



Comparative Analysis of Dust Storm Frequency Modeling Using the XGBoost Machine Learning Model and the Element-Free Galerkin Method (Case study: Ilam Province)

Mohammad Ansari Ghojghar¹✉  | Paria Pourmohammad² 

1. Corresponding Author, Department of Reclamation of Arid and Mountainous Regions Engineering, Faculty of Natural Resources, University of Tehran, Iran. Ansari.ghojghar@ut.ac.ir

2. Department of Reclamation of Arid and Mountainous Regions Engineering, Faculty of Natural Resources, University of Tehran, Iran. Paria.pormohamad7@ut.ac.ir

Article Info	ABSTRACT
Article type: Research Article	Dust storm modeling plays a crucial role in improving prediction accuracy, mitigating environmental risks, and developing effective response strategies. It enhances the understanding of dynamic system interactions and facilitates more reliable forecasts. In this context, machine learning models capable of analyzing complex datasets, alongside the Galerkin numerical method with its high-precision simulation capabilities, provide a robust foundation for more efficient solutions. This study compares the modeling of dust storm day frequency using the XGBoost machine learning model and the mesh-free Galerkin method across eight stations over a 40-year statistical period (1981–2020). In both models, Combination 1 was identified as the optimal configuration for predicting the FDS index. The results indicate that XGBoost outperforms the Galerkin method, exhibiting lower RMSE and MAE values and R and NS values closer to one. However, the difference between the two models was not statistically significant at the 95% and 99% confidence levels. Given its lower computational cost and reduced processing demand, the Galerkin numerical method remains a viable alternative for modeling the FDS index. The findings of this research can contribute to more accurate dust storm predictions, the development of early warning systems, informed crisis management strategies, and efforts to minimize the environmental impacts of dust storms.
Article history:	
Received: March. 14, 2025	
Revised: May. 1, 2025	
Accepted: May. 4, 2025	
Published online: Aug. 2025	
Keywords: Numerical Models, Horizontal Visibility, Forecasting, Decision Tree, The Frequency Index of Dust Storm Days.	
Cite this article: Ansari Ghojghar, M., & Pourmohammad, P. (2025) Comparative Analysis of Dust Storm Frequency Modeling Using the XGBoost Machine Learning Model and the Element-Free Galerkin Method (Case study: Ilam Province), <i>Iranian Journal of Soil and Water Research</i>, 56 (6), 1421-1436. https://doi.org/10.22059/ijswr.2025.392122.669904 © The Author(s). Publisher: The University of Tehran Press. DOI: https://doi.org/10.22059/ijswr.2025.392122.669904	



EXTENDED ABSTRACT

Introduction

Dust storms, recognized as pervasive and destructive meteorological phenomena in arid and semi-arid regions, intensify desertification and drought, deplete water resources, increase soil salinity, cause traffic accidents in critical zones, and exacerbate respiratory and other health-related issues. These effects collectively impose substantial economic, social, and environmental costs. Such phenomena typically lead to severe visibility reduction, with horizontal visibility often dropping below one kilometer. Monitoring the spatial and temporal variations of dust storms is crucial for risk forecasting, mitigation, and the effective management of this challenge. Mapping land sensitivity to dust generation in Ilam Province using data mining models indicated that approximately 23% and 18% of the area fall under highly sensitive and very highly sensitive categories, respectively. Despite the increasing importance of dust storms and their extensive social, economic, and environmental repercussions, a review of the literature reveals limited research in this field. Addressing these gaps, this study explores the performance of machine learning models and the Element-Free Galerkin algorithm for modeling dust storms. The integration of numerical models with machine learning algorithms to predict the Frequency of Dust Storm Days presents an innovative and transformative approach. Employing the advanced gradient boosting algorithm for dust storm modeling proves to be a highly efficient strategy, offering several advantages. XGBoost demonstrates exceptional computational power and the ability to handle complex data, enabling precise and reliable predictions of storm intensity and timing. Moreover, its capacity to identify nonlinear relationships among variables provides deeper insights into the factors influencing storm occurrences. Its rapid processing and computational efficiency make it particularly suitable for real-time applications, such as developing early warning systems. These benefits collectively contribute to enhanced disaster management, reduced environmental and economic losses, and the formulation of more effective policies for combating dust storms. This novel approach not only significantly improves prediction accuracy but also provides deeper insights, paving the way for new advancements in managing and mitigating the impacts of dust storms. Additionally, employing the Element-Free Galerkin algorithm for dust storm modeling allows for the precise simulation of complex atmospheric flows with reduced computational demands and costs compared to machine learning approaches. The elimination of grid meshing requirements provides greater flexibility in analyzing regions with intricate geometries. This method's strong capability to minimize numerical errors and model nonlinear interactions further enhances prediction accuracy. Its computational stability and ease of implementation make it an excellent choice for advanced analyses under dynamic atmospheric conditions. These features facilitate the development of more effective strategies for managing the adverse effects of dust storms.

Method

This research focuses on a comparative analysis of dust storm frequency modeling using the XGBoost machine learning model and the mesh-free Galerkin method. The study aims to predict the frequency of dust storm days at eight meteorological stations in Ilam Province (Mehran, Dehloran, Doviraj, Ilam, Ivan- Gharb, Arkavaz-e Malek, Cham-e Gaz, and Kahreh-Heliyan) over a 40-year statistical period (1981–2020). In this study, hourly data on horizontal visibility of less than 1000 meters were employed to detect dust storms for all stations.

XGBoost

This is a machine learning algorithm that belongs to the ensemble learning methods. By combining boosting and regularization techniques, this algorithm constructs a robust and accurate predictive model. In boosting, several weak models (typically decision trees) are trained sequentially, with each new model correcting the errors of the previous one. The algorithm identifies the best split by evaluating all possible division points and uses a sketch algorithm to speed up computations. Specifically, at each training step, gradient descent is applied to update the model parameters and reduce the loss function. The second-order Taylor expansion is used to approximate the loss function and calculate the updated parameters. Furthermore, XGBoost utilizes various optimization strategies, such as column subsampling, row subsampling, and feature importance evaluation, to improve the model's performance and stability. These strategies allow XGBoost to handle high-dimensional and sparse data more effectively, increase the interpretability of the model, and enhance its overall performance and robustness.

EFG algorithm

The element-free Galerkin method, similar to the finite element method, uses formulations based on the principle of virtual work or weighted residual approaches. These formulations allow for the derivation of the

weak form of differential equations, offering greater stability than the strong form. To correctly apply essential boundary conditions, this method typically uses Lagrange multipliers or penalty methods.

Results

Dust storms exert considerable impacts on air quality, public health, and natural ecosystems, necessitating meticulous analysis and robust predictive models. Advanced machine learning approaches, such as XGBoost, harness large-scale datasets and excel in detecting complex patterns, offering powerful tools for forecasting dust storm behavior. In parallel, numerical methods like the Element-Free Galerkin approach provide accurate and stable outcomes, leveraging advanced mathematical formulations, reduced computational costs, and simplified relationships. The integration of these modern and classical methodologies offers a comprehensive framework to enhance analysis and decision-making in dust storm management. In this study, the performance of the XGBoost machine learning model and the Element-Free Galerkin algorithm in modeling dust storms was rigorously evaluated using statistical indices, including R, RMSE, MAE, and NS. The analysis was conducted across eight meteorological stations in Ilam Province (Mehran, Dehloran, Doyrej, Ilam, Ivan Gharb, Arkwaz Malek, Chamgaz, and Kahreh Helian). Modeling accuracy improves with an increasing frequency of dust storm days, with Mehran station demonstrating the highest concordance between observed and predicted values. Based on the results of t-test analysis, it can be concluded that the mean time series of observed and predicted values are equal for both the XGBoost and EFG models. XGBoost, known for its ability to process large and diverse datasets, enhances prediction accuracy while minimizing computational time. Meanwhile, EFG offers a cost-effective and computationally efficient alternative. Combining these methods not only refines predictive accuracy but also enables detailed scenario analysis and practical solutions for risk mitigation, early warnings, and sustainable development initiatives.

Author Contributions

All authors contributed equally to the conceptualization of the article and writing of the original and subsequent drafts.

Data Availability Statement

Data available on request from the authors.

Acknowledgements

The authors would like to thank the reviewers and editor for their critical comments that helped to improve the paper. The authors gratefully acknowledge the support and facilities provided by the Department of Reclamation of arid and mountainous regions, Faculty of Natural Resources, University of Tehran, Iran.

Ethical considerations

The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Conflict of interest

The author declares no conflict of interest.

بررسی مقایسه مدل سازی فراوانی روزهای همراه با طوفان گرد و غبار با استفاده از مدل یادگیری ماشین XGBoost و روش بدون المان گالرکین (مطالعه موردی: استان ایلام)

محمد انصاری قوجقار^۱ | پریا پورمحمد^۲

۱. نویسنده مسئول، گروه مهندسی احیاء مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: Ansari.ghojghar@ut.ac.ir

۲. گروه مهندسی احیاء مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: Paria.pormohamad7@ut.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	مدل سازی طوفان های گرد و غبار به بهبود پیش بینی این پدیده، کاهش ریسک های زیست محیطی و تدوین راهبردهای مناسب برای مقابله با اثرات آن کمک می کند و امکان درک بهتر تعاملات نظام های پویا و ارائه پیش بینی های معتبرتر را فراهم می سازد. در این راستا، استفاده از مدل های یادگیری ماشین با قابلیت تحلیل داده های پیچیده و روش عددی گالرکین با توانایی شبیه سازی دقیق، زمینه ارائه راهکارهای کارآمدتر را مهیا می کند. بدین ترتیب پژوهش حاضر به بررسی مقایسه مدل سازی فراوانی روزهای همراه با طوفان گرد و غبار با استفاده از مدل یادگیری ماشین XGBoost و روش بدون المان گالرکین در هشت ایستگاه در طول دوره آماری ۴۰ ساله (۲۰۲۰-۱۹۸۱) پرداخته است. در هر دو مدل، ترکیب یک به عنوان ترکیب بهینه جهت پیش بینی شاخص FDSD برگزیده شد. نتایج نشان داد که مدل یادگیری ماشین XGBoost با مقدار معیارهای خطای RMSE و MAE کمتر و همچنین مقادیر R و NS نزدیک به واحد، با اختلاف ناچیزی، نسبت به روش بدون المان گالرکین، به عنوان مدل برگزیده جهت مدل سازی طوفان های گرد و غبار در استان ایلام انتخاب شده است. این در حالی است که این اختلاف در سطح اطمینان ۹۵ و ۹۹ درصد معنادار نبود و با توجه به هزینه ها و حجم محاسبات کمتر مدل عددی گالرکین، می توان از الگوریتم عددی گالرکین به عنوان یک روش مناسب جهت مدل سازی شاخص FDSD استفاده نمود. نتایج این پژوهش می تواند در پیش بینی دقیق تر طوفان های گرد و غبار، طراحی سیستم های هشدار، تدوین سیاست های مدیریت بحران و کاهش خسارت های محیط زیستی این طوفان ها، کاربرد داشته باشد.
واژه های کلیدی: مدل های عددی، دید افقی، پیش بینی، درخت تصمیم، شاخص FDSD.	

استناد: انصاری قوجقار؛ محمد، پورمحمد؛ پریا. (۱۴۰۴) بررسی مقایسه مدل سازی فراوانی روزهای همراه با طوفان گرد و غبار با استفاده از مدل یادگیری ماشین XGBoost و روش بدون المان گالرکین (مطالعه موردی: استان ایلام)، مجله تحقیقات آب و خاک ایران، ۵۶ (۶)، ۱۴۳۶-۱۴۳۱.



© نویسندگان.

<https://doi.org/10.22059/ijswr.2025.392122.669904>

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2025.392122.669904>

مقدمه

طوفان های گرد و غبار به عنوان یکی از پدیده های شایع و مخرب جوی در مناطق خشک و نیمه خشک هستند که تحت تأثیر سرعت بالای باد و آشفته گی جریان هوا در سطح خاک های بدون پوشش گیاهی و حساس به فرسایش رخ داده و با تشدید فرآیند بیابان زایی و خشکسالی، کاهش منابع آبی، افزایش شوری خاک، ایجاد تصادفات جاده ای در مناطق بحرانی و افزایش مشکلات تنفسی و سایر بیماری ها، خسارات قابل توجهی در ابعاد اقتصادی، اجتماعی و زیست محیطی به بار می آورند (Goudie, 2009; Guo et al., 2018; Miller et al., 2008; Stuu et al., 2014). بادهای شدید به عنوان مهم ترین عامل این پدیده شناخته می شوند، اما ویژگی های سطح زمین، از جمله پوشش گیاهی، زبری سطح، بافت و رطوبت خاک، نقش مهمی در جابه جایی ذرات شن و انتشار گرد و غبار ایفا می کنند (Singer et al., 2003). این پدیده به طور معمول میدان دید را به شدت کاهش می دهد و در بسیاری از موارد دید افقی به کمتر از یک کیلومتر می رسد (Dayan et al., 2007). در شرایط بحرانی، غلظت ذرات معلق موجود در هوا ممکن است از ۶۰۰۰ میلی گرم بر مترمکعب فراتر رود، که این مقدار می تواند تأثیرات زیست محیطی و بهداشتی قابل توجهی به دنبال داشته باشد (Song et al., 2007). ذرات معلق که در طول این طوفان ها به هوا بلند می شوند، توانایی جابجایی مسافت های طولانی را دارند و گاهی می توانند بیش از ۱۰۰۰ کیلومتر از منبع خود فاصله بگیرند تا در نهایت در مناطق دوردست ته نشین شوند (O'Hara et al., 2006). این ویژگی باعث می شود طوفان های گرد و غبار نه تنها بر مناطق منبع، بلکه بر مناطق پایین دست نیز اثرات زیان بار داشته باشند. منابع تولید ذرات گرد و غبار به دو گروه اصلی منابع طبیعی و منابع انسانی تقسیم می شوند (Alizadeh-Choobari et al., 2014). فعالیت های انسانی، نظیر تغییرات در کاربری زمین از طریق کشاورزی، جنگل زدایی، چرای بیش از حد، مدیریت نادرست منابع آبی (مانند احداث سدهای متعدد) و ... در سطح زمین، به عنوان عوامل کلیدی در ایجاد گرد و غبار شناخته می شوند (Rashki et al., 2012). مناطق مختلف ایران به دلیل قرارگیری در کمربند بیابانی و منطقه خشک و نیمه خشک دنیا، ساختار سطحی خاص، عبور سیستم های همدیدی متعدد، نزدیکی به بیابان ها و منابع عمده تولید گرد و غبار، به طور مداوم تحت تأثیر طوفان های گرد و غبار قرار می گیرند. شدت، مدت، گستردگی عمودی، مسیر حرکت و منابع تولید گرد و غبار در نقاط مختلف کشور متفاوت است. مشکلات و اثرات ناشی از گرد و غبار به ویژه در مناطقی که نزدیک به منابع اصلی تولید گرد و غبار هستند، شدیدتر است. با توجه به ویژگی هایی مانند اندازه ذرات، زمان وقوع، گستردگی عمودی و حرکت افقی گرد و غبار، اثرات آن در مناطق مختلف متفاوت است. از این رو، شناسایی تغییرات مکانی و زمانی گرد و غبار نقش حیاتی در پیش بینی و کاهش خطرات و همچنین مدیریت مؤثر این پدیده دارد (Araghizadeh et al., 2022). نتایج تهیه نقشه حساسیت اراضی استان ایلام به تولید گرد و غبار با استفاده از مدل های داده کاوی نشان داد که حدوداً ۲۳ و ۱۸ درصد مساحت استان ایلام به ترتیب در مناطق با حساسیت زیاد و خیلی زیاد قرار دارد (پورهاشمی، ۱۴۰۲). بررسی پیشینه تحقیق نشان می دهد که علی رغم اهمیت روزافزون طوفان های گرد و غبار و مشکلات اجتماعی، اقتصادی، محیط زیستی فراوان آن، مطالعات چندانی در این حوزه صورت نگرفته است. از طرفی با توجه به اینکه استان ایلام یکی از مناطق پر چالش در حوزه طوفان گرد و غبار می باشد لذا مدلسازی این پدیده مخرب زیست محیطی می تواند در کاهش خسارات اجتماعی، اقتصادی، سلامت و بهداشت، محیط زیست و ... مؤثر واقع شود. در این راستا، در این مطالعه برای رفع چالش های موجود، به بررسی عملکرد مدل های یادگیری ماشین و الگوریتم بدون امان گالرکین جهت مدل سازی طوفان های گرد و غبار پرداخته شده است. ترکیب مدل های عددی با الگوریتم های یادگیری ماشین برای پیش بینی شاخص فراوانی روزهای همراه با طوفان گرد و غبار^۱ (FDS)، رویکردی بدیع و تحول آفرین به شمار می آید. مدل سازی طوفان های گرد و غبار با استفاده از الگوریتم تقویت گرادیانی پیشرفته^۲ (XGBoost) رویکردی کارآمد است که مزایای قابل توجهی را به همراه دارد. این الگوریتم با قدرت پردازش بالا و توانایی تحلیل داده های پیچیده، پیش بینی هایی دقیق و قابل اعتماد از شدت و زمان بندی طوفان ها ارائه می دهد. توانایی XGBoost در شناسایی روابط غیرخطی میان متغیرها، امکان درک عمیق تری از عوامل مؤثر بر وقوع طوفان ها را فراهم می کند. علاوه بر این، سرعت بالا و کارایی محاسباتی آن، این روش را برای کاربردهای بلادرنگ، مانند توسعه سیستم های هشدار، ایده آل می سازد. در مجموع، این مزایا به بهبود مدیریت بحران، کاهش خسارات زیست محیطی و اقتصادی، و تدوین سیاست های مؤثرتر برای مقابله با طوفان های گرد و غبار منجر می شود. این روش نوین نه تنها دقت پیش بینی ها را به طور چشمگیری ارتقا می بخشد، بلکه با ارائه بینش های عمیق تر، افق های تازه ای را در مدیریت و کاهش اثرات طوفان های گرد و غبار می گشاید. علاوه

1 Frequency of Dust Storm Days

2 Extreme Gradient Boosting

براین، مدل سازی طوفان های گرد و غبار با استفاده از الگوریتم بدون المان گالرکین^۱ (EFG) امکان شبیه سازی دقیق تر جریان های پیچیده جوی را با توجه به حجم محاسبات و هزینه های کمتر نسبت به الگوریتم های یادگیری ماشین^۲ فراهم می کند. این روش با حذف نیاز به تقسیم بندی شبکه، انعطاف پذیری بیشتری در تحلیل مناطق با هندسه های پیچیده ارائه می دهد. همچنین، توانایی بالای آن در کاهش خطاهای عددی و مدل سازی تعاملات غیرخطی، دقت پیش بینی ها را بهبود می بخشد. پایداری محاسباتی و سهولت اجرای این روش، آن را به گزینه ای مناسب برای تحلیل های پیشرفته در شرایط متغیر جوی تبدیل کرده است. این ویژگی ها به توسعه راهبردهای کارآمدتر در مدیریت اثرات طوفان های گرد و غبار کمک می کند.

روش شناسی پژوهش

منطقه مورد مطالعه

در پژوهش حاضر به بررسی مقایسه مدل سازی فراوانی روزهای همراه با طوفان گرد و غبار با استفاده از مدل یادگیری ماشین XGBoost و روش بدون المان گالرکین به منظور پیش بینی فراوانی روزهای همراه با طوفان های گرد و غبار در هشت ایستگاه هواشناسی استان ایلام (مهران، دهلران، دویرج، ایلام، ایوان غرب، ارکواز ملک، چم گز و کهره هلیان) با طول دوره آماری ۴۰ سال (۲۰۲۰-۱۹۸۱) پرداخته شده است (شکل ۱). بنابر تعریف سازمان جهانی هواشناسی^۳ (WMO)، روز همراه با طوفان های گرد و غبار روزی است که در آن، حداقل در یکی از هشت سینوپ، یکی از کدهای مرتبط با گرد و غبار (۰۶، ۰۷، ۰۸، ۰۹، ۳۰، ۳۱، ۳۲، ۳۳، ۳۴، ۳۵ و ۹۸) در بخش گزارش هوای حاضر اعلام شده باشد (جدول ۱). مشروط بر این که، داده های قدرت دید افقی متناظر با کد اعلام شده، کمتر از ۱۰۰۰ متر ثبت شده باشد (O'Loingsigh et al., 2014). لذا در این پژوهش، از داده های ساعتی مربوط به (قدرت دید افقی کمتر از ۱۰۰۰ متر) به منظور تشخیص طوفان های گرد و غبار، استفاده شده است (انصاری قوجقار و همکاران، ۱۴۰۱).

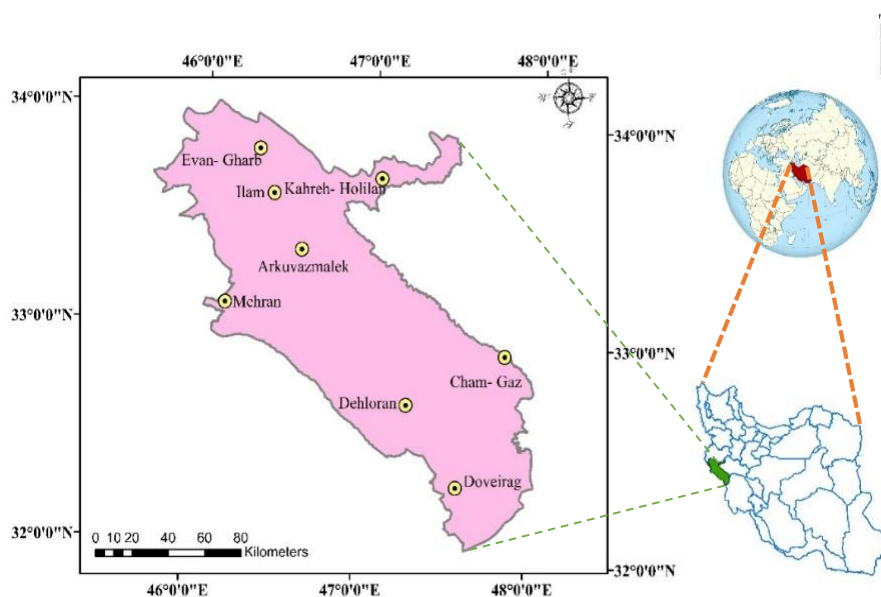
جدول ۱. کدهای سازمان جهانی هواشناسی مرتبط با فرسایش بادی و پدیده های گرد و غبار (O'Loingsigh et al., 2014).

کد	توضیحات
۰۶	رخداد طوفان گرد و خاک گسترده در خارج از ایستگاه هواشناسی
۰۷	خاک یا شن برخاسته از زمین در محدوده ایستگاه هواشناسی
۰۸	مشاهده گردباد در محدوده ایستگاه هواشناسی یا خارج از آن طی ساعت های دیدبانی یا ساعت گذشته
۰۹	طوفان شن یا خاک در زمان دیدبانی در خارج از ایستگاه یا طی ساعت گذشته در محدوده ایستگاه هواشناسی
۳۰	طوفان خفیف یا متوسط شن یا خاک با روند کاهشی طی ساعت گذشته- با سرعت باد ۱۵ متر بر ثانیه یا بیش تر- کاهش میدان دید به کمتر از ۱۰۰۰ ولی بیش تر از ۲۰۰ متر
۳۱	طوفان خفیف یا متوسط شن یا خاک بدون تغییر شدت طی ساعت گذشته- با سرعت باد ۱۵ متر بر ثانیه یا بیش تر- کاهش میدان دید به کمتر از ۱۰۰۰ ولی بیش تر از ۲۰۰ متر
۳۲	طوفان خفیف یا متوسط شن یا خاک با افزایش شدت طی ساعت گذشته- با سرعت باد ۱۵ متر بر ثانیه یا بیش تر- کاهش میدان دید به کمتر از ۱۰۰۰ ولی بیش تر از ۲۰۰ متر
۳۳	طوفان شدید شن یا خاک همراه با روند کاهشی طی ساعت گذشته- سرعت باد ۱۵ متر بر ثانیه یا بیش تر- کاهش دید کمتر از ۲۰۰ متر
۳۴	طوفان شدید شن یا خاک همراه بدون تغییر شدت طی ساعت گذشته- سرعت باد ۱۵ متر بر ثانیه یا بیش تر- کاهش دید کمتر از ۲۰۰ متر
۳۵	طوفان شدید شن یا خاک همراه با افزایش شدت طی ساعت گذشته- سرعت باد ۱۵ متر بر ثانیه یا بیش تر- کاهش دید کمتر از ۲۰۰ متر
۹۸	رعد و برق- بدون بارش- همراه با طوفان شن یا خاک

1 Element- Free Galerkin Method

2 Machine Learning

3 World Meteorological Organization



شکل ۱- معرفی ایستگاه‌های مورد مطالعه در پژوهش حاضر

الگوریتم تقویت گرادیانی پیشرفته (XGBoost)

درخت تصمیم‌گیری تقویت‌شده با گرادیان^۱ (GBDT) مدلی از درخت تصمیم است که بر اساس ایده یادگیری تقویتی ترکیبی^۲ ساخته شده است. XGBoost یک الگوریتم بهینه‌سازی برای GBDT و یک الگوریتم یادگیری ماشین است که به دسته‌بندی الگوریتم‌های یادگیری مجموعه‌ای^۳ تعلق دارد (Chen et al., 2023). این الگوریتم با ترکیب دو روش تقویت^۴ و منظم‌سازی^۵ یک مدل پیش‌بینی قوی و دقیق ایجاد می‌کند. در روش تقویت، چندین مدل ضعیف (معمولاً درخت تصمیم) به‌طور متوالی آموزش داده می‌شود و هر مدل جدید اشتباهات مدل قبلی را اصلاح می‌کند (Suhendra et al., 2023). منظم‌سازی از طریق کاهش نرخ یادگیری و تنظیم اهمیت ویژگی‌ها انجام می‌شود که این امر باعث جلوگیری از بیش‌برازش^۶ و بهبود توانایی تعمیم مدل می‌شود (Maulana et al., 2023). از این طریق، بسط تیلور، تابع هدف را تقریب زده و بهینه‌سازی می‌کند تا توابع ضرر پیچیده و غیریکپارچه به صورت کارآمد بهینه شوند (Chen and Guestrin, 2016). این الگوریتم بر پایه نظریه درخت‌های طبقه‌بندی و رگرسیون بنا شده است و به دلیل این ساختار، در حل مسائل رگرسیون و طبقه‌بندی بسیار کارآمد عمل می‌کند (Ding et al., 2020). در هر مرحله از فرایند آموزش، الگوریتم XGBoost یک درخت تصمیم ایجاد می‌کند تا مجموع تابع ضرر و عبارت منظم‌سازی را به حداقل برساند. عبارت منظم‌سازی به منظور کنترل پیچیدگی مدل مورد استفاده قرار می‌گیرد. با تنظیم وزن مربوط به عبارت منظم‌سازی، این الگوریتم قادر است ضمن کاهش پیچیدگی مدل، توانایی تعمیم‌دهی^۷ آن را بهبود بخشد، در حالی که عملکرد کلی مدل همچنان حفظ می‌شود (Luo et al., 2024). در الگوریتم XGBoost، تابع ضرر معمولاً به صورت جمع کل تفاوت بین پیش‌بینی‌های مدل و مقادیر واقعی به صورت روابط ۱ و ۲ بیان می‌شود.

$$(y_i, \tilde{y}_i) = (y_i - \tilde{y}_i)^2 \quad \text{رابطه ۱}$$

$$\Omega(f) = \gamma \times T + \frac{1}{2} \lambda \times \sum_{j=1}^T w_j^2 \quad \text{رابطه ۲}$$

1 Gradient Boosting Decision Tree

2 Hybrid Reinforcement Learning Approach

3 Ensemble Learning Algorithms

4 Boosting

5 Regularization

6 Overfitting

7 generalization

در رابطه‌های ۱ و ۲، y_i ماتریس مقادیر واقعی، $\tilde{y}_i$ ماتریس مقادیر پیش‌بینی شده مدل، T تعداد گره‌ها، w_j وزن گره‌ها، γ و λ پارامترهای منظم‌سازی هستند که پیچیدگی درخت و میزان جریمه برای وزن‌های گره‌های برگ را کنترل می‌کنند. الگوریتم XGBoost درخت را به‌طور مرحله به مرحله می‌سازد. در هر مرحله ویژگی‌ها و تقسیماتی را انتخاب می‌کند که مقدار تابع ضرر در آن مرحله به حداقل برسد. این الگوریتم با بررسی تمام نقاط ممکن برای تقسیم، بهترین تقسیم را یافته و برای سرعت بخشیدن به محاسبات از الگوریتم اسکچ کمیت وزنی^۱ استفاده می‌کند (Luo et al., 2024). در واقع در هر مرحله از آموزش از الگوریتم نزول گرادیان برای به‌روزرسانی پارامترهای مدل و کاهش تابع ضرر استفاده می‌شود. به‌طور خاص، از بسط دوم تیلور برای تقریب تابع ضرر و محاسبه پارامترهای به‌روزرسانی شده استفاده می‌کند. علاوه بر این، XGBoost از استراتژی‌های بهینه‌سازی مختلف مانند زیرنمونه‌گیری ستون‌ها، زیرنمونه‌گیری ردیف‌ها و ارزیابی اهمیت ویژگی‌ها برای بهبود عملکرد و پایداری مدل بهره می‌برد (Dhaliwal et al., 2022). این استراتژی‌ها به XGBoost کمک می‌کنند تا در مواجهه با داده‌های با ابعاد بالا و پراکنده عملکرد بهتری داشته باشد و مدل را قابل فهم‌تر کند و همچنین پایداری آن را بهبود بخشد.

روش بدون المان گالرکین (EFG)

این روش تکامل یافته روش عناصر پخش شده^۲ (DEM) است که توسط بلیشکو^۳ ارائه شده است (Belytschko et al., 1994). از مشخصه‌های اصلی روش EFG می‌توان به استفاده از روش حداقل مربعات^۴ (MLS) جهت ساخت توابع پایه گالرکین، استفاده از فرم ضعیف معادله انتگرالی مقید^۵ به عنوان یک ابزار برای تبدیل معادلات دیفرانسیل به یک سیستم معادلات جبری و ایجاد شبکه پس‌زمینه جهت انتگرال‌گیری به منظور ساخت ماتریس‌های سیستم در روش‌های عددی، اشاره کرد (Liu et al., 2002). روش MLS به عنوان ابزاری جهت تولید توابع تقریب مناسب، در فضای حل، استفاده می‌شود تا خطای تقریب به حداقل برسد. این فرایند به‌ویژه در مسائل پیچیده یا زمانی که داده‌های کم‌دقت وجود دارند، بسیار مفید است. روش گالرکین بدون المان، مشابه روش اجزای محدود^۶ (FEM)، از فرمول‌بندی‌های مبتنی بر اصل کار مجازی یا رویکردهای باقی‌مانده وزنی استفاده می‌کند. این فرمول‌بندی‌ها به دست آوردن فرم ضعیف معادلات دیفرانسیل، که نسبت به فرم قوی از پایداری بیشتری برخوردار است را ممکن می‌سازند (Liu et al., 2002). به همین دلیل، در روش گالرکین، مشابه روش اجزای محدود، ابتدا معادلات دیفرانسیل مرتبط با مسئله به فرم انتگرالی تبدیل می‌شوند و سپس از طریق انتگرال‌گیری عددی در نقاط گوسی در دامنه مسئله، این فرم به یک سیستم معادلات جبری تبدیل می‌گردد. در روش EFG، به دلیل این‌که توابع شکل فاقد ویژگی تابع کرونگر دلتا^۷ هستند، برای اعمال صحیح شرایط مرزی ضروری معمولاً از ضرایب لاگرانژ^۸ یا روش پنالتی^۹ استفاده می‌شود (Tourei et al., 2024).

معیارهای ارزیابی

در پژوهش حاضر، به منظور ارزیابی دقت و کارایی مدل‌ها، از ضریب همبستگی^{۱۰} (R)، ریشه میانگین مربعات خطا^{۱۱} (RMSE)، میانگین قدر مطلق خطا^{۱۲} (MAE) و ضریب نش-ساتکلیف^{۱۳} (NS) براساس روابط ۳ تا ۶، استفاده شده است. بهترین مقدار برای این چهار نمایه، به ترتیب، یک، صفر، صفر و یک است.

- 1 Weighted Quantile Sketch
- 2 Diffuse Element Method
- 3 Belytschko
- 4 Moving Least Squares
- 5 Constrained Galerkin Integral Equation
- 6 Finite Element Method
- 7 Kronecker Delta Function
- 8 Lagrange Multipliers
- 9 Penalty Method
- 10 Correlation
- 11 Root Mean Square Error
- 12 Mean Absolute Error
- 13 Nash- Sutcliffe

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - f_i)^2} \quad \text{رابطه ۳}$$

$$NS = 1 - \frac{\sum_{i=1}^n (O_i - f_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad \text{رابطه ۴}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - f_i| \quad \text{رابطه ۵}$$

$$R = \frac{\frac{1}{n} \sum_{i=1}^n (O_i - \bar{O})(f_i - \bar{f})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - \bar{f})^2}} \quad \text{رابطه ۶}$$

در رابطه های بالا، O_i مقادیر مشاهده شده در زمان i ، f_i مقادیر پیش بینی شده در زمان i ، $\bar{f}$ میانگین مقادیر پیش بینی شده، $\bar{O}$ میانگین مقادیر مشاهده شده و n تعداد داده ها می باشد. به همین ترتیب، مدلی که RMSE و MAE پایین تری داشته باشد و مقادیر R و NS آن بالاتر و به عدد یک نزدیک تر باشد، به عنوان مدل منتخب جهت مدل سازی طوفان های گرد و غبار در تحقیق حاضر انتخاب می شود.

نتایج مدل یادگیری ماشین XGBoost

جدول ۲، نتایج مدل سازی شاخص FDSD در هشت ایستگاه سینوپتیک استان ایلام در بازه زمانی ۴۰ ساله (۱۹۸۱-۲۰۲۰) را نشان می دهد. با توجه به شاخص های ارزیابی استفاده شده، تمامی ایستگاه ها در ترکیب های اول و دوم نتایج بهتری را به نسبت سایر ترکیب ها از خود نمایش دادند. بدین ترتیب استفاده از یک یا دو فصل پیشین به منظور مدل سازی شاخص فراوانی روزهای همراه با طوفان گرد و غبار، در همه ایستگاه ها، نتایج پیش بینی را بهبود می بخشد. به عنوان مثال در ایستگاه مهران، مقدار معیارهای R و RMSE به ترتیب از ۰/۷۲۱ و ۰/۳۰۳ روز در ترکیب سوم، به ۰/۷۲۳ و ۰/۳۰۱ روز تغییر یافته است. در ایستگاه های دوبرج، ایلام و ایوان غرب نیز، مقادیر NS و R به ترتیب از ۰/۷۸۴، ۰/۷۷۹، ۰/۷۷۶ و ۰/۷۱۲، ۰/۷۰۹ و ۰/۷۰۶ در ترکیب چهارم، به ۰/۷۸۷، ۰/۷۸۳، ۰/۷۷۸ و ۰/۷۱۵، ۰/۷۱۲ و ۰/۷۰۸ افزایش یافته است. لذا می توان نتیجه گرفت که در این بین، ترکیب شماره یک به عنوان ترکیب بهینه جهت پیش بینی و مدل سازی شاخص FDSD با استفاده از مدل یادگیری ماشین XGBoost، انتخاب شده است. با توجه به نتایج معیارهای ارزیابی در ایستگاه کهره هلیان، با کم ترین فراوانی روزهای همراه با طوفان گرد و غبار، با کاسته شدن از تعداد روزهای گرد و غباری، دقت پیش بینی کاهش می یابد. نتایج این قسمت با نتایج Karimzadeh و همکاران (۲۰۲۴) و Tourei و همکاران (۲۰۲۴) مطابقت دارد.

نتایج مدل سازی شاخص FDSD با الگوریتم بدون المان گالرکین

پیاده سازی الگوریتم گالرکین جهت پیش بینی شاخص فراوانی روزهای همراه با طوفان های گرد و غبار در استان ایلام، با نرم افزار R انجام شده است و نتایج مدل سازی در جدول ۳ نمایش داده شده است. مقادیر خطای RMSE و MAE پس از استفاده از ترکیب های سوم و چهارم با داشتن سیر افزایشی، کاهش دقت در مدل سازی طوفان های گرد و غبار را به نمایش گذاشتند. به عنوان مثال در ایستگاه ایلام، مقادیر ارزیابی R، MAE، RMSE و NS به ترتیب از ۰/۷۰۱، ۰/۲۶۳، ۰/۳۲۷ و ۰/۷۵۱ در ترکیب فصلی شماره چهارم، به ۰/۲۶۳، ۰/۳۲۱ و ۰/۷۵۳ در نتیجه استفاده از دو فصل گذشته جهت مدل سازی شاخص FDSD تغییر یافته است. این در حالی است که مقادیر خطای RMSE و MAE، پس از استفاده از یک فصل گذشته به جای دو فصل، به ۰/۳۱۹ و ۰/۲۶۲ کاهش یافته است. همچنین، می توان نتیجه گرفت که عملکرد EFG به منظور پیش بینی طوفان های گرد و غبار با افزایش شاخص FDSD نسبت مستقیمی دارد. نتایج این بخش با مطالعات انصاری قوجقار و همکاران (۱۴۰۱) و Maulana و همکاران (۲۰۲۳) مطابقت دارد.



جدول ۲- نتایج آماری داده‌های ورودی مدل یادگیری ماشین XGBoost به‌منظور پیش‌بینی شاخص FDSD

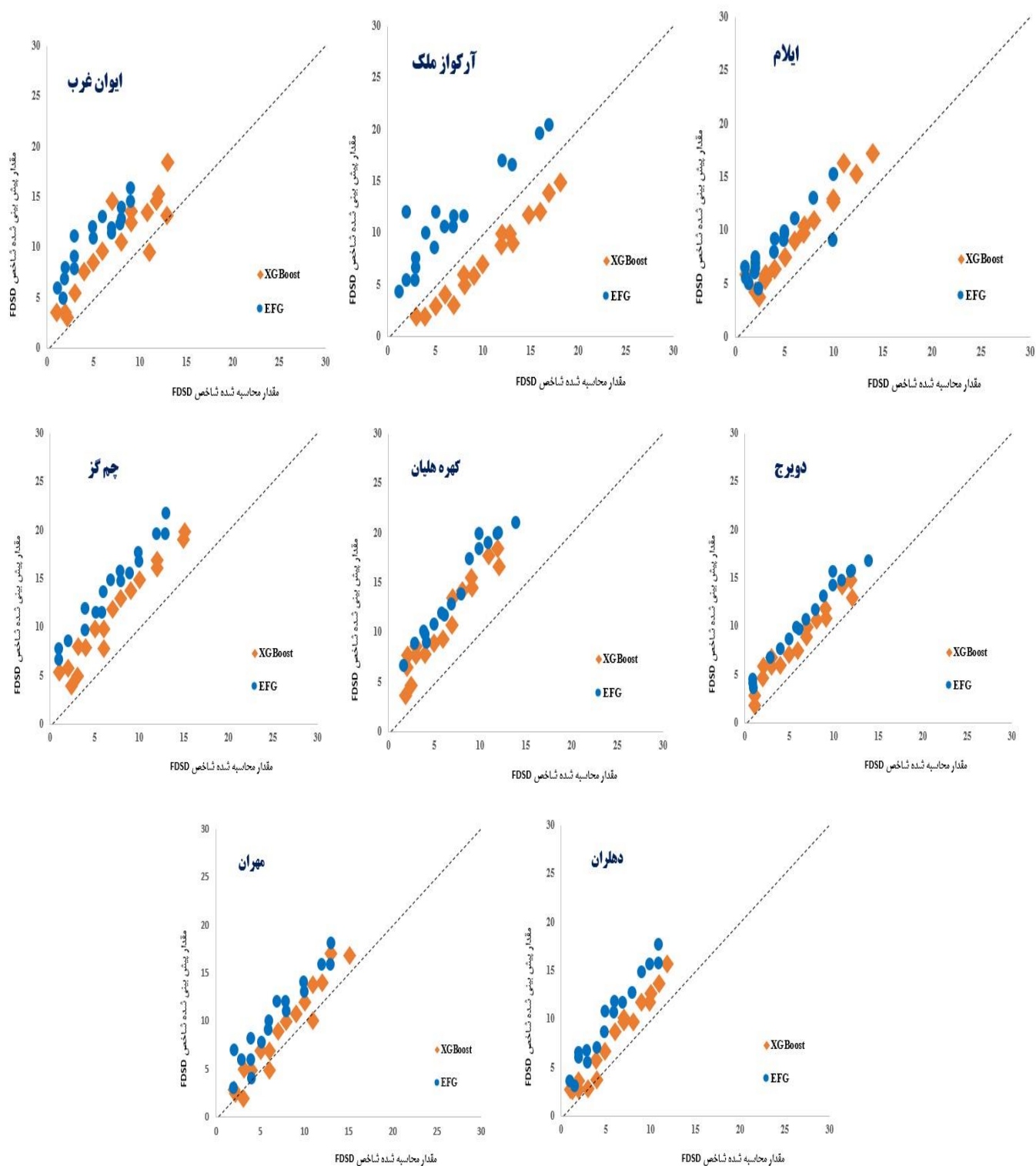
ایستگاه	شماره ترکیب	مجموعه داده آموزش				مجموعه داده آزمون			
		NS	RMSE	MAE	R	NS	RMSE	MAE	R
مهران	۱	۰/۸۲۲	۰/۲۸۴	۰/۲۱۱	۰/۷۵۳	۰/۷۹۳	۰/۳۰۱	۰/۲۳۷	۰/۷۲۳
	۲	۰/۸۲۱	۰/۲۸۳	۰/۲۱۲	۰/۷۵۳	۰/۷۹۲	۰/۳۰۲	۰/۲۳۹	۰/۷۲۱
	۳	۰/۸۲۱	۰/۲۸۵	۰/۲۱۳	۰/۷۵۲	۰/۷۹۱	۰/۳۰۳	۰/۲۳۹	۰/۷۲۱
	۴	۰/۸۲۱	۰/۲۸۵	۰/۲۱۴	۰/۷۵۱	۰/۸۰۲	۰/۳۰۴	۰/۲۴۱	۰/۷۱۹
دهلران	۱	۰/۸۲۱	۰/۲۸۶	۰/۲۱۵	۰/۷۴۹	۰/۸۰۱	۰/۳۰۴	۰/۲۴۲	۰/۷۱۸
	۲	۰/۸۱۹	۰/۲۸۷	۰/۲۱۶	۰/۷۴۸	۰/۷۹۱	۰/۳۰۴	۰/۲۴۳	۰/۷۱۸
	۳	۰/۸۱۸	۰/۲۸۸	۰/۲۱۷	۰/۷۴۶	۰/۷۸۹	۰/۳۰۵	۰/۲۴۴	۰/۷۱۷
	۴	۰/۸۱۷	۰/۲۸۹	۰/۲۱۸	۰/۷۴۵	۰/۷۸۸	۰/۳۰۶	۰/۲۴۵	۰/۷۱۶
دویرج	۱	۰/۸۱۶	۰/۲۹۱	۰/۲۱۹	۰/۷۴۴	۰/۷۸۷	۰/۳۰۶	۰/۲۴۶	۰/۷۱۵
	۲	۰/۸۱۵	۰/۲۹۲	۰/۲۱۹	۰/۷۴۹	۰/۷۸۶	۰/۳۰۶	۰/۲۴۷	۰/۷۱۴
	۳	۰/۸۱۴	۰/۲۹۳	۰/۲۱۹	۰/۷۵۱	۰/۷۸۵	۰/۳۰۷	۰/۲۴۸	۰/۷۱۳
	۴	۰/۸۱۳	۰/۲۹۴	۰/۲۲۱	۰/۷۵۲	۰/۷۸۴	۰/۳۰۸	۰/۲۴۹	۰/۷۱۲
ایلام	۱	۰/۸۱۳	۰/۲۹۵	۰/۲۲۲	۰/۷۴۵	۰/۷۸۳	۰/۳۰۹	۰/۲۵۱	۰/۷۱۲
	۲	۰/۸۱۳	۰/۲۹۵	۰/۲۲۳	۰/۷۴۴	۰/۷۸۲	۰/۳۱۱	۰/۲۵۲	۰/۷۱۱
	۳	۰/۸۱۳	۰/۲۹۶	۰/۲۲۴	۰/۷۴۳	۰/۷۸۱	۰/۳۱۲	۰/۲۵۳	۰/۷۱۰
	۴	۰/۸۱۲	۰/۲۹۷	۰/۲۲۵	۰/۷۴۲	۰/۷۷۹	۰/۳۱۳	۰/۲۵۴	۰/۷۰۹
ایوان غرب	۱	۰/۸۱۱	۰/۲۹۸	۰/۲۲۶	۰/۷۴۱	۰/۷۷۸	۰/۳۱۴	۰/۲۵۵	۰/۷۰۸
	۲	۰/۸۱۰	۰/۲۹۸	۰/۲۲۷	۰/۷۳۹	۰/۷۷۷	۰/۳۱۵	۰/۲۵۶	۰/۷۰۸
	۳	۰/۸۰۹	۰/۲۹۹	۰/۲۲۸	۰/۷۳۸	۰/۷۷۶	۰/۳۱۶	۰/۲۵۷	۰/۷۰۷
	۴	۰/۸۰۹	۰/۳۰۱	۰/۲۲۹	۰/۷۳۷	۰/۷۷۶	۰/۳۱۷	۰/۲۵۸	۰/۷۰۶
آرکواز ملک	۱	۰/۸۰۸	۰/۳۰۱	۰/۲۳۱	۰/۷۳۶	۰/۷۷۶	۰/۳۱۸	۰/۲۵۹	۰/۷۰۵
	۲	۰/۸۰۷	۰/۳۰۱	۰/۲۳۲	۰/۷۳۵	۰/۷۷۵	۰/۳۱۹	۰/۲۶۱	۰/۷۰۴
	۳	۰/۸۱۱	۰/۳۰۲	۰/۲۳۳	۰/۷۳۴	۰/۷۷۴	۰/۳۲۱	۰/۲۶۲	۰/۷۰۳
	۴	۰/۸۰۶	۰/۳۰۳	۰/۲۳۴	۰/۷۳۴	۰/۷۷۳	۰/۳۲۲	۰/۲۶۳	۰/۷۰۲
چم‌گز	۱	۰/۸۰۶	۰/۳۰۴	۰/۲۳۵	۰/۷۳۳	۰/۷۷۲	۰/۳۲۳	۰/۲۶۴	۰/۷۰۲
	۲	۰/۸۰۵	۰/۳۰۵	۰/۲۳۶	۰/۷۳۲	۰/۷۷۱	۰/۳۲۴	۰/۲۶۵	۰/۷۰۲
	۳	۰/۸۰۵	۰/۳۰۶	۰/۲۳۶	۰/۷۳۱	۰/۷۷۰	۰/۳۲۵	۰/۲۶۶	۰/۷۰۲
	۴	۰/۸۰۴	۰/۳۰۷	۰/۲۳۶	۰/۷۲۹	۰/۷۶۹	۰/۳۲۶	۰/۲۶۶	۰/۷۰۲
کهره هلیان	۱	۰/۸۰۹	۰/۳۰۸	۰/۲۳۷	۰/۷۲۸	۰/۷۶۸	۰/۳۲۷	۰/۲۶۷	۰/۷۰۱
	۲	۰/۸۰۸	۰/۳۰۹	۰/۲۳۸	۰/۷۲۷	۰/۷۶۷	۰/۳۲۸	۰/۲۶۸	۰/۷۰۱
	۳	۰/۸۰۷	۰/۳۱۱	۰/۲۳۹	۰/۷۲۶	۰/۷۶۷	۰/۳۲۹	۰/۲۶۹	۰/۷۰۱
	۴	۰/۸۰۶	۰/۳۱۲	۰/۲۴۱	۰/۷۲۵	۰/۷۶۶	۰/۳۳۱	۰/۲۷۱	۰/۷۰۱

جدول ۳- نتایج آماری داده های ورودی روش گالرکین به منظور پیش بینی شاخص FDSD

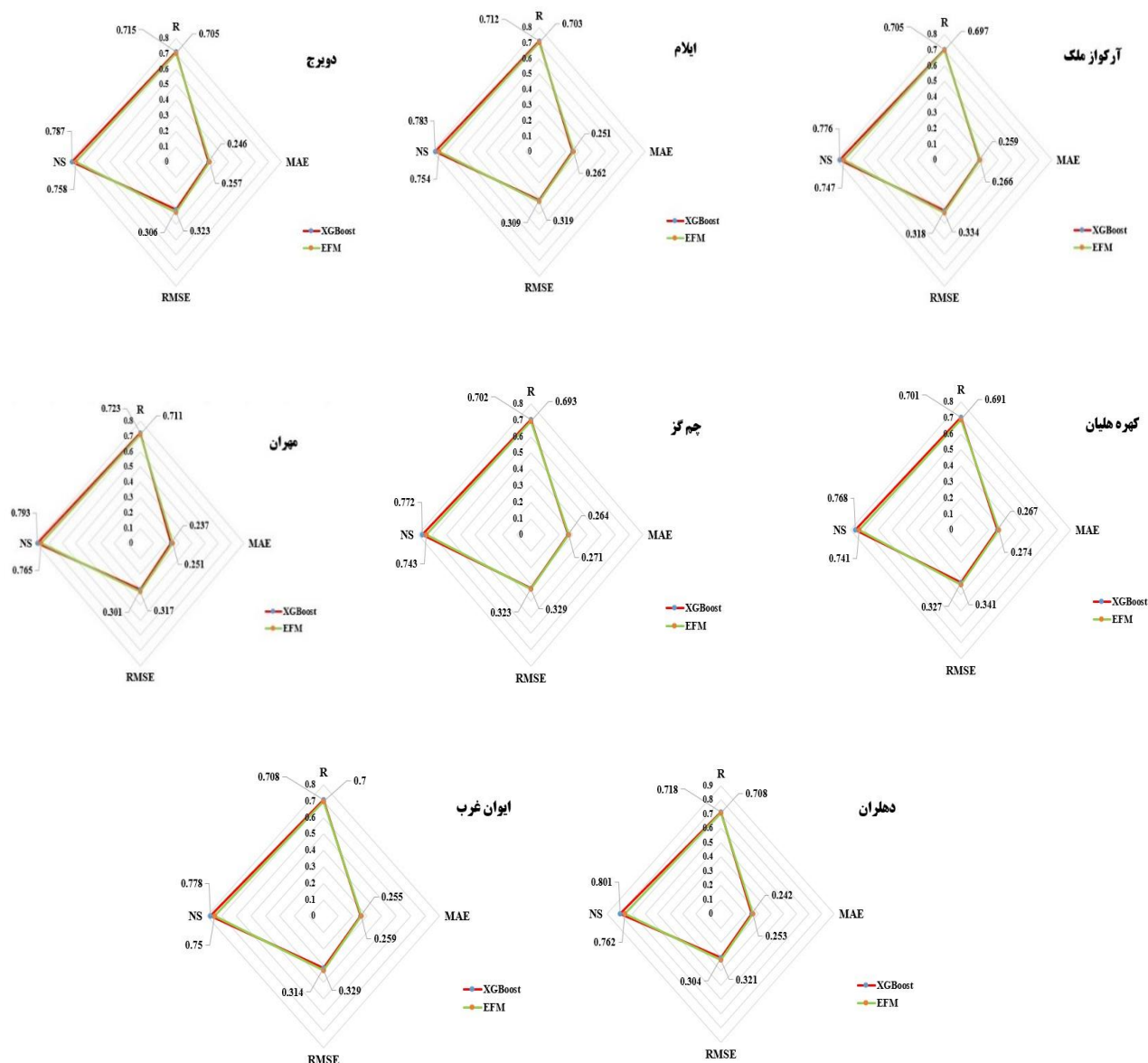
ایستگاه	شماره ترکیب	مجموعه داده آموزش				مجموعه داده آزمون			
		NS	RMSE	MAE	R	NS	RMSE	MAE	R
مهران	۱	۰/۸۱۳	۰/۲۷۶	۰/۲۳۱	۰/۷۴۳	۰/۷۶۵	۰/۳۱۷	۰/۲۵۱	۰/۷۱۱
	۲	۰/۸۱۲	۰/۲۷۶	۰/۲۳۲	۰/۷۴۳	۰/۷۶۵	۰/۳۱۸	۰/۲۵۱	۰/۷۱۱
	۳	۰/۸۱۲	۰/۲۷۷	۰/۲۳۳	۰/۷۴۳	۰/۷۶۴	۰/۳۱۸	۰/۲۵۱	۰/۷۰۹
	۴	۰/۸۱۵	۰/۲۷۸	۰/۲۳۴	۰/۷۴۲	۰/۷۶۳	۰/۳۱۹	۰/۲۵۲	۰/۷۰۹
دهلران	۱	۰/۸۱۴	۰/۲۷۹	۰/۲۳۵	۰/۷۴۱	۰/۷۶۲	۰/۳۲۱	۰/۲۵۳	۰/۷۰۸
	۲	۰/۸۱۴	۰/۲۸۱	۰/۲۳۶	۰/۷۴۰	۰/۷۶۱	۰/۳۲۱	۰/۲۵۴	۰/۷۰۷
	۳	۰/۸۱۳	۰/۲۸۲	۰/۲۳۶	۰/۷۳۹	۰/۷۶۰	۰/۳۲۱	۰/۲۵۵	۰/۷۰۶
	۴	۰/۸۱۲	۰/۲۸۳	۰/۲۳۷	۰/۷۳۸	۰/۷۵۹	۰/۳۲۲	۰/۲۵۶	۰/۷۰۶
دویرج	۱	۰/۸۱۲	۰/۲۸۴	۰/۲۳۷	۰/۷۳۸	۰/۷۵۸	۰/۳۲۳	۰/۲۵۷	۰/۷۰۵
	۲	۰/۸۱۱	۰/۲۸۵	۰/۲۳۹	۰/۷۳۷	۰/۷۵۷	۰/۳۲۴	۰/۲۵۸	۰/۷۰۴
	۳	۰/۸۱۰	۰/۲۸۶	۰/۲۴۱	۰/۷۳۶	۰/۷۵۶	۰/۳۲۵	۰/۲۵۹	۰/۷۰۴
	۴	۰/۸۰۹	۰/۲۸۷	۰/۲۴۲	۰/۷۳۵	۰/۷۵۵	۰/۳۲۶	۰/۲۶۱	۰/۷۰۳
ایلام	۱	۰/۸۰۸	۰/۲۸۸	۰/۲۴۳	۰/۷۳۵	۰/۷۵۴	۰/۳۱۹	۰/۲۶۲	۰/۷۰۳
	۲	۰/۸۰۷	۰/۲۸۹	۰/۲۴۴	۰/۷۳۴	۰/۷۵۳	۰/۳۲۱	۰/۲۶۳	۰/۷۰۲
	۳	۰/۸۰۶	۰/۲۸۹	۰/۲۴۵	۰/۷۳۳	۰/۷۵۲	۰/۳۲۲	۰/۲۶۳	۰/۷۰۲
	۴	۰/۸۰۵	۰/۲۸۹	۰/۲۴۵	۰/۷۳۲	۰/۷۵۱	۰/۳۲۷	۰/۲۶۳	۰/۷۰۱
ایوان غرب	۱	۰/۸۰۴	۰/۲۹۱	۰/۲۴۵	۰/۷۳۱	۰/۷۵۰	۰/۳۲۹	۰/۲۵۹	۰/۷۰۰
	۲	۰/۸۰۳	۰/۲۹۱	۰/۲۴۶	۰/۷۳۱	۰/۷۴۹	۰/۳۳۱	۰/۲۶۱	۰/۶۹۹
	۳	۰/۸۰۲	۰/۲۹۲	۰/۲۴۶	۰/۷۲۹	۰/۷۴۹	۰/۳۳۲	۰/۲۶۴	۰/۶۹۹
	۴	۰/۸۰۲	۰/۲۹۳	۰/۲۴۷	۰/۷۲۶	۰/۷۴۸	۰/۳۳۳	۰/۲۶۵	۰/۶۹۸
آرکواز ملک	۱	۰/۸۰۱	۰/۲۹۴	۰/۲۴۸	۰/۷۲۵	۰/۷۴۷	۰/۳۳۴	۰/۲۶۶	۰/۶۹۷
	۲	۰/۸۰۱	۰/۲۹۴	۰/۲۴۹	۰/۷۲۴	۰/۷۴۶	۰/۳۳۵	۰/۲۶۷	۰/۶۹۶
	۳	۰/۸۰۳	۰/۲۹۵	۰/۲۵۱	۰/۷۲۳	۰/۷۴۵	۰/۳۳۶	۰/۲۶۸	۰/۶۹۵
	۴	۰/۸۰۴	۰/۲۹۶	۰/۲۵۲	۰/۷۲۳	۰/۷۴۴	۰/۳۳۷	۰/۲۶۹	۰/۶۹۴
چم گز	۱	۰/۸۰۵	۰/۲۹۷	۰/۲۵۳	۰/۷۲۲	۰/۷۴۳	۰/۳۲۹	۰/۲۷۱	۰/۶۹۳
	۲	۰/۸۰۵	۰/۲۹۷	۰/۲۵۴	۰/۷۲۱	۰/۷۴۳	۰/۳۳۱	۰/۲۷۱	۰/۶۹۳
	۳	۰/۸۰۵	۰/۲۹۷	۰/۲۵۵	۰/۷۲۰	۰/۷۴۲	۰/۳۳۹	۰/۲۷۲	۰/۶۹۳
	۴	۰/۸۰۳	۰/۲۹۹	۰/۲۵۶	۰/۷۱۹	۰/۷۴۱	۰/۳۴۱	۰/۲۷۳	۰/۶۹۲
کهره هلیان	۱	۰/۸۰۲	۰/۳۰۱	۰/۲۵۲	۰/۷۱۸	۰/۷۴۱	۰/۳۴۱	۰/۲۷۴	۰/۶۹۱
	۲	۰/۸۰۱	۰/۳۰۲	۰/۲۵۷	۰/۷۱۷	۰/۷۳۹	۰/۳۴۱	۰/۲۷۵	۰/۶۹۱
	۳	۰/۸۰۱	۰/۳۰۳	۰/۲۵۸	۰/۷۱۷	۰/۷۳۸	۰/۳۴۲	۰/۲۷۶	۰/۶۹۰
	۴	۰/۷۹۹	۰/۳۰۴	۰/۲۵۹	۰/۷۱۶	۰/۷۳۷	۰/۳۴۳	۰/۲۷۶	۰/۶۹۰

جدول ۴- آزمون مقایسه میانگین سری زمانی مشاهداتی و پیش بینی شده شاخص FDSD

میانگین	آماره t	مشاهداتی			
		EFG	XGBoost	EFG	XGBoost
مهران	۰/۹۰۸	۰/۸۸۵	۰/۸۷۷	۱/۷۹۲	۱/۷۸۸
دهلران	۰/۹۰۲	۰/۸۷۶	۰/۸۷۳	۱/۷۸۳	۱/۷۷۴
دویرج	۰/۸۹۴	۰/۸۶۹	۰/۸۶۶	۱/۷۶۴	۱/۷۶۱
ایلام	۰/۸۸۵	۰/۸۵۳	۰/۸۴۹	۱/۷۵۶	۱/۷۵۱
ایوان غرب	۰/۸۷۳	۰/۸۴۱	۰/۸۳۶	۱/۷۴۲	۱/۷۳۹
آرکواز ملک	۰/۸۵۴	۰/۸۳۴	۰/۸۳۷	۱/۷۳۶	۱/۷۳۷
چم گز	۰/۸۳۲	۰/۸۲۷	۰/۸۲۲	۱/۷۳۴	۱/۷۲۵
کهره هلیان	۰/۸۰۷	۰/۸۰۷	۰/۸۰۱	۱/۷۰۷	۱/۷۰۳



شکل ۲- مقایسه مقادیر مشاهده شده و پیش بینی شده شاخص FDSD در استان ایلام



شکل ۳- مقایسه شاخص های ارزیابی مدل بهینه جهت مدل سازی طوفان های گرد و غبار

نتیجه گیری

مدل سازی طوفان های گرد و غبار به عنوان یکی از چالش های محیط زیستی و اقلیمی، اهمیت ویژه ای در مدیریت بحران، پیشگیری از خسارات و برنامه ریزی های کلان دارد. این پدیده با تأثیر بر کیفیت هوا، سلامت انسان و اکوسیستم های طبیعی، نیازمند تحلیل دقیق و پیش بینی های معتبر می باشد. در این راستا، مدل های یادگیری ماشین با بکارگیری حجم زیادی از داده ها و توانایی شناسایی الگوهای پیچیده، ابزاری قدرتمندی جهت پیش بینی رفتار این طوفان ها می باشند. این مدل ها با الگوریتم های پیچیده محاسباتی و افزایش دقت، امکان تحلیل دقیق و کارآمد را فراهم می کنند. از سوی دیگر، روش های عددی مانند گالرکین نیز به دلیل ارائه فرمول بندی های قوی و ضعیف و استفاده از اصول ریاضیاتی پیشرفته، به ویژه در مدل سازی های عددی، کاهش هزینه ها، حجم محاسباتی اندک و ساده سازی روابط، توانایی ارائه نتایج دقیق و پایدار را دارند. این ترکیب از فناوری های نوین و روش های عددی کلاسیک، بستری مناسب برای بهبود تحلیل ها و تصمیم گیری ها در حوزه مدیریت طوفان های گرد و غبار فراهم می آورد. بدین ترتیب، در پژوهش حاضر از شاخص های ارزیابی R ، MAE ، $RMSE$ و NS به منظور مقایسه عملکرد مدل یادگیری ماشین $XGBoost$ و الگوریتم بدون المان گالرکین جهت مدل سازی طوفان های گرد و غبار در هشت ایستگاه (مهران، دهلران، دویرج، ایلام، ایوان غرب، ارکواز ملک، چم گز و کهره هلیان) در استان ایلام استفاده شده است. جدول شماره ۴، میانگین مقادیر مشاهداتی و پیش بینی شده شاخص FSD در این استان را با استفاده از آزمون t نشان

می‌دهد. براساس نتایج این آزمون، می‌توان نتیجه گرفت که میانگین سری‌های زمانی مشاهده‌شده و پیش‌بینی شده برای هر دو مدل XGBoost و EFG برابر می‌باشد. در شکل ۲، بهبود عملکرد مدل‌سازی و افزایش دقت روش‌های استفاده‌شده، متناسب با افزایش شاخص فراوانی روزهای همراه با طوفان گرد و غبار به‌خوبی نمایش داده شده است. ایستگاه مهران با بیش‌ترین فراوانی روزهای گرد و غباری در منطقه مورد مطالعه، بهترین و بیش‌ترین تطابق موجود میان مقادیر مشاهده‌شده و پیش‌بینی شده را دارد. ارتباط بین مقادیر مشاهداتی و پیش‌بینی شده در هر هشت ایستگاه در شکل ۲، توسط نیم‌ساز ربع اول (منحنی $F(x) = x$) تبیین می‌شود. لذا با توجه موارد بیان شده می‌توان نتیجه گرفت که مدل یادگیری ماشین XGBoost با بیش‌ترین تطابق نسبت به این منحنی و با اختلاف اندکی نسبت به روش EFG، عملکرد بهتر و راندمان بالاتری نسبت به روش بدون المان گالرکین در مدل‌سازی طوفان‌های گرد و غبار دارد. همان‌طور که در شکل ۳ مشخص است، مدل یادگیری ماشین XGBoost با بیش‌ترین مقدار ضریب همبستگی و کم‌ترین خطا نسبت به روش گالرکین، به‌عنوان مدل بهینه جهت مدل‌سازی طوفان‌های گرد و غبار در استان ایلام انتخاب شده است. روش گالرکین، با توانایی بالا در شبیه‌سازی دقیق پدیده‌های فیزیکی و ارائه فرمول‌بندی‌های عددی پایدار، می‌تواند جزئیات پیچیده‌ای از فرآیندهای مرتبط با طوفان‌های گرد و غبار را مدل‌سازی کند. این روش به‌ویژه در درک تأثیر متقابل عوامل مختلف محیطی مانند سرعت باد، رطوبت خاک، و تراکم ذرات معلق نقش کلیدی دارد و امکان ارائه سناریوهای دقیق برای تحلیل شرایط بحرانی را فراهم می‌کند. از سوی دیگر، مدل یادگیری عمیق XGBoost با توانایی پردازش داده‌های حجیم و متنوع، ابزاری قدرتمند برای پیش‌بینی رفتار طوفان‌های گرد و غبار و شناسایی الگوهای پنهان در داده‌ها ارائه می‌دهد. این مدل با بهبود دقت پیش‌بینی‌ها و کاهش زمان محاسباتی، به مدیران و سیاست‌گذاران اجازه می‌دهد تا تصمیم‌گیری‌های بهتری در زمینه اقدامات پیشگیرانه و برنامه‌ریزی بلندمدت انجام دهند. از سوی دیگر، با توجه به حجم محاسباتی و هزینه‌های کمتر الگوریتم بدون المان گالرکین در مقایسه با مدل یادگیری ماشین XGBoost و همچنین اختلاف اندک معیارهای نیکویی برازش این دو روش، می‌توان نتیجه گرفت که الگوریتم بدون المان EFG یکی از روش‌های باصرفه، کم هزینه و کارآمد نسبت به مدل‌های حجیم و پیچیده یادگیری ماشین به منظور مدل‌سازی و پیش‌بینی فراوانی روزهای همراه با طوفان گرد و غبار محسوب می‌شود. ترکیب این روش‌ها نه تنها باعث افزایش دقت پیش‌بینی‌ها و شبیه‌سازی‌ها می‌شود، بلکه امکان تحلیل سناریوهای پیچیده و ارائه راهکارهای عملی برای کاهش مخاطرات را نیز فراهم می‌کند. در حوزه مدیریت بحران و کاهش مخاطرات، این ترکیب می‌تواند با مدل‌سازی شاخص FDSD مناطق پرخطر و ارائه هشدارهای زودهنگام، به جلوگیری از خسارات جبران‌ناپذیر جانی و مالی کمک کند. همچنین، این رویکرد می‌تواند در طراحی سیاست‌های کارآمد برای کاهش انتشار گرد و غبار، مدیریت منابع طبیعی و بهبود کیفیت هوا تأثیر قابل توجهی داشته باشد.

"هیچ‌گونه تعارض منافع توسط نویسندگان وجود ندارد"

منابع

- انصاری قوجقار، محمد؛ بذرافشان، جواد و عراقی نژاد، شهاب. (۱۴۰۱). ارزیابی کارایی فرامدل‌های هیبریدی یادگیری ماشین و باکس جنکینز به‌منظور مدل‌سازی طوفان‌های گرد و غبار (مطالعه موردی: استان خوزستان). *تحقیقات آب و خاک ایران*. (۸) ۵۳. doi: 1695-1714. 10.22059/ijswr.2022.346694.669335
- پورهایمی، سیما. (۱۴۰۲). تهیه نقشه حساسیت اراضی استان ایلام به تولید گرد و غبار با استفاده از مدل‌های داده‌کاوی. (E.E.R). (۳) ۱۴، ۸۳-۱۰۱.

REFERENCES

- Alizadeh-Chooabari, O., Zawar-Reza, P., & Sturman, A. (2014). The "wind of 120 days" and dust storm activity over the Sistan Basin. *Atmospheric research*, 143, 328-341.
- Ansari Ghoghjar, M., Bazrafshan, J., & Araghinejad, S. (2022). Evaluating the Efficiency of Hybrid Metamodels of Machine Learning and Box Jenkins in Order to Model Dust Storms (Case Study: Khuzestan Province). *Iranian Journal of Soil and Water Research*, 53(8), 1695-1714. [In Persian]
- Araghizadeh, M., & Masoodian, S. A. (2022). Comparative study of aerosol optical depth satellite data with earth's observations. *Desert*, 27(1), 153-166.
- Belytschko, T., Lu, Y. Y., & Gu, L., 1994. Element-free Galerkin methods. *International Journal for Numerical Methods in Engineering*, 37(22), 2292-2562.
- Chen, C., Yin, C., Wang, Y., Zeng, J., Wang, S., Bao, Y., ... & Liu, X. (2023). XGBoost-based machine learning test improves the accuracy of hemorrhage prediction among geriatric patients with long-term

- administration of rivaroxaban. *BMC geriatrics*, 23(1), 418.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Dayan, U., Ziv, B., Shoob, T., & Enzel, Y. (2008). Suspended dust over southeastern Mediterranean and its relation to atmospheric circulations. *International Journal of Climatology*, 28(7), 915-924.
- Dhaliwal, J. K., Panday, D., Saha, D., Lee, J., Jagadamma, S., Schaeffer, S., & Mengistu, A. (2022). Predicting and interpreting cotton yield and its determinants under long-term conservation management practices using machine learning. *Computers and Electronics in Agriculture*, 199, 107107.
- Ding, X., Jiang, T., Xue, W., Li, Z., & Zhong, Y. (2020). A new method of human gesture recognition using Wi-Fi signals based on XGBoost. In *2020 IEEE/CIC International Conference on Communications in China (ICCC Workshops)* (pp. 237-241). IEEE.
- Givehchi, R., Arhami, M., & Tajrishy, M. (2013). Contribution of the Middle Eastern dust source areas to PM10 levels in urban receptors: Case study of Tehran, Iran. *Atmospheric environment*, 75, 287-295.
- Goudie, A. S. (2009). Dust storms: Recent developments. *Journal of environmental management*, 90(1), 89-94.
- Guo, L., Fan, B., Zhang, F., Jin, Z., & Lin, H. (2018). The clustering of severe dust storm occurrence in China from 1958 to 2007. *Journal of Geophysical Research: Atmospheres*, 123(15), 8035-8046. <https://doi.org/10.1029/2018JD029042>.
- Karimzadeh, S., Mohammadi, A., Salahuddin, U., Carvalho, A., & Lourenço, P. B. (2024). Backbone ground motion model through simulated records and XGBoost machine learning algorithm: An application for the Azores plateau (Portugal). *Earthquake Engineering & Structural Dynamics*, 53(2), 668-693.
- Kurosaki, Y., & Mikami, M. (2003). Recent frequent dust events and their relation to surface wind in East Asia. *Geophysical Research Letters*, 30(14).
- Liu, G. R., Chen, X. L., & Reddy, J. N. (2002). Buckling of symmetrically laminated composite plates using the element-free Galerkin method. *International Journal of Structural Stability and Dynamics*, 2(03), 281-294.
- Liu, Y., Wang, G., Hu, Z., Shi, P., Lyu, Y., Zhang, G., ... & Liu, L. (2020). Dust storm susceptibility on different land surface types in arid and semiarid regions of northern China. *Atmospheric research*, 243, 105031.
- Luo, S., El, X., & Li, X. (2024). Data Preprocessing Method for Landslide Displacement Prediction Based on XG Boost. In *2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS)* (pp. 745-750). IEEE.
- Maulana, A., Noviandy, T. R., Suhendra, R., Earlia, N., Sofyan, H., Subianto, M., & Idroes, R. (2023). Performance Analysis and Feature Extraction for Classifying the Severity of Atopic Dermatitis Diseases. *2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)*, 226-231. <https://doi.org/10.1109/COSITE60233.2023.10249760>
- Miller, S. D., Kuciauskas, A. P., Liu, M., Ji, Q., Reid, J. S., Breed, D. W., ... & Mandoos, A. A. (2008). Haboob dust storms of the southern Arabian Peninsula. *Journal of Geophysical Research: Atmospheres*, 113(D1).
- Moridnejad, A., Karimi, N., & Ariya, P. A. (2015). A new inventory for middle east dust source points. *Environmental monitoring and assessment*, 187, 1-11.
- Nayroles, B., Touzot, G., & Villon, P., 19921992. Generalizing the finite element method: Diffuse approximation and diffuse elements. *Computational Mechanics*, 1010(Δ).
- O'Hara, S. L., Clarke, M. L., & Elatrash, M. S. (2006). Field measurements of desert dust deposition in Libya. *Atmospheric Environment*, 40(21), 3881-3897.
- O'Loingsigh, T., McTainsh, G. H., Tews, E. K., Strong, C. L., Leys, J. F., Shinkfield, P., & Tapper, N. J. (2014). The Dust Storm Index (DSI): a method for monitoring broadscale wind erosion using meteorological records. *Aeolian Research*, 12, 29-40.
- Pourhashemi, S. (2024). Preparing a map of the sensitivity of the lands of Ilam province to dust production using data mining models. *Environmental Erosion Research Journal*, 14(3), 83-101. [In Persian]
- Rashki, A., Kaskaoutis, D. G., Rautenbach, C. D., Eriksson, P. G., Qiang, M., & Gupta, P. (2012). Dust storms and their horizontal dust loading in the Sistan region, Iran. *Aeolian Research*, 5, 51-62.
- Singer, A. R. I. E. H., Zobeck, T., Poberezsky, L., & Argaman, E. (2003). The PM10 and PM2.5 dust generation potential of soils/sediments in the Southern Aral Sea Basin, Uzbekistan. *Journal of Arid Environments*, 54(4), 705-728.
- Song, Z., Wang, J., & Wang, S. (2007). Quantitative classification of northeast Asian dust events. *Journal of Geophysical Research: Atmospheres*, 112(D4).



- Stuut, J. B. W. (2014). Subaquatic dust deposits. *Mineral Dust: A Key Player in the Earth System*, 443-462.
- Suhendra, R., Suryadi, S., Husdayanti, N., Maulana, A., Noviandy, T. R., Sasmita, N. R., Subianto, M., Earlia, N., Niode, N. J., & Idroes, R. (2023). Evaluation of Gradient Boosted Classifier in Atopic Dermatitis Severity Score Classification. *Heca Journal of Applied Sciences*, 1(2), 54-61. <https://doi.org/10.60084/hjas.v1i2.85>.
- Tourei, A., Pak, A., & Iranmanesh, M. A. (2024). Numerical modeling of land subsidence induced by groundwater extraction considering unsaturated effects and using element-free Galerkin (EFG) method. *Sharif Journal of Civil Engineering*, (), -. doi: 10.24200/j30.2023.59625.3256