذرت، آفتابگردان به صورت دیم (۶۵٪ اراضی) و آبی (۳۵٪ اراضی) در منطقه کشت می‌شوند. در کمپین اسپارک ۲۰۰۴، چندین پارامتر پوشش گیاهی همچون کسر پوشش گیاهی^۴، کلروفیل، میزان آب برگ، میزان ماده خشک و LAI، هم‌زمان با گذر ماهواره کریس - پروبا^۵ گردآوری شد. شکل ۱ منطقه مطالعاتی و پراکنش نمونه‌های گردآوری‌شده را نشان می‌دهد.

جدول ۱. مشخصات ماهواره کریس- پروبا

ارتفاع سکو	۶۰۰ کیلومتر	زاویه دید	۵۵+ درجه روبه‌رونگر ^۱ ، ۵۵- درجه عقب‌نگر ^۱ ، ۳۶+ درجه روبه‌رونگر، ۳۶- درجه عقب‌نگر و نادیر ^۱
پهنای نوار (swath)	۱۴ کیلومتر	تعداد باند طیفی	۶۲ باند (۳۴ متر) ۱۹ باند (۱۸ متر)
قدرت تفکیک رادیومتری	۱۲ بیت	قدرت تفکیک طیفی	از ۴۰۰ تا ۱۰۵۰ نانومتر (پهنای باندها متغیر ۱.۳ تا ۱۲ نانومتر)
قدرت تفکیک مکانی	۱۸ متر ۳۴ متر	زمان گردش به‌دور زمین	خورشیدآهنگ، ۹۷ دقیقه

رگرسیون بردار پشتیبان است (Fawagreh et al., 2014; Mutanga et al., 2012). یکی از مهم‌ترین عوامل تأثیرگذار در مدل‌سازی جنگل تصادفی مشخص کردن پارامترهای مدل است. از جمله این موارد، می‌توان به تعداد درخت‌هایی که در فرایند جنگل تصادفی مشارکت می‌کنند و تعداد متغیرهایی که، به صورت تصادفی، در ایجاد یک گره برای هر درخت استفاده می‌شود اشاره کرد (Yu, et al., 2011).

۲-۲-۲- شبکه عصبی مصنوعی

شبکه عصبی مصنوعی یکی از رایج‌ترین روش‌های ناپارامتریک غیرخطی است که در مسائل رگرسیونی و طبقه‌بندی کاربرد متداولی دارد (Kumar et al., 2015; Verrelst et al., 2018). شبکه عصبی مصنوعی از سیستم‌های یادگیری طبیعی الهام گرفته شده است که در آنها مجموعه‌ای پیچیده از نورون‌های بهم‌متصل در کار یادگیری دخیل‌اند (LeCun et al., 1988; Rumelhart et al., 1986). شبکه‌های چندلایه از یک لایه ورودی، یک لایه خروجی و یک یا چند لایه بین آنها (لایه پنهان)^۶، که با ضرایبی به داده‌های ورودی و نتایج خروجی متصل‌اند، تشکیل یافته‌اند. انجام عملیات

۲-۲-۱- الگوریتم‌های بازبایی متغیرهای بیوفیزیکی پوشش گیاهی

۲-۱-۲- جنگل تصادفی

جنگل تصادفی الگوریتمی ناپارامتریک و جزء روش‌های یادگیری ترکیبی^۱ برای رگرسیون و طبقه‌بندی است که بریمن^۲ (۲۰۰۱) آن را توسعه داده است. برای شکل گرفتن جنگلی تصادفی، به مجموعه‌ای از درخت‌های تصمیم مستقل و مجزا نیاز است. در الگوریتم جنگل تصادفی، تعادل مناسبی بین بیاس و واریانس مدل شکل می‌گیرد و حفظ این تعادل سبب جلوگیری از بیش‌تخمینی و کم‌تخمینی^۳ در مدل می‌شود. در جنگل تصادفی، نمونه به صورت بگینگ^۴ انتخاب می‌شود. در این روش، زیرمجموعه‌هایی از مجموعه داده‌های اصلی، به صورت تصادفی و با جایگذاری، انتخاب و به هریک از درخت‌های تصمیم یک زیرمجموعه داده می‌شود. این بدان معنی است که احتمال دارد یک نمونه چندین بار در فرایند مدل‌سازی استفاده شود. هر درخت تصمیم‌گیری، تا رسیدن به شرط اتمام رشد، به تقسیم‌شدن ادامه می‌دهد (Breiman, 2001; Fawagreh et al., 2014). در جنگل تصادفی، داده‌های آموزشی دو قسمت می‌شوند و اغلب حدود ۶۳٪ از کل داده آموزشی، به شکل ورودی، به مدل وارد می‌شود و باقی‌مانده داده‌ها، به صورت نمونه‌های خارج از کیسه (OOB)^۵، برای آزمون مدل به کار می‌روند تا مدل دچار بیش‌برازشی نشود. عملکرد OOB در جنگل تصادفی مشابه ارزیابی متقاطع در

1. Ensemble
2. Breiman
3. Underfitting
4. Bootstrap Aggregating (bagging)
5. Out Of Bag
6. Hidden Layers

c مقدار تنظیم‌کننده شناخته می‌شود که مقدار ثابتی دارد. همچنین، ξ_i^+ , ξ_i^- قيود بالا و پایین خروجی سیستم را تعیین می‌کنند. رگرسیون بردار پشتیبان از رابطه‌ای خطی برای تخمین داده‌های جدید استفاده می‌کند اما، اگر رابطه خطی عملکرد مناسبی از خود نمایش ندهد، رگرسیون بردار پشتیبان بردارهای ورودی را به یک فضا با ابعاد بالاتر نگاشت می‌دهد. ولی نگاشت این داده‌ها به فضایی با ابعاد بالاتر مشکلاتی ایجاد می‌کند که، برای رفع آن، از یک تابع کرنل^۲ استفاده می‌شود. انتخاب تابع کرنل موضوعی مهمی است که باید، با توجه به نوع مسئله، آن را بررسی کرد (Erästä, 2001).

از مهم‌ترین و پرکاربردترین این کرنل‌ها می‌توان به کرنل خطی، درجه دوم، گاوسی و چندجمله‌ای اشاره کرد و معمولاً تابع کرنل گاوسی شعاعی، برای پیش‌بینی، عملکرد بهتری نشان می‌دهد (Cherkassky & Ma, 2004).

برای عملکرد مناسب الگوریتم رگرسیون بردار پشتیبان، پارامترهای مدل، از جمله مقادیر c و پارامتر کرنل مورد استفاده، باید بهینه شوند که معمولاً برای این بهینه‌سازی از ارزیابی مقاطع بهره گرفته می‌شود (باعث جلوگیری از بیش‌برازش می‌شود). از مزایای این روش، می‌توان به گیرنیفتادن در ماکزیمم و مینیمم محلی (در صورتی که در شبکه عصبی مصنوعی چنین نیست) و عملکرد مناسب در داده‌های دارای ابعاد بالا اشاره کرد.

۲-۲-۴- فرایند گاوسی

فرایند گاوسی یک روش توسعه‌یافته یادگیری ماشین است که از قوانین بیزین پیروی می‌کند. این روش در مسائل رگرسیون (Williams & Rasmussen, 1996)، طبقه‌بندی (Kuss & Rasmussen, 2005) و کاهش ابعاد به کار می‌رود (Lawrence, 2005). فرایند گاوسی را تونی اوهگن^۳ در آمار مطرح کرد (۱۹۷۸) که در

محاسباتی در شبکه عصبی در نوروها صورت می‌گیرد (Hagan et al., 1996). از مواردی که می‌تواند در عملکرد شبکه عصبی تأثیر به‌سزایی داشته باشد تعداد لایه پنهان شبکه عصبی، تعداد نوروها در لایه پنهان، تابع فعال‌سازی، مقداردهی اولیه وزن‌ها، تابع آموزش شبکه، نرخ یادگیری مدل، پارامترهای تعمیم‌پذیری برای جلوگیری از بیش‌برازش مدل هستند. این پارامترها باید، پیش از استفاده از روش شبکه عصبی به‌منظور مدل‌سازی، بهینه شوند.

۲-۲-۳- رگرسیون بردار پشتیبان

یکی از محبوب‌ترین الگوریتم‌های متداول، در چند سال اخیر، روش ماشین بردار پشتیبان بوده است. این روش ابتدا برای مسائل طبقه‌بندی استفاده شد اما بعدتر، به دلیل عملکرد خوب آن، برای حل مسائل رگرسیون نیز توسعه یافت (رگرسیون بردار پشتیبان) و از آن استقبال شد (Smola & Schölkopf, 1998; Vapnik, 2013).

رگرسیون بردار پشتیبان یک روش یادگیری نظارت‌شده است. این الگوریتم بر مبنای تئوری یادگیری آماری استوار است که از اصل کمینه‌سازی خطای ساختاری (SRM)^۱ بهره می‌جوید و منجر به یک جواب بهینه کلی می‌شود (Vapnik, 2013). در مدل رگرسیون بردار پشتیبان، از مجموعه توابع به فرم $f(x) = w^T \cdot x + b$ برای پیش‌بینی استفاده می‌شود که w وزن بردار x و b مقدار بایاس است (Smola & Schölkopf, 2004). تابع (معادله) هدف کلی، که مدل قصد کمینه‌کردن آن را دارد، عبارت است از:

$$\phi(w, \xi) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^l (\xi_i^- + \xi_i^+) \quad \text{رابطه (۱)}$$

$$\text{subject to} \begin{cases} y_i - (\langle w, x_i \rangle + b) \leq \varepsilon + \xi_i^- \\ (\langle w, x_i \rangle + b) - b \leq \varepsilon + \xi_i^+ \\ \xi_i^-, \xi_i^+ \geq 0 \end{cases} \quad \text{رابطه (۲)}$$

1. Structural Risk Minimization
2. Kernel function
3. O'Hagan

V ضریب مقیاس و B تعداد ویژگی را نشان می‌دهد که، در این تحقیق، منظور تعداد باندهای تصویر است. σ_b به کنترل روابط هر باند طیفی با دیگر باندها اختصاص داده شده است (Williams & Rasmussen, 2006).^۲

۲-۳- اجرا و تنظیم پارامترهای مدل

در این تحقیق، از چهار روش جنگل تصادفی، شبکه عصبی، رگرسیون بردار پشتیبان و فرایند گاوسی برای بازیابی LAI، با استفاده از تصویر ابرطیفی کریس- پروبا با زاویه اوج نادیر، استفاده شد. بنا به توصیه‌های صورت گرفته در مطالعات قبلی، سه باند اول تصویر کریس - پروبا، به دلیل وجود نویز بالا، از این تحقیق کنار گذاشته شد و ۵۹ باند طیفی باقی‌مانده (Mousivand, 2015)، با قدرت تفکیک مکانی ۳۶ متر برای بازیابی LAI، به کار رفتند.

در این تحقیق، به منظور اجرای الگوریتم جنگل تصادفی، تعداد بهینه برای پارامترهای تأثیرگذار همچون تعداد درخت، تعداد برگ و متغیرهای تصادفی هر گره ($mtry$) به صورت تجربی انتخاب شد. تعداد درخت (پارامتر $ntree$)، از ۱۰۰ تا ۱۰۰۰ درخت، و تعداد برگ، از ۳ تا ۱۰ برگ، مورد آزمون قرار گرفت. در نهایت، برای ریشه میانگین مربعات خطا (RMSE) تعداد بهینه درخت ۳۰۰ و تعداد بهینه برگ ۵ به دست آمد و برای مدل‌سازی استفاده شد. تعداد متغیرهای تصادفی به کاررفته در هر گره درخت ($mtry$) نیز ۲۰ در نظر گرفته شد.

فرایند یادگیری در شبکه عصبی شامل تعدیل ضرایب با کمینه‌سازی تابع هزینه، با استفاده از الگوریتم پس‌انتشار است (Atkinson & Tatnall, 1997). برای انجام دادن تحقیق حاضر،

جامعه علمی زمین آمار با نام کریجینگ^۱ شناخته می‌شود. با این حال، به علت پیچیدگی محاسباتی و نبود ابزار مناسب، تا قبل از قرن ۲۱ از آن استقبال نشد (Williams & Rasmussen, 2006). فرایند گاوسی جزء خانواده کرنل- مینا دسته‌بندی می‌شود و مجموعه‌ای از متغیرهای تصادفی را، با توزیع گاوسی چندمتغیره توأمان، در نظر می‌گیرد.

این مدل بین ورودی $X \in R^B$ (تعداد باندهای تصویر) و خروجی $y \in R$ (میزان LAI) رابطه‌ای شبیه به مدل رگرسیون بردار پشتیبان عمل می‌کند. (Camps-Valls et al., 2016). با این حال، این روش دو مزیت مهم نسبت به روش رگرسیون بردار پشتیبان دارد. اول اینکه در این روش، علاوه بر میانگین متغیر پیش‌بینی شده، مدل همچنین تخمینی از میزان واریانس آن بدست می‌دهد. باید توجه داشت که، برای پیش‌بینی میانگین، از ترکیب خطی بین ورودی و خروجی بهره برده می‌شود؛ در حالی که پیش‌بینی واریانس فقط براساس داده‌های ورودی است و تفاوت در نتایج به دست آمده را می‌توان به تفاوت در نوع کرنل و اطلاعات ورودی به تابع تقریب مرتبط دانست. دومین مزیت این است که می‌توان از توابع کرنل بسیار کارآمد و مناسبی در مدل فرایند گاوسی استفاده کرد که همه ابرپارامترهای مورد استفاده در کرنل، در مراحل یادگیری مدل، با به حداکثر رساندن احتمال حاشیه در مجموعه داده‌های آموزشی بهینه می‌شوند (Camps-Valls et al., 2009; Verrelst et al., 2012). در این روش نیز، همانند رگرسیون بردار پشتیبان، می‌توان از توابع کرنل متفاوتی استفاده کرد. یکی از پرکاربردترین کرنل‌ها گاوسی - شعاعی است که عبارت است از:

$$K(x_i, x_j) = V \exp \left(- \sum_{b=1}^B \frac{(x_i^{(b)} - x_j^{(b)})^2}{2\sigma_b^2} \right) \quad \text{رابطه (۳)}$$

1. Kriging

۲. برای به دست آوردن جزئیات بیشتر درباره الگوریتم فرایند گاوسی، نک. Williams & Rasmussen, 2006 و برای اجرای

این الگوریتم نیز، رک، پایگاه داده

<http://www.gaussianprocess.org/gpml/>

در اجرای فرایند گاوسی، همانند الگوریتم رگرسیون بردار پشتیبان، اعتبارسنجی متقاطع ۱۰ فولدی به کار رفت. در این تحقیق کرنل گاوسی، که بهترین عملکرد را در بین سایر کرنل‌ها داشت، به‌منزله تابع کواریانس انتخاب شد. مقادیر بهینه V و σ_b نیز مشخص و استفاده شد.

۲-۴- ارزیابی روش‌ها

روش‌ها برای ارزیابی LAI با به‌کارگیری شاخص‌های آماری ریشه میانگین مربعات خطا، ضریب تعیین (R^2) و ریشه میانگین مربعات خطای نرمال شده (NRMSE)^۷ ارزیابی شدند. فرمول‌های مربوط به RMSE و R^2 در ادامه توضیح داده شده است.

$$RMSE = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n}} \quad \text{رابطه (۴)}$$

$$R^2 = \frac{\sum_i^n (\hat{y}_i - \bar{y})^2}{\sum_i^n (y_i - \bar{y})^2} \quad \text{رابطه (۵)}$$

در معادلات بالا، \hat{y}_i و y_i نشان‌دهنده LAI برداشت‌شده زمینی، میانگین LAI برداشت‌شده زمینی و مقدار LAI تخمین‌زده شده از طریق مدل‌اند. NRMSE از تقسیم RMSE بر دامنه تغییرات داده‌های اندازه‌گیری‌شده زمینی به دست می‌آید. برای محاسبه این شاخص‌ها، از اعتبارسنجی متقاطع ۱۰ فولدی استفاده شد که نمونه‌های هر فولد به صورت تصادفی انتخاب شدند. در فرایندی تکرارکننده، مدل با به‌کارگیری ۹ فولد کالیبره شد و یک فولد باقی‌مانده برای اعتبارسنجی به کار رفت. به منظور ارزیابی درستی تخمین روش‌های متفاوت میانگین RMSE، NRMSE و R^2 محاسبه شده برای ۱۰ فولد اعتبارسنجی

الگوریتم‌های (GSC) Gradient Scaled Conjugate، (LM) Gradient Descent، (Levenberg-Marquardt)، (GD) و (GDA) Gradient Descent Adaptive در نرم‌افزار متلب آزمایش شدند (اما، به دلیل اختصار مقاله، همه نتایج به دست آمده نشان داده نشده است). ترکیب توابع تانژانت سیگموئید^۱ و تبدیلات خطی قابلیت بالایی در برآزش^۲ انواع گوناگون توابع دارد (از این رو، لایه‌های پنهان از نورون‌های تانژانت سیگموئید هایدربولیک^۳ و لایه خروجی از نورون منفرد خطی^۴ ساخته شد. ریشه میانگین مربعات خطا^۵ بین متغیر هدف و خروجی شبکه، به صورت تابع هزینه، به کار رفت و تعداد لایه‌های پنهان و نورون‌های هر لایه، به صورت تجربی و با دستیابی به کمترین خطا، تعیین شد. به منظور حصول نتایج پایدار برای مدل، از اعتبارسنجی متقاطع ۱۰ فولدی استفاده شد و برای هر فولد، ۵۰ تکرار صورت گرفت. در هر تکرار، نمونه‌های آموزشی و ارزیابی و تست با نسبت‌های ۷۰، ۱۵، ۱۵٪ و به حالت تصادفی انتخاب شدند. میانگین مقادیر تخمینی تکرارها به‌منزله رقم نهایی تخمین تعیین شد و به تبع آن، میانگین معیارهای ارزیابی تکرارها نتیجه آن فولد در نظر گرفته شد. این روند در مورد همه فولدها انجام شد. به‌کارگیری تابع گرادیان نزولی با نرخ یادگیری تطبیقی پس انتشار (GDA)^۶ و تنظیم شبکه به صورت لایه‌های پنهان و چهار نورون کمترین خطا را به همراه داشت. نتایج به دست آمده در بخش بعدی این تحقیق آورده شده است.

در الگوریتم رگرسیون بردار پشتیبان پارامترهای مدل، از جمله مقادیر c و پارامتر کرنل، با استفاده از اعتبارسنجی متقاطع ۱۰ فولدی کالیبره و ضرایب مدل تعیین شد. کرنل گاوسی نیز، به‌منزله نوع کرنل به کاررفته در این مدل، مشخص شد. شایان ذکر است که، در تمامی موارد، تست‌های گوناگونی برای حصول اطمینان از عملکرد پارامترهای انتخابی در روش‌های متفاوت انجام شد و در نهایت، پارامترهای بهینه برای هر مدل تعیین شد.

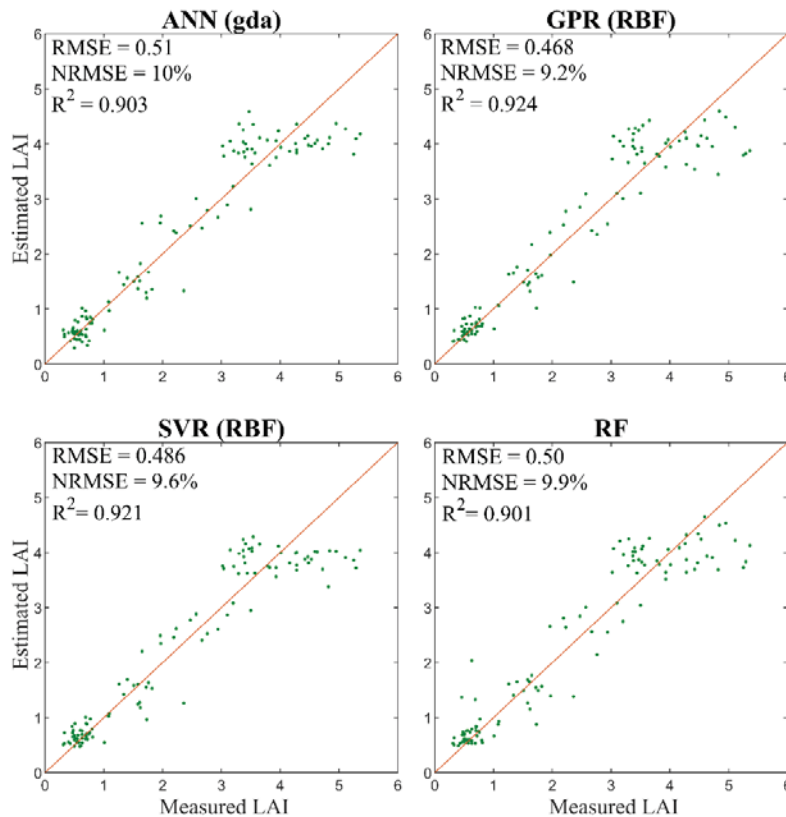
1. Tangent Sigmoid
2. Fitting
3. Hyperbolic
4. Single Linear
5. Root Mean Square Error (RMSE)
6. Gradient Descent with Adaptive Learning Rate Backpropagation (GDA)
7. Normalized Root Mean Square Error

شکل ۲ مقایسه نتایج هر روش با نمونه‌های اندازه‌گیری شده زمینی LAI را نمایش می‌دهد. در نگاه اول، هر چهار روش عملکردی مناسب و نزدیک به هم در بازیابی LAI داشته‌اند و مقادیر RMSE نزدیک به ۰.۵، ضریب تعیین بالاتر از ۰.۹۰ و NRMSE در حدود ۰.۱۰ به دست آمده است. باین حال، عملکرد روش‌های فرایند گاوسی و رگرسیون بردار پشتیبان اندکی بهتر از دو روش دیگر بوده است؛ در مجموع، بهترین عملکرد متعلق به روش فرایند گاوسی با RMSE نزدیک به ۰.۴۷ و ضریب تعیین ۰.۹۲ و NRMSE در حدود ۰.۰۹ است. دلیل عملکرد مناسب تمامی روش‌ها در این تحقیق را شاید بتوان با توانایی مدل‌سازی مطلوب این شیوه‌ها، پراکندگی مناسب نمونه‌های اندازه‌گیری شده و نیز، انتخاب و تنظیم پارامترهای مناسب در هر شیوه و بهینه‌سازی فرایند یادگیری روش‌ها توضیح داد.

مقاطع محاسبه شد. اجرای ۱۰ فولدی اعتبارسنجی متقاطع، با توجه به تعداد نمونه‌ها و در نظر گرفتن مصالحه‌ای بین هزینه محاسباتی و صحت مورد انتظار، صورت گرفت تا تقریب مناسبی از نتایج به دست آید.

۳- نتایج

در این بخش، نتایج مدل‌سازی چهار روش به کاررفته در این تحقیق بیان می‌شود. ابتدا، صحت نتایج روش‌های شبکه عصبی مصنوعی، فرایند گاوسی، رگرسیون بردار پشتیبان و جنگل تصادفی در بازیابی LAI آمده و عملکرد هر یک از روش‌ها در دامنه‌های متفاوت LAI بررسی شده است. سپس، برای نشان دادن پایداری و دقت و صحت هر روش، نقشه انحراف معیار نتایج و نیز، نقشه LAI خروجی با NDVI و تصویر رنگی کاذب منطقه مقایسه شده است.

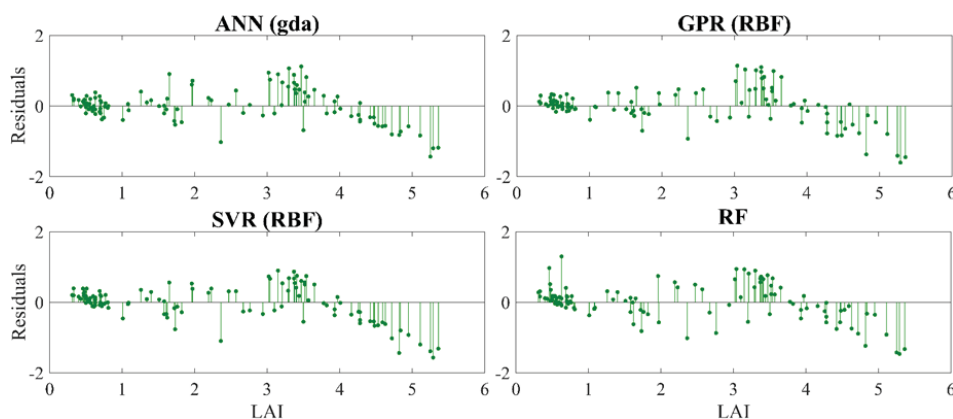


شکل ۲. نمودارهای مقایسه‌ای LAI های اندازه‌گیری شده زمینی و LAI های تخمینی، با استفاده از روش‌های یادگیری ماشین

برای بازیابی پارامترهای گیاهی (بهتر از ۰.۱۰٪) از سوی کمیسیون پایش جهانی برای امنیت و محیط (GMES) ۳ مطابق دارد (Mousivand, 2015). از آنجا که مقدار RMSE حساسیت بالایی به مقادیر LAI دارد، NRMSE امکان بهتری، برای مقایسه با سایر مدل‌ها و سایر مطالعات انجام‌شده، در اختیار قرار می‌دهد. میزان NRMSE در حدود ۱۰٪ به‌منزلهٔ مقداری قابل قبول برای ارزیابی و اعتبارسنجی مدل‌های بازیابی مورد استناد قرار می‌گیرد.

نقشهٔ LAI منطقه با میانگین‌گیری مقادیر LAI برای هر پیکسل در ۱۰ فولد اعتبارسنجی متقاطع محاسبه شد و به همین ترتیب، انحراف معیار هر پیکسل در این فولدها، به‌منزلهٔ معیاری برای عدم اطمینان LAI خروجی، در نظر گرفته شد. نقشهٔ LAI تولیدشده با استفاده از چهار روش همراه با تصویر ماهواره‌ای کریس- پروبا منطقه و شاخص NDVI در شکل ۴ نشان داده شده است و شکل ۵ نیز نقشهٔ انحراف معیار هر روش را به‌نمایش می‌گذارد. در اینجا، شاخص NDVI به‌صورت عامل تشخیصی برای اطلاع از عملکرد هر روش در خارج از محدودهٔ مقادیر اندازه‌گیری‌شدهٔ زمینی (مانند مناطق دارای پوشش گیاهی تنک و بدون پوشش گیاهی) استفاده شده است.

شکل ۳ میزان انحراف مقادیر LAI بازیابی‌شده از مقادیر اندازه‌گیری‌شدهٔ زمینی را به‌نسبت خط یک‌به‌یک نشان می‌دهد. اعداد منفی و مثبت، به‌ترتیب، بیانگر کم‌برازش^۱ و بیش‌برازش^۲ مدل در مقایسه با مقادیر واقعی‌اند. بررسی شکل ۳ مشخص می‌کند که، با چشم‌پوشی از تفاوت‌های اندک، تمامی روش‌ها الگوی تقریباً یکسانی را، برای بازیابی مقادیر متفاوت LAI، نشان می‌دهند؛ بدین‌صورت که عملکرد تمامی روش‌ها در مقادیر پایین LAI (مقادیر کمتر از ۲) بهتر است و زمانی که LAI به‌سمت مقادیر بزرگ میل می‌کند، صحت تخمین تمامی روش‌ها کاهش می‌یابد. هر چهار روش در مقادیر LAI بالاتر از ۴ دچار کم‌برازش می‌شوند؛ درحالی‌که مقادیر بین ۲ تا ۳ با بیش‌برازشی این روش‌ها همراه بوده است. بیش‌برازشی مدل‌های متفاوت در مطالعات پیشین هم گزارش شده است (Mousivand, 2015; Verrelst, 2012) و ممکن است ناشی از تغییرات اندک طیفی در مقادیر بالای LAI، عدم اطمینان اندازه‌گیری‌های زمینی در مقادیر بالا و نیز، اشباع‌شدگی مدل باشد (Mousivand et al., 2014). نکتهٔ جالب دیگر میزان خطای نسبی بهتر از ۱۰٪ برای همهٔ روش‌هاست که با عدم اطمینان مطلوب تعیین شده

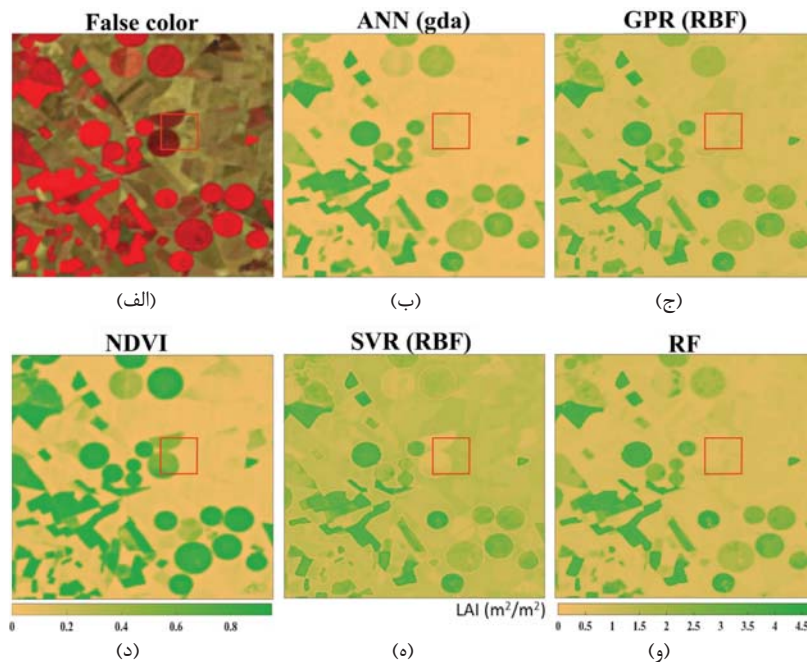


شکل ۳. میزان انحراف مقادیر LAI بازیابی‌شده از مقادیر اندازه‌گیری‌شدهٔ زمینی به‌نسبت خط ۱-۱

1. Understimation
2. Overestimation
3. Global Monitoring for Environment and Security

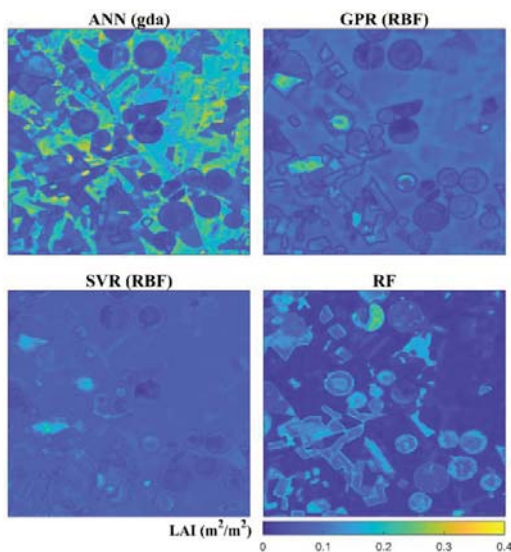
با بررسی دقیق شاخص NDVI همراه با تصویر ماهواره‌ای و همچنین، نقشه کاربری و پوشش اراضی منطقه، مشخص شد که مقادیر کمتر از ۰.۱، در واقع، خاک لخت بدون پوشش گیاهی‌اند و مقادیر بیشتر از ۰.۱ پوشش گیاهی تنک تا متراکم را مشخص می‌کنند. برای نمونه، مربع قرمز در شکل ۴ قسمتی از یک مزرعه آفتابگردان، در اوایل دوران رشد، و خاک لخت در مجاورت این مزرعه را مشخص می‌کند. پراکنش اندازه‌گیری‌های زمینی در این مزرعه، با میانگین مقادیر LAI در حدود ۰.۴، در شکل ۲ نشان داده شده است. با نگاهی به عملکرد مدل‌های متفاوت از نظر انحراف معیار حاصل‌شده و نقشه LAI برای محدوده مشخص‌شده در تصویر و نیز در کل منطقه، مشخص می‌شود که روش رگرسیون بردار پشتیبان، در تخمین مقادیر پایین LAI (کمتر از ۰.۳)، عملکرد ضعیفی داشته و این مدل نتوانسته است، در جداسازی خاک لخت از پوشش گیاهی تنک، موفق عمل کند. دو روش فرایند گاوسی و جنگل تصادفی عملکرد بهتری در این

به دلیل فقدان مقادیر اندازه‌گیری‌شده زمینی برای پیکسل‌های با چنین شرایطی، از این شاخص برای ارزیابی عملکرد مدل‌های متفاوت در این مناطق بهره گرفته شده است. نویسندگان این مقاله از مشکلات شاخص‌های پوشش گیاهی، همچون NDVI، آگاه‌اند و این شاخص، در اینجا، با احتیاط و صرفاً برای مقایسه استفاده شده است. شاخص NDVI رابطه‌ای لگاریتمی با LAI دارد و با افزایش مقادیر LAI، این شاخص به آستانه اشباع‌شدگی می‌رسد. با توجه به اینکه بین مقادیر پایین LAI و NDVI رابطه‌ای نسبتاً خطی وجود دارد، استفاده از این شاخص، برای سنجش توانایی مدل‌ها در چنین شرایطی، توجیه‌پذیر است. رابطه بین LAI و NDVI در شکل ۴ مشاهده می‌شود. علاوه‌بر این، نقشه عدم اطمینان (شکل ۵) می‌تواند معیاری برای دقت برآزش مدل در تهیه نقشه LAI، در تراکم‌های متفاوت پوشش گیاهی برای پیکسل‌های نامعلوم و نیز پایداری مدل در تخمین این مقادیر، به‌شمار رود.



شکل ۴. تصویر رنگ کاذب (الف)؛ نقشه پیوسته LAI شبکه عصبی مصنوعی (ب)؛ نقشه پیوسته LAI فرایند گاوسی (ج)؛ نقشه NDVI محاسبه‌شده از تصویر کریس- پروبا (د)؛ نقشه پیوسته LAI رگرسیون بردار پشتیبان (ه)؛ نقشه پیوسته LAI جنگل تصادفی (و)

روش‌های کرنل-مبنا از توزیع نمونه‌ها باشد که مقادیر تخمینی را به بازه داده‌های ورودی محدود می‌کنند (Rivera-Caicedo et al., 2017).



شکل ۵. نقشه انحراف معیار مربوط به اعتبارسنجی مقاطع ۱۰ فولدی

۴- بحث و جمع‌بندی

الگوریتم‌های یادگیری ماشین برای رگرسیون، در زمینه بازبایی پارامترهای بیوفیزیکی و بیوشیمیایی گیاهان از داده‌های سنجنش از دوری، در زمره روش‌های کارآمد ناپارامتریک غیرخطی در رویکرد آماری به‌شمار می‌روند. سرعت و دقت مناسب روش‌های یادگیری ماشین برای رگرسیون سبب کاربرد روزافزون آنها در بازبایی متغیرهای پوشش گیاهی از داده‌های سنجنش از دوری شده است. با این حال، هنوز مطالعه جامعی درباره عملکرد بهتر یک الگوریتم، در مقایسه با سایر الگوریتم‌ها، صورت نگرفته است. در این تحقیق، چهار روش شبکه عصبی مصنوعی، فرایند گاوسی، رگرسیون بردار پشتیبان و جنگل تصادفی برای بازبایی LAI در منطقه باراکس اسپانیا به کار رفت که همراه با استفاده از تصویر ماهواره کریس - پروبا و داده‌های زمینی گردآوری شده در کمپین SPARC2004 بود. برای به‌دست آوردن خروجی نهایی LAI از اعتبارسنجی

زمینه داشته‌اند؛ با این حال، به نظر می‌رسد که روش شبکه عصبی مصنوعی بالاترین تناسب را با نقشه NDVI دارد؛ به‌ویژه در مورد مناطق دارای پوشش گیاهی تنک و یا بدون پوشش گیاهی. از سوی دیگر، روش شبکه عصبی مصنوعی انحراف معیار بالایی را در قسمت‌های بدون پوشش گیاهی نشان می‌دهد که دقت پایین این روش در تکرارهای متفاوت را می‌رساند. میانگین LAI‌های به‌دست آمده به واقعیت زمینی نزدیک‌تر است که صحت بیشتر این روش را نشان می‌دهد. به عبارت دیگر، با در نظر گرفتن نقشه عدم اطمینان، اگرچه دقت حاصل در مورد مناطق بدون پوشش گیاهی، با استفاده از روش شبکه عصبی مصنوعی، کمتر به نظر می‌رسد؛ مقایسه با نقشه NDVI و تصویر رنگ کاذب نشان می‌دهد نتایج به‌دست آمده با این روش صحت بیشتری دارد.

با مقایسه شکل‌های ۴ و ۵، مشخص می‌شود که روش جنگل تصادفی دقت و صحت نسبتاً بالایی در بازبایی مقادیر LAI برای مناطق با پوشش گیاهی تنک و بدون پوشش گیاهی داشته؛ در حالی که دقت این روش، در مورد مناطق دارای پوشش گیاهی متوسط و متراکم، کمتر از روش‌های دیگر بوده است. روش رگرسیون بردار پشتیبان، در بازبایی مناطق بدون پوشش و یا با پوشش گیاهی تنک، دچار کم‌برازشی شده و همان‌گونه که اشاره کردیم، دقت و صحتی کمتر از سایر روش‌ها داشته است. هرچند این روش، در بازبایی مناطق دارای پوشش گیاهی متراکم و متوسط، عملکردی تقریباً مشابه با سایر روش‌ها داشته است. صحت بازبایی در روش فرایند گاوسی، در مورد نواحی بدون پوشش و یا دارای پوشش گیاهی تنک، مشابه روش جنگل تصادفی بوده اما، با توجه به شکل ۵، می‌توان دریافت که دقت آن به مراتب کمتر از روش جنگل تصادفی بوده است. این در حالی است که صحت و دقت روش فرایند گاوسی، در مورد نواحی با پوشش گیاهی متراکم، شبیه به دیگر روش‌ها بوده است. این نکته ممکن است ناشی از توزیع نامناسب دامنه نمونه‌های زمینی در مزارع تنک (با مقادیر LAI در بازه ۰ و ۱) و تأثیرپذیری

نتایج را بهبود بخشد (Verrelst et al., 2016).
 باین حال، بازیابی از طریق مجموعه‌ای باندی تا کنون
 مبحثی باز در بین محققان بوده و روش مناسبی برای
 آن پیشنهاد نشده است.

همان‌طور که در بخش ۲-۲-۱ توضیح داده شد،
 توسعه الگوریتم جنگل تصادفی روش درخت‌های
 تصمیم‌گیری را بهبود بخشیده و مشکل این روش‌ها را،
 در مواجهه با هم‌خطی بین باندهای تصاویر ابرطیفی،
 حل کرده است (Rivera-Caicedo et al., 2017).
 جنگل تصادفی، از راه ترکیب^۳ درخت‌های تصمیم
 مجزا^۴ با زیرمجموعه‌های متفاوت از ویژگی‌ها (باندها)
 که تصادفی انتخاب می‌شوند، طرح‌های ترکیبی ویژه‌ای^۵
 را در پیش‌بینی نهایی به کار می‌برد. قابلیت انتخاب
 ویژگی‌های مؤثر در این الگوریتم سبب می‌شود که
 جنگل تصادفی، در داده‌های دارای متغیرهای مستقل
 زیاد همچون داده‌های ابرطیفی، عملکرد مناسب‌تری
 داشته باشد و به نوبت حساسیت کمتری نشان بدهد و
 بدین ترتیب، از مشکل بیش‌برازشی جلوگیری شود
 (Belgiu & Drăguț, 2016). به کارگیری زیرمجموعه‌ای
 از ویژگی‌های مؤثر در الگوریتم جنگل تصادفی این
 روش را در تحلیل داده‌های طیف‌سنجی توانا تر کرده
 است. همان‌طور که در نقشه عدم اطمینان (انحراف معیار)
 شکل ۵ مشاهده می‌شود، این الگوریتم، در تخمین
 مقادیر خارج از دامنه نمونه‌ها، پایدارتر از دیگر روش‌ها
 بوده و به‌خصوص برای تخمین مقادیر پایین LAI و
 خاک، بهتر از فرایند گاوسی و رگرسیون بردار پشتیبان
 عمل کرده است (شکل ۴). باین حال، توانایی روش شبکه
 عصبی مصنوعی در برقراری روابط پیچیده بین
 متغیرهای مستقل (در این تحقیق، باندها) و همچنین،

مقاطع ۱۰ فولدی استفاده شد تا از بیش‌برازشی مدل‌ها
 جلوگیری شود. میانگین ارقام تخمینی نیز به‌منزله
 تخمین نهایی محاسبه شد. این کار، علاوه‌بر
 اعتبارسنجی هر فولد از داده‌ها، این امکان را فراهم آورد
 که نقشه عدم اطمینان برای کل تصویر مورد نظر
 محاسبه شود. نظر به محدودیت نمونه‌های
 اندازه‌گیری شده زمینی در هر تحقیق، کسب اطلاع از
 میزان دقت و عدم اطمینان خروجی اهمیت ویژه‌ای در
 استفاده و انتخاب مدل مناسب برای بازیابی پارامتر
 دارد. نتایج این تحقیق نشان می‌دهد که هر چهار روش
 یادگیری ماشین توانایی مناسبی برای تخمین LAI
 منطقه دارند و صحت نتایج، در مقایسه با نمونه‌های
 اندازه‌گیری شده زمینی، این مورد را تأیید می‌کند.
 باین همه، نتایج نشان می‌دهد که روش فرایند گاوسی
 اندکی بهتر از سه روش دیگر در بازیابی نمونه‌های LAI
 بوده است. به توانایی ویژه روش فرایند گاوسی، در
 مطالعات دیگر در زمینه ویژگی‌های پوشش گیاهی،
 اشاره شده است (Verrelst et al., 2015). در تحقیق
 حاضر، به‌قصد بررسی صحت پیش‌بینی هر یک از
 روش‌ها برای سایر پیکسل‌ها (خارج از دامنه نمونه‌ها) و
 تطابق مقادیر تخمینی با پوشش منطقه، نقشه‌های
 خروجی LAI با نقشه NDVI و تصویر رنگی کاذب
 مقایسه شد. با مقایسه نقشه‌های تهیه شده LAI و نقشه
 NDVI، مشاهده می‌شود که روش‌های کرنل - مبنای
 فرایند گاوسی و رگرسیون بردار پشتیبان، به‌رغم اینکه
 صحت بالایی در تخمین نمونه‌های زمینی دارند و در
 نقشه انحراف معیار دقت بیشتری را در مورد پیکسل‌های
 خاک نشان می‌دهند، در برآورد پیکسل‌های با مقادیر
 پایین تر از دامنه داده‌های زمینی، عملکردی ضعیف‌تر از
 دیگر روش‌ها دارند. مطالعات پیشین حاکی از این است
 که روش‌های فرایند گاوسی و رگرسیون بردار پشتیبان
 در تحلیل داده‌های طیف‌سنجی و تصاویر ابرطیفی،
 به دلیل هم‌خطی^۱ شدید بین ویژگی‌ها (باندها)، با
 مشکل مواجه می‌شوند. ترکیب روش‌های کاهش ابعاد^۲
 و نیز، انتخاب باندها در کاربرد این روش‌ها ممکن است

1. Collinearity
2. Dimension Reduction
3. Ensemble
4. Individual Decision Trees
5. Particular Combination Schemes

می‌تواند، در شرایطی که امکان برداشت نمونه زمینی چندانی وجود ندارد، به کار رود. علاوه بر این، نتایج تحقیق دیگر بار توانایی مدل‌های یادگیری ماشین را، در بازیابی متغیرهای گیاهی، نشان داد. پیشنهاد می‌شود که، در مطالعات آتی، کارایی الگوریتم‌های یادگیری ماشین در بازیابی متغیرهای گیاهی، بدون استفاده از نمونه‌های زمینی و صرفاً با استفاده از داده‌های شبیه‌سازی شده مستخرج از مدل‌های انتقال تابش، ارزیابی شود. همچنین، پیشنهاد می‌شود کاربرد روش‌های انتخاب باندهای بهینه و ترکیب آنها با الگوریتم‌های یادگیری ماشین، به منظور ارتقای عملکرد این الگوریتم‌ها، سنجیده شود.

۵- منابع

- Asner, G.P., 1998, **Biophysical and Biochemical Sources of Variability in Canopy Reflectance**, Remote Sensing of Environment, 64(3), PP. 234-253.
- Atkinson, P.M. & Tatnall, A.R., 1997, **Introduction Neural Networks in Remote Sensing**, International Journal of Remote Sensing, 18(4), PP. 699-709 .
- Baptista, F.D., Rodrigues, S. & Morgado-Dias, F., 2013, **Performance Comparison of ANN Training Algorithms for Classification**, Paper presented at the 2013 IEEE 8th International Symposium on Intelligent Signal Processing.
- Baret, F. & Buis, S., 2008, **Estimating Canopy Characteristics from Remote Sensing Observations: Review of Methods and Associated Problems**, In Advances in Land Remote Sensing (PP. 173-201), Springer.

قدرت درونیابی و برون‌یابی^۱ (Duveiller et al., 2011) موجب شده است که، در تخمین مقادیر خارج از دامنه و جداسازی پیکسل‌های خاک از پوشش گیاهی، عملکرد مناسبی داشته باشد.

قدرت انطباق‌پذیری^۲ شبکه عصبی مصنوعی کارایی این روش را در تحلیل انواع داده‌ها تضمین می‌کند و در زمینه تخمین ویژگی‌های گیاهی^۳ نیز گسترش فزاینده‌ای پیدا کرده است (Verrelst et al., 2018). معماری شبکه عصبی در میزان پایداری مدل، مدت زمان محاسباتی و دقت مدل‌سازی بسیار تأثیرگذار است. تعیین توابع لایه‌های ورودی و خروجی و تعداد نورون‌های لایه پنهان و همچنین، نوع تابع آموزش ارکان اصلی معماری شبکه را تشکیل می‌دهد؛ به طوری که توابع آموزش متفاوت، در شرایط یکسان، ممکن است نتایج متفاوتی به دست دهد (Baptista et al., 2013). در این تحقیق استفاده از تابع آموزش GDA، در مقایسه با سایر توابع، در بهبود صحت نتایج تأثیر داشته است. استفاده از راهبرد اعتبارسنجی متقاطع و تکرار مدل‌سازی با انتخاب‌های تصادفی نمونه‌های آموزش (رک، بخش ۲-۳) نیز در پایداری مدل تأثیر به‌سزایی داشته است. دستاوردهای تحقیقات پیشین نیز نتایج این مقاله را تأیید می‌کند. برای نمونه، در سال‌های اخیر، توانایی روش‌های شبکه عصبی مصنوعی، در مقایسه با روش‌های پارامتریک شاخص - مینا و ناپارامتریک، در مدل‌سازی و پیش‌بینی متغیرهای پوشش گیاهی، در مقالات گوناگونی نشان داده شده است (Wang et al., 2013). برخی مطالعات نیز، در تخمین ویژگی‌های پوشش گیاهی با تصاویر ابرطیفی، به برتری نسخه‌های پیشرفته شبکه عصبی مصنوعی بر سایر الگوریتم‌های یادگیری ماشین اشاره کرده‌اند (Pham et al., 2017). نتایج این تحقیق نیز تأکید می‌کند که روش شبکه عصبی مصنوعی کارایی مناسب‌تری برای بازیابی متغیرهای پوشش گیاهی، در مجموع، بر روی دامنه نمونه‌های اندازه‌گیری شده و همچنین، مقادیر حاشیه‌ای دارد. نتایج این تحقیق

1. Extrapolation
2. Adaptability
3. Vegetation Properties Mapping

- Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., . . . & Samain, O., 2007, **LAI, fAPAR and fCover CYCLOPES global Products Derived from VEGETATION: Part 1: Principles of the Algorithm**, Remote Sensing of Environment, 110(3), PP. 275-286.
- Belgiu, M. & Drăguț, L., 2016, **Random Forest in Remote Sensing: A Review of Applications and Future Directions**, ISPRS Journal of Photogrammetry and Remote Sensing, 114, PP. 24-31.
- Breiman, L., 2001, **Random Forests**, Machine learning, 45(1), PP. 5-32.
- Camps-Valls, G., 2009, **Machine Learning in Remote Sensing Data Processing**, Paper presented at the 2009 IEEE International Workshop on Machine Learning for Signal Processing.
- Camps-Valls, G., Gómez-Chova, L., Muñoz-Mari, J., Vila-Francés, J., Amoros, J., del Valle-Tascon, S. & Calpe-Maravilla, J., 2009, **Biophysical Parameter Estimation with Adaptive Gaussian Processes**, Paper presented at the 2009 IEEE International Geoscience and Remote Sensing Symposium.
- Camps-Valls, G., Verrelst, J., Munoz-Mari, J., Laparra, V., Mateo-Jimenez, F. & Gomez-Dans, J., 2016, **A Survey on Gaussian Processes for Earth-Observation Data Analysis: A Comprehensive Investigation**, IEEE Geoscience and Remote Sensing Magazine, 4(2), PP. 58-78.
- Cherkassky, V. & Ma, Y., 2004, **Practical Selection of SVM Parameters and Noise Estimation for SVM Regression**, Neural Networks, 17(1), PP. 113-126.
- Clevers, J., 2014, **Beyond NDVI: Extraction of Biophysical Variables from Remote Sensing Imagery**, In Land Use and Land Cover Mapping in Europe (PP. 363-381), Springer.
- Darvishzadeh, R., 2008, **Hyperspectral Remote Sensing of Vegetation Parameters Using Statistical and Physical Models**.
- Demuth, H. & Beale, M., 1998, **User's Guide for Neural Network Toolbox for Use with MATLAB**, The Mathworks Inc., Natick, 3.
- Durbha, S.S., King, R.L. & Younan, N.H., 2007, **Support Vector Machines Regression for Retrieval of Leaf Area Index from Multiangle Imaging Spectroradiometer**, Remote Sensing of Environment, 107(1-2), PP. 348-361.
- Duveiller, G., Weiss, M., Baret, F. & Defourny, P., 2011, **Retrieving Wheat Green Area Index During the Growing Season from Optical Time Series Measurements Based on Neural Network Radiative Transfer Inversion**, Remote Sensing of Environment, 115(3), PP. 887-896.
- Erästö, P., 2001, **Support Vector Machines-Backgrounds and Practice**.
- Fawagreh, K., Gaber, M.M. & Elyan, E., 2014, **Random Forests: From Early Developments to Recent Advancements**, Systems Science & Control Engineering: An Open Access Journal, 2(1), PP. 602-609.
- Hagan, M.T., Demuth, H.B., Beale, M.H. & De Jesús, O., 1996, **Neural Network Design** (Vol. 20), Pws Pub. Boston.
- Kimes, D.S., Knyazikhin, Y., Privette, J., Abuelgasim, A. & Gao, F., 2000, **Inversion Methods for Physically-Based Models**, Remote Sensing Reviews, 18(2-4), PP. 381-439.
- Kumar, P., Gupta, D.K., Mishra, V.N. & Prasad, R., 2015, **Comparison of Support Vector Machine, Artificial**

- Neural Network, and Spectral Angle Mapper Algorithms for Crop Classification Using LISS IV Data**, International Journal of Remote Sensing, 36(6), PP. 1604-1617.
- Kuss, M. & Rasmussen, C.E., 2005, **Assessing Approximate Inference for Binary Gaussian Process Classification**, Journal of Machine Learning Research, 6(Oct), PP. 1679-1704.
- Lawrence, N., 2005, **Probabilistic Non-Linear Principal Component Analysis with Gaussian Process Latent Variable Models**, Journal of Machine Learning Research, 6(Nov), PP. 1783-1816.
- LeCun, Y., Touresky, D., Hinton, G. & Sejnowski, T., 1988, **A Theoretical Framework for Back-Propagation**, Paper presented at the Proceedings of the 1988 Connectionist Models Summer School.
- Menenti, M., Rast, M., Baret, F., van den Hurk, B., Knorr, W., Mauser, W., . . . & Verstraete, M., 2003, **Understanding Vegetation Response to Climate Variability from Space: Recent Advances towards the SPECTRA Mission**, Paper presented at the EGS-AGU-EUG Joint Assembly.
- Meroni, M., Colombo, R. & Panigada, C., 2004, **Inversion of a Radiative Transfer Model with Hyperspectral Observations for LAI Mapping in Poplar Plantations**, Remote Sensing of Environment, 92(2), PP. 195-206.
- Moreno, J.F., Baret, F., Leroy, M., Menenti, M., Rast, M. & Shaepman, M., 2003, **Retrieval of Vegetation Properties from Combined Hyperspectral/Multiangular Optical Measurements: Results from the DAISEX Campaigns**, Paper presented at the IGARSS 2003, 2003 IEEE International Geoscience and Remote Sensing Symposium, Proceedings (IEEE Cat. No. 03CH37477).
- Mousivand, A., 2015, **Retrieval of Vegetation Properties Using Top of Atmosphere Radiometric Data: A Multi-Sensor Approach**.
- Mousivand, A., Menenti, M., Gorte, B. & Verhoef, W., 2014, **Global Sensitivity Analysis of the Spectral Radiance of a Soil-Vegetation System**, Remote Sensing of Environment, 145, PP. 131-144.
- Mutanga, O., Adam, E. & Cho, M.A., 2012, **High Density Biomass Estimation for Wetland Vegetation Using WorldView-2 Imagery and Random Forest Regression Algorithm**, International Journal of Applied Earth Observation and Geoinformation, 18, PP. 399-406.
- O'Hagan, A., 1978, **Curve Fitting and Optimal Design for Prediction**, Journal of the Royal Statistical Society: Series B (Methodological), 40(1), PP. 1-24.
- Pham, T.D., Yoshino, K. & Bui, D.T., 2017, **Biomass Estimation of Sonneratia Caseolaris (L.) Engler at a Coastal Area of Hai Phong City (Vietnam) Using ALOS-2 PALSAR Imagery and GIS-Based Multi-Layer Perceptron Neural Networks**, GIScience & Remote Sensing, 54(3), PP. 329-353.
- Rivera-Caicedo, J.P., Verrelst, J., Muñoz-Marí, J., Camps-Valls, G. & Moreno, J., 2017, **Hyperspectral Dimensionality Reduction for Biophysical Variable Statistical Retrieval**, ISPRS Journal of Photogrammetry and Remote Sensing, 132, PP. 88-101.
- Rumelhart, D.E., Hinton, G.E. & McClelland, J.L., 1986, **A General Framework for Parallel Distributed**

- Processing, Parallel Distributed Processing: Explorations in the Microstructure of Cognition**, 1(45-76), P. 26.
- Sellers, P., Dickinson, R., Randall, D., Betts, A., Hall, F., Berry, J., . . . & Nobre, C., 1997, **Modeling the Exchanges of Energy, Water, and Carbon between Continents and the Atmosphere**, Science, 275(5299), PP. 502-509.
- Smola, A.J. & Schölkopf, B., 1998, **On a Kernel-Based Method for Pattern Recognition, Regression, Approximation, and Operator Inversion**, Algorithmica, 22(1-2), PP. 211-231.
- Smola, A.J. & Schölkopf, B., 2004, **A Tutorial on Support Vector Regression**, Statistics and Computing, 14(3), PP. 199-222.
- Vapnik, V., 2013, **The Nature of Statistical Learning Theory**, Springer Science & Business Media.
- Verger, A., Baret, F. & Camacho, F., 2011, **Optimal Modalities for Radiative Transfer-Neural Network Estimation of Canopy Biophysical Characteristics: Evaluation over an Agricultural Area with CHRIS/PROBA Observations**, Remote Sensing of Environment, 115(2), PP. 415-426.
- Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J. & Moreno, J., 2012, **Retrieval of Vegetation Biophysical Parameters Using Gaussian Process Techniques**, IEEE Transactions on Geoscience and Remote Sensing, 50(5), PP. 1832-1843.
- Verrelst, J., Malenovsky, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., . . . & Moreno, J., 2018, **Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods**, Surveys in Geophysics, PP. 1-41.
- Verrelst, J., Rivera, J.P., Gitelson, A., Delegido, J., Moreno, J. & Camps-Valls, G., 2016, **Spectral Band Selection for Vegetation Properties Retrieval Using Gaussian Processes Regression**, International Journal of Applied Earth Observation and Geoinformation, 52, PP. 554-567.
- Verrelst, J., Rivera, J.P., Veroustraete, F., Muñoz-Marí, J., Clevers, J.G., Camps-Valls, G. & Moreno, J., 2015, **Experimental Sentinel-2 LAI Estimation Using Parametric, Non-Parametric and Physical Retrieval Methods—A Comparison**, ISPRS Journal of Photogrammetry and Remote Sensing, 108, PP. 260-272.
- Verstraete, M.M., Pinty, B. & Myneni, R.B., 1996, **Potential and Limitations of Information Extraction on the Terrestrial Biosphere from Satellite Remote Sensing**, Remote Sensing of Environment, 58(2), PP. 201-214.
- Wang, F., Huang, J., Wang, Y., Liu, Z., Peng, D. & Cao, F., 2013, **Monitoring Nitrogen Concentration of Oilseed Rape from Hyperspectral Data Using Radial Basis Function**, International Journal of Digital Earth, 6(6), PP. 550-562.
- Watson, D.J., 1947, **Comparative Physiological Studies on the Growth of Field Crops: I. Variation in Net Assimilation Rate and Leaf Area between Species and Varieties, and within and between Years**, Annals of Botany, 11(41), PP. 41-76.
- Weiss, M., Baret, F., Smith, G., Jonckheere, I. & Coppin, P., 2004, **Review of Methods for in Situ Leaf Area Index**

- (LAI) Determination: Part II. Estimation of LAI, Errors and Sampling**, Agricultural and Forest Meteorology, 121(1-2), PP. 37-53.
- Williams, C.K. & Rasmussen, C.E., 1996, **Gaussian Processes for Regression**, Paper presented at the Advances in Neural Information Processing Systems.
- Williams, C.K. & Rasmussen, C.E., 2006, **Gaussian Processes for Machine Learning** (Vol. 2), MIT Press Cambridge, MA.
- Yi, G., Shi, J. & Choi, T., 2011, **Penalized Gaussian Process Regression and Classification for High-Dimensional Nonlinear Data**, Biometrics, 67(4), PP. 1285-1294.
- Yu, X., Hyypä, 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(1), PP. 28-37.
- Yuan, H., Yang, G., Li, C., Wang, Y., Liu, J., Yu, H., ... & Yang, X., 2017, **Retrieving Soybean Leaf Area Index from Unmanned Aerial Vehicle Hyperspectral Remote Sensing: Analysis of RF, ANN, and SVM Regression Models**, Remote Sensing, 9(4), P. 309.