

The Journal of Spatial Planning & Geomatics

Research Paper

Evaluation of Machine Learning Algorithms for Dust Source Susceptibility Mapping by Integrating Remote Sensing and Environmental Parameters (Case Study: Kara-Bogaz- Gol Basin)

Farideh Bamerinejad¹, Ali Mohammadian Behbahani*², Chooghi
Bairam Komaki³, Ramin Papi⁴, Ali Darvishi Boloorani⁵

1. Ph.D. Student in Desert Management and Control, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran.
2. Associate Professor in Arid zone Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran.
3. Associate Professor in Arid zone Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran.
4. PhD. in Remote Sensing and Geographic Information Systems, National Cartographic Center (NCC), Tehran, Iran.
5. Professor in Remote Sensing and GIS, University of Tehran, Tehran, Iran.

Received: 2025/07/15
Revised: 2025/10/28
Accepted: 2025/12/02

ABSTRACT

Dust storms, characterized by their capacity to transport aeolian sediments across extensive distances from their origins, represent significant threats to human societies and are linked with considerable detrimental impacts on public health, ecological systems, and the economic stability of communities. In light of the swift dispersal and extensive distribution patterns of dust particulates, coupled with their transportational dynamics via wind currents, numerous occurrences remain elusive to detection and monitoring, thereby necessitating the elucidation and spatial delineation of their origin regions. The Kara-Bogaz Gol Basin, alongside the Karakum Desert, has emerged as one of the principal contributors to dust emissions affecting Golestan Province in recent years. This study critically assesses the efficacy of sophisticated machine learning algorithms in pinpointing dust emission origins within the Kara-Bogaz Gol Basin. In the present investigation, a robust analytical framework founded on the amalgamation of remote sensing data and machine learning methodologies was employed. Environmental datasets, encompassing nine distinct parameters—namely, dusty days, soil moisture, soil texture, precipitation, wind velocity, Normalized Difference Vegetation Index (NDVI), Digital Elevation Model (DEM), air temperature, and land cover—were systematically extracted and processed utilizing the Google Earth Engine platform across the temporal scope of 2003-2023. A total of 340 dust emission sources were discerned through visual interpretation of MODIS satellite imagery, serving as training data for the machine learning algorithms. The findings revealed the following accuracies in identifying areas with high dust emission potential: Random Forest 91.8%, Artificial Neural Network 70.9%, XGBoost 89.9%, Gradient Boosting 87.9%, Bagged CART 89.9%, and LightGBM 91.8%. Notably, Random Forest and LightGBM exhibited superior performance in the identification of dust sources. An examination of explainability techniques indicated that three variables—vegetation index (27% contribution), soil moisture (23%), and DEM (19%)—exerted the most substantial influence on the prediction of dust-prone areas. The primary contribution of this research resides in the formulation of a hybrid machine learning framework adept at identifying regions susceptible to dust emissions and quantifying the role of environmental parameters in the genesis of these storms.

Keywords:

Dust sources, Machine Learning, Central Asia, Kara-Bogaz Gol, Aerosol Optical Depth, Remote sensing.

*Corresponding Author: Associate Professor in Arid zone Management, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran.

<https://orcid.org/0000-0002-4813-8151>
ali.mohammadian@gau.ac.ir



Copyright© 2025, TMU Press. This open-access article is published under the terms of the Creative Commons Attribution- NonCommercial 4.0 International License which permits Share (copy and redistribute the material in any medium or format) and Adapt (remix, transform, and build upon the material) under the Attribution-NonCommercial terms.

D

Introduction

ust storms emanating from the Kara-Bogaz-Gol Basin present substantial environmental and public health challenges within the northeastern region of Iran. This investigation addresses the pressing necessity for precise identification of dust sources by establishing an advanced machine learning framework that synthesizes remote sensing data alongside environmental parameters. The study emphasizes the mitigation of the limitations inherent in conventional dust monitoring methodologies through innovative computational strategies, thereby offering a thorough analysis of dust emission patterns in this ecologically vulnerable area from 2003 to 2023.

Research Method

The research utilized a multi-tiered analytical methodology that integrated satellite remote sensing with advanced machine learning techniques. Data were processed through Google Earth Engine to derive nine environmental variables, comprising dusty days, NDVI, soil moisture, and wind velocity. A visual interpretation technique successfully identified 340 dust hotspots, which were subsequently employed as training data for six distinct machine learning algorithms: Random Forest, Artificial Neural Networks, XGBoost, Gradient Boosting, Bagged CART, and LightGBM. The performance of the models was meticulously assessed utilizing the Kolmogorov-Smirnov test alongside 5-fold cross-validation, while SHAP values and Sobol sensitivity analysis contributed to the interpretability of the findings.

Results & discussion

The comprehensive analysis elucidated that both Random Forest and LightGBM attained exceptional performance metrics, with accuracies recorded at 91.8% for each model. The principal findings indicated that NDVI contributed 27%, soil moisture contributed 23%, and DEM accounted for 19%, thereby identifying these as the most pivotal factors affecting dust emissions. The investigation discerned distinct spatial patterns, where regions characterized by lower elevation displayed heightened susceptibility to dust, attributable to the synergistic effects of vegetation depletion and soil aridity. These findings are consistent with, yet also enhance, prior global dust models, particularly in delineating region-specific interactions between meteorological phenomena and terrestrial conditions. Furthermore, the SHAP analysis yielded innovative perspectives on nonlinear relationships, exemplified by the threshold effects of wind velocity on dust generation across various land cover types.

Conclusion

This investigation delineates a notable enhancement in the identification of dust sources through the pioneering amalgamation of machine learning methodologies and remote sensing technologies. The established framework yields considerable advancements in predictive accuracy and interpretability when juxtaposed with traditional approaches, bearing significant ramifications for environmental stewardship in arid zones. The methodology employed in the study exhibits the potential for modification and application in other dust-affected regions globally, while the resultant findings furnish a robust scientific foundation for the formulation of targeted strategies aimed at dust mitigation. Prospective avenues for research encompass the integration of high-resolution datasets and the extension of temporal analyses to evaluate the repercussions of climate change on the frequency and intensity of dust storms.

ارزیابی الگوریتم‌های یادگیری ماشین برای نقشه‌برداری از حساسیت منابع گردوغبار با تلفیق سنجش‌ازدور و پارامترهای محیطی (مطالعه موردی: حوضه خلیج قره‌بغاز)

فریده بامری‌نژاد^۱، علی محمدیان بهبهانی^{۲*}، چوقی بایرام کمکی^۳، رامین پایی^۴، علی درویشی بلورانی^۵

۱. دانشجوی دکتری مدیریت و کنترل بیابان، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، گرگان، ایران.
۲. دانشیار گروه مدیریت مناطق بیابانی، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، گرگان، ایران (نویسنده مسئول).
۳. دانشیار گروه مدیریت مناطق بیابانی، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، گرگان، ایران.
۴. دکتری سنجش از دور و سامانه اطلاعات جغرافیایی، سازمان نقشه‌برداری کشور، تهران، ایران.
۵. استاد گروه سنجش از دور و سامانه اطلاعات جغرافیایی، دانشگاه تهران، تهران، ایران.

چکیده

طوفان‌های گردوغبار به دلیل قابلیت انتقال طولانی مدت رسوبات بادی به مناطق دور از منشأ آن‌ها، خطرات متعددی را برای جوامع انسانی به همراه دارند و با اثرات منفی زیادی بر سلامت انسان، محیط‌زیست و اقتصاد جوامع همراه هستند. با توجه به ماهیت پخش و پراکنش سریع ذرات گردوغبار و جابه‌جایی به‌وسیله باد، بسیاری از رخدادهای گردوغباری به‌راحتی قابل تشخیص و پایش نیستند و به همین جهت نیاز به شناسایی و توصیف رفتار مکانی کانون‌های آن‌ها وجود دارد. محدوده خلیج قره‌بغاز علاوه بر صحرای قره‌قوم یکی از کانون‌های بحران برداشت گردوغبار ورودی به استان گلستان در سال‌های اخیر بوده است، این پژوهش به ارزیابی کارایی الگوریتم‌های پیشرفته یادگیری ماشین در شناسایی کانون‌های تولید گردوغبار در حوضه خلیج قره‌بغاز می‌پردازد. در این پژوهش، از یک چارچوب تحلیلی جامع مبتنی بر تلفیق داده‌های سنجش‌ازدور و روش‌های یادگیری ماشین استفاده شده است. مجموعه داده‌های محیطی شامل ۹ پارامتر روزهای گردوغباری، رطوبت خاک، بافت خاک، بارش، سرعت باد، شاخص نرمال‌شده تفاوت پوشش گیاهی، توپوگرافی، دمای هوا و پوشش زمین با استفاده از پلتفرم گوگل ارشانه‌چین برای بازه زمانی (۲۰۰۳-۲۰۲۳) استخراج و پردازش شدند. تعداد ۳۴۰ کانون مولد گردوغبار از طریق رویکرد تفسیر چشمی تصاویر ماهواره‌ای مادیس شناسایی شده، به‌عنوان نقاط آموزشی الگوریتم‌های یادگیری ماشین مورد استفاده و ارزیابی قرار گرفتند. نتایج نشان داد مدل جنگل تصادفی با ۹۱/۸ درصد، شبکه عصبی مصنوعی با ۷۰/۹ درصد، تقویت گرادینانی پیشرفته با ۸۹/۹ درصد، تقویت گرادینانی با ۸۷/۹ درصد، الگوریتم درختان طبقه‌بندی و رگرسیون با ۸۹/۹ درصد و ماشین تقویت گرادینانی با ۹۱/۸ درصد مناطق با پتانسیل بالای گردوغبار را تشخیص داده‌اند که مدل‌های جنگل تصادفی و ماشین تقویت گرادینانی با LightGBM بهترین عملکرد را در شناسایی کانون‌های گردوغبار داشته‌اند. بررسی روش‌های توضیح‌پذیری نشان داد که سه متغیر شاخص نرمال‌شده تفاوت پوشش گیاهی با سهم (۲۷٪)، رطوبت خاک (۲۳٪) و ارتفاع از سطح دریا (۱۹٪) بیشترین تأثیر را در پیش‌بینی مناطق مستعد گردوغبار دارند.

تاریخ دریافت: ۱۴۰۴/۰۴/۲۴
آخرین ویرایش: ۱۴۰۴/۰۸/۰۶
تاریخ پذیرش: ۱۴۰۴/۰۹/۱۱

واژگان کلیدی:

منابع گردوغبار، یادگیری ماشین، آسیای میانه، خلیج قره‌بغاز، عمق نوری آتروسول، سنجش‌ازدور.

ali.mohammadian@gau.ac.ir

* نویسنده مسئول

کپی‌رایت © ۲۰۲۵، انتشارات دانشگاه تربیت مدرس (TMU Press). این مقاله به‌صورت دسترسی آزاد منتشر شده و تحت مجوز بین‌المللی Creative Commons Attribution-NonCommercial 4.0 قرار دارد. بر اساس این مجوز، شما می‌توانید این مطلب را در هر قالب و رسانه‌ای کپی، بازنشر و بازآفرینی کنید و یا آن را ویرایش و بازسازی نمایید، به شرط آنکه نام نویسنده را ذکر کرده و از آن برای مقاصد غیرتجاری استفاده کنید.

۱. مقدمه

طوفان‌های گردوغبار به‌طور وسیع در مناطق خشک و نیمه‌خشک جهان رخ می‌دهد. منابع اصلی گردوغبار جهان اغلب در فرورفتگی‌های^۱ توپوگرافی در مناطق خشک و نیمه‌خشک با نرخ بارندگی کم قرار دارند که مواد آبرفتی را از طریق رودخانه‌ها دریافت می‌کنند. بر این اساس، بزرگ‌ترین منابع جهانی طوفان‌های گردوغبار در نیمکره شمالی، عمدتاً در امتداد کمربند عرض جغرافیایی میانه، از سواحل غربی شمال آفریقا به سمت خاورمیانه، آسیای مرکزی و در نهایت مغولستان واقع شده است (Middleton, 2017; Papi, et al., 2022).

وقوع طوفان‌های گردوغبار اساساً تابعی از شرایط جغرافیایی و اقلیمی است. بنابراین، منابع گردوغبار واقع در مناطق جغرافیایی مختلف ممکن است منشأ و انواع متفاوتی داشته باشند. علاوه بر این، پارامترهای محیطی مختلف مانند سرعت باد، تبخیر و تعرق، بارش، خواص فیزیکی و شیمیایی خاک (مانند رطوبت خاک، بافت، و مواد آلی)، پوشش گیاهی و توپوگرافی به‌عنوان عوامل مؤثر طوفان‌های گردوغبار شناخته می‌شوند (Papi et al., 2022). این محرک‌ها به‌صورت تعاملی در تشکیل گردوغبار نقش دارند. به‌عنوان مثال، منابع آب بر پوشش گیاهی تأثیر می‌گذارد و به‌عنوان یک عامل اساسی در انتشار طوفان گردوغبار شناخته می‌شود (Hamidi, 2020). کاهش رطوبت خاک باعث کاهش چسبندگی بین ذرات شده و به همین دلیل خاک‌های خشک در مقایسه با خاک‌های مرطوب، در مقابل باد فرسایش‌پذیری بیشتری دارند. این امر یکی از دلایلی است که طوفان‌های گردوغبار با اقلیم‌های خشک مرتبط هستند (پاپی و همکاران، ۱۴۰۱). کاهش رطوبت سطحی خاک و در نتیجه آن کاهش پوشش گیاهی در کانون‌های خارجی واقع در کشورهای غرب ایران باعث افزایش وقوع گردوغبار در کل منطقه جنوب غرب و حتی مرکز ایران می‌شود (مطیعی و همکاران، ۱۴۰۱). شناسایی منابع و محرک‌های طوفان‌های گردوغبار مؤثرترین رویکرد برای مدیریت پیامدهای این پدیده است. محرک‌های مؤثر طیف وسیعی از پارامترها از جمله تضعیف پوشش گیاهی (Li et al., 2020)، شکل‌گیری بسترهای خشک‌شده پهنه‌های آبی (Darvishi Boloorani et al., 2021)، سرعت باد (Ebrahimi-Khusfi et al., 2020) هستند. در این میان، فعالیت‌های انسانی در تغییرات کاربری زمین و منابع آب نیز می‌تواند تأثیرات این محرک‌ها را تشدید کند (Papi et al., 2022).

در سال‌های اخیر، از روش‌های یادگیری ماشین به‌صورت گسترده‌ای برای پیش‌بینی رفتار کانون‌های گردوغبار و تهیه نقشه‌های پتانسیل فرسایش‌پذیری استفاده شده است. لری^۲ و همکاران (۲۰۱۶)، برای اولین بار پتانسیل بالای الگوریتم‌های یادگیری ماشین را در طبقه‌بندی و شناسایی کانون‌های گردوغبار نشان دادند. نبوی^۳ و همکاران (۲۰۱۸)، پنج الگوریتم یادگیری ماشین شامل رگرسیون خطی^۴، چندگانه، جنگل تصادفی^۵، رگرسیون چندجمله‌ای تطبیقی

^۱ - Depression

^۲ - Lary

^۳ - Nabavi

^۴ - Linear regression

^۵ - Random Forest

چندمتغیره^۱، ماشین بردار پشتیبان^۲ و شبکه عصبی مصنوعی^۳ را برای پیش‌بینی عمق نوری آئروسول (AOD)^۴ در غرب آسیا معرفی کردند. غلامی^۵ و همکاران (۲۰۲۰)، شش الگوریتم یادگیری ماشین را برای بررسی حساسیت زمین به انتشار گردوغبار در جنوب شرق ایران به کار بردند. غلامی و همکاران (۲۰۲۱)، از الگوریتم‌های یادگیری ماشین برای تهیه نقشه‌های مکانی کانون‌های گردوغبار و ارزیابی قابلیت تفسیر این نقشه‌ها در آسیای میانه ارائه کردند. پورهایمی^۶ و همکاران (۲۰۲۲)، مطالعه‌ای عوامل مؤثر بر گردوغبار بررسی و نقشه حساسیت در استان خوزستان در ایران و استان‌های بصره و میسان در عراق انجام دادند. در این تحقیق، از تصاویر ماهواره‌ای و تکنیک‌های یادگیری ماشین از جمله خطوط رگرسیون تطبیقی چندمتغیره، جنگل تصادفی و رگرسیون لجستیک^۷ جهت شناسایی منبع گردوغبار و نقشه حساسیت استفاده شد. ۱۵۲ کانون گردوغبار برای یک دوره ۲۰۲۰-۲۰۰۵ در منطقه مورد مطالعه شناسایی شد که ۷۰ درصد آن‌ها به‌عنوان داده آموزشی جهت آموزش مدل‌ها و ۳۰ درصد برای اعتبارسنجی استفاده شد. اعتبارسنجی مدل‌ها توسط منحنی ROC^۸ انجام گرفت. نتایج مدل‌ها نشان داد که کاربری اراضی بیشترین تأثیر را در ایجاد کانون‌های گردوغبار داشته است (Pourhashemi, 2022).

به دلیل قرار گرفتن ایران در منطقه خشک و نیمه‌خشک جهان، میزان وقوع گردوغبار در این منطقه به‌طور قابل‌توجهی افزایش یافته است. همچنین، نزدیکی به کشورهای عراق، سوریه، عربستان، ترکمنستان و افغانستان که مناطق وسیعی از بیابان را در بردارند، نقش مؤثری در این پدیده داشته است (عبدالشاه نژاد و همکاران، ۱۳۹۹). نتایج بررسی فعالیت‌های گردوخاک فصلی و شبیه‌سازی عددی پدیده گردوغبار در شمال شرق ایران و بررسی فصلی مقادیر AOD نشان داد که بیابان‌های آرال‌قوم، قزل‌قوم، قره‌قوم و قره‌باز به‌عنوان کانون‌های اصلی گردوغبار فرامرزی و سرخس به‌عنوان منبع داخلی در شمال شرق هستند (زرین و همکاران، ۱۴۰۱). با توجه به بررسی منابع، می‌توان اذعان کرد که علیرغم ردیابی فرآیندهای حمل‌ونقل طوفان‌های گردوغبار، ارزیابی، پایش و شناسایی منابع گردوغبار با استفاده از مجموعه داده‌ها و رویکردهای مختلف، هنوز هیچ مطالعه یکپارچه و جامعی از منابع گردوغبار در حوضه خلیج قره‌باز انجام نشده است. بر این اساس، مطالعه حاضر به منظور رفع این خلاء پژوهشی با در نظر گرفتن پارامترهای محیطی و کانون‌های فعال گردوغبار برای شناسایی منابع گردوغبار با پتانسیل بالا در این حوضه انجام گردید.

مطالعه حاضر با هدف شناسایی منابع انتشار گرد و غبار در حوضه خلیج قره‌باز با استفاده از یک رویکرد ترکیبی که تفسیر چشمی تصاویر ماهواره‌ای را با مدل‌های یادگیری ماشین ترکیب می‌کند، انجام شده است. برای دستیابی به این هدف، الگوریتم‌های پیشرفته یادگیری ماشین شامل جنگل تصادفی، شبکه عصبی مصنوعی، تقویت گرادیانی

^۱- Multivariate Adaptive Regression Splines

^۲- Support Vector Machine

^۳- Artificial Neural Network

^۴- Aerosol Optical Depth

^۵- Gholami

^۶- Pourhashemi

^۷- Logistic Regression

^۸- Receiver Operating Characteristics

پیشرفته^۱، تقویت گرادیانی^۲، الگوریتم درختان طبقه‌بندی و رگرسیون CART^۳ و ماشین تقویت گرادیان LightGBM^۴ برای شناسایی مناطق مستعد انتشار گردوغبار به کار گرفته شدند. داده‌های ماهواره‌ای مادیس^۵ و متغیرهای اقلیمی و محیطی به عنوان ورودی استفاده شدند. در این چارچوب، از روش تفسیر چشمی تصاویر ماهواره‌ای (VISI)^۶ برای استخراج نمونه‌های آموزشی استفاده شد و سپس با استفاده از مدل‌های یادگیری ماشین به کل منطقه مورد مطالعه تعمیم داده شدند. عملکرد این مدل‌ها با استفاده از پلات کولموگروف-اسمیرنوف^۷ ارزیابی شد. علاوه بر این، توضیح‌پذیری مدل‌ها با استفاده از تکنیک‌های مختلف تجزیه و تحلیل ویژگی، مانند اهمیت ویژگی^۸، مقادیر SHAP^۹ و تجزیه و تحلیل حساسیت سوبل^{۱۰} انجام شد. با وجود پیشرفت‌های اخیر، ادغام VISI با یادگیری ماشین در مقیاس منطقه‌ای نسبتاً کمتر مورد بررسی قرار گرفته است. نوآوری این تحقیق در بهره‌گیری از این رویکرد یکپارچه برای افزایش دقت شناسایی منابع گردوغبار است و در نتیجه به مدیریت و برنامه‌ریزی زیست‌محیطی مؤثرتر کمک می‌کند.

۲. مواد و روش‌ها

۲-۱. منطقه مورد مطالعه

در حاشیه شرقی دریای خزر، خلیج قره‌بغاز با مساحتی نزدیک به ۳۷۰۵۷۹ کیلومتر مربع با طول جغرافیایی از ۵۸ درجه و ۱۴ دقیقه و ۱۳ ثانیه تا ۵۸ درجه و ۴۱ دقیقه و ۲۹ ثانیه شرقی عرض جغرافیایی از ۴۷ درجه و ۲۵ دقیقه و ۶ ثانیه تا ۴۷ درجه و ۴۵ دقیقه و ۳۶ ثانیه شمالی قرار گرفته و در مجاورت سرزمین ترکمنستان واقع شده است. این منطقه به دلیل تأثیر قابل توجه اقلیم بیابانی، شاهد تبخیر شدید آب است. افزایش میزان تبخیر و همچنین ساخت و بهره‌برداری از سدهای متعدد بر روند تغذیه آب، تأثیرات منفی قابل توجهی بر پدیده تغییرات مرفولوژیکی گذاشته است و این موارد موجب تشدید بحران کم‌آبی و تغییرات اکولوژیکی در این منطقه شده است (اسپهد، ۱۳۸۴). بخش‌هایی از گردوغبار منتشرشده حوضه دریای آرال و به‌ویژه از صحرای قره‌قوم از طریق این حوضه به محدوده دریای خزر وارد می‌شود و همچنین وجود پهنه‌های وسیع از اراضی بایر و خشک و تپه‌های ماسه‌ای در این منطقه در کنار خشکی حاشیه‌های خلیج قره‌بغاز شرایط انتشار گردوغبار را از این منطقه و ورود آن به استان گلستان را فراهم کرده است. منطقه مورد بررسی از نظر توپوگرافی بسیار متنوع است، به طوری که ارتفاع از حدود -۱۸۵ متر در دشت‌های پست تا ۱۸۶۵ متر در کوهستان‌های مرتفع تغییر می‌کند (شکل ۱).

¹ - Extreme Gradient Boosting

² - Gradient Boosting

³ - Bagged Classification & Regression Trees

⁴ - Light Gradient Boosting Machine

⁵ - MODIS

⁶ - Visual Interpretation of Satellite Imagery

⁷ - Kolmogorov-Smirnov

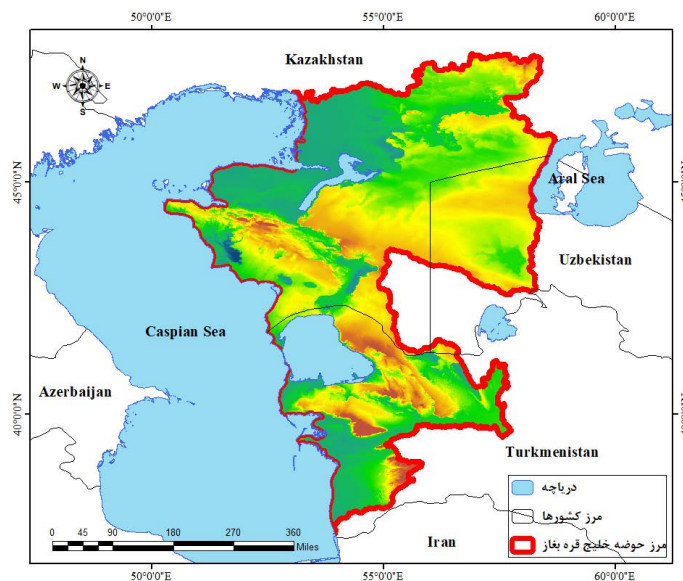
⁸ - Feature Importance

⁹ - Shapley Additive ExPlanations

¹⁰ - Sobol Sensitivity Analysis

شکل ۱. موقعیت جغرافیایی حوضه خلیج قره‌بغاز

Figure 1. Geographical location of the Kara-Bogaz-Gol Basin



۲-۲. داده‌ها

در این پژوهش مؤثرترین پارامترهای محیطی (جدول ۱) در انتشار گردوغبار انتخاب شده و با استفاده از تکنیک‌های تشخیص تغییرات و داده‌های سری زمانی ۲۰۰۳ تا ۲۰۲۳ از سامانه گوگل ارث‌انجین^۱ (GEE) فراخوانی شدند. تعداد ۹ پارامتر محیطی شامل، روزهای گردوغباری، رطوبت خاک، بافت خاک، بارش، سرعت باد، شاخص نرمال‌شده تفاوت پوشش گیاهی (NDVI)^۲، توپوگرافی، دمای هوا و پوشش زمین انتخاب شد. درنهایت جهت انجام مدل‌سازی همه لایه‌های اطلاعاتی تهیه‌شده پارامترها با قدرت تفکیک مکانی ۱۰۰۰ متر همسان‌سازی گردید.

جدول ۱. مشخصات داده‌های مکانی برای پارامترهای محیطی مؤثر بر گردوغبار

Table 1. Spatial data specifications of environmental parameters affecting dust

پارامتر	بازه زمانی	قدرت تفکیک مکانی	منبع
روزهای گردوغباری	۲۰۰۳ تا ۲۰۲۳	۱ کیلومتر	MODIS/Terra+Aqua
رطوبت خاک	۲۰۰۳ تا ۲۰۲۳	۴ کیلومتر	Terraclimate
بافت خاک	۲۰۲۳	۲۵۰ متر	OpenLandMap Soil Texture Class
بارش	۲۰۰۳ تا ۲۰۲۳	۴ کیلومتر	Terraclimate

^۱ - Google Earth Engine

^۲ - Normalized Difference Vegetation Index

Terraclimate	۴ کیلومتر	۲۰۰۳ تا ۲۰۲۳	سرعت باد
MODIS/Terra	۱ کیلومتر	۲۰۰۳ تا ۲۰۲۳	NDVI
NASA SRTM Digital Elevation 30m	۳۰ متر	۲۰۲۳	توپوگرافی
Terraclimate	۴ کیلومتر	۲۰۰۳ تا ۲۰۲۳	دما
MODIS/Terra+Aqua	۱ کیلومتر	۲۰۲۳	پوشش زمین

۲-۱-۲. همبستگی بین متغیرها

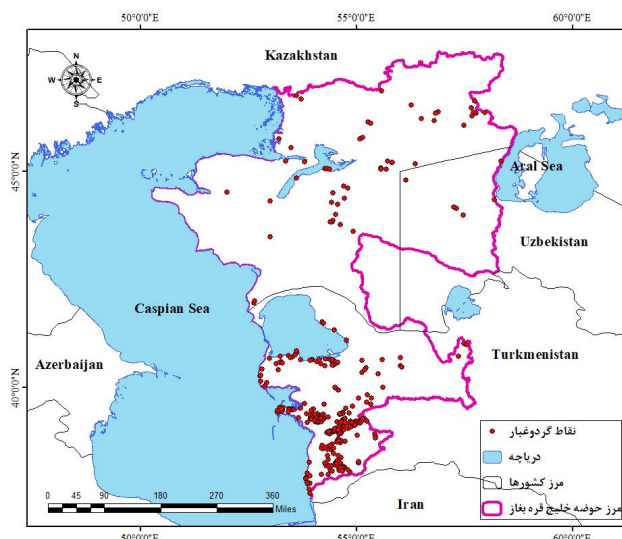
جهت محاسبه همبستگی بین متغیرها از بررسی هم خطی دو پارامتر عامل تورم واریانس (VIF) و (TC)^۱ استفاده شد و هیچگونه هم خطی بین پارامترها مشاهده نگردید و در نتیجه هیچ کدام از پارامترها حذف نشد.

۲-۳. شناسایی کانون‌های مولد گردوغبار با تفسیر چشمی تصاویر مادیس

تصاویر مادیس در حال حاضر بهترین منبع داده با وضوح مکانی و زمانی بالا برای شناسایی چشمی کانون‌های فعال گردوغبار (هات اسپات‌های گردوغبار) در نظر گرفته می‌شوند (Darvishi Boloorani et al., 2022). برای این منظور در این پژوهش جهت شناسایی کانون‌های فعال شکل‌گیری گردوغبار از قابلیت‌های تصاویر مادیس سامانه تحت وب ناسا^۲ در ترکیب رنگی طبیعی از سال ۲۰۰۳ تا ۲۰۲۳ در سایت <https://worldview.earthdata.nasa.gov/> به‌طور چشمی تفسیر شده که تعداد ۳۴۰ کانون رخداد گردوغبار شناسایی و به‌عنوان نقاط آموزشی معرفی شدند (شکل ۲).

شکل ۲. کانون‌های مولد گردوغبار شناسایی شده با استفاده از تفسیر چشمی تصاویر مادیس

Figure 2. Dust hotspots identified using visual interpretation of MODIS imagery



¹ - Tolerance

² - NASA's EOSDIS web-based application

فرآیند شناسایی منبع گردوغبار از طریق تفسیر چشمی تصاویر ماهواره‌ای در چهار مرحله اصلی انجام می‌شود، (شکل ۳) شماتیکی از مراحل را به‌عنوان نمونه نشان می‌دهد: ابتدا در مرحله شروع (a)، با عبور سرعت باد از آستانه فرسایش، ذرات گردوغبار از سطح زمین بلند می‌شوند؛ سپس در مرحله حداکثر کشند (b)، ابر گردوغبار به حداکثر غلظت و گستردگی فضایی خود می‌رسد؛ پس از آن، در مرحله جدا شدن (c)، انتشار گردوغبار متوقف شده و ابر از منبع جدا می‌شود؛ و در نهایت در مرحله پراکندگی (d)، ابر به تدریج محو می‌گردد (Darvishi Boloorani et al., 2023).

شکل ۳. چهار مرحله از یک رویداد گردوغبار (از ۱۵ تا ۱۸ فوریه ۲۰۰۵) برای شناسایی کانون‌های گردوغبار توسط تفسیر چشمی تصاویر

مادیس در منطقه مورد مطالعه (<https://worldview.earthdata.nasa.gov>)

Figure 3. Four stages of a dust event (from 15 to 18 February 2005) explored for hotspot dust sources identification (<https://worldview.earthdata.nasa.gov/>) in the study area by VISI



۲-۴. روش‌شناسی پژوهش

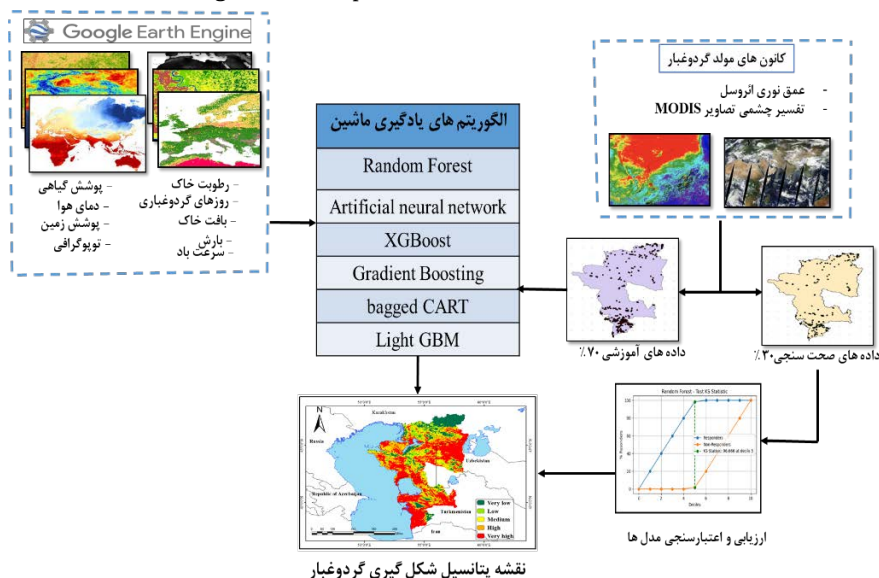
به‌منظور بررسی و تعیین مناطق حساس به انتشار گردوغبار، تحقیق حاضر براساس گام‌های ذیل انجام گردید. شکل ۴ مدل مفهومی و فرایند انجام پژوهش را نشان می‌دهد. در این مطالعه، ابتدا عوامل مؤثر بر مناطق منشأ گردوغبار شناسایی و تجزیه و تحلیل شدند و سپس با استفاده از مدل‌های یادگیری ماشین، مناطق مستعد گردوغبار پیش‌بینی شدند. برای

آمایش فضا و ژئوماتیک

ارزیابی عملکرد مدل‌ها از آزمون کولموگروف-اسمیرنوف استفاده شد. علاوه بر این، برای افزایش قابلیت تفسیر مدل‌ها، از تکنیک‌های مختلفی مانند اهمیت ویژگی، مقادیر SHAP و تحلیل حساسیت سوبل استفاده شد.

شکل ۴. مدل مفهومی و فرایند انجام پژوهش

Figure 4. Conceptual Model and Research Process



۲-۴-۱. مدل جنگل تصادفی

الگوریتم جنگل تصادفی یک الگوریتم یادگیری ماشین برای طبقه‌بندی و رگرسیون است. مدل جنگل تصادفی با به‌کارگیری رویکرد یادگیری ترکیبی، خروجی تعداد زیادی درخت تصمیم را برای طبقه‌بندی ادغام می‌کند. پیش‌بینی نهایی از طریق رأی اکثریت درختان به دست می‌آید. برای کاهش همبستگی بین درختان و جلوگیری از بیش‌برازش، در ساخت هر درخت از دو مکانیسم تصادفی‌سازی استفاده می‌شود. نمونه‌گیری از داده‌های آموزشی و انتخاب تصادفی زیرمجموعه‌ای از ویژگی‌ها در هر گره پارامترهای مدل با استفاده از جستجوی شبکه‌ای و با معیار ارزیابی دقت بهینه‌سازی شدند که بر اساس آن تعداد درختان ۵۰۰ عدد، تعداد ویژگی‌های کاندید در هر تقسیم معادل جذر تعداد کل ویژگی‌ها (معادل ۱۰ ویژگی)، و حداقل نمونه در گره برگ ۵ نمونه تعیین گردید. این پیکربندی بهینه موازنه‌ای مطلوب بین دقت و تعمیم‌پذیری مدل در کار طبقه‌بندی ایجاد کرد (Breiman, 2001).

۲-۴-۲. مدل شبکه عصبی مصنوعی

شبکه‌های عصبی مدل‌های غیرخطی چندمتغیره هستند، آن‌ها از نورون‌های پردازشی در سه لایه شامل یک لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی تشکیل شده‌اند (Nabavi et al., 2018). الگوریتم یادگیری شبکه عصبی مصنوعی مورد استفاده در این پژوهش از نوع پرسپترون چندلایه^۱ مبتنی بر تئوری پس انتشار است. در این مطالعه، به

^۱ -Multi-layer perceptron (MLP)

منظور آموزش ساختار شبکه چندلایه تعریف شده، پارامترهای محیطی به عنوان نورون ورودی، و کلاس گردوغبار به عنوان نورون خروجی قرار گرفت. خروجی نهایی مدل را می‌توان به صورت معادله ۱ بدست آمده است.

رابطه ۱

$$\gamma_{jk} = F_K \left(\sum_{i=1}^{NK-1} W_{ijk\gamma_i} (K-1) + \beta_{jk} \right)$$

که در آن γ_{jk} نورون مربوط به خروجی j ام از لایه K ام است، β_{jk} وزن بایاس برای نورون j ام در لایه K است؛ $W_{ijk\gamma_i}$ و F_K پارامترهای برازش مدل هستند. $W_{ijk\gamma_i}$ وزن مشترک برای هر اتصال شبکه است که به صورت تصادفی در ابتدای مرحله آموزش انتخاب می‌شود. F_K به عنوان تابع انتقال یا تابع فعال‌سازی شناخته می‌شود که عمدتاً یک تابع غیرخطی است اما می‌تواند خطی، گاوسی، دوقطبی سیگموئید و غیره نیز باشد. این الگوریتم یادگیری بر اساس نظریه انتشار معکوس است.

۲-۴-۳. مدل تقویت گرادیان پیشرفته XGBoost

الگوریتم تقویت گرادیان پیشرفته یک مدل درخت گروهی مبتنی بر چارچوب تقویت گرادیان پیشرفته است که به ویژه برای حل مسائل طبقه‌بندی مفید است. این الگوریتم یک نسخه اصلاح شده و بهبود یافته از تقویت گرادیان است (Chen et al., 2015). محاسبات با XGBoost همچنین بسیار سریع‌تر از بسته‌های آماده است (Chang et al., 2018). برای توسعه یک یادگیرنده قوی، الگوریتم تقویت گرادیان پیشرفته تمام پیش‌بینی‌های مجموعه‌ای از یادگیرنده‌های ضعیف را ترکیب می‌کند (Fan et al., 2018).

۲-۴-۴. مدل تقویت گرادییانی

تقویت گرادیان از جمله روش‌های یادگیری ماشین گروهی برای مسائل رگرسیون و طبقه‌بندی است. در این الگوریتم، به داده‌ها وزن اختصاص داده می‌شود. روش یادگیری و مدلی برای آموزش داده‌ها که از ترکیب مجموعه‌ای از یادگیرنده‌های ضعیف ساخته می‌شوند. الگوریتم تقویت گرادییانی از تکنیک گرادیان نزولی به عنوان الگوریتم بهینه‌سازی خود برای به حداقل رساندن تابع زیان استفاده می‌کند که عملکرد درخت‌های تصمیم را ارزیابی می‌کند (Akashah & Petronas, 2020).

۲-۴-۵. مدل درخت‌های طبقه‌بندی و رگرسیون Bagged CART

این الگوریتم با تقسیم زیرمجموعه‌های مجموعه داده‌ها و استفاده از همه ویژگی‌ها به عنوان پیش‌بینی برای ایجاد دو زیرگروه برای هر تقسیم مکرر، با شروع از کل مجموعه داده‌ها، ایجاد می‌شود. این الگوریتم واحدهای نمونه آموزشی را به صورت بازگشتی با استفاده از تعداد از پیش تعیین شده‌ای از متغیرها تقسیم می‌کند (Res & Evren, 2022). این الگوریتم با در نظر گرفتن کل مجموعه داده به عنوان گره ریشه، به صورت بازگشتی با استفاده از شاخص جینی^۱ اقدام به تقسیم داده‌ها می‌کند. این شاخص، میزان ناخالصی یا درهم‌ریختگی کلاس‌ها را در یک گره اندازه‌گیری می‌کند. در

^۱ - Gini Index

هر مرحله، ویژگی و مقدار آستانه‌ای انتخاب می‌شوند که کاهش ناخالصی جینی را در دو گره بیشینه کنند (Tangirala, 2020).

۲-۴-۶. مدل ماشین تقویت گرادیان LightGBM

الگوریتم ماشین تقویت گرادیان یک مدل داده مبتنی بر $GBDT^1$ است که توسط مایکروسافت در سال ۲۰۱۷ ارائه شده است. مانند سایر الگوریتم‌های تقویتی، این الگوریتم یادگیرنده‌های ضعیف را برای تشکیل یک یادگیرنده قوی ترکیب می‌کند و علاوه بر اینکه دقت پیش‌بینی را کاهش نمی‌دهد، سرعت پیش‌بینی را نیز تا حد زیادی افزایش می‌دهد و استفاده از حافظه را کاهش می‌دهد. این الگوریتم از یک الگوریتم هیستوگرام بهبود یافته استفاده می‌کند (Ju et al., 2019).

۲-۴-۷. بهینه‌سازی هایپرپارامترهای مدل‌ها

در این پژوهش از روش Grid Search CV به‌عنوان تکنیک بهینه‌سازی استفاده شده است. فضای پارامترها به‌صورت سیستماتیک تعریف و جستجو شده است. این فرآیند با تلفیق اعتبارسنجی متقاطع انجام شده تا از بیش‌برازش جلوگیری شود. در این روش، فضای پارامترها به‌صورت سیستماتیک تعریف شده و با معیارهای ارزیابی دقیق، بهترین ترکیب پارامترها انتخاب می‌گردد. این فرآیند به مدل اجازه می‌دهد تا ضمن اجتناب از بیش‌برازش، بهترین تعمیم‌پذیری را روی داده‌های جدید داشته باشد (Belete & Huchaiah, 2021).

۲-۴-۸. ارزیابی مدل‌های یادگیری ماشین

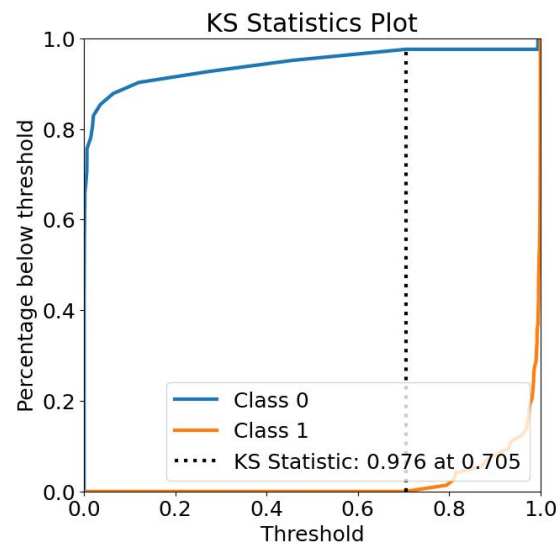
یکی از مراحل مهم بعد از مدلسازی، ارزیابی و اعتبارسنجی مدلسازی است. در این مطالعه برای ارزیابی و اعتبارسنجی الگوریتم‌ها از پلات کولموگروف-اسمیرنوف استفاده گردید و ۳۰ درصد از کانون‌های شکل‌گیری گردوغبار به‌طور تصادفی جهت ارزیابی صحت نقشه منابع گردوغبار استفاده شد. پلات کولموگروف-اسمیرنوف می‌تواند به‌عنوان یک معیار ارزیابی برای یک طبقه‌بندی‌کننده باینری^۳ استفاده شود. این پلات به درک اینکه آیا دو مجموعه از نمونه‌ها به توزیع احتمال مشابهی تعلق دارند یا خیر، کمک می‌کند و از این رو میزان تفکیک‌پذیری بین دو کلاس در مجموعه داده‌ها را اندازه‌گیری می‌کند. به‌عنوان نمونه در نمودار شماره ۱، طول فاصله بین کلاس ۱ و صفر در پلات نشان‌دهنده تفکیک خوب بین دو کلاس است، از این رو طبقه‌بندی‌کننده کلاس را با صحت بیشتری پیش‌بینی می‌کند.

¹ - Gradient Boosting Decision Tree

²- Overfitting

³- Binary

نمودار ۱. پلات کولموگروف-اسمیرنوف
Diagram 1. Kolmogorov-Smirnov (K-S) plot



۹-۴-۲. توضیح‌پذیری الگوریتم‌های یادگیری ماشین

در این مطالعه، توضیح‌پذیری مدل‌های یادگیری ماشین با هدف شناسایی عوامل کلیدی مؤثر بر شکل‌گیری گردوغبار و افزایش شفافیت تصمیم‌گیری مدل‌ها انجام شد. برای این منظور، از روش‌های مقادیر شیلی، حساسیت سوبل و اهمیت ویژگی استفاده گردید اگر چه این سه روش هدف یکسانی (شناسایی عوامل مؤثر در گردوغبار) دارند، اما هر کدام مکانیسم و جنبه متفاوتی از مدل را ارزیابی می‌کنند. در واقع این روش توضیح می‌دهد که هر متغیر چگونه و در چه جهتی خروجی مدل را «تحت تأثیر» قرار می‌دهد (Rodríguez & Jürgen, 2020). تحلیل حساسیت سوبل نشان می‌دهد مدل به تغییرات کدام متغیرها «حساس» است (Zouhri et al., 2021) و اهمیت ویژگی نیز نشان می‌دهد که مدل از کدام متغیرها بیشتر «استفاده» کرده است (Kaneko, 2023).

الف. روش SHAP

مقادیر شیلی از جمله پرکاربردترین رویکردها برای تفسیر مدل‌های یادگیری ماشین هستند که به صورت کمی سهم هر ویژگی را در پیش‌بینی مدل بیان می‌کنند (Liu et al., 2021; Luo et al., 2022). روش توضیحات افزایشی شیلی یک چارچوب شهودی برای توضیح تأثیر ویژگی‌های مختلف بر پیش‌بینی‌های مدل ارائه می‌دهد و یک الگوریتم اصولی برای ارائه توضیحات محلی (یعنی توضیحات برای پیش‌بینی‌های فردی) از خروجی‌های مدل ارائه می‌دهد (Lundberg & Lee, 2017). این روش پیش‌بینی‌های مدل را با استفاده از نظریه بازی مشارکتی تفسیر می‌کند (Molnar, 2020). در SHAP، مقادیر شیلی با در نظر گرفتن هر متغیر مستقل به عنوان یک بازیکن که با دیگران همکاری می‌کند تا سهمی از «پرداخت» را دریافت کند، محاسبه می‌شوند، که به عنوان تفاوت بین پیش‌بینی مدل خاص و میانگین همه پیش‌بینی‌ها تعریف می‌شود. سهم هر بازیکن بر اساس سهم او تعیین می‌شود و این تخصیص با استفاده از مقادیر شیلی به دست

آمایش فضا و ژئوماتیک

می‌آید (Antonini et al., 2024). این روش مبتنی بر تئوری بازی‌ها است و فاکتورهای استفاده شده از نظر اهمیت رتبه‌بندی شده‌اند. نظریه بازی‌ها، تأثیر هر متغیر را در سطح نمونه‌های فردی تحلیل می‌کند.

ب. تحلیل حساسیت سوئل

تحلیل حساسیت سوئل به طور گسترده در مهندسی و سایر زمینه‌ها برای ارزیابی تأثیر پارامترهای ورودی بر خروجی‌های مدل‌های پیچیده استفاده می‌شود. این روش مبتنی بر واریانس قادر به محاسبه شاخص‌های حساسیت مرتبه اول و کل است و در نتیجه امکان شناسایی و اولویت‌بندی عوامل مؤثر در مدل را فراهم می‌کند (Nossent et al., 2011). این روش برای بررسی تعامل بین متغیرها استفاده می‌شود و مقادیر حساسیت اولیه^۱ (S1) و حساسیت کلی^۲ (ST) را نشان می‌دهد. که حساسیت اولیه تأثیر مستقیم متغیر و حساسیت کلی تأثیر متغیر همراه در تعاملات با دیگر متغیرها را مشخص می‌کنند.

ج. تحلیل اهمیت ویژگی‌ها

اهمیت ویژگی به میزان تأثیر هر متغیر در مجموعه داده‌ها بر دقت و عملکرد مدل اشاره دارد و سهم هر ویژگی را در قابلیت پیش‌بینی مدل نشان می‌دهد (Megantara & Ahmad, 2020). تکنیک‌های مختلفی برای ارزیابی اهمیت ویژگی‌ها توسعه داده شده است که محققان را قادر می‌سازد تا «جعبه سیاه» الگوریتم‌های یادگیری ماشین را بررسی کرده و ویژگی‌هایی را که برای دستیابی به پیش‌بینی‌های دقیق حیاتی هستند، شناسایی کنند (Musolf et al., 2022).

۲-۴-۱۰. اعتبارسنجی متقاطع

در این پژوهش، فرآیند ارزیابی و بهینه‌سازی مدل‌ها با رعایت استانداردهای روش‌شناسی انجام شده و از روش اعتبارسنجی متقاطع^۳ K-Fold با K=5 به صورت سیستماتیک استفاده شده است (Tougui et al, 2021). این انتخاب مبتنی بر مطالعات متعدد روش‌شناسی بوده که نشان می‌دهد K=5 تعادل مناسبی بین bias و واریانس ایجاد می‌کند (Nti, 2021).

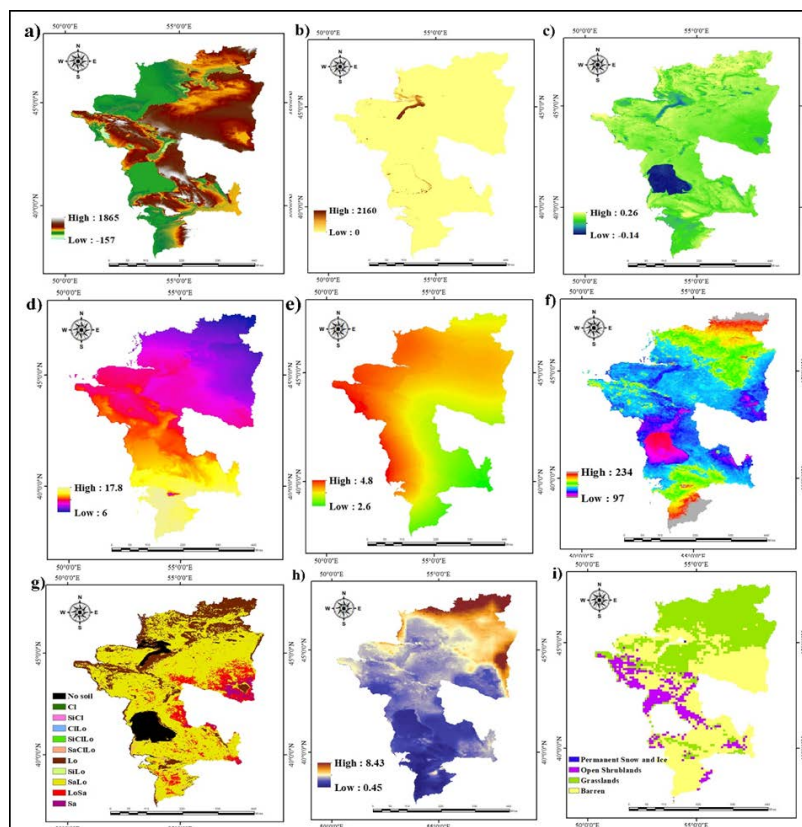
۳. نتایج

نخستین مرحله در اجرای مدل‌ها، تعیین داده‌های ورودی و هدف است. مطابق با شکل ۵، در این پژوهش میانگین پارامترهای محیطی شامل مدل رقومی ارتفاعی از سطح دریا، روزهای گردوغباری، دما، سرعت باد، بارش، شاخص نرمال شده تفاوت پوشش گیاهی، بافت خاک، رطوبت خاک و پوشش زمین در بازه زمانی ۲۰۰۳ تا ۲۰۲۳ به عنوان ورودی مدل و تابع هدف پیش‌بینی مناطق دارای پتانسیل گردوغبار برای حوضه خلیج قره‌بغاز استخراج گردیدند.

1- First sensitivity
2- Overall sensitivity
3- Cross-Validation

شکل ۵. پارامترهای محیطی موثر بر وقوع گردوغبار: a: مدل رقومی ارتفاعی از سطح دریا برحسب متر؛ b: روزهای گردوغباری؛ c: NDVI؛ d: دما بر حسب درجه سانتی-گراد؛ e: سرعت باد بر حسب متر بر ثانیه؛ f: بارش برحسب میلی‌متر؛ g: بافت خاک؛ h: رطوبت خاک بر حسب میلی‌متر و i: پوشش زمین

Figure 5. Key environmental drivers of dust event :a) DEM (m); b) Dusty days; c) NDVI; d) Temperature (°C); e) Wind speed (m/s); f) Precipitation (mm); g) Soil texture; h) Soil moisture (mm); i) Land cover



۳-۱. منابع مستعد انتشار گردوغبار با استفاده از الگوریتم‌های یادگیری ماشین

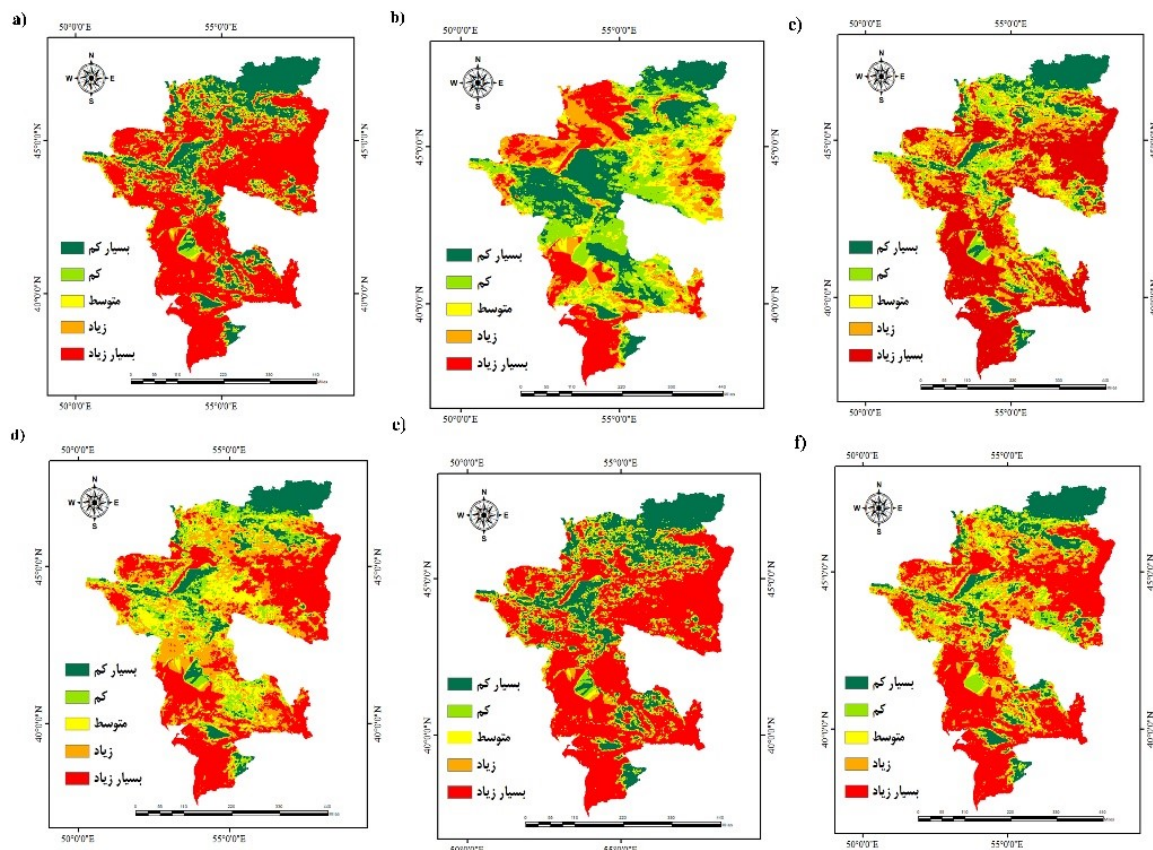
در این مطالعه، ۷۰ درصد از کانون‌های مولد گردوغبار به عنوان داده‌های آموزشی انتخاب شدند و عملکرد الگوریتم‌های یادگیری ماشین جنگل تصادفی، شبکه عصبی مصنوعی، تقویت گرادیانی پیشرفته، تقویت گرادیانی، الگوریتم درخت‌های طبقه‌بندی و رگرسیون Bagged CART و ماشین تقویت گرادیان LightGBM مورد ارزیابی قرار گرفتند. پس از فرآیند آموزش هر یک از مدل‌های طبقه‌بندی، روابط معناداری بین مقادیر پارامترهای ورودی در پیکسل‌های مربوط به کانون‌های گردوغبار شناسایی گردید. در ادامه، مدل‌های آموزش‌دیده بر اساس الگوهای استخراج‌شده که در حقیقت بازنمایی ارتباطات بین متغیرهای پیش‌بین و داده‌های آموزشی هستند برای حوضه خلیج قره‌بغاز پیاده‌سازی شدند. پس از مدل‌سازی نقشه‌های رستری تولید شده وارد نرم‌افزار سامانه اطلاعات جغرافیایی شده و با استفاده از روش شکست-های طبیعی^۱ (Chen et al., 2013) به ۵ کلاس حساسیت اراضی نسبت به گردوغبار خیلی کم، کم، متوسط، زیاد و خیلی زیاد تقسیم‌بندی گردید (شکل ۶). کلاس‌های زیاد و خیلی زیاد عمدتاً شامل نقاطی هستند که به عنوان کانون

^۱ - Natural Breaks (Jenks)

گردوغبار شناخته می‌شوند. نواحی از حوضه که در طبقات حساسیت پایین‌تری قرار بگیرند کمتر پتانسیل ایجاد گردوغبار را دارند.

شکل ۶. نقشه منابع مستعد انتشار گردوغبار با استفاده از الگوریتم‌های یادگیری ماشین (a: جنگل تصادفی؛ b: شبکه عصبی مصنوعی؛ c: تقویت گرادیانی پیشرفته؛ d: تقویت گرادیانی؛ e: درختان طبقه‌بندی و رگرسیون Bagged CART و f: ماشین تقویت گرادیان LightGBM)

Figure 6. Mapping of Dust Emission-Prone Sources Using Machine Learning Algorithms. a) Random Forest; b) Artificial Neural Network (ANN); c) XGBoost; d) Gradient Boosting; e) Bagged CART; f) LightGBM



۲-۳. ارزیابی و اعتبارسنجی الگوریتم‌های یادگیری ماشین

در این پژوهش، عملکرد مدل‌های مبتنی بر یادگیری ماشین از طریق ارزیابی شاخص‌های برازش و قدرت پیش‌بینی مورد تحلیل قرار گرفت. از آنجایی که مدل‌های مورد استفاده در این پژوهش کلاس‌های باینری را پیش‌بینی می‌کند، از پلات کولموگروف-اسمیرنوف برای ارزیابی عملکرد آن‌ها استفاده شد. این پلات بیشترین تفاوت بین توابع توزیع تجمعی دو کلاس را اندازه‌گیری می‌کند. پلات از ۰ (همپوشانی کامل) تا ۱ (جدایی کامل) متغیر است (Hassan et al., 2025). پلات KS بالا ۰٫۹۱ در شکل ۷ نشان‌دهنده کلاس‌های به‌خوبی جدا شده توسط مدل‌ها است. مدل جنگل تصادفی با ۹۱/۸ درصد، شبکه عصبی مصنوعی با ۷۰/۹ درصد، تقویت گرادیانی پیشرفته با ۸۹/۹ درصد، تقویت گرادیانی با ۸۷/۹ درصد، الگوریتم درختان طبقه‌بندی و رگرسیون Bagged CART با ۸۹/۹ درصد و ماشین تقویت گرادیان

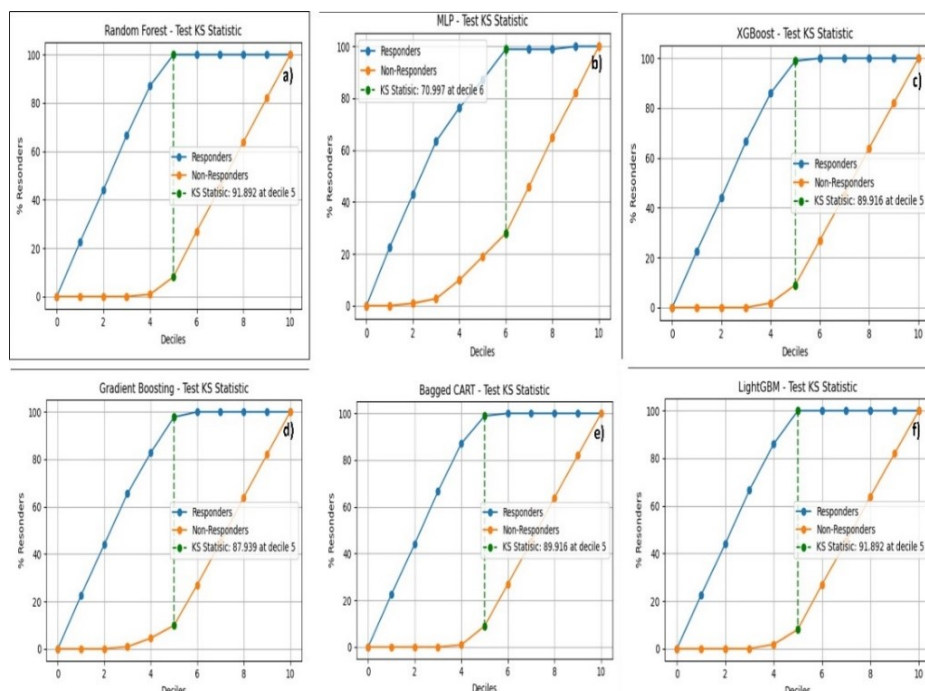
آمایش فضا و ژئوماتیک

LightGBM با ۹۱/۸ درصد مناطق با پتانسیل بالای گردوغبار را تشخیص داده‌اند. مدل جنگل تصادفی و ماشین تقویت گرادیان LightGBM به‌طور قابل توجهی از انتخاب تصادفی برای هر دو کلاس بهتر عمل می‌کند که نشان‌دهنده عملکرد طبقه‌بندی رضایت‌بخش آن‌ها است.

شکل ۷. نتایج ارزیابی مدل‌های یادگیری ماشین: a) جنگل تصادفی؛ b) شبکه عصبی مصنوعی؛ c) تقویت گرادیان پیشرفته؛ d) تقویت گرادیان؛ e) درختان طبقه‌بندی و رگرسیون CART و f) ماشین تقویت گرادیان (LightGBM) با استفاده از روش ارزیابی KS Statistic

Figure 7. Evaluation results of machine learning models using the KS Statistic method

a) Random Forest; b) Artificial Neural Network; c) XGBoost; d) Gradient Boosting; e) Bagged CART; f) LightGBM



۳-۳. روش تحلیل مشارکت متغیرها

شرایط آب‌وهوایی خشک، بادهای شدید و وضعیت شکننده سطح زمین از مهم‌ترین عوامل ایجاد طوفان‌های گردوغبار هستند (Wang et al., 2023). در نتیجه تعیین سهم تمام عوامل مؤثر بر توزیع منابع گردوغبار برای کاهش پیامدهای زیست‌محیطی بسیار حائز اهمیت است. نمودار میله‌ای (a) (شکل ۸)، مربوط به پیش‌بینی منشأ گردوغبار با استفاده از مقادیر SHAP است که اهمیت و تأثیر هر ویژگی را در خروجی مدل نشان می‌دهد. بر اساس این نمودار، متغیرهای روزهای گردوغباری و بافت خاک با SHAP مثبت بالا بیشترین تأثیر را بر پیش‌بینی مدل دارند و به‌عنوان مهم‌ترین عوامل شناسایی شده‌اند؛ به این صورت که افزایش تعداد روزهای گردوغباری و تخریب تدریجی ساختار بافت خاک خطر گردوغبار را افزایش می‌دهد. مدل رقومی ارتفاع با SHAP منفی، مناطق پائین‌تر (مثل دشت‌ها) حساسیت بالاتری دارند. پس از آن‌ها، دما و سرعت باد قرار گرفته‌اند که تأثیر مختلطی دارند (بعضی نقاط مثبت، بعضی منفی) و نقش قابل توجهی در پیش‌بینی منشأ گردوغبار ایفا می‌کنند. سایر متغیرها مانند رطوبت خاک، شاخص نرمال‌شده تفاوت پوشش گیاهی، بارش و پوشش زمین نیز تأثیر کمتری دارند. مقادیر SHAP نشان می‌دهند که چگونه هر یک از این ویژگی‌ها،

آمایش فضا و ژئوماتیک

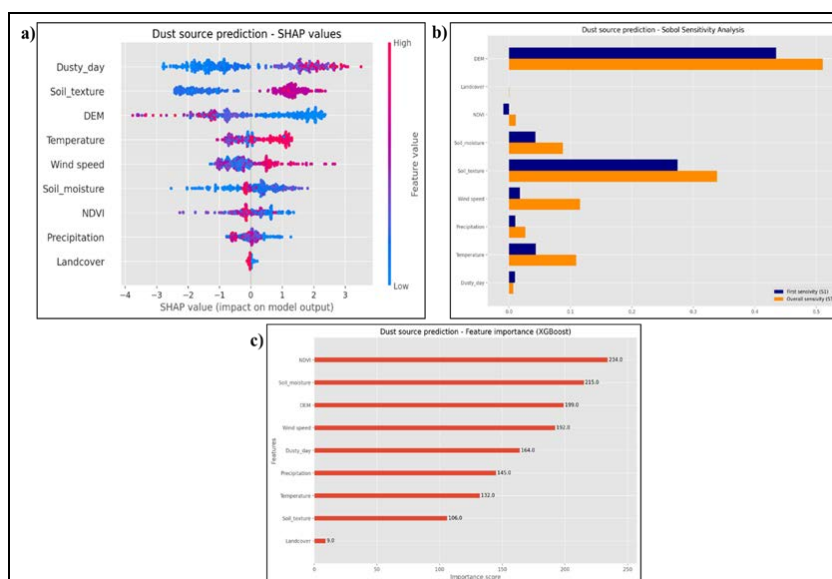
بسته به مقدارشان (بالا یا پایین)، خروجی مدل را به سمت پیش‌بینی منشأ گردوغبار سوق می‌دهند. به‌طور کلی، این تحلیل به درک بهتر عوامل کلیدی مؤثر در ایجاد گردوغبار و اولویت‌بندی آن‌ها کمک می‌کند.

نمودار (b) (شکل ۸)، مربوط به تحلیل حساسیت سوبل برای پیش‌بینی منشأ گردوغبار است که سهم هر متغیر در تغییرات خروجی مدل را بررسی می‌کند. بر اساس این نمودار، مدل رقومی ارتفاع بیشترین حساسیت را در مدل دارد و به عنوان مؤثرترین عامل در پیش‌بینی منشأ گردوغبار شناسایی شده است. در ادامه پوشش زمین و شاخص نرمال شده تفاوت پوشش گیاهی با مقادیر حساسیت منفی قرار دارند که نشان‌دهنده نقش بازدارنده آن‌ها در تشکیل گردوغبار است. پس از آن، بافت خاک با $S1=0.8$ و $ST=1.6$ بیش‌ترین تأثیر را در مدل دارد. سایر متغیرها مانند رطوبت خاک، سرعت باد، بارندگی، دما و روزهای گردوغباری حساسیت کمتری دارند. مقادیر حساسیت اولیه و حساسیت کلی نشان می‌دهند که مدل رقومی ارتفاع نه تنها در $S1$ ، بلکه در تعامل با سایر متغیرها ST نیز بیشترین تأثیر را بر خروجی مدل دارد. این تحلیل تأکید می‌کند که ارتفاع منطقه و پوشش زمین از عوامل کلیدی در تعیین منشأ گردوغبار هستند و باید در مدل‌سازی و مدیریت این پدیده مورد توجه ویژه قرار گیرند.

نمودار ارائه‌شده (c) (شکل ۸)، اهمیت نسبی ویژگی‌های مؤثر در پیش‌بینی منبع گردوغبار را نشان می‌دهد. بر همین اساس مشخص می‌شود که شاخص نرمال‌شده تفاوت پوشش گیاهی بیشترین نقش را در پیش‌بینی مدل دارد، به‌طوری‌که کاهش شاخص نرمال‌شده تفاوت پوشش گیاهی احتمال تشکیل گردوغبار را افزایش می‌دهد. پس از آن رطوبت خاک و سپس ارتفاع از سطح دریا به ترتیب مهم‌ترین عوامل هستند و حساسیت مناطق را به گردوغبار افزایش می‌دهند. ویژگی‌هایی مانند سرعت باد، روزهای گردوغباری، بارش و دما نیز نقش معناداری دارند، در مقابل، بافت خاک و پوشش زمین کمترین تأثیر را دارند. این نتایج بر اهمیت کنترل عوامل طبیعی مانند حفظ پوشش گیاهی، افزایش رطوبت خاک و ارتفاع منطقه (هموار بودن اراضی) در کاهش گردوغبار تأکید می‌کند.

شکل ۸. اهمیت نسبی متغیرها

Figure 8. Relative importance of variables



۴. بحث

این پژوهش با استفاده از الگوریتم‌های یادگیری ماشین و همچنین چند روش تفسیر مدل به بررسی عوامل مؤثر بر وقوع گردوغبار و پیش‌بینی مناطق مستعد پرداخته است. یافته‌ها نشان می‌دهد متغیرهای محیطی مختلفی در تشکیل گردوغبار نقش دارند که از میان آن‌ها شاخص NDVI به عنوان عامل کاهش‌دهنده گردوغبار شناسایی شد. این نتیجه با مطالعات وانگ^۱ و همکاران (۲۰۲۳) که کاهش پوشش گیاهی را عامل تشدیدکننده گردوغبار می‌دانند، همخوانی دارد. همچنین رطوبت خاک به عنوان یک عامل مؤثر شناخته شد که این یافته نیز با نتایج گودی و^۲ میدلتون (۲۰۰۶) مطابقت نشان می‌دهد. در بخش دیگری از نتایج، ارتفاع از سطح دریا به عنوان عامل مهم دیگری شناسایی شد که در آن مناطق پایین‌تر مانند دشت‌ها حساسیت بیشتری به گردوغبار نشان دادند. این یافته با مطالعات حسن^۳ و همکاران (۲۰۲۵) در حوضه قره‌بغاز همسو است. از نظر عملکرد مدل‌های یادگیری ماشین، مدل‌های جنگل تصادفی و ماشین تقویت گرادیان LightGBM بهترین عملکرد را در پیش‌بینی مناطق مستعد گردوغبار نشان دادند که این نتیجه با یافته‌های لی^۴ و همکاران (۲۰۲۰) که از این مدل‌ها برای پدیده‌های محیطی استفاده کرده بودند، همخوانی دارد. نتایج حاصل از تحلیل اهمیت ویژگی‌ها نشان داد که متغیرهای شاخص نرمال‌شده تفاوت پوشش گیاهی، رطوبت خاک و ارتفاع از سطح دریا بیشترین تأثیر را در پیش‌بینی مدل دارند. این یافته با مطالعه چن^۵ و همکاران (۲۰۱۳) که از روش اهمیت ویژگی در مدل‌های جنگل تصادفی استفاده کرده بودند، همسو است. آن‌ها نیز به نقش کلیدی پوشش گیاهی و رطوبت خاک در کنترل گردوغبار اشاره داشتند. کاربرد تکنیک SHAP در این پژوهش، تفسیر جامع‌تری از روابط غیرخطی بین متغیرها ارائه داد. نتایج SHAP نشان داد که روزهای گردوغباری و بافت خاک بیشترین تأثیر مثبت را بر خروجی مدل دارند. این یافته با تحقیق لوندبرگ^۶ و همکاران (۲۰۱۷) که از SHAP برای تفسیر مدل‌های پیچیده استفاده کرده بودند، همخوانی دارد. مقایسه روش‌های تفسیر مدل در این پژوهش نشان داد که اگرچه روش‌های مختلف در شناسایی عوامل کلیدی تا حدی همپوشانی دارند، اما هر کدام جنبه‌های متفاوتی از روابط را آشکار می‌کنند.

۵. نتیجه‌گیری

این مطالعه با هدف ارزیابی الگوریتم‌های یادگیری ماشین برای شناسایی منابع بالقوه گردوغبار در حوضه خلیج قره‌بغاز به عنوان بخشی از کانون‌های خارجی ورود گردوغبار به استان‌های شمال شرقی کشور ایران انجام شد. با استفاده از داده‌های ۹ پارامتر محیطی، مدل‌های مختلف یادگیری ماشین شامل جنگل تصادفی، شبکه عصبی مصنوعی، تقویت گرادیانی پیشرفته، تقویت گرادیانی، الگوریتم درختان طبقه‌بندی و رگرسیون CART و ماشین تقویت گرادیان LightGBM مورد بررسی قرار گرفتند. یافته‌های این پژوهش نشان می‌دهد که ترکیب روش‌های یادگیری ماشین با

^۱ - Wang

^۲ - Goudie & Middleton

^۳ - Hassan

^۴ - Li

^۵ - Chen

^۶ - Lundberg

داده‌های سنجش از دور، رویکردی مؤثر برای شناسایی کانون‌های تولید گردوغبار در مناطق خشک محسوب می‌شود. بر اساس نتایج به‌دست آمده، مدل‌های گروهی به ویژه جنگل تصادفی و ماشین تقویت گرادیان LightGBM با صحت بالای ۹۰٪ توانستند بهترین عملکرد را در پیش‌بینی مناطق مستعد گردوغبار داشته باشند. تحلیل اهمیت متغیرها نشان داد که شاخص نرمال‌شده تفاوت پوشش گیاهی، رطوبت خاک و ارتفاع از سطح دریا به ترتیب بیشترین تأثیر را در شکل‌گیری کانون‌های گردوغبار دارند. در مقایسه با روش‌های سنتی، رویکرد پیشرفته ارائه شده در این تحقیق می‌تواند دقت شناسایی کانون‌های گردوغبار را به میزان قابل توجهی افزایش دهد. این بهبود عمدتاً ناشی از بهره‌گیری از الگوریتم‌های یادگیری ماشین پیشرفته، انتخاب بهینه ویژگی‌ها و استفاده از داده‌های با وضوح بالا بوده است. با این حال، همچنان چالش‌هایی در زمینه در نظر گرفتن تمامی عوامل مؤثر بر شکل‌گیری گردوغبار، به ویژه عوامل انسانی، تغییرات کاربری اراضی و خشکیدگی اکوسیستم‌های آبی وجود دارد که لازم است در مطالعات آینده مورد توجه قرار گیرد.

این پژوهش از چند جهت حائز اهمیت است: (۱) چارچوب روش‌شناسی جامعی برای شناسایی مناطق مستعد گردوغبار ارائه می‌دهد که قابلیت تطبیق با سایر مناطق را دارد. (۲) تحلیل‌های حساسیت‌سنجی انجام شده می‌تواند به درک بهتر مکانیسم‌های شکل‌گیری گردوغبار کمک کند. (۳) نتایج این تحقیق می‌تواند مبنای علمی مناسبی برای تصمیم‌گیری‌های مدیریتی و برنامه‌ریزی‌های محیط‌زیستی در منطقه باشد. برای مطالعات آینده، پیشنهاد می‌شود اثرات تغییرات اقلیمی و فعالیت‌های انسانی با جزئیات بیشتری مورد بررسی قرار گیرد و از داده‌های با وضوح بالاتر برای بهبود دقت مدل‌ها استفاده شود.

Reference

- Abdolshahnejad, M., Khosravi, H., Nazari Samani, A. A., Zehtabian, G. R., & Alambaigi, A. (2020). Determining the conceptual framework of dust risk based on evaluating resilience (Case study: Southwest of Iran). *Strategic Research Journal of Agricultural Sciences and Natural Resources*, 5(1), 33-44. (In Persian)
- Akashah, E., & Petronas, U. T. (2020). A Study on Gradient Boosting Algorithms for Development of AI Monitoring and Prediction Systems. *International Conference on Computational Intelligence (ICCI)* 10–16. <https://doi.org/10.1109/ICCI51257.2020.9247843>
- Antonini, A. S., Tanzola, J., Asiain, L., Ferracutti, G. R., Castro, S. M., Bjerg, E. A., & Luján, M. (2024). Applied Computing and Geosciences Machine Learning model interpretability using SHAP values: Application to Igneous Rock Classification task. *Applied Computing and Geosciences*, 100178. <https://doi.org/10.1016/j.acags.2024.100178>
- Belete, D. M., & Huchaiah, M. D. (2021). Grid search in hyperparameter optimization of machine learning models for prediction of HIV / AIDS test results. *International Journal of Computers and Applications*. Volume 44. <https://doi.org/10.1080/1206212X.2021.1974663>
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

- Chang, Y., Chang, K., & Wu, G. (2018). Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Applied Soft Computing Journal*. <https://doi.org/10.1016/j.asoc.2018.09.029>
- Chen, J., Li, G., Xiao, B., Wen, Z., Lv, M., & Chen, C. (2015). Assessing the transferability of support vector machine model for estimation of global solar radiation from air temperature. *Energy Conversion and Management*, 89, 318–329. <https://doi.org/10.1016/j.enconman.2014.10.004>
- Chen, J., Yang, S., Li, H., Zhang, B., & Lv, J. (2013). Research on Geographical Environment Unit Division Based on the Method of Natural Breaks (Jenks). *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Archives)* <https://doi.org/10.5194/isprsarchives-XL-4-W3-47-2013>
- Darvishi Boloorani, A., Papi, R., Soleimani, M., Karami, L., Amiri, F., & Neysani Samany, N. (2021). Water bodies changes in Tigris and Euphrates basin has impacted dust storms phenomena. *Aeolian Research*, 100698. <https://doi.org/10.1016/j.aeolia.2021.100698>
- Darvishi Boloorani, A., Neysani Samany, N., Papi, R., & Soleimani, M. (2022). Dust source susceptibility mapping in Tigris and Euphrates basin using remotely sensed imagery. *Catena*, 209(P1), 105795. <https://doi.org/10.1016/j.catena.2021.105795>
- Darvishi Boloorani, A., Papi, R., Soleimani, M., & Al-hemoud, A. (2023). Remote Sensing Applications : Society and Environment Visual interpretation of satellite imagery for hotspot dust sources identification. *Remote Sensing Applications: Society and Environment*, 100888. <https://doi.org/10.1016/j.rsase.2022.100888>
- Ebrahimi-Khusfi, Z., Mirakbari, M., Ebrahimi-Khusfi, M., & Taghizadeh-Mehrjardi, R. (2020). Impacts of vegetation anomalies and agricultural drought on wind erosion over Iran from 2000 to 2018. *Applied Geography*, 102330. <https://doi.org/10.1016/j.apgeog.2020.102330>
- Espahbod, M. R. (2005). Morphotectonic changes and Caspian Sea level fluctuations with a special focus on the Kara-Bogaz-Gol Bay using satellite images. *21st Symposium on Earth Sciences, Geological Survey and Mineral Explorations of Iran*. . (In Persian)
- Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., & Xiang, Y. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates : A case study in China. *Energy Conversion and Management*, 102–111. <https://doi.org/10.1016/j.enconman.2018.02.087>
- Gholami, H., Mohamadifar, A., & Collins, A. L. (2020). Spatial mapping of the provenance of storm dust: Application of data mining and ensemble modelling. *Atmospheric Research*, 233, 104716. <https://doi.org/10.1016/j.atmosres.2019.104716>
- Gholami, H., Mohammadifar, A., Malakooti, H., Esmaeilpour, Y., Golzari, S., Mohammadi, F., ... & Collins, A. L. (2021). Integrated modelling for mapping spatial sources of dust in central Asia-An important dust source in the global atmospheric system. *Atmospheric Pollution Research*, 12(9), 101173. <https://doi.org/10.1016/j.apr.2021.101173>
- Goudie, A. S., & Middleton, N. J. (2006). Desert dust in the global system. Springer.
- Hamidi, M. (2020). The key role of water resources management in the Middle East dust events. *Catena*, 104337. <https://doi.org/10.1016/j.catena.2019.104337>
- Hassan, I., Nahid, N., Islam, M., Hossain, S., & Schuller, B. (2025). Automated Autism Assessment with Multimodal Data and Ensemble Learning : A Scalable and Consistent Robot-Enhanced Therapy Framework. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, PP, 1. <https://doi.org/10.1109/TNSRE.2025.3546519>
- Ju, Y., Sun, G., Chen, Q., Zhang, M., Zhu, H., & Rehman, M. (2019). A Model Combining Convolutional Neural Network and LightGBM Algorithm for Ultra-Short-Term Wind Power Forecasting. *IEEE Access*, 7, 28309–28318. <https://doi.org/10.1109/ACCESS.2019.2901920>

- Kaneko, H. (2023). Interpretation of Machine Learning Models for Data Sets with Many Features Using Feature Importance. *ACS Omega*, 8 (25), 23218-23225. <https://doi.org/10.1021/acsomega.3c03722>
- Li, J., Garshick, E., Al-Hemoud, A., Huang, S., & Koutrakis, P. (2020). Impacts of meteorology and vegetation on surface dust concentrations in Middle Eastern countries. *Science of the Total Environment*, 712, 136597. <https://doi.org/10.1016/j.scitotenv.2020.136597>
- Liu, X., Song, H., Lei, T., Liu, P., Xu, C., Wang, D., Yang, Z., Xia, H., Wang, T., & Zhao, H. (2021). Effects of natural and anthropogenic factors and their interactions on dust events in Northern China. *Catena*, 104919. <https://doi.org/10.1016/j.catena.2020.104919>
- Lundberg, S. M., & Lee, S. (2017). A Unified Approach to Interpreting Model Predictions. Section 2, 1–10.
- Luo, X., Jiang, Y., & Xiao, X. (2022). Feature Inference Attack on Shapley Values. *31st International Conference on Neural Information Processing Systems (NIPS'17)*, 2233–2247. <https://doi.org/10.1145/3548606.3560573>
- Molnar, C. (2020). Interpretable machine learning. Lulu. com.
- Megantara, A. A., & Ahmad, T. (2020). Feature Importance Ranking for Increasing Performance of Intrusion Detection System. *3rd International Conference on Computer and Informatics Engineering (IC2IE)* 37–42. <https://doi.org/DOI: 10.1109/IC2IE50715.2020.9274570>
- Middleton, N. J. (2017). Desert dust hazards: A global review. *Aeolian Research*, 24, 53–63. <https://doi.org/10.1016/j.aeolia.2016.12.001>
- Musolf, A. M., Holzinger, E. R., Malley, J. D., & Bailey, J. E. (2022). What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. *Human Genetics*, 141(9), 1515–1528. <https://doi.org/10.1007/s00439-021-02402-z>
- Motiee, H., Abduh Kolahchi, A., Aminian, R. (2014). Assessment of soil moisture using remote sensing ECV Database and its correlation with dust events - South and West of Iran *Iranian Journal of Soil and Water Research*, Volume 53, 1531-1544. (In Persian)
- Nabavi, S. O., Haimberger, L., Abbasi, R., & Samimi, C. (2018). Prediction of aerosol optical depth in West Asia using deterministic models and machine learning algorithms. *Aeolian Research*, 69–84. <https://doi.org/10.1016/j.aeolia.2018.10.002>
- Nossent, J., Elsen, P., & Bauwens, W. (2011). Environmental Modelling & Software Sobol' sensitivity analysis of a complex environmental model. *Environmental Modelling and Software*, 26(12), 1515–1525. <https://doi.org/10.1016/j.envsoft.2011.08.010>
- Nti, I. K. (2021). Performance of Machine Learning Algorithms with Different K Values in K-fold Cross- Validation. *International Journal of Information Technology and Computer Science*, 6, 61-71. <https://doi.org/10.5815/ijitcs.2021.06.05>
- Papi, R., Attarchi, S., Darvishi Boloorani, A., & Neysani Samany, N. (2022). Characterization of Hydrologic Sand and Dust Storm Sources in the Middle East. *Sustainability*, 14(22). <https://doi.org/10.3390/su142215352>. (In Persian)
- Papi, R., Kakroodi, A. A., Soleimani, M., Karami, L., Amiri, F., & Alavipanah, S. K. (2022). Identifying sand and dust storm sources using spatial-temporal analysis of remote sensing data in Central Iran. *Ecological Informatics*, 101724. <https://doi.org/10.1016/j.ecoinf.2022.101724>
- Pourhashemi, S. Zanganeh Asadi, M.A., Boroughani, M., & Azadi, H. (2022). Dust Source and Susceptibility Map in Iran and Iraq: Application of Remote Sensing and Machine Learning Techniques. *A preprint of Environmental Science and Pollution Research* 1–27. <https://doi.org/10.21203/rs.3.rs-1613035/v1>
- Res, A. M., & Evren, B. (2022). The estimating of hypothyroidism with the bagged CART model based on clinical dataset and identify of risk factors. *Annals of Medical Research*, 29(10), 1189–1193. <https://doi.org/10.5455/annalsmedres.2022.10.302>

- Rodríguez, R., & Jürgen, P. (2020). Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 34(10), 1013–1026. <https://doi.org/10.1007/s10822-020-00314-0>
- Tangirala, S. (2020). Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm. *International Journal of Advanced Computer Science and Applications*, 11(2), 612–619. <https://doi.org/10.14569/IJACSA.2020.0110277>
- Tougui, I., Jilbab, A., & Mhamdi, J. El. (2021). Impact of the Choice of Cross-Validation Techniques on the Results of Machine Learning- Based Diagnostic Applications. *Healthcare Informatics Research*, 27(3), 189–199. <https://doi.org/10.4258/hir.2021.27.3.189>
- Wang, W., Samat, A., Abuduwaili, J., De Maeyer, P., & Van de Voorde, T. (2023). Machine learning-based prediction of sand and dust storm sources in arid Central Asia. *International Journal of Digital Earth*, 16(1), 1530–1550. <https://doi.org/10.1080/17538947.2023.2202421>
- Zouhri, W., Homri, L., & Dantan, J. (2021). Handling the impact of feature uncertainties on SVM: A robust approach based on Sobol sensitivity analysis. *Expert Systems with Applications*, 189, 115691. <https://doi.org/10.1016/j.eswa.2021.115691>
- Zarin, A., Salehabadi, N., Mofidi, A., & Dadashi Roudbari, A. B. (2022). Seasonal review of dust storms in northeastern Iran and numerical simulation of dust events using the WRF-CHEM model. *Journal of Earth and Space Physics*, 48(2), 421–440. (In Persian)