

Application of Remote Sensing and Machine Learning for the Identification of Dust Sources in Khuzestan Province

Hesam, S. M.¹  | Sabzehzari, M.² 

1. Forecasting, and Crisis Management of Atmospheric Hazards, Khuzestan Province Meteorological Department, Ahvaz, Iran.

2. Khuzestan Province Meteorological Department, Ahvaz, Iran.

Corresponding Author E-mail: hesam.mohsen@gmail.com

(Received: 21 July 2025, Revised: 3 Aug 2025, Accepted: 18 Oct 2025, Published online: 16 Dec 2025)

Summary

Khuzestan Province, due to its arid climate and unique geographical location, is one of the main regions affected by dust storms in Iran. This study utilized satellite data and the Random Forest machine learning algorithm to identify dust source areas in the province during the period from January 26 to April 26, 2025. Sentinel-2, Sentinel-5P, and SMAP satellite data were used to calculate the Normalized Difference Vegetation Index (NDVI), Absorbing Aerosol Index (AAI), and surface soil moisture respectively. These indices monitored vegetation cover, the presence of dust-related aerosols, and soil dryness.

Google Earth Engine (GEE) was employed for data processing. The steps included defining the study area (Khuzestan, with a surface area of 64,057 km², based on FAO/GAUL/2015 data), filtering satellite images with cloud cover less than 10%, calculating the average values of NDVI, AAI, and soil moisture, and integrating these layers into a composite image. A Random Forest classifier with 50 decision trees was used to categorize regions into source and non-source classes. Sample points included 200 source points (e.g., Hoor-Al-Azim wetland) and 200 non-source points (urban and agricultural areas). Of these points, 70% were used for model training, 15% for validation, and 15% for final testing and evaluation. NDVI, Soil Moisture and AAI time series were also analyzed to assess aerosol variation trends, and model accuracy was evaluated using a confusion matrix. The results obtained from the validation data (overall accuracy = 0.95, F1-score = 0.95, Cohen's Kappa = 0.90) and the final test (overall accuracy = 0.967, F1-score = 0.967, Cohen's Kappa = 0.933) indicate a very strong and stable performance of the mode. The results indicated that the southern and western parts of Khuzestan—especially areas around Hoor-Al-Azim, southeastern Ahvaz, and the regions of Mahshahr, Hendijan, Omidiyeh, Susangerd, and Hoveyze are major dust sources. The classification map (Figure 8) highlighted these areas in red and non-source areas in green. The NDVI map (Figure 2) showed low vegetation coverage (values less than 0) in the southwest, particularly near Hoor-Al-Azim, indicating dry, bare-soil conditions. The AAI index (Figure 3) confirmed high aerosol values (0.2 to 0.5) in the central and southern regions, especially around Ahvaz and Hoor-Al-Azim. Soil moisture values (Figure 4) were also low in the south and central areas (below 0.1 cm³/cm³), indicating high potential for wind erosion. Key contributing factors to dust source activation included drought, reduced rainfall, the drying of the Jarrahi and Karun rivers, and desiccation of the southern parts of Hoor-Al-Azim.

This study successfully identified dust source areas in Khuzestan and is consistent with previous research, such as that by Rangzan et al. (1393). The inclusion of soil moisture as a new variable improved identification accuracy. The Random Forest algorithm provided reliable performance, due to its ability to model nonlinear relationships and handle heterogeneous data. However, limitations included the lack of field data for validation and the influence of atmospheric conditions on satellite observations.

Future studies are encouraged to incorporate field observations for validation and increase the number of sample points. Moreover, more advanced methods such as deep learning and the integration of satellite data with meteorological models (e.g., wind speed and direction), could improve dust storm forecasting accuracy. This approach especially with the flexible coding environment of Google Earth Engine offers an effective tool for monitoring and managing dust sources and can be applied in wetland restoration and soil stabilization efforts in desert regions.

Keywords: Dust source, Remote sensing, Machine learning Random Forest, Khuzestan.

Cite this article: Hesam, S. M., & Sabzehzari, M. (2025). Application of Remote Sensing and Machine Learning for the Identification of Dust Sources in Khuzestan Province. *Journal of the Earth and Space Physics*, 51(3), 703- 717. DOI: <http://doi.org/10.22059/jesphys.2025.398558.1007704>

E-mail: (2) sabzeh218@gmail.com



© Authors Retain the Copyright and Full Publishing Rights.

Publisher: University of Tehran Press.

DOI: <http://doi.org/10.22059/jesphys.2025.398558.1007704>

Print ISSN: 2538-371X
Online ISSN: 2538-3906

کاربرد سنجش از دور و یادگیری ماشین برای تشخیص کانون‌های گرد و خاک در استان خوزستان

سید محسن حسام^۱ | محمد سبزه زاری^۲

۱. پیش‌بینی و مدیریت بحران مخاطرات جوی، اداره کل هواشناسی استان خوزستان، اهواز، ایران.

۲. اداره کل هواشناسی استان خوزستان، اهواز، ایران.

رایانامه نویسنده مسئول: hesam.mohsen@gmail.com

(دریافت: ۱۴۰۴/۴/۳۰، بازنگری: ۱۴۰۴/۵/۱۲، پذیرش نهایی: ۱۴۰۴/۷/۲۶، انتشار آنلاین: ۱۴۰۴/۹/۲۵)

چکیده

استان خوزستان به دلیل شرایط اقلیمی خشک و موقعیت جغرافیایی ویژه، یکی از مناطق اصلی وقوع طوفان‌های گرد و خاک در ایران است. این مطالعه با استفاده از داده ماهواره‌ای شامل شاخص نرمال‌شده تفاوت پوشش گیاهی (NDVI)، شاخص هواویز جذبی (AAI)، رطوبت سطحی خاک و الگوریتم یادگیری ماشین جنگل تصادفی (Random Forest) به شناسایی کانون‌های گرد و خاک در خوزستان طی بازه زمانی ۲۶ ژانویه تا ۲۶ آوریل ۲۰۲۵ پرداخته است. داده سنتینل-۲، سنتینل-۵، و اسمپ برای محاسبه شاخص‌ها استفاده شدند. با استفاده از پلتفرم گوگل ارث انجین (GEE)، میانگین شاخص نرمال‌شده تفاوت پوشش گیاهی، میانگین شاخص هواویز جذبی، میانگین رطوبت سطحی خاک در بازه یک ماهه، یک تصویر ترکیبی ساخته شد. تحلیل سری زمانی شاخص نرمال‌شده تفاوت پوشش گیاهی، شاخص هواویز جذبی و رطوبت سطحی خاک افزایش فعالیت هواویزها را در ماه‌های خشک‌تر تأیید کرد. با تحلیل تصاویر ماهواره یومت ست-۲۰۰ نقطه به‌عنوان کانون گرد و خاک و ۲۰۰ نقطه به‌عنوان نقاط کانونی انتخاب شد که ۷۰ درصد آنها جهت آموزش مدل، ۱۵ درصد جهت اعتبارسنجی و ۱۵ درصد برای آزمون و ارزیابی نهایی استفاده شد. مدل طبقه‌بندی جنگل تصادفی با ۵۰ درصد دقت تصمیم مورد استفاده قرار گرفت. نتایج حاصل از داده‌های اعتبارسنجی (دقت کلی ۰/۹۵، امتیاز F1 ۰/۹۵ و ضریب کاپا ۰/۹) و آزمون نهایی (دقت کلی ۰/۹۶۷، امتیاز F1 ۰/۹۶۷ و ضریب کاپا ۰/۹۳۳) نشان‌دهنده عملکرد بسیار قوی و پایدار مدل است. نتایج نشان داد که مناطق جنوب و غرب خوزستان، به‌ویژه اطراف هورالعظیم، به‌عنوان کانون‌های اصلی گرد و خاک شناسایی شدند. این روش می‌تواند به مدیریت و کاهش اثرات طوفان‌های گرد و خاک کمک کند.

واژه‌های کلیدی: کانون گرد و خاک، سنجش از دور، یادگیری ماشین، جنگل تصادفی.

۱. مقدمه

سنجده مودیس و PM10 را در خوزستان گزارش کردند
($R^2=0.78$).

بحرینی و همکاران (۱۳۹۷) در مطالعه‌ای در منطقه خشک
بردخون با استفاده از شاخص نرمال‌شده تفاوت
پوشش گیاهی، شاخص پوشش گیاهی بهبودیافته (EVI) و
شاخص پوشش گیاهی تنظیم‌شده برای خاک (SAVI) نشان
دادند که شاخص نرمال‌شده تفاوت پوشش گیاهی با ضریب
تعیین $R^2=0.56$ بیشترین همبستگی را با داده‌های میدانی
درصد تاج پوشش گیاهی اندازه‌گیری‌شده در ۲۹ پلات
نمونه‌برداری‌شده دارد.

طوفان‌های گرد و خاک از مشکلات زیست‌محیطی جدی در
مناطق خشک و نیمه‌خشک، به‌ویژه در خوزستان، هستند. این
پدیده‌ها بر سلامت عمومی، کشاورزی، و زیرساخت‌ها
تأثیرات منفی دارند. شناسایی دقیق کانون‌های گرد و خاک
برای کاهش اثرات این پدیده ضروری است. سنجش از دور
و یادگیری ماشین ابزارهای قدرتمندی برای پایش و تحلیل
این پدیده‌ها فراهم می‌کنند.

مطالعات متعددی از روش‌های سنجش از دور برای شناسایی
کانون‌های گرد و خاک انجام شده است. رنگزن و همکاران
(۱۳۹۳) همبستگی زیاد بین داده ضخامت نوری هواویزهای

استناد: حسام، سید محسن و سبزه زاری، محمد (۱۴۰۴). کاربرد سنجش از دور و یادگیری ماشین برای تشخیص کانون‌های گرد و خاک در استان خوزستان. مجله فیزیک زمین و فضا،

DOI: <http://doi.org/10.22059/jesphys.2025.398558.1007704>

رایانامه: (۲) sabzeh218@gmail.com

الگوریتم ماشین بردار پشتیبان، پوشش گیاهی را با دقت ۹۹/۱۳٪ طبقه‌بندی کردند. زنگنه (۱۳۹۸) با یادگیری عمیق، تغییرات پوشش گیاهی را با دقت بیش از ۹۹٪ پیش‌بینی کرد. این مطالعات پتانسیل بالای یادگیری ماشین در تحلیل داده ماهواره‌ای را نشان می‌دهند.

احمدی بیرگانی (۲۰۲۵) با مطالعه دریاچه ارومیه، رویکردی نوین برای اولویت‌بندی منابع گرد و خاک ارائه داد. در این پژوهش، با استفاده از داده‌های ضخامت نوری هواویزا سنجنده مودیس و الگوریتم مایاک در بستر گوگل‌ارث انجین، مناطق غیر فعال، فعال و به‌شدت فعال تعریف شد. روند داده‌های ضخامت نوری هواویزا طی سال‌های ۲۰۰۱ تا ۲۰۲۴ محاسبه و نواحی اولویت‌دار شناسایی شد. نتایج نشان داد که از سال ۲۰۰۸ نواحی فعال در شرق دریاچه ارومیه شکل گرفته و تا سال ۲۰۱۹ به بخش‌های جنوبی و مرکزی گسترش یافته‌اند. همچنین امکان تفکیک منابع دائمی از منابع فصلی یا موقتی را فراهم شد.

در راستای توسعه روش‌های نوین پهنه‌بندی کانون‌های گرد و خاک، بروغنی و همکاران (۱۴۰۱) با به‌کارگیری مدل‌های یادگیری ماشین شامل گرادیان تقویتی (Gradient Boosting Model)، جنگل تصادفی و مدل افزودنی کلی (Generalized Additive Model)، اقدام به پیش‌بینی آسیب‌پذیری مناطق مستعد گرد و خاک در کویر مرکزی کردند. این پژوهش با استفاده از تصاویر سنجنده مودیس و شاخص‌های مختلف مانند شاخص تفاوت دمای درخشندگی باند ۳۱ و ۳۲ (BTD3132) و شاخص نرمال‌شده تفاوت گرد و خاک (NDDI)، موفق به شناسایی ۱۳۵ کانون گرد و خاک شد. نتایج نشان داد مدل گرادیان تقویتی با دقت ۶۴/۲٪ بهترین عملکرد را در بین الگوریتم‌ها داشته و عوامل کاربری اراضی و خاک‌شناسی به‌عنوان مؤثرترین پارامترهای مؤثر بر تشکیل کانون‌های گرد و خاک شناسایی شدند. این مطالعه به‌عنوان نمونه‌ای موفق از تلفیق سنجش از دور و هوش مصنوعی در مدیریت پدیده گرد و خاک می‌تواند مبنای مناسبی برای تصمیم‌گیری‌های مدیریتی در مناطق خشک باشد.

در مطالعه حاضر با استفاده از محصولات ماهواره‌ای شاخص

رایگانی و همکاران (۱۳۹۸) با تلفیق داده ماهواره‌ای لندست ۸ و مودیس، روشی جامع برای شناسایی مناطق مستعد تولید گرد و خاک ارائه دادند. در این پژوهش، با ترکیب شاخص‌های پوشش گیاهی، رطوبت خاک و ویژگی‌های زمین‌شناسی در قالب ارزیابی چند معیاره، نقشه‌های پتانسیل بادفرسایی تهیه شد. سپس با تحلیل مسیرهای جریان هوا به کمک مدل‌های اسپلیت و پایش تغییرات بلندمدت، کانون‌های فعال گرد و خاک با دقت زیاد شناسایی شد. این مطالعه نشان داد که رویکردهای تلفیقی سنجش از دور و داده زمینی می‌توانند چارچوب مناسبی برای پایش و مدیریت کانون‌های ریزگرد ارائه دهند.

تحلیل شاخص ضخامت نوری هواویزا نیز ابزاری کلیدی برای مکان‌یابی منابع اصلی تولید گرد و خاک به‌شمار می‌آید. در این راستا، گندهام و همکاران (۲۰۲۳) با استفاده از محصولات سنجنده مودیس و الگوریتم مایاک (MAIAC) و تحلیل فراوانی رخدادهای ضخامت نوری هواویزهای زیاد، کانون‌های فعال گرد و خاک در شبه‌جزیره عربستان را شناسایی کرده و ارتباط آنها را با ویژگی‌های زمین‌شناسی و کاربری اراضی بررسی کرده‌اند.

در حوزه مدل‌سازی مکانی، جعفری و همکاران (۲۰۲۲) در ایران از ترکیب شاخص‌های سنجش از دور نظیر شاخص نرمال‌شده تفاوت پوشش گیاهی، رطوبت خاک، سرعت باد و ضخامت نوری هواویزا با الگوریتم‌های یادگیری ماشین مانند جنگل تصادفی، مدل خطی تعمیم‌یافته (Generalized Linear Model)، تحلیل درخت طبقه‌بندی (Linear Model Classification)، شبکه عصبی مصنوعی (Tree Analysis)، رگرسیون تطبیقی چند متغیره اسپلاین (Neural Network)، حد اکثر (Multivariate Adaptive Regression Splines)، حد اکثر آنترپی (Maximum Entropy) و الگوریتم‌های آنسامبل برای تهیه نقشه‌های آسیب‌پذیری به انتشار گرد و خاک بهره گرفته‌اند. نتایج این مطالعات نشان می‌دهد که مدل‌های آنسامبل به‌دلیل پایداری و قابلیت تفسیر ویژگی‌ها، عملکرد خوبی در شناسایی مناطق مستعد انتشار دارند.

یادگیری ماشین نیز در تحلیل داده سنجش از دور کاربرد گسترده‌ای دارد. رشیدی و همکاران (۱۴۰۲) با استفاده از

نرمال شده تفاوت پوشش گیاهی، شاخص هواویز جذبی، رطوبت خاک و الگوریتم جنگل تصادفی، به شناسایی کانون های گرد و خاک در خوزستان پرداخته شده است. همچنین، کد استفاده شده در پلتفرم گوگل ارث انجین امکان تغییر بازه های زمانی را فراهم می کند، که به هواشناسان اجازه می دهد با بررسی میانگین ها در بازه های کوتاه تر (مثلاً هفتگی یا ماهانه)، تغییرات سریع تر در وضعیت کانون های گرد و خاک را رصد کنند و پیش بینی های دقیق تری انجام دهند.

۲. روش پژوهش

۲-۱. منطقه مورد مطالعه

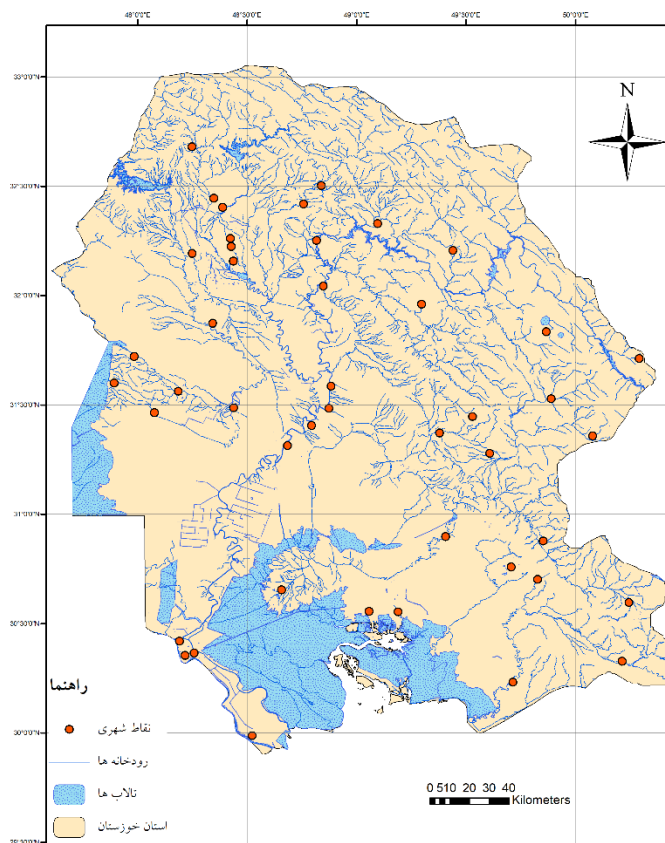
استان خوزستان در جنوب غربی ایران، بین طول های جغرافیایی ۴۷/۶۶ تا ۵۰/۵۲ درجه شرقی و عرض های جغرافیایی ۲۹/۹۲ تا ۳۲/۹۹ درجه شمالی قرار دارد و مساحتی حدود ۶۴,۰۵۷ کیلومتر مربع را پوشش می دهد. این استان به دلیل شرایط اقلیمی خشک و نیمه خشک، مجاورت با تالاب

هورالعظیم و همجواری با بیابان های عراق، به ویژه در فصل های خشک، مستعد وقوع طوفان های گرد و خاک است. محدوده جغرافیایی استان با استفاده از مجموعه داده FAO/GAUL/2015 در محیط گوگل ارث انجین تعریف شد (شکل ۱).

۲-۲. روش کار

در این مطالعه از پلتفرم گوگل ارث انجین برای پردازش و تحلیل داده ماهواره ای استفاده شد. گوگل ارث انجین یک پلتفرم ابری است که امکان دسترسی به داده ماهواره ای و ابزارهای پردازش مکانی را فراهم می کند. مراحل کار و برنامه نویسی در این پلتفرم به شرح زیر است:

الف) تعریف منطقه مورد مطالعه و بازه زمانی ب) بارگذاری و فیلتر داده ماهواره ای ج) محاسبه شاخص ها، پردازش، میانگین گیری و تحلیل سری زمانی د) ترکیب داده و تولید تصویری ترکیبی ه) ایجاد نقاط نمونه و آموزش مدل و) طبقه بندی و نمایش نتایج ز) اعتبارسنجی و آزمون نهایی.



شکل ۱. موقعیت رودخانه ها، تالاب ها و نقاط شهری استان خوزستان.

۲-۱. داده‌های مورد استفاده

داده سنتینل-۲ از مجموعه COPERNICUS/S2_SR با تفکیک‌پذیری مکانی ۱۰ متر و تفکیک‌پذیری زمانی ۵ روز در استوا و ۲ تا ۳ روز در عرض‌های میانی، برای محاسبه شاخص نرمال‌شده تفاوت پوشش گیاهی استفاده شدند. شاخص نرمال‌شده تفاوت پوشش گیاهی با فرمول زیر محاسبه می‌شود:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

که در آن NIR شدت تابش در طول موج ۸۴۲ نانومتر (باند B8) و Red شدت تابش در طول موج ۶۶۵ نانومتر (باند B4) هستند. این شاخص در این مطالعه برای شناسایی مناطق با پوشش گیاهی کم یا بدون پوشش گیاهی استفاده شد.

داده شاخص هواویز جذبی از مجموعه COPERNICUS/S5P/NRTI/L3_AER_AI استخراج شدند. این مجموعه داده شامل شاخص هواویز جذب‌کننده از ماهواره سنتینل ۵ است که برای پایش ذرات معلق در هوا مانند گرد و خاک و دود با تفکیک‌پذیری مکانی ۳ کیلومتر و تفکیک‌پذیری زمانی روزانه استفاده می‌شود. مقادیر زیاد این شاخص نشان‌دهنده حضور هواویزهای جذبی مانند گرد و خاک است.

داده رطوبت خاک از مجموعه ناسا NASA/SMAP/SPL4SMGP/007 با تفکیک‌پذیری مکانی ۹ کیلومتر و تفکیک‌پذیری زمانی ۳ ساعته استفاده شدند. این مجموعه داده از ماهواره اسمپ تهیه شده و شامل اندازه‌گیری‌های رطوبت خاک سطحی با واحد cm^3/cm^3 است که به معنای نسبت حجم آب به حجم کل خاک است. این واحد، کسری از حجم خاک را نشان می‌دهد که توسط آب اشغال شده است. به عبارت دیگر، بیانگر میزان آبی است که در واحد حجم خاک (معمولاً در لایه‌های سطحی خاک، تا عمق چند سانتی‌متری) وجود دارد. رطوبت خاک در لایه سطحی (۰-۵ سانتی‌متر) برای بررسی خشکی خاک و ارتباط آن با گرد و خاک استفاده می‌شود. رطوبت کم خاک با افزایش پتانسیل تولید گرد و خاک مرتبط است.

برای تولید داده‌های آموزشی موردنیاز مدل، از محصولات

ماهواره‌ی متئوسات (Meteosat) استفاده شد. بدین ترتیب که نقاط مربوط به کلاس کانونی گرد و خاک بر اساس تصاویر محصول گرد و خاک (Dust RGB) استخراج شدند. در مقابل، برای تعریف کلاس غیر کانونی از تصاویر رنگ طبیعی (Natural Colour RGB) استفاده شد. بدین ترتیب داده‌های دو کلاس بر اساس ویژگی‌های طیفی مشخص و قابل تفکیک از یکدیگر تعیین شدند. با توجه به این که مساحت مناطق کانونی گرد و خاک در استان مورد مطالعه کمتر از مساحت مساحت مناطق غیر کانونی است، داده‌ها به‌طور طبیعی نامتوازن هستند. الگوریتم‌های یادگیری ماشین جنگل تصادفی، در شرایط نامتوازن، تمایل دارند کلاس غالب (مناطق غیر کانونی) را پیش‌بینی کرده و کلاس اقلیت (مناطق کانونی) را نادیده بگیرند (چن و همکاران، ۲۰۰۴). از آنجا که هدف اصلی پژوهش شناسایی دقیق کانون‌های گرد و خاک است، داده‌ها در مرحله آموزش مدل به‌صورت متوازن در نظر گرفته شدند تا الگوریتم بتواند الگوهای مرتبط با کلاس اقلیت را به‌درستی بیاموزد. این رویکرد در بسیاری از مطالعات مشابه نیز به‌عنوان روشی معتبر برای مقابله با مشکل عدم توازن کلاس‌ها به کار رفته است.

۲-۲. پیش‌پردازش و ترکیب داده:

برای ترکیب داده‌های ماهواره‌ای شاخص نرمال‌شده تفاوت پوشش گیاهی، شاخص هواویز و رطوبت خاک، ابتدا پیش‌پردازش‌هایی برای همگن‌سازی زمانی و مکانی انجام شد. داده‌های شاخص نرمال‌شده تفاوت پوشش گیاهی، پس از اعمال فیلتر حذف پیکسل‌های ابری و محاسبه شاخص، به‌صورت میانگین ماهانه برای دوره مورد مطالعه (در بازه زمانی ۲۶ ژانویه تا ۲۶ آوریل ۲۰۲۵) تجمیع شدند. داده‌های شاخص هواویز جذبی، به میانگین ماهانه تبدیل شدند. همچنین، داده‌های رطوبت خاک، ابتدا به میانگین روزانه و سپس به میانگین ماهانه تجمیع شدند. این فرآیند میانگین‌گیری ماهانه با هدف کاهش نویز و تغییرات کوتاه‌مدت و ایجاد سازگاری زمانی بین مجموعه داده‌ها انجام شد، در حالی که پیکسل‌های نامعتبر (مانند داده‌های تحت تأثیر ابر) از محاسبات حذف

شدند.

پس از پیش‌پردازش، داده‌ها به تفکیک‌پذیری مکانی مشترک ۹ کیلومتر مطابق با داده‌های رطوبت خاک، همسان‌سازی شدند؛ داده‌های شاخص نرمال‌شده تفاوت پوشش گیاهی و هواویز جذبی با روش میانگین‌گیری مکانی تجمع باز نمونه‌برداری شدند. در نهایت، سه لایه شاخص نرمال‌شده تفاوت پوشش گیاهی، شاخص هواویز جذبی، و رطوبت خاک در یک تصویر ترکیبی چند لایه ادغام شدند، به‌طوری‌که هر متغیر به‌عنوان یک باند جداگانه حفظ شد.

۲-۳. مبانی نظری

یادگیری ماشین شاخه‌ای از هوش مصنوعی است که هدف آن توسعه الگوریتم‌ها و مدل‌هایی است که می‌توانند بر اساس داده‌های ورودی، الگوها و روابط پنهان را شناسایی کرده و بدون نیاز به برنامه‌نویسی صریح، پیش‌بینی یا طبقه‌بندی انجام دهند. در این رویکرد، داده‌های آموزشی شامل نمونه‌هایی با ویژگی‌های مختلف به مدل داده می‌شود و مدل با استفاده از این داده‌ها قوانین عمومی را یاد می‌گیرد. سپس می‌توان از مدل آموزش‌دیده برای پیش‌بینی وضعیت نمونه‌های جدید استفاده کرد. یکی از مهم‌ترین مزایای یادگیری ماشین توانایی آن در کار با داده‌های حجیم، پیچیده، چند متغیره و متنوع مانند داده‌های ماهواره‌ای و محیطی است (بیشاپ، ۲۰۰۶).

در میان روش‌های یادگیری ماشین، الگوریتم جنگل تصادفی به‌عنوان یکی از دقیق‌ترین و پرکاربردترین روش‌ها برای مسائل طبقه‌بندی و رگرسیون شناخته می‌شود. این الگوریتم مبتنی بر درخت‌های تصمیم (Decision Trees) عمل می‌کند. درخت تصمیم ساختاری است که با استفاده از تقسیمات متوالی داده‌ها بر اساس ویژگی‌های مختلف، نمونه‌ها را به دسته‌های مشخص طبقه‌بندی می‌کند. با این حال، درخت تصمیم به تنهایی ممکن است دچار بیش‌برازش (Overfitting) شود و عملکرد آن در داده‌های جدید کاهش یابد (هان، ۲۰۱۱).

برای رفع این مشکل، الگوریتم جنگل تصادفی مجموعه‌ای از صدها یا هزاران درخت تصمیم را به‌صورت تصادفی تولید

می‌کند. در هر درخت، هم داده‌های آموزشی و هم متغیرهای ورودی به‌طور تصادفی نمونه‌برداری می‌شوند. در نهایت، خروجی مدل بر اساس رأی‌گیری اکثریت (در مسائل طبقه‌بندی) یا میانگین‌گیری (در مسائل رگرسیون) از نتایج همه درخت‌ها تعیین می‌شود. این سازوکار باعث می‌شود که جنگل تصادفی دقت بالاتری داشته باشد، نسبت به نویز داده‌ها مقاوم‌تر باشد، و مشکل بیش‌برازش به‌میزان زیادی کاهش یابد (لیا، ۲۰۰۲).

الگوریتم جنگل تصادفی به‌طور خاص به دلایل زیر استفاده شد:

مقاومت در برابر بیش‌برازش: بریمن (۲۰۰۱) نشان داد جنگل تصادفی با استفاده از چندین درخت تصمیم و میانگین‌گیری نتایج، از بیش‌برازش جلوگیری می‌کند، به‌ویژه در داده سنجش از دور که ممکن است نویز زیادی داشته باشند. مدیریت داده چندمتغیره: بیائو و اسکورت (۲۰۱۶) در پژوهشی توانایی این الگوریتم را مدل‌سازی روابط غیرخطی بین متغیرها (شاخص پوشش گیاهی نرمال‌شده، شاخص هواویز جذبی، و رطوبت خاک) را نشان داد. کارایی بالا در طبقه‌بندی: در مطالعه بلجیو و دراگوت (۲۰۱۶) نشان داده شده که جنگل تصادفی در طبقه‌بندی داده‌های سنجش از دور، دقت زیادی (اغلب بیشتر ۹۰٪) ارائه می‌دهد. انعطاف‌پذیری: کاتلر و همکاران (۲۰۰۷) نشان دادند که این الگوریتم نیازی به فرضیات اولیه در مورد توزیع داده ندارد و برای داده ناهمگن مانند داده ماهواره‌ای مناسب است.

۳. اندازه‌گیری، مشاهده و محاسبه

۳-۱. طبقه‌بندی با یادگیری ماشین

الگوریتم جنگل تصادفی با ۵۰ درخت تصمیم برای طبقه‌بندی مناطق به دو کلاس کانونی (گرد و خاک) و غیرکانونی استفاده شد. نقاط نمونه شامل ۲۰۰ نقطه کانونی (مانند هورالعظیم) و ۲۰۰ نقطه غیرکانونی (مناطق شهری یا کشاورزی) بود. نقاط کانونی بر اساس تحلیل تاریخیچه‌ای محصول گرد و خاک و نقاط غیرکانونی بر اساس تصاویر محصول رنگ طبیعی ماهواره‌ای متئوست انتخاب شدند. ۷۰

مناطق کانونی و غیر کانونی)، امکان ارزیابی دقیق‌تر کارایی مدل را فراهم می‌سازد.

۳-۲-۱. دقت مثبت

نسبت نمونه‌هایی که مدل به درستی به‌عنوان مثبت (کانون گرد و خاک) را شناسایی کرده است به کل نمونه‌هایی که به‌عنوان مثبت پیش‌بینی شده‌اند. این شاخص نشان می‌دهد پیش‌بینی‌های مثبت مدل تا چه حد معتبر هستند.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

۳-۲-۲. بازخوانی

نسبت نمونه‌های مثبت شناسایی شده درست به کل نمونه‌های مثبت واقعی. این معیار نشان می‌دهد چه میزان از کانون‌های واقعی گرد و خاک توسط مدل شناسایی شده‌اند.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

۳-۲-۳. امتیاز F1

میانگین هارمونیک دقت مثبت و بازخوانی است که توازن بین این دو شاخص را نشان می‌دهد. این معیار در شرایط داده‌های نامتوازن اهمیت بیشتری دارد.

$$\text{F1-Score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 2 \quad (4)$$

۳-۲-۴. دقت کلی

نسبت کل نمونه‌های درست شناسایی شده (در تمام کلاس‌ها) به کل نمونه‌ها. این معیار تصویری کلی از صحت عملکرد مدل ارائه می‌دهد.

$$\text{Overall Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

۳-۲-۵. ضریب کاپا

معیاری پیشرفته‌تر برای سنجش توافق بین نتایج مدل و داده‌های واقعی است که اثر توافق تصادفی را حذف می‌کند. این شاخص نسبت به عدم تعادل در داده‌ها حساس‌تر بوده و در مقایسه با دقت کلی ارزیابی دقیق‌تری ارائه می‌دهد.

$$\kappa = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \quad (6)$$

که در آن $\text{Pr}(a)$ نسبت توافق مشاهده شده، یعنی درصد نمونه‌هایی که مدل به درستی پیش‌بینی کرده است و $\text{Pr}(e)$ نسبت توافق مورد انتظار بر اساس احتمال تصادفی، یعنی میزان توافقی که حتی بدون دقت مدل نیز انتظار می‌رفت.

درصد نقاط (۱۴۰ نقطه کانونی و ۱۴۰ نقطه غیر کانونی) برای آموزش مدل، ۱۵ درصد (۳۰ نقطه کانونی و ۳۰ نقطه غیر کانونی) برای اعتبار سنجی و ۱۵ درصد (۳۰ نقطه کانونی و ۳۰ نقطه غیر کانونی) برای آزمون و ارزیابی نهایی استفاده شد.

۳-۲. معیارهای ارزیابی

برای ارزیابی عملکرد مدل‌های طبقه‌بندی، از ماتریس درهم‌ریختگی (Confusion Matrix) که در شکل ۲ نمایش داده شده است، استفاده می‌شود که شامل چهار مقدار اصلی است.

		مقدار واقعی	
		مثبت	منفی
مقدار پیش‌بینی	مثبت	TP	FP
	منفی	FN	TN

شکل ۲. ماتریس درهم‌ریختگی.

مثبت درست (TP): تعداد نمونه‌هایی که به‌درستی به‌عنوان کانون گرد و خاک شناسایی شدند.

مثبت نادرست (FP): تعداد نمونه‌هایی که اشتباهاً به‌عنوان کانون شناسایی شدند.

منفی درست (TN): تعداد نمونه‌هایی که درست به‌عنوان غیر کانون شناسایی شدند.

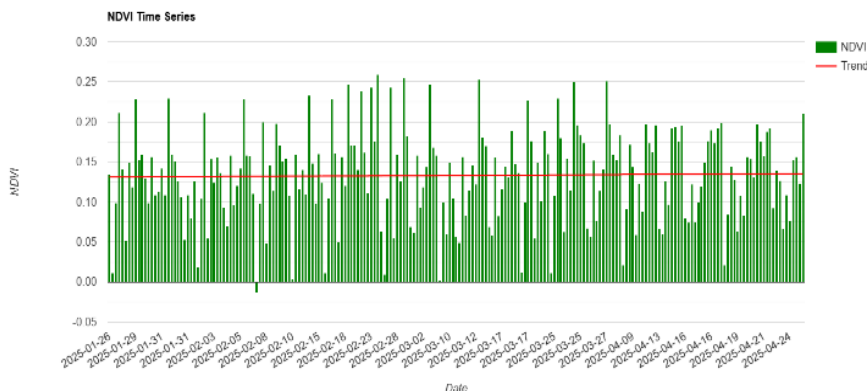
منفی نادرست (FN): تعداد نمونه‌هایی که اشتباهاً به‌عنوان غیر کانون شناسایی شدند.

به‌منظور ارزیابی دقت نتایج طبقه‌بندی و اطمینان از صحت شناسایی کانون‌های گرد و خاک، از شاخص‌های آماری مختلف استفاده شد. این شاخص‌ها شامل دقت مثبت (Precision)، بازخوانی (Recall)، امتیاز F1 (F1-Score)، دقت کلی (Overall Accuracy) و ضریب کاپا (Cohen's Kappa) هستند که هر کدام جنبه متفاوتی از عملکرد مدل را اندازه‌گیری می‌کنند. استفاده از این معیارها به‌ویژه در شرایطی که داده‌ها نامتوازن باشند (مانند تفاوت چشمگیر میان

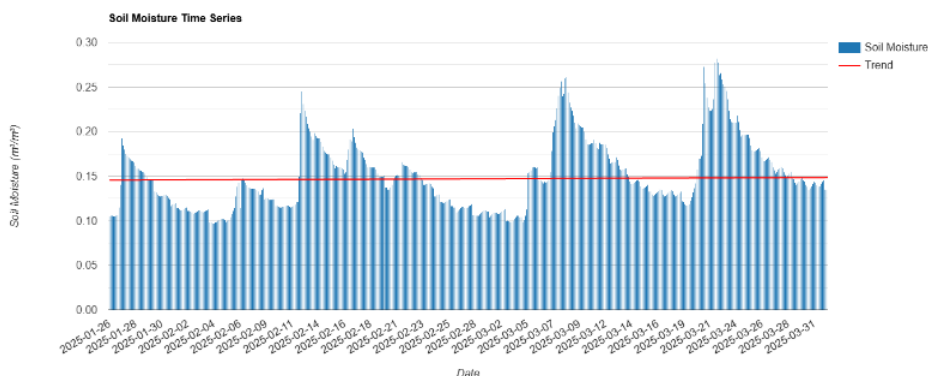
مقدار K بین ۰ و ۱ متغیر است؛ $K=1$ نشان‌دهنده توافق کامل و $K=0$ نشان‌دهنده توافقی در حد تصادف است.

۳-۳. سری زمانی شاخص‌های پوشش گیاهی، هواویز جذبی و رطوبت خاک داده‌های سری زمانی شاخص نرمال‌شده تفاوت

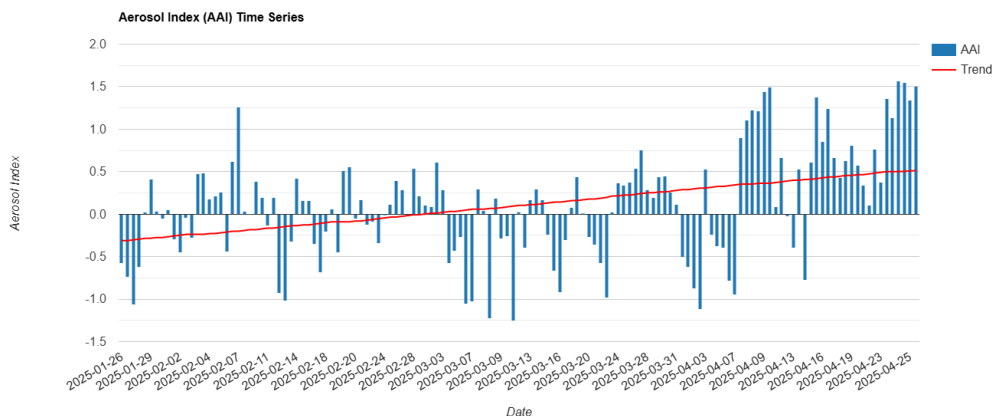
پوشش گیاهی، رطوبت خاک و شاخص هواویز جذبی در بازه زمانی ۲۶ ژانویه تا ۲۴ آوریل ۲۰۲۵ تحلیل شدند تا ارتباطات احتمالی بین پوشش گیاهی، شرایط هیدرولوژیکی و بارگذاری هواویزها اتمسفری بررسی شود. داده‌ها به صورت مشاهدات روزانه ارائه شده‌اند و زمان‌هایی که داده وجود نداشت، حذف شدند.



شکل ۳. سری زمانی شاخص نرمال‌شده تفاوت پوشش گیاهی در بازه زمانی مورد مطالعه.



شکل ۴. سری زمانی رطوبت خاک در بازه زمانی مورد مطالعه.



شکل ۵. سری زمانی شاخص هواویز جذبی در بازه زمانی مورد مطالعه.

بعد (تا حدود ۱/۵) به جذب هواویز رو به رشد اشاره دارد که احتمالاً به طوفان‌های گرد و خاک مرتبط است. دوره‌های رطوبت خاک کم با کاهش شاخص نرمال‌شده تفاوت پوشش گیاهی همزمان است که نشان می‌دهد دسترسی محدود به آب بر پوشش گیاهی تأثیر می‌گذارد. به‌طور همزمان، افزایش شاخص هواویز جذبی در مارس تا آوریل با نوسانات رطوبت خاک همپوشانی دارد که ممکن است نشان‌دهنده نقش خاک خشک در جابجایی گرد و خاک و افزایش غلظت هواویزها باشد.

۴. بحث

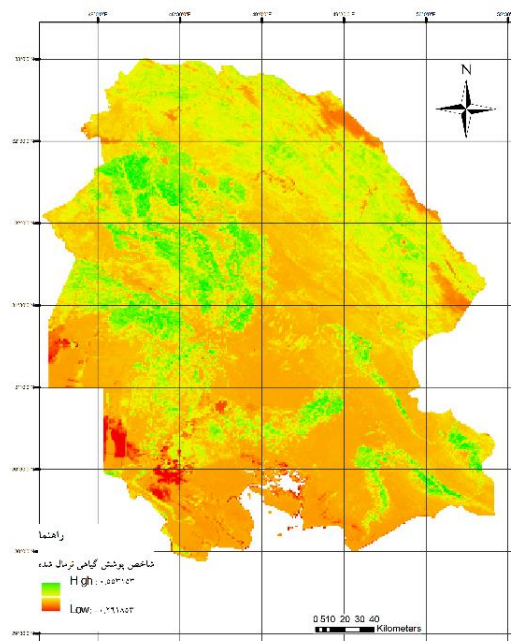
۴-۱. تحلیل شاخص‌های نرمال‌شده تفاوت پوشش گیاهی، هواویز جذبی و رطوبت خاک
نتایج شاخص نرمال‌شده تفاوت پوشش گیاهی در شکل ۱، شاخص هواویز جذبی در شکل ۲ و رطوبت خاک در شکل ۳ نمایش داده شده است.

میانگین شاخص نرمال‌شده تفاوت پوشش گیاهی (شکل ۶) برای استان خوزستان، که از داده سنتینل ۲- در بازه زمانی ۲۶ ژانویه تا ۲۶ آوریل ۲۰۲۵ استخراج شده است، نمایانگر توزیع پوشش گیاهی در این منطقه است. در این تصویر، رنگ‌های سبز نشان‌دهنده مناطق با شاخص پوشش گیاهی نرمال‌شده زیاد (بیش از ۰/۴) و پوشش گیاهی متراکم، عمدتاً در شمال و شرق خوزستان (مانند دزفول، شوشتر، و ایذه) هستند که به دلیل وجود ارتفاعات و زمین‌های کشاورزی این الگو مشاهده می‌شود. رنگ‌های زرد و نارنجی، نمایانگر شاخص پوشش گیاهی نرمال‌شده متوسط تا کم (بین ۰ و ۰/۴)، در بخش‌های مرکزی و جنوبی، از جمله اطراف اهواز و رامشیر، نشان‌دهنده مناطق با پوشش گیاهی کم یا بیابانی هستند. نواحی قرمز در جنوب غربی، به‌ویژه نزدیک هورالعظیم، با شاخص نرمال‌شده تفاوت پوشش گیاهی کمتر از ۰ (مناطق بدون پوشش گیاهی یا خاک برهنه)، به‌عنوان کانون‌های بالقوه گرد و خاک شناسایی می‌شوند. این الگو با شرایط اقلیمی خشک و تالاب‌های خشک‌شده در این مناطق همخوانی دارد و تأییدکننده نقش شاخص پوشش گیاهی نرمال‌شده در شناسایی مناطق مستعد فرسایش بادی است.

سری زمانی شاخص نرمال‌شده تفاوت پوشش گیاهی (شکل ۳) رفتار نوسانی حول میانگین تقریبی ۰/۱۵ را نشان می‌دهد، با مقادیر در بازه‌ای از ۰ تا حدود ۰/۲۵، به‌عنوان معیاری برای تراکم پوشش گیاهی، هیچ روند کلی مشخصی را نشان نمی‌دهد، همان‌طور که خط روند قرمز افقی در حدود ۰/۱۵ تثبیت شده است. در اوایل دوره (اواخر ژانویه تا اواسط فوریه)، شاخص نرمال‌شده تفاوت پوشش گیاهی معمولاً کمتر (۰/۰۵ تا ۰/۱۵) است که ممکن است نشان‌دهنده دسترسی محدود به آب باشد. قله‌های بعدی در مارس (تا ۰/۲۵) نشان‌دهنده سبز شدن موقت پوشش گیاهی است که احتمالاً ناشی از بارش‌های گذرا یا بهبود شرایط رشد است، و پس از آن کاهش در آوریل رخ داده است. این نوسانات نشان‌دهنده پوشش گیاهی پایدار اما تحت فشار در محیط‌های نیمه‌خشک است.

سری زمانی رطوبت خاک (شکل ۴)، که به‌صورت واحد حجمی (m^3/m^3) اندازه‌گیری شده، کاهش اولیه ملایمی از اوج‌های اولیه حدود ۰/۲ در اواخر ژانویه به یک سطح ثابت در حدود ۰/۱۵ تا اواسط فوریه را نشان می‌دهد، با نوسانات بعدی (بازه: ۰/۱ تا ۰/۲۵). خط روند تخت در حدود ۰/۱۵ نشان‌دهنده عدم تغییر قابل توجه در طول بازه مورد مطالعه است، اگرچه افزایش‌های کوتاه‌مدت (مثلاً قله‌های میانه مارس که از ۰/۲۵ فراتر می‌رود) ممکن است با رویدادهای بارندگی که لایه‌های سطحی خاک را پر می‌کنند، مرتبط باشد. مقادیر کمتر در آوریل (۰/۱ تا ۰/۱۵) به خشک شدن تدریجی اشاره دارد که می‌تواند فرسایش را تشدید کرده یا جذب آب گیاهان را کاهش دهد.

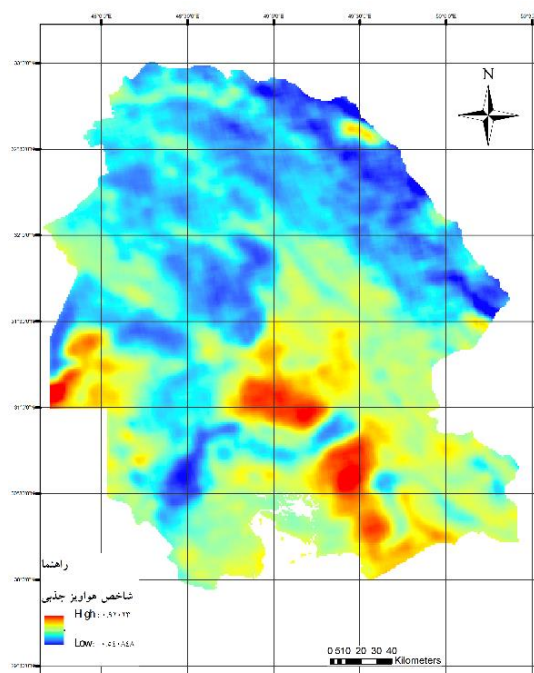
سری زمانی شاخص هواویز جذبی (شکل ۵) یک روند مثبت ملایم دارد که از حدود ۰/۲- در ژانویه به ۰/۲ تا اواخر آوریل افزایش یافته، در حالی که نوسانات زیادی دارد. مقادیر مثبت زمانی شاخص هواویز جذبی نشان‌دهنده حضور آئروسول‌های جاذب مانند گرد و خاک است، در حالی که مقادیر منفی نشان‌دهنده ذرات غیرجاذب مانند سولفات‌ها هستند. کاهش‌های اولیه منفی ممکن است شرایط اتمسفری شفاف‌تر یا غالب بودن هواویزهای پراکنده‌کننده را منعکس کند. با این حال، افزایش قله‌های مثبت از مارس به



شکل ۶. میانگین شاخص پوشش گیاهی نرمال شده در بازه زمانی ۲۶ ژانویه تا ۲۶ آوریل ۲۰۲۵.

به ویژه اطراف اهواز و هورالعظیم، مقادیر زیادتر شاخص هواویز جذبی (بین ۰/۲ تا ۰/۵) را نشان می دهند که با حضور هواویزهای ناشی از گرد و خاک مرتبط است. این الگو با مناطق خشک و بیابانی جنوب غربی خوزستان همخوانی دارد و تأیید می کند که شاخص هواویز جذبی می تواند به عنوان شاخصی کلیدی برای شناسایی کانون های گرد و خاک و پایش کیفیت هوا در این منطقه عمل کند.

تصویر میانگین شاخص هواویز جذبی برای استان خوزستان (شکل ۷) به دست آمده است، توزیع هواویزها در این منطقه را نشان می دهد. در این تصویر، رنگ آبی نمایانگر مقادیر کم شاخص هواویز جذبی (کمتر از ۰/۲) هستند که در شمال و شرق خوزستان، مانند دزفول و ایذه، مشاهده می شود و نشان دهنده وجود هوای پاک تر یا پوشش گیاهی متراکم تر است. رنگ های زرد، نارنجی و قرمز در مرکز و جنوب،



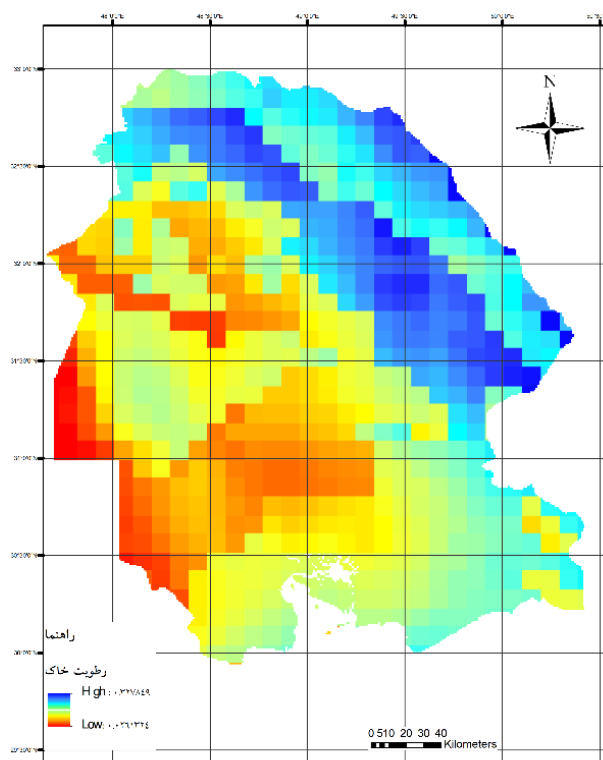
شکل ۷. میانگین شاخص هواویز جذبی در بازه زمانی ۲۶ ژانویه تا ۲۶ آوریل ۲۰۲۵.

انتظار می‌رود مناطق با شاخص نرمال‌شده تفاوت پوشش گیاهی کم، شاخص هواویز جذبی زیاد، و رطوبت خاک کم به‌عنوان کانون‌های گرد و خاک شناسایی شوند.

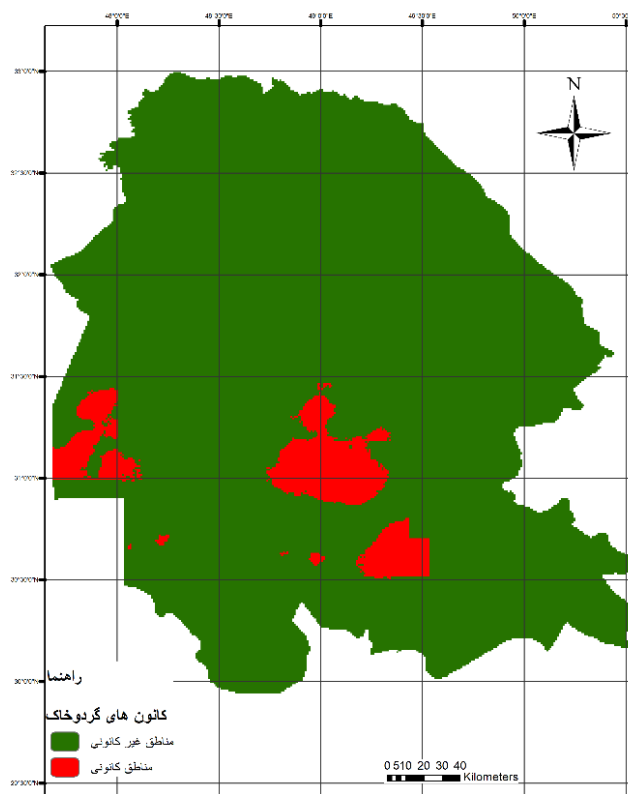
۴-۲. طبقه‌بندی کانون‌های استان خوزستان

نقشه طبقه‌بندی (شکل ۹) نشان می‌دهد که مناطق وسیعی در جنوب، مرکز و غرب خوزستان، به‌ویژه جنوب شرق شهر اهواز و هورالعظیم (جنوب‌غربی اهواز)، به‌عنوان کانون‌های گرد و خاک شناسایی شدند. این مناطق شامل بخش‌هایی از شهرهای اهواز، رامشیر (کانون جنوب شرق اهواز)، بندرماهشهر، امیدیه، هندیجان، سوسنگرد و هویزه (کانون هور هویزه) هستند که با رنگ قرمز مشخص شده‌اند. مناطق سبز نشان‌دهنده نواحی غیرکانونی، عمدتاً در شمال و شرق خوزستان، هستند که پوشش گیاهی بهتری دارند.

تصویر میانگین رطوبت سطحی خاک (شکل ۸) برای استان خوزستان، که از داده اسمپ در بازه زمانی ۲۶ ژانویه تا ۲۶ آوریل ۲۰۲۵ استخراج شده است، توزیع رطوبت در این منطقه را نشان می‌دهد. در این تصویر، رنگ‌های قرمز و نارنجی نمایانگر رطوبت کم (کمتر از $0.1 \text{ cm}^3/\text{cm}^3$) هستند که در بخش‌های جنوبی و مرکزی، به‌ویژه اطراف هورالعظیم، اهواز، و رامشیر، غالب است و نشان‌دهنده شرایط خشک و مستعد فرسایش بادی است. رنگ‌های زرد و سبز، که رطوبت متوسط تا زیاد (بین 0.1 تا $0.3 \text{ cm}^3/\text{cm}^3$) را نشان می‌دهند، در شمال و شرق خوزستان، مانند دزفول و ایذه، دیده می‌شوند که با وجود پوشش گیاهی و ارتفاعات بیشتر همخوانی دارد. این الگو تأیید می‌کند که رطوبت خاک کم در مناطق جنوبی نقش مهمی در ایجاد کانون‌های گرد و خاک دارد و با ترکیب این داده با شاخص نرمال‌شده تفاوت پوشش گیاهی و شاخص هواویز جذبی، می‌توان دقت شناسایی این مناطق را افزایش داد.



شکل ۸. میانگین رطوبت خاک ماهواره اسمپ در بازه زمانی ۲۶ ژانویه تا ۲۶ آوریل ۲۰۲۵.



شکل ۹. نقشه کانون‌های گرد و خاک در خوزستان (رنگ سبز برای مناطق غیر کانونی، رنگ قرمز برای مناطق کانون گرد و خاک).

شناسایی صحیح نقاط کانونی است. امتیاز F1 برابر با ۰/۹۵ و دقت کلی ۰/۹۵ به دست آمده است، که عملکرد مدل را در سطح بسیار خوب نشان می‌دهد. ضریب کاپا ۰/۹ بود، که بیانگر توافق زیاد بین پیش‌بینی مدل و داده‌های واقعی و فراتر از پیش‌بینی تصادفی است.

جدول ۱. نتایج اعتبارسنجی مدل.

۰/۹۳۵۴۸۴	دقت مثبت
۰/۹۶۶۶۶۷	بازخوانی
۰/۹۵۰۸۲	امتیاز F1
۰/۹۵	دقت کلی
۰/۹	ضریب کاپا

جدول ۲ نتایج آزمون و ارزیابی نهایی مدل را نشان می‌دهد. دقت مثبت و بازخوانی هر دو برابر با ۰/۹۶۷ بودند و امتیاز F1 نیز همین مقدار به دست آمده، که نشان‌دهنده ثبات عملکرد مدل در داده‌های جدید و نادیده است. دقت کلی در آزمون نهایی ۰/۹۶۷ و ضریب کاپا ۰/۹۳۳ به دست آمد، که تأییدکننده قابلیت اعتماد و قدرت تعمیم مدل است.

جنوب شرق اهواز یکی از نواحی مستعد گرد و خاک در خوزستان است. تداوم خشکسالی، کاهش بارندگی و افت جریان‌های سطحی رودخانه‌های جراحی و کارون، فرایند بیابان‌زایی را در این منطقه تشدید کرده است. کاهش چشمگیر بارش در سال‌های گذشته سبب فعال‌شدن دوباره آن شده است. کانون محدوده‌ی ماهشهر و هندیجان در برگرفته مناطق رسی وسیع و فاقد پوشش گیاهی است و در گذشته به صورت آبگیر بوده است. خشکسالی و کاهش بارندگی و عدم وجود سیلاب‌های رودخانه‌ای باعث خشک شدگی شدید این محدوده شده است. همچنین کمبود آب ورودی به تالاب هورالعظیم و کاهش بارندگی باعث خشک شدن بخش‌های جنوبی تالاب شده است.

۳-۴. ارزیابی مدل جنگل تصادفی

نتایج اعتبارسنجی مدل بر اساس ماتریس درهم‌ریختگی در جدول ۱ آمده است. دقت مثبت برابر با ۰/۹۳۵ و بازخوانی برابر با ۰/۹۶۷ بود، که نشان‌دهنده توانایی بالای مدل در

تغییر بازه‌های زمانی را فراهم می‌کند. با تنظیم بازه‌های کوتاه‌تر (مانند ماهانه یا هفتگی)، هواشناسان می‌توانند میانگین‌های به‌روز شاخص‌ها را محاسبه کرده و تغییرات سریع در وضعیت کانون‌های گرد و خاک را رصد کنند، که این امر به پیش‌بینی به‌موقع طوفان‌ها و برنامه‌ریزی‌های مدیریت بحران کمک شایانی می‌کند.

نتایج این مطالعه با یافته‌های رنگرن و همکاران (۱۳۹۳) هم‌خوانی دارد که همبستگی زیادی بین هواویزها و گرد و خاک در خوزستان گزارش کردند. شناسایی هورالعظیم به‌عنوان کانون اصلی گرد و خاک با مطالعات میدانی پیشین نیز سازگار است. استفاده از رطوبت خاک به‌عنوان متغیر جدید، دقت شناسایی را بهبود بخشید، زیرا خاک خشک‌تر پتانسیل بیشتری برای فرسایش بادی دارد.

این یافته‌ها با گزارش سازمان زمین‌شناسی و اکتشافات معدنی ایران در تحقیق نظری سامانی و همکاران (۱۳۹۹) هم‌راستا است، که کانون‌های اصلی را شامل جنوب و جنوب شرق اهواز (۱۱۲ هزار هکتار)، بندر امام-امیدیه (۸۸ هزار هکتار)، هورالعظیم (۵۳ هزار هکتار)، ماهشهر-هندیجان (۳۳/۵ هزار هکتار)، شمال و شرق خرمشهر (۲۸ هزار هکتار)، شرق هندیجان (۱۸ هزار هکتار)، و شرق اهواز (۱۵ هزار هکتار) معرفی کرده بود. هرچند مساحت دقیق کانون‌ها در مطالعه ما به‌صورت عددی محاسبه نشده، اما تمرکز بر هورالعظیم و مناطق جنوبی و غربی با گزارش سازمان زمین‌شناسی هم‌خوانی دارد و نشان‌دهنده پایداری این مناطق به‌عنوان کانون‌های گرد و خاک است. دقت زیاد الگوریتم جنگل تصادفی نسبت به روش‌های سنتی سازمان زمین‌شناسی، که عمدتاً مبتنی بر داده میدانی و نقشه‌برداری بودند، بیانگر پتانسیل بالای ترکیب سنجش از دور و یادگیری ماشین است. انتخاب ۲۰۰ نقطه نمونه به‌برای هر کلاس، تنوع مکانی خوبی را پوشش داد و دقت کلی مدل در اعتبارسنجی را به ۰/۹۵ و در آزمون و ارزیابی نهایی ۰/۹۶۷ رساند، که نشان‌دهنده اهمیت تعداد کافی نقاط در بهبود عملکرد مدل است.

با این حال، این مطالعه محدودیت‌هایی نیز داشت. عدم دسترسی به داده میدانی برای اعتبارسنجی در پلتفرم گوگل ارث انجین، ممکن است دقت واقعی مدل را تحت تأثیر

جدول ۲. نتایج آزمون و ارزیابی نهایی مدل.

۰/۹۶۶۶۶۷	دقت مثبت
۰/۹۶۶۶۶۷	بازخوانی
۰/۹۶۶۶۶۷	امتیاز F1
۰/۹۶۶۶۶۷	دقت کلی
۰/۹۳۳۳۳۳	ضریب کاپا

۵. نتیجه‌گیری

این مطالعه با موفقیت کانون‌های گرد و خاک در خوزستان را با استفاده از شاخص نرمال‌شده تفاوت پوشش گیاهی، شاخص هواویز جذبی، و رطوبت خاک و الگوریتم جنگل تصادفی شناسایی کرد. مناطق جنوب و غرب خوزستان، به‌ویژه اطراف هورالعظیم، به‌عنوان کانون‌های اصلی گرد و خاک تأیید شدند. این روش می‌تواند به‌عنوان ابزاری مؤثر برای پایش و مدیریت طوفان‌های گرد و خاک در مناطق خشک استفاده شود. جنگل تصادفی به‌دلیل توانایی در مدیریت داده ناهمگن و غیرخطی، عملکرد بسیار خوبی از خود نشان داد. این الگوریتم توانست روابط پیچیده بین شاخص پوشش گیاهی نرمال‌شده، شاخص هواویز جذبی، و رطوبت خاک را مدل‌سازی کند و به شناسایی دقیق کانون‌های گرد و خاک کمک کند. بر اساس ارزیابی‌های انجام‌شده بر روی مدل جنگل تصادفی برای شناسایی کانون‌های گرد و خاک در استان خوزستان، نتایج حاصل از داده‌های اعتبارسنجی (دقت کلی ۰/۹۵، امتیاز F1 ۰/۹۵۱، ضریب کاپا ۰/۹۶۷) و آزمون نهایی (دقت کلی ۰/۹۶۷، امتیاز F1 ۰/۹۶۶۷، ضریب کاپا ۰/۹۳۳) نشان‌دهنده عملکرد بسیار قوی و پایدار مدل است. تفاوت اندک (حدود ۳- درصد) بین معیارهای عملکرد در اعتبارسنجی و آزمون نهایی، همراه با مقادیر بالای ضریب کاپا، حاکی از آن است که مدل به‌خوبی تعمیم یافته و بیش‌برازش قابل توجهی رخ نداده است. این پایداری عملکرد، به‌ویژه در شناسایی دقیق کانون‌های گرد و خاک، نشان‌دهنده قابلیت اطمینان مدل برای استفاده در کاربردهای عملیاتی و پیش‌بینی‌های محیطی است. این مقاله ابزاری کاربردی برای پیش‌بینی طوفان‌های گرد و خاک در اختیار هواشناسان قرار می‌دهد، به‌ویژه با قابلیت انعطاف‌پذیری کد در پلتفرم گوگل ارث انجین، که امکان

قرار داده باشد. همچنین، داده ماهواره‌ای ممکن است تحت تأثیر شرایط جوی (مانند ابرناکی) یا خطاهای پیش‌پردازش قرار گیرند. پیشنهاد می‌شود در مطالعات آینده از داده میدانی برای اعتبارسنجی استفاده شود و تعداد نقاط نمونه حتی بیشتر افزایش یابد تا تنوع بیشتری از شرایط محیطی پوشش داده شود.

علاوه بر این، می‌توان از روش‌های پیشرفته‌تر یادگیری ماشین، مانند یادگیری عمیق، برای بهبود دقت استفاده کرد، اگرچه این روش‌ها به داده آموزشی بیشتری نیاز دارند. همچنین، ترکیب داده ماهواره‌ای با مدل‌های هواشناسی (مانند سرعت باد و جهت آن) می‌تواند به پیش‌بینی دقیق‌تر طوفان‌های گرد و خاک کمک کند. در نهایت، این روش می‌تواند به عنوان ابزاری برای سیاست‌گذاری و مدیریت منابع در مناطق مستعد گرد و خاک به‌ویژه در زمینه احیای تالاب‌ها و تثبیت خاک در مناطق بیابانی مورد استفاده قرار گیرد.

مراجع

بحرینی، ف.؛ پناهی، ف.؛ جعفری، م. و ملکیان، آ. (۱۳۹۷). شناسایی مناطق آسیب‌پذیر پوشش گیاهی به خشکسالی با استفاده از سنجش از دور (مطالعه موردی: استان بوشهر). *مجله منابع طبیعی ایران (مدیریت مرتع و آبخیزداری)*، ۷۱ (۲)، ۳۴۱-۳۵۴.

بروغنی، م.؛ میرچولی، ف. و محمدی، م. (۱۴۰۱). پهنه‌بندی آسیب‌پذیری گرد و غبار با استفاده از تصاویر ماهواره‌ای و مدل‌های یادگیری ماشین. *مطالعات جغرافیایی مناطق*

random forest to learn imbalanced data (Research Report). University of California, Berkeley.
Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792.
Gandham, H., Dasari, H. P., Saharwardi, M. S., Karumuri, A., & Hoteit, I. (2023). *Dust sources over the Arabian Peninsula. Environmental Research Letters*, 18(9), Article 094053.
Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques (3rd ed.)*. Elsevier.
Jafari, R., Amiri, M., Asgari, F., & Tarkesh, M. (2022). Dust source susceptibility mapping based on remote sensing and machine learning techniques. *Ecological Informatics*, 72, 101872.

خشک، ۱۳ (۴۷)، ۱-۱۳.

رایگانی، ب.؛ براتی قهفرخی، س. خوشنوا، ا. (۱۳۹۸). شناسایی کانون‌های تولید گرد و غبار به کمک داده‌های دورسنجی: یک رویکرد جامع. *مرتع و آبخیزداری (منابع طبیعی ایران)*، ۷۲ (۱)، ۸۳-۱۰۵.

رنگزن، ک.؛ زراسوندی، ع. ر.؛ عبدالحانی، ع. و مجردی، ب. (۱۳۹۳). مدل‌سازی آلودگی هوا با استفاده از تصاویر سنجنده مودیس: مطالعه موردی توده‌های گرد و غبار استان خوزستان. *زمین‌شناسی کاربردی پیشرفته*، ۴ (۴)، ۳۸-۴۵.

رشیدی، م. ج.؛ فلاح، ا. و شتابی جویباری، ش. (۱۴۰۲). ارزیابی و مقایسه کرنل‌های الگوریتم ماشین بردار پشتیبان برای تشخیص گونه‌های کنار و گون با بهره‌گیری از تصاویر رقومی هوایی UltraCam-D (مطالعه موردی سازند میشان منطقه تنگ ارم استان بوشهر). *در بیست و ششمین همایش انجمن زمین‌شناسی ایران، ارومیه*.

زنگنه، ا. (۱۳۹۸). *پیش‌بینی مکانی-زمانی داده‌های سنجش از دور با استفاده از یادگیری عمیق*. پایان‌نامه کارشناسی ارشد. به راهنمایی هدی مشایخی و سعیده فردوسی. *شاهروود: دانشگاه صنعتی شاهرود*.

نظری سامانی، ع. ا.؛ زهتاییان، غ. و علم‌یگی، ا. (۱۳۹۹). تعیین چارچوب مفهومی ریسک گرد و غبار بر مبنای سنجش تاب‌آوری (بررسی موردی: جنوب غرب کشور). *پژوهش‌های راهبردی در علوم کشاورزی و منابع طبیعی*، ۵ (۱)، ۳۳-۴۴.

Ahmady-Birgani, H. (2025). *Conceptualizing dust emission areas and hotspots over the Aeolian landforms via remote-sensing aerosol algorithms: case study, Lake Urmia, a major hypersaline lake in the Middle East*. *Geoscience Letters*, 12, Article 28.
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, 24-31.
Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197-227.
Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
Chen, C., Liaw, A., & Breiman, L. (2004). Using

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.