

Original Article



Optimization of the Results of the ML-Based GMDH Algorithm in order to Increase the Accuracy of Walnut Pollen Detection and Horizontal Optical Depth through the TLBO Algorithm

Mehdi Amiri^{1*}, Farzad Amiri², Mohammad Hossein Pourasad³, Seyfollah Soleimani⁴

Affiliation

1. Prof. and Director of Information Technology, Dep. of Information Technology, Allameh Tabarsi Applied Science and Technology Center, University of Applied Science, Tehran, Iran
2. Assistant Prof. of Dep. of industrial Engineering, Faculty of Engineering Management, Kermanshah University of Technology, Kermanshah, Iran
3. Ph.D. Student of Dep. of Health Information Technology, Faculty of Paramedical, Kermanshah University of Medical Sciences, Kermanshah, Iran
4. Assistant Prof., Dep. of Computer Engineering, Faculty of Engineering, Arak University, Arak, Iran

Citation: Amiri, M., Amiri, F., Pourasad, M. H., Soleimani, S., Optimization of the Results of the ML-Based GMDH Algorithm in order to Increase the Accuracy of Walnut Pollen Detection and Horizontal Optical Depth through the TLBO Algorithm, Iran J Remote Sens GIS. 16(1):49-66.

ABSTRACT

Introduction: As one of the most essential needs of living beings, clean air quality has been threatened by natural and human activities. In recent years, dust storms have been increasing spatially and temporally, causing numerous damages to social, economic, and environmental health for the residents of the southern and southwestern regions of Iran. In the present study, MODIS sensor data were used to investigate dust storms and detect horizontal optical depth.

Materials and Methods: The advantages of MODIS sensor data include high spectral and temporal resolution. Additionally, meteorological station data were collected based on the study period. After preprocessing the data and preparing field observations, the necessary features for modeling were extracted using the differential method between selected bands of each MODIS sensor image, along with features extracted from ground-based meteorological station sensors. After further investigations and evaluations and using the viewpoints of meteorological experts, 36 differential features from various MODIS image bands and six features from ground-based meteorological station data, totaling 42 features, were extracted. Subsequently, using feature selection techniques, the best features were identified. A novel method named ML-Based GMDH, which improves the GMDH neural network by altering partial functions with machine learning models, was employed to detect dust concentration and horizontal optical depth. To achieve optimal accuracy, the hyper-parameters of this model were heuristically tuned using the TLBO optimization algorithm. Additionally, machine learning methods such as Basic GMDH, SVM, MLP, MLR, RF, and their ensemble models were implemented to compare with the main approach. According to the results, the TLBO-tuned ML-Based GMDH method provided superior accuracy in detecting dust concentration compared to the aforementioned machine-learning methods.

Results and Discussion: The SVM-PSO method was selected as the best method in the feature selection phase, the RF method was chosen as the best method among basic classification methods, and the Ensemble SVM and Ensemble RF methods were selected as the best methods in the ensemble and classification phase. It was also observed that using the ensemble approach led to a desirable improvement in horizontal optical depth classification. In the second approach, a method titled ML-Based GMDH, which improves the GMDH neural network by altering partial functions with machine learning algorithms, was used for estimating dust concentration. Additionally, to achieve suitable accuracy, the hyper-parameters of this model were finely tuned using the TLBO optimization algorithm. The results showed that this method provided appropriate accuracy in estimating dust concentration and horizontal optical depth, outperforming the best-selected methods from the first approach.

Keywords: Dust, Horizontal Optical Detection, Remote Sensing, Feature selection, Machine Learning, MODIS, TLBO Algorithm, GMDH Neural Network.

* Corresponding Author: mehdiamiri@tabarsi-uast.com
DOI: <https://doi.org/10.48308/gisj.2022.102900>

Received: 2022.02.05
Accepted: 2022.11.12



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license <https://creativecommons.org/licenses/by/4.0/>.

بهینه‌سازی نتایج الگوریتم ML-Based GMDH به منظور افزایش دقت تشخیص گردوغبار و عمق دید افقی از طریق الگوریتم TLBO

مهدی امیری^{۱*}، فرزاد امیری^۲، محمدحسین پوراسد^۳، سیفاله سلیمانی^۴



سمت

۱. استاد گروه فناوری اطلاعات، مرکز علمی کاربردی علوم و فنون علامه طبرسی، دانشگاه جامع علمی کاربردی، تهران، ایران
۲. استادیار گروه مهندسی صنایع، دانشکده مدیریت مهندسی، دانشگاه صنعتی کرمانشاه، کرمانشاه، ایران
۳. دانشجوی دکتری گروه فناوری اطلاعات سلامت، دانشکده پیراپزشکی، دانشگاه علوم پزشکی کرمانشاه، کرمانشاه، ایران
۴. استادیار گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه اراک، اراک، ایران

چکیده

سابقه و اهداف: کیفیت هوای پاک، به‌منزله یکی از ضروری‌ترین نیازهای موجودات زنده، برآثر فعالیت‌های طبیعی و انسانی به‌مخاطره افتاده است. در سال‌های اخیر، طوفان‌های گردوغبار از لحاظ مکانی و زمانی همواره در حال افزایش بوده و سبب آسیب‌های بی‌شمار در حوزه سلامت اجتماعی، اقتصادی و زیست‌محیطی، برای ساکنان مناطق جنوب و جنوب‌غرب ایران، شده است. در پژوهش حاضر، به‌منظور بررسی طوفان‌های گردوغبار و تشخیص عمق دید افقی، داده‌های سنجنده مادیس به‌کار رفته است.

مواد و روش‌ها: از مزایای داده‌های سنجنده مادیس می‌توان به توان تفکیک طیفی و زمانی بالا اشاره کرد. همچنین داده‌های ایستگاه‌های هواشناسی با توجه به بازه زمانی مورد مطالعه جمع‌آوری شده است. پس از پیش‌پردازش داده‌ها و آماده‌سازی مشاهدات میدانی، به‌منظور استخراج ویژگی‌های مورد نیاز برای انجام دادن مدل‌سازی‌ها، از طریق روش تفاضلی بین باندهای منتخب هر تصویر داده‌های سنجنده مادیس، به‌همراه ویژگی‌های استخراج‌شده از سنسورهای ایستگاه‌های هواشناسی زمینی استفاده شده است. با بررسی‌های بیشتر و ارزیابی‌های صورت‌گرفته و استفاده از دیدگاه‌های خبرگان هواشناسی، ۳۶ ویژگی تفاضلی از باندهای گوناگون تصاویر مادیس و شش ویژگی از داده‌های ایستگاه‌های هواشناسی زمینی، یعنی در مجموع ۴۲ ویژگی، استخراج شده است. در ادامه، از طریق تکنیک‌های انتخاب ویژگی، بهترین ویژگی‌ها شناسایی و با به‌کارگیری روشی جدید با نام ML-Based GMDH، که حاصل بهبود شبکه عصبی GMDH از طریق تغییر توابع جزئی با مدل‌های یادگیری ماشین است، برای تشخیص غلظت گردوغبار و دید افقی استفاده شد. برای دستیابی به دقت مناسب نیز ابرپارامترهای این مدل به‌صورت ابتکاری، با استفاده از الگوریتم بهینه‌سازی TLBO، تنظیم شدند. در ادامه، روش‌های یادگیری ماشین Basic GMDH SVM، MLP، MLR، RF و مدل گروهی آنها نیز، برای مقایسه با رویکرد اصلی، اجرایی شد؛ طبق نتایج، روش ML-Based GMDH تنظیم‌شده با TLBO با ایجاد بهبود در قیاس با روش‌های یادگیری ماشین ذکرشده، دقت بهتری را در تشخیص غلظت گردوغبار فراهم کرده است.

نتایج و بحث: روش SVM-PSO به‌منزله روش برتر در مرحله انتخاب ویژگی، روش RF به‌منزله روش برتر در میان روش‌های پایه دسته‌بندی و روش‌های Ensemble SVM و Ensemble RF به‌منزله روش‌های برتر در مرحله گروهی و دسته‌بندی انتخاب شدند. همچنین مشاهده شد، با استفاده از رویکرد گروهی، بهبود مطلوبی در تشخیص دسته دید افقی پدید آمد. در رویکرد دوم، روشی با عنوان ML-Based GMDH که حاصل بهبود شبکه عصبی GMDH از طریق تغییر توابع جزئی با مدل‌های یادگیری ماشین است، استفاده شد که کاربرد آن در تقریب غلظت گردوغبار است. همچنین، برای دستیابی به دقت مناسب، ابرپارامترهای این مدل با الگوریتم بهینه‌سازی TLBO با دقت بسیار بالا تنظیم شدند. نتایج حاصل نشان دادند این روش، با ایجاد بهبود در مقایسه با بهترین روش‌های انتخابی از رویکرد اول، دقت مناسبی را در تقریب غلظت گردوغبار و عمق دید افقی فراهم کرده است.

واژه‌های کلیدی: گردوغبار، تشخیص دید افقی، سنجش از دور، یادگیری ماشین، الگوریتم TLBO، شبکه عصبی GMDH.

استناد: امیری، م.، امیری، ف.، پوراسد، م. ح.، سلیمانی، س.، بهینه‌سازی نتایج الگوریتم ML-Based GMDH به‌منظور افزایش دقت تشخیص گردوغبار و عمق دید افقی از طریق الگوریتم TLBO، نشریه سنجش از دور و GIS ایران، سال ۱۶، شماره ۱، بهار ۱۴۰۳: ۴۹-۶۶.



۱- مقدمه

رخداد گردوغبار در مناطق خشک و نیمه‌خشک جهان در کیفیت هوای منطقه، چرخه‌های آب و انرژی و اکوسیستم تأثیر می‌گذارد. به‌دلایلی همچون عمر کوتاه پدیده گردوغبار، چندانگانه‌بودن آن و واکنش‌های شدید با شرایط هواشناختی، آشکارسازی این پدیده به‌منزله یکی از پارامترهای تأثیرگذار در کیفیت هوای شهری، همواره با چالش روبه‌روست. با توجه به اهمیت این پدیده و مشکلات ناشی از گسترش وسیع مکانی آن در سال‌های اخیر و همچنین عواقب این پدیده در سلامت، بهداشت، کشاورزی، هوانوردی و دیگر موارد، تحقیق در زمینه رخداد گردوغبار بیش‌ازپیش ضروری به‌نظر می‌رسد (Zhao et al., 2010). نظارت بر کیفیت هوا معمولاً با استفاده از ایستگاه‌های هواشناسی پیشرفته و تجهیزات استقرار یافته در زمین انجام می‌شود و دامنه نظارتی این ایستگاه‌ها بسیار محدود است. این تجهیزات پارامترهای گوناگونی همچون دما، عمق دید، فشار هوا، بارش را در محل ایستگاه و متناسب با اهداف راه‌اندازی ایستگاه، اندازه‌گیری می‌کنند (RezaPour et al., 2010). مدل‌های عددی هواشناسی به‌تنهایی قادر به ردیابی و آشکارسازی طوفان‌های گردوغبار نیستند و اغلب خطا دارند. به‌علت مشاهدات اقلیمی و زیست‌محیطی محدود در این نواحی، لزوم استفاده از داده‌های سنجش از دور در کنار مشاهدات زمینی و مدل‌های عددی، بیش‌ازپیش احساس می‌شود. بی‌تردید بارزسازی رخدادهای، پایش سمت و سرعت حرکت آنها، تخمین غلظت ریزگردها و درنهایت، طبقه‌بندی آنها می‌تواند در کاهش صدمات ناشی از رخداد گردوغبار نقش مهمی ایفا کند. یکی از روش‌های بسیار مؤثر، در زمینه پدیده گردوغبار، استفاده از روش‌های سنجش از دور است. فتاوری سنجش از دور به‌دلیل وسعت منطقه مورد پایش و صرف وقت و هزینه کمتر، همواره یکی از کارآمدترین روش‌ها در مطالعه هم‌زمان رویدادهای گوناگون اقلیمی، جوی و فرایندهای هیدرولوژیکی مدنظر محققان بوده است (Taghavi et al., 2012).

آشکارسازی طوفان‌های گردوغبار با تکیه بر رفتار منحصربه‌فرد ذرات گردوغبار در طول موج‌های مرئی^۱ تا مادون قرمز حرارتی^۲، به‌ویژه با راه‌اندازی ماهواره‌های هواشناسی و محیطی، مورد توجه بسیار واقع شده است. ذرات گردوغبار قادرند پرتوهای خورشیدی و زمینی را در هر دو محدوده طیفی مرئی و مادون قرمز، تغییر دهند. به‌دلیل ویژگی‌های نوری ذرات گردوغبار، تابندگی مشاهداتی ماهواره‌ها می‌تواند مبین مشخصات طیفی این رخداد باشد که از مولکول‌های هوا، ابرها و سطح زیرین متمایز است (Zhao, 2012). در حال حاضر، استفاده از روش سنجش از دور یکی از مناسب‌ترین و دقیق‌ترین روش‌ها برای بررسی این پدیده اقلیمی است زیرا، علاوه بر توانایی در فراهم کردن متغیرهای اولیه برای شبیه‌سازی مدل، در ارزیابی صحت و دقت مدل‌ها نیز کاربرد دارد (Liu & Park, 2007). مشاهدات ماهواره‌ای محدودیت‌های ایستگاه‌های زمینی را برطرف می‌کنند و می‌توان آنها را به‌منزله مکمل ایستگاه‌های زمینی به‌کار برد. استفاده از تصاویر ماهواره‌ای، به‌دلیل سهولت و بی‌نیازی به مراجعه مستقیم به پدیده‌ها و امکان دورسنجی مطالب و موضوعات متفاوت طبیعی، اهمیت بسیار یافته است. در این میان، استخراج برخی خصوصیات و ویژگی‌ها از تصاویر بسیار کمک‌کننده است. با توجه به مطالب بیان‌شده، با استفاده از پتانسیل تصاویر ماهواره‌ای و به‌کارگیری روش‌های گوناگون پردازش تصاویر و هوش مصنوعی، می‌توان گام‌های مؤثری در زمینه نظارت بر کیفیت هوا برداشت. اطلاعات حاصل از روش‌های پردازش تصاویر ماهواره‌ای مزایایی بیشتر از اطلاعات ایستگاه‌های هواشناسی زمینی دارند. مزیت دیگر استفاده از تصاویر ماهواره‌ای این است که می‌تواند بر بسیاری از آلاینده‌ها نظارت هم‌زمان داشته باشد. همچنین دارای امکان نظارت ساده‌تر آنلاین است و ارزیابی سریع و پیوسته‌ای در

1. visible

2. Thermal inferared

استفاده تحلیل همبستگی، مهم‌ترین متغیرهای مستقل (ویژگی) را از میان متغیرهای در نظر گرفته شده شناسایی کردند. سپس با استفاده از یک مدل MLR، میزان غلظت PM2.5 را از روی متغیرهای مستقل تقریب کردند. پارک^۹ و همکاران (۲۰۱۹) نیز، برای دستیابی به تقریب مناسب از غلظت PM2.5، ترکیبی از ویژگی‌ها و اطلاعات گوناگون را به‌کار بردند. این پژوهشگران، علاوه بر AOD و داده‌های هواسنجی، از داده‌هایی درباره انتشار انواع گازها و داده‌های جمعیتی و صنعتی به‌منزله متغیرهای تعریف‌کننده استفاده کردند. سپس با استفاده از یک مدل RF، غلظت PM2.5 را تقریب زدند. چن^{۱۰} و همکاران (۲۰۱۹) میزان تقریبی غلظت PM2.5 در شهرهای امریکا را به‌دست آوردند. آنها برای دستیابی به دقت مناسب، در کنار تصاویر AOD ضبط‌شده از ماهواره، تصاویر AOD سطح زمین را به‌کار بردند. سپس با استفاده از مدل MLR، مقدار غلظت PM2.5 را از روی ویژگی‌های تهیه‌شده، به‌صورت تقریبی به‌دست آوردند. امیری^{۱۱} و همکاران (۲۰۲۰)، از طریق شبکه عصبی مصنوعی و روش تفاضلی، گردوغبار را براساس تصاویر ماهواره مادیس^{۱۲} تشخیص دادند و ویژگی‌های آن را استخراج کردند. رحمتی^{۱۳} و همکاران (۲۰۲۰) از ترکیب شبکه عصبی فازی تطبیقی (ANFIS) و سه الگوریتم متاهیوربستیک عنکبوت، فرهنگی و تفاضل تکاملی برای پیش‌بینی طوفان گردوغبار استفاده کردند. فن و همکاران (۲۰۲۰)، با استفاده از تصاویر هوایی بالاتر از سطح اتمسفر و به‌کارگیری شبکه عصبی عمیق، میزان غلظت PM2.5 را تقریب زدند. برخلاف بیشتر مطالعات که بر استفاده از AOD تمرکز داشته‌اند، نوآوری

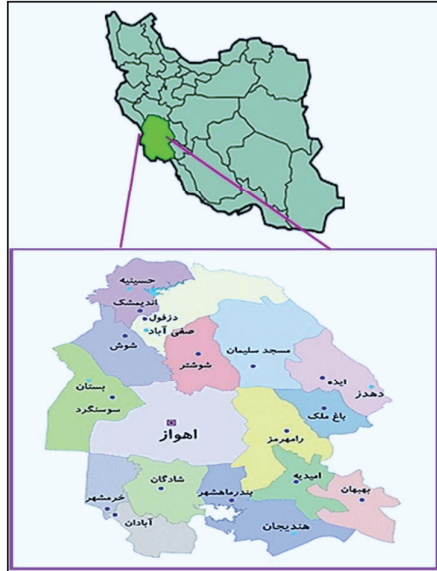
اختیار قرار می‌دهد. این موارد از جمله برتری‌های استفاده از تصاویر ماهواره‌ای بر ایستگاه‌های هواشناسی زمینی است (Guo et al., 2009). بنابراین اهمیت پدیده گردوغبار در مناطق یادشده و تأثیر آن در سلامت افراد، محیط‌زیست، کشاورزی، بارش و دیگر موارد و ازسوی دیگر، پتانسیل‌های بالای تصاویر ماهواره‌ای در آشکارسازی و پایش این پدیده لزوم انجام‌شدن این پژوهش را می‌رساند.

۲- مرور مطالعات گذشته

در مطالعات سالیان اخیر، محققان اندکی میزان گردوغبار را تخمین زده‌اند. از جمله، صمدی و درویشی بلورانی^۱ (۲۰۱۴) فرایندی خودکار را برای استخراج توده‌های گردوغبار، با استفاده از باندهای مادون‌قرمز حرارتی و انعکاسی، در منطقه غرب ایران، طی سال‌های ۲۰۰۰ تا ۲۰۱۱ توسعه دادند. جی و فن^۲ (۲۰۱۹) رویکردی بلندمدت را برای پیش‌بینی طوفان گردوغبار در شمال چین مطرح کردند. آنها میزان طوفان‌های گردوغبار در بهار را به میزان نمو نباتی^۳ در تابستان گذشته، میزان پوشش دریا با یخ در زمستان گذشته و نیز مقدار نوسان قطبی ارتباط دادند. سپس، با استفاده از مدل‌های آماری، پیش‌بینی انجام دادند. یون^۴ و همکاران (۲۰۱۹)، با استفاده از شبکه عصبی LSTM، میزان غلظت گردوغبار را پیش‌بینی کردند. لی و جانگ^۵ (۲۰۱۹) برای تقریب میزان غلظت PM2.5 در یکی از شهرستان‌های کشور چین از ترکیب تصاویر سنسور از دوری، اطلاعات هواسنجی و یادگیری ماشین استفاده کردند. آنها هم‌زمان چهار روش یادگیری ماشین (MLR)^۶، رگرسیون بردار پشتیبان (SVR)^۷، رگرسیون اسپلاین تطبیقی چندمتغیره (MARS) و RF را ارزیابی و با مقایسه دقت تقریب، روش مناسب را انتخاب کردند. وانگ^۸ و همکاران (۲۰۱۹)، برای دستیابی به دقت مناسب در تقریب غلظت PM2.5، ترکیب اطلاعات AOD، غلظت آلودگی‌های گازی (GASs) و داده‌های هواسنجی را بررسی کردند. آنها، با

1. Samadi & Darvishi Boloorani
2. Ji & Fan
3. Plant Growth
4. Yoon
5. Li & Zhang
6. Machine Learning Regression
7. Support Vector Regression
8. Wang
9. Park
10. Chen
11. Amiri
12. Modis
13. Rahmati

۱ نقشه منطقه مورد مطالعه را تصویر کرده است.



شکل ۹. نقشه منطقه مورد مطالعه

۴- داده‌ها و روش‌ها

در این بخش، ابتدا داده‌های مورد استفاده معرفی شده و سپس، درباره روش‌هایی که در قالب رویکرد کلی به‌کار رفته‌اند، توضیح مختصری بیان شده است.

۴-۱- داده‌ها

داده‌های مورد استفاده در این پژوهش شامل ۱۳ مشاهده مابین تاریخ‌های ۲۰۱۸/۱/۱ تا ۲۰۱۸/۱/۲۸، دریافتی از طریق سنجنده مادیس و مشاهدات دقیق ایستگاه‌های سینوپتیکی زمینی‌اند. سنجنده مادیس روی دو ماهواره ترا^۱ و آکوا^۲ قرار دارد. داده‌های این سنجنده را می‌توان برای کسب اطلاعات بسیاری، از جمله دما و رطوبت جوی، پوشش ابر و ویژگی‌های آن، ویژگی‌های هواویزها، دمای سطح خشکی و دریا،

پژوهش فن و همکاران (۲۰۲۰) در استفاده از تصاویر بالاتر از سطح اتمسفر است. این درحالی است که، در حوزه‌های مشابهی مانند پیش‌بینی غلظت PM_{2.5}، مطالعات گسترده‌ای انجام شده است (برای نمونه، Zamim et al., 2019; Wang et al., 2019; Shao et al., 2020; Li et al., 2020; Harba et al., 2020; Sahu et al., 2020; Feng et al., 2020; Fan et al., 2020). امیری و سلیمانی^۱ (۲۰۲۱) ساختار پایه شبکه عصبی (GMDH) را توسعه دادند. آنها این کار را با استفاده از مدل‌های رایج یادگیری ماشین، انجام دادند. این محققان (۲۰۲۲)، با روش ترکیبی گروهی^۲، تصاویر ماهواره‌ای را با هدف آشکارسازی گردوغبار، پردازش کردند. مدل آنها شامل ترکیب روش‌های انتخاب ویژگی با مدل‌های یادگیری ماشین می‌شود.

۳- منطقه مورد مطالعه

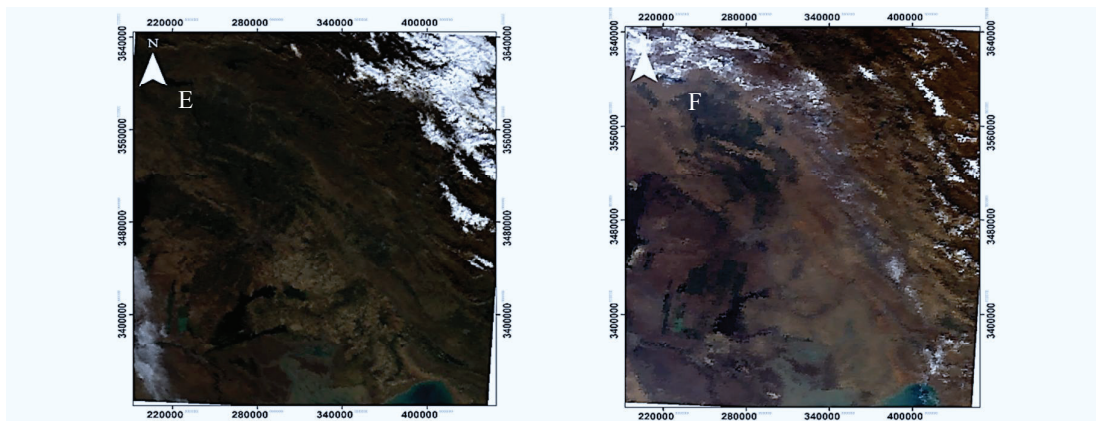
به‌منظور بررسی میزان گردو غبار موجود در هوا، استان خوزستان در کشور ایران به‌منزله منطقه مورد مطالعه این پژوهش انتخاب شد. خوزستان یکی از استان‌هایی است که به‌علت هم‌جواری با بیابان‌های عراق و عربستان، در بیشتر روزهای سال با مشکل ریزگردها روبه‌روست. میزان آسیب‌های ناشی از ریزگردها به‌حدی است که زیرساخت‌های خطوط انتقال برق و انرژی با مشکل جدی مواجه شده‌اند. استان خوزستان، با مساحتی حدود ۶۴۰۵۷ کیلومترمربع، بین طول جغرافیایی ۴۷ درجه و ۴۱ دقیقه تا ۵۰ درجه و ۳۹ دقیقه شرقی از نصف‌النهار گرینویچ و ۲۹ درجه و ۵۸ دقیقه تا ۳۳ درجه و ۴ دقیقه عرض شمالی از خط استوا، در جنوب‌غرب ایران قرار گرفته است. این استان از شمال با استان‌های ایلام، لرستان، چهارمحال و بختیاری، از شرق با کهگیلویه و بویراحمد، از جنوب با بوشهر و خلیج فارس و از غرب با کشور عراق هم‌مرز است. شمال و شرق استان خوزستان را سلسله‌جبال زاگرس محصور کرده و هرچه به‌سمت جنوب‌غرب پیش برویم، از ارتفاع آن کاهش می‌یابد؛ تا در نواحی جنوبی‌تر، به‌صورت تپه‌ماهورهایی نمایان می‌شود. شکل

1. Soleimani
2. Ensemble
3. Terra
4. Aqua

مادیس و داده‌های ایستگاه‌های سینوپیک، شامل میزان دید افقی، سرعت باد، جهت باد، رطوبت هوا، فشار سطح دریا، نقطه شبنم، دما، استفاده شده است. در تحقیق حاضر، تولید ویژگی برپایه روش تفاضلی گذاشته شد. مقدار تفاضل دو باند تصویری از یکدیگر نشان‌دهنده وضعیت دو باند نسبت به یکدیگر است. هر ویژگی تولیدشده، عددی با علامت مثبت یا منفی است که علامت آن رفتار صعودی یا نزولی بین دو باند مورد استفاده در تولید آن ویژگی را بیان می‌کند. مقدار ویژگی نیز معرف میزان تفاوت رفتار طیفی عارضه بین دو باند مورد استفاده در تولید شاخص است. به‌منظور ایجاد تمایز بین عوارض گوناگون، تمامی حالت‌های تفاضلی بین باندهای مزبور ایجاد شد

آتش‌سوزی‌های طبیعی و مصنوعی، توزیع و عمق یخ و برف، رنگ اقیانوس، شاخص‌های گیاهی، به‌کار برد. دو تصویر از تصویرهای رنگی مرئی مادیس، مورد استفاده در این پژوهش، از منطقه مورد مطالعه در شکل ۲ نشان داده شده که متعلق به دو روز پنجم و ششم ژانویه ۲۰۱۸ است.

داده‌های مادیس در طول موج ۰/۴ میکرومتر تا ۱۴/۵ میکرومتر دریافت می‌شوند. سنجنده مادیس در ۳۶ باند طیفی متفاوت تصویربرداری می‌کند. با توجه به اهداف پژوهش، از محصول سطح یک استفاده شده که تصاویر مختصات دارند. مطابق جدول ۱، در این پژوهش از هفت باند انعکاسی (باندهای ۱ تا ۷) و شش باند حرارتی (۲۳-۲۰ و ۳۲-۳۱)، متعلق به تصاویر



شکل ۲. تصویر ترکیب رنگی واقعی مادیس از منطقه مورد مطالعه در پنجم و ششم ژانویه ۲۰۱۸

جدول ۳. ویژگی‌های ارزیابی شده در روش‌های انتخاب ویژگی

نام شاخص	تعداد شاخص	نوع شاخص
باند انعکاسی (Reflective band)	۲۱	ماهواره
باند حرارتی (thermal bands)	۱۵	ماهواره
فشار سطح دریا (SLP)	۱	زمینی
نقطه شبنم (dew_point)	۱	زمینی
سرعت باد (wind_speed)	۱	زمینی
جهت باد (wind_direction)	۱	زمینی
رطوبت (humidity)	۱	زمینی
دما (temp)	۱	زمینی
مجموع	۴۲	ترکیبی

۴-۲-۲- روش پیشرو پی‌درپی^۲

روش پیشرو پی‌درپی (SFS) از شیوه‌های انتخاب ویژگی است؛ براساس آن از میان ویژگی‌های موجود، با بررسی تمامی آنها، ویژگی‌ای به‌منزله و ویژگی نخست انتخاب می‌شود که بیشترین دقت را در سیستم طبقه‌بندی دارد.

۴-۲-۳- روش آدابوست^۳

روش آدابوست با درختان تصمیم‌گیری کوچک به‌کار می‌رود؛ بدین‌صورت که درخت اول ایجاد می‌شود و از عملکرد آن روی هر نمونه آموزشی، برای سنجش میزان توجه درخت بعدی به نمونه‌ها استفاده می‌شود. در این الگوریتم می‌توان، با استفاده از مجموعه‌ای از طبقه‌بندی‌کننده‌های ضعیف، طبقه‌بندی قدرتمندی به‌دست آورد.

۴-۲-۴- جنگل تصادفی^۴

این روش طبقه‌بندی نوعی شیوه یادگیری گروهی برای طبقه‌بندی است.

۴-۲-۵- شبکه عصبی پرسپترون چندلایه^۵

شبکه‌های پرسپترون چندلایه از انواع متداول شبکه‌های عصبی‌اند.

۴-۲-۶- رگرسیون خطی چندگانه^۶

این روش به‌منظور مدل‌کردن رابطه خطی بین متغیرهای مستقل و وابسته، با استفاده از داده‌های مشاهده‌شده، به‌کار می‌رود.

۴-۲-۷- ماشین بردار پشتیبان

در این شیوه، با فرض جداپذیری خطی کلاس‌ها از هم، ابرصفحاتی به‌دست می‌آیند که قادرند کلاس‌ها را از یکدیگر جدا کنند.

B_x نشان‌دهنده باند و x شماره هر باند است). با استفاده از هفت باند انعکاسی، ۲۱ ویژگی به‌روش تفاضلی (تولید ویژگی به‌روش پیشنهادی) ایجاد شد. در محدوده حرارتی باندهای مادیس، پانزده ویژگی به‌شیوه تفاضلی، از باندهای ۲۰ تا ۲۳ و همچنین باند ۳۱ و باند ۳۲، به‌کار رفت. برای نمونه، در رابطه (۱)، B_1-B_4 معرف تفاضل بین باندهای اول و چهارم است که به‌منزله ویژگی شماره ۳، در فهرست کلی ویژگی‌ها، به‌منظور ارزیابی مدل استخراج ویژگی ثبت شده است.

$$B_1 - B_2, \dots, B_1 - B_7, B_2 - B_3, \dots, B_2 - B_7, B_3 - B_4, \dots, B_3 - B_7, B_4 - B_5, \dots, B_4 - B_7, B_5 - B_6, \dots,$$

تعداد ویژگی‌ها در روش پیشنهادی به‌صورت رابطه (۱) محاسبه می‌شود. F گویای تعداد ویژگی و B نشان‌دهنده تعداد باندهای مورد استفاده است. برای نمونه، تعداد ویژگی‌های ۷ باند انعکاسی که حاصل تفاضل تمامی باندها از یکدیگر است، بدین‌صورت محاسبه می‌شود:

$$F = \binom{B}{2} \Rightarrow F = \binom{7}{2} = 21 \quad (1)$$

در این پژوهش، براساس مطالعات پیشین، میزان غلظت گردوغبار در چهار سطح دسته‌بندی شده است؛ به‌نحوی که میزان دید افقی در بین این سطوح تغییری اساسی خواهد داشت. این چهار سطح عبارت‌اند از:

(۱) دید بیشتر از ۱۰۰۰۰ متر (کد ۱)؛

(۲) دید بین ۱۰۰۰۰ تا ۳۰۰۰ متر (کد ۲)؛

(۳) دید بین ۳۰۰۰ تا ۱۰۰۰ متر (کد ۳)؛

(۴) دید کمتر از ۱۰۰۰ متر (کد ۴).

۴-۲- روش‌ها

روش‌های یادگیری ماشین مورد استفاده در این پژوهش، برای انتخاب ویژگی و دسته‌بندی پیکسل‌های تصاویر، بدین‌قرار است که همگی اجرا و عملی شده‌اند.

۴-۲-۱- الگوریتم بهینه‌سازی ازدحام ذرات^۱

این الگوریتم رویکردی فراابتکاری و مبتنی‌بر جمعیت است. در ساخت آن، از رفتار پرندگان و ماهی‌ها که در یافتن غذا فعالیت گروهی دارند، الگوبرداری شده است.

1. Particle Swarm Optimization (PSO)
2. Sequential Forward Selection(SFS)
3. Adaboost
4. Random Forest (RF)
5. Multi Layer Perceptron (MLP)
6. Multi Linear Regression (MLR)

۴-۲-۸- ماشین یادگیری افراطی^۱

ماشین یادگیری افراطی یک قالب یکپارچه را با انواع انتقال ویژگی‌های وسیعی که می‌توان در لایه مخفی استفاده کرد، فراهم می‌کند و می‌توان آن را مستقیماً در دست‌بندی چنددسته‌ای و رگرسیون به‌کار برد.

چندجمله‌ای نسبتاً ساده، مدل پیچیده‌تری برای نگاشت از فضای ورودی به فضای خروجی پیکربندی می‌شود. در شکل ۳، GMDH با سه لایه و چهار ورودی نشان داده شده است.

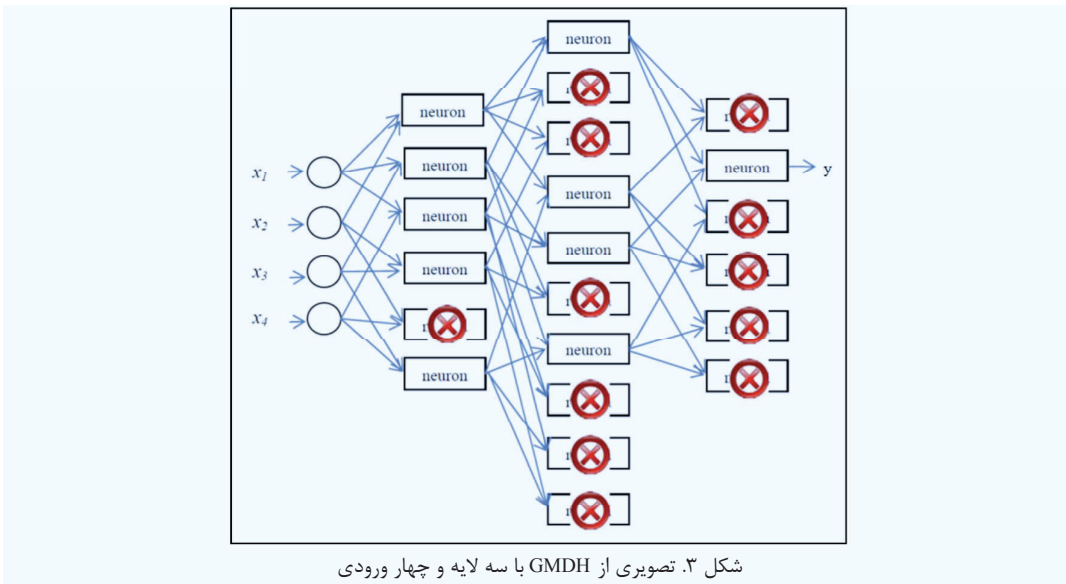
۴-۲-۹- شبکه عصبی مدل‌سازی گروهی داده‌ها^۲

در این شبکه، مجموعه‌ای از نورون‌ها^۳ حضور دارند که از به‌هم‌پیوستن چند جفت متفاوت از یک چندجمله‌ای درجه دوم به‌وجود می‌آیند. الگوریتم‌های GMDH براساس روش استدلالی مشخص می‌شوند؛ در این روش، مرتب‌سازی روی مدل‌های چندجمله‌ای به‌تدریج پیچیده‌شونده انجام می‌شود و با استفاده از معیار خارجی تعیین شده، بهترین راه‌حل انتخاب می‌شود. در ساختار توابع نگاشت مورد استفاده در هر یک از نورون‌های GMDH، توابع چندجمله‌ای (اغلب از درجه ۲) هستند که با روش کمترین مربعات^۴ برازش می‌شوند. درواقع، با ترکیبات چندلایه‌ای از نگاشت‌های ایجادشده از توابع

۱۰-۲-۴- روش ML-based GMDH تنظیم‌شده با

الگوریتم TLBO^۵

در ML-based GMDH، توابع بخشی چندجمله‌ای در مدل GMDH با مدل‌های یادگیری ماشین جایگزین می‌شوند. به‌عبارت دیگر، با حفظ ساختار اصلی GMDH، تابعی که برای نگاشت از مقادیر ورودی زوجی به‌کار می‌رود پیچیده‌تر از حالت پایه می‌شود و این امکان را ایجاد می‌کند تا نگاشت‌های پیچیده‌تر نیز مدل شوند. درموردی که الگوهای پیچیده‌ای میان مقادیر زوجی ورودی و خروجی وجود داشته باشد، استفاده از مدل‌های یادگیری ماشین این امکان را فراهم می‌کند که بلوک‌های سازنده GMDH حاوی تقریب‌های دقیق‌تری باشند. در ساختار پایه GMDH، با



1. Extreme Learning Machine (ELM)
2. Group Method of Data Handling (GMDH)
3. Neuron
4. Least Squares
5. Teaching-learning-based optimization

میانگین مربعات خطای برازش در میان نورون‌های هر لایه‌اند. α نیز اهرم انتخاب است و پارامتری تنظیمی از شبکه GMDH محسوب می‌شود. تعداد لایه‌های شبکه و بیشینه مجاز نورون در هر لایه (متغیری دیگر برای کنترل پیچیدگی مدل) نیز که میزان پیچیدگی GMDH را کنترل می‌کنند، دیگر پارامترهای تنظیمی آن به‌شمار می‌روند.

M تعداد مشاهدات و ML_{pq} مدل یادگیری ماشین آموزش دیده روی مشاهدات $\{(y_i, X_{ip}, X_{iq}); i \in Train\}$ است. همچون مدل GMDH پایه و مکانیسم انتخاب نورون‌های هر لایه، در ML-based GMDH نیز، براساس مقادیر MSE حاصل از ارزیابی مدل‌های یادگیری ماشین پس از آموزش، نورون‌های هر لایه انتخاب و حذف می‌شوند. در شکل ۴، ML-GMDH based با چهار متغیر ورودی، دو لایه میانی و توابع جزئی از مدل یادگیری ماشین (MLP)، با یک لایه میانی حاوی سه نورون، نمایش داده شده است.

مدل‌های گوناگون یادگیری ماشین را می‌توان به‌منزله توابع بخشی در مدل ML-based GMDH به‌کار برد. در این پژوهش، چهار مدل رایج MLP، SVR، RF و ELM به‌منزله گزینه‌های مدل یادگیری ماشین (ML) مورد استفاده، به‌منزله توابع جزئی در نظر گرفته شد. در ML-based GMDH، ابرپارامترها^۱ با این نام‌ها وجود خواهند داشت که در ادامه، از طریق الگوریتم بهینه‌سازی TLBO محاسبه و تنظیم می‌شوند.

- (۱) اهرم انتخاب (α) که آستانه انتخاب نورون‌ها را در هر لایه تعیین می‌کند؛
- (۲) تعداد لایه‌های شبکه (N-layer)؛
- (۳) بیشینه مجاز تعداد نورون در هر لایه (Max-Neurons)؛
- (۴) نوع مدل یادگیری ماشین به‌منزله توابع بخشی (از میان MLP، SVR، RF و ELM)؛
- (۵) درصدی از مشاهدات که برای آموزش استفاده می‌شوند (Ptrain).

استفاده یک دو جمله‌ای، مقادیر خروجی هر نورون حاصل می‌شود اما در ML-based GMDH، درون هر نورون یک بار آموزش مدل یادگیری ماشین (نوع مدل ML اختیاری است) با استفاده از بخش آموزش مشاهدات $\{(y_i, X_{ip}, X_{iq}); i \in Train\}$ که در آن X_{ip}, X_{iq} زوج ورودی‌های هر نورون و y تابع نگاشت شده است، انجام می‌شود و با این اقدام، فهرستی از وزن‌ها و پارامترهای ML شکل می‌گیرد. سپس، با ارزیابی خروجی‌های مدل ML آموزش دیده، میزان خطای تقریب برای مشاهدات اعتبارسنجی محاسبه می‌شود. هر تابع بخشی (نورون) از ML-based GMDH به‌صورت یک black box در نظر گرفته می‌شود و همانند ساختار GMDH پایه، خروجی نورون‌های هر لایه نیز به‌منزله ورودی‌های لایه بعدی انگاشته می‌شود. بدیهی است میزان میانگین مربعات خطا، در تقریب مشاهدات اعتبارسنجی برای هر نورون از ML-based GMDH نیز، با استفاده از رابطه (۲) محاسبه می‌شود.

$$MSE_{validation} = \frac{\sum_{i \in Validation} (y_i - ML_{pq}(X_{ip}, X_{iq}))^2}{M}$$

Validation = ((1 - P_{train}) * M)
number of observations selected randomly

رابطه (۲)

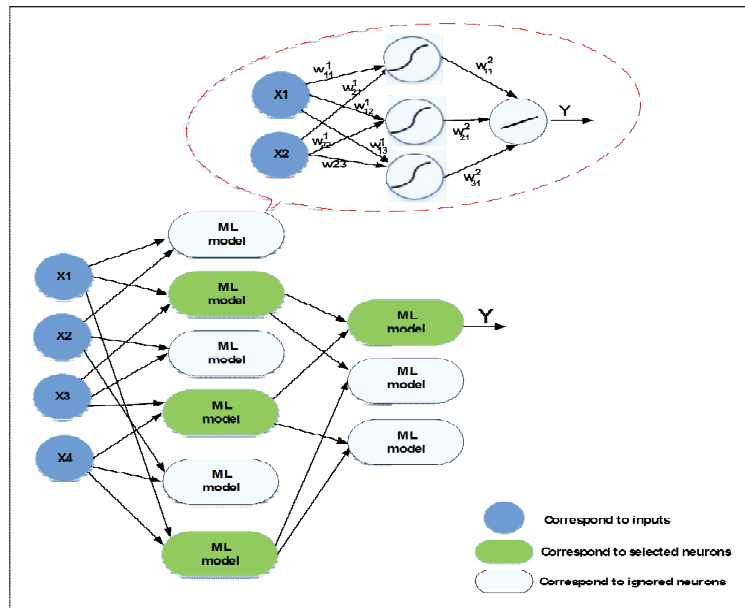
برای جلوگیری از افزایش محاسبات، با به‌کارگیری مکانیسم انتخاب طبیعی، برخی نورون‌های هر لایه از شبکه GMDH کنار گذاشته می‌شود؛ به‌صورتی که، با مقایسه مجموع مربعات خطای برازش هر نورون با یک مقدار آستانه، برخی نورون‌ها و خروجی آنها از ساختار شبکه حذف می‌شوند. مقدار آستانه طبق رابطه (۳) محاسبه می‌شود.

$$T_l = \alpha MSE_{validation}^{\min}_l + (1 - \alpha) MSE_{validation}^{\max}_l$$

رابطه (۳)

۱ نشان‌دهنده لایه أم شبکه، $MSE_{validation}^{\min}_l$ و $MSE_{validation}^{\max}_l$ به ترتیب، کمترین و بیشترین مقدار

1. Hyper Parameters



شکل ۴. نمونه‌ای از ML-based GMDH با چهار متغیر ورودی، دو لایه میانی و توابع جزئی MLP

۳-۴- روش اجرا

به‌منظور ارزیابی میزان گردوغبار موجود در هوا و تشخیص عمق دید افقی، در این پژوهش، مراحل اولیه که در ادامه نام برده می‌شود انجام شد.

۳-۴-۱- انتخاب ویژگی

در این بخش، با استفاده از روش‌های PSO، SFS و آدابوست، زیرمجموعه‌ای بهینه از ویژگی‌های استخراج‌شده از داده‌های مورد مطالعه انتخاب شد.

(جدول ۱) و نتایج آن در یک مدل SVM قرار گرفت تا دسته‌بندی مناسبی از ویژگی‌های انتخابی حاصل شود. به‌این ترتیب سه مدل با نام‌های SVM-PSO، SVM-SFS و SVM-Adaboost به‌دست آمد. برای ارزیابی دقت به‌دست‌آمده از روش‌ها، معیارهای رایج ارزیابی طبقه‌بندی به کار رفت. معیارهای ارزیابی مورد استفاده روش‌های حاضر در این بخش، برای انتخاب ویژگی، در جدول ۲ نشان داده شده‌اند.

جدول ۲. معیارهای ارزیابی به‌کاررفته در روش‌های انتخاب ویژگی

معیار	نحوه محاسبه
صحت	$Accuracy = (TP + TN) / (TP + TN + FP + FN)$
دقت	$Precision = TP / (TP + FP)$
حساسیت	$Sensitivity = TP / (TP + FN)$
تشخیص	$Specificity = TN / (TN + FP)$
F-score	$F-score = 2 \times (Precision \times Sensitivity) / (Precision + Sensitivity)$



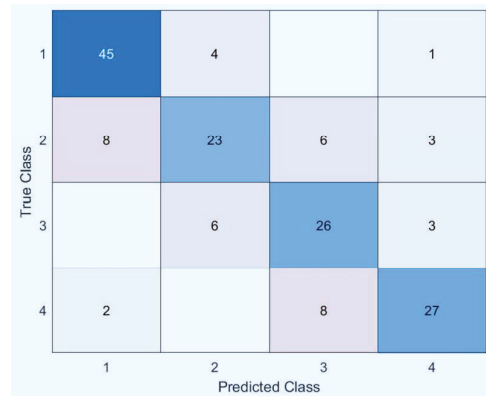
شکل ۶. نتایج به‌دست‌آمده از روش SVM-SFS

پس از آنکه معیارهای ارزیابی برای روش‌های انتخاب ویژگی گزینش شدند، ویژگی‌های جمع‌آوری شده وارد مدل‌ها شد تا، با تکیه بر معیارهای ارزیابی، بتوان بهترین روش انتخاب ویژگی و به‌صورت متناظر، بهترین زیرمجموعه ویژگی‌ها را انتخاب کرد. در همین راستا، روش‌ها یکی پس از دیگری مدل‌سازی و نتایج آنها در به‌دست‌آوردن معیارهای ارزیابی، با یکدیگر مقایسه شد. شکل ۵ ماتریس درهم‌ریختگی^۱ حاصل از دسته‌بندی به‌دست‌آمده پس از انتخاب ویژگی با روش SVM-PSO را نشان می‌دهد.

ویژگی‌های انتخاب‌شده با این روش نیز در جدول ۴ بیان شده‌اند.

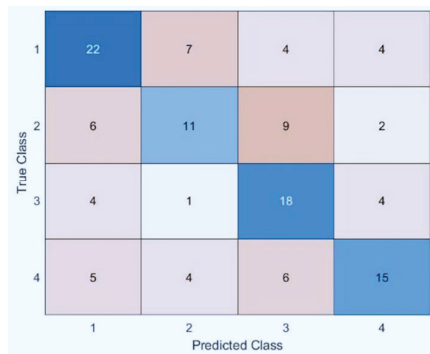
جدول ۴. ویژگی‌های انتخاب‌شده با روش SVM-SFS

نام روش	ویژگی‌های انتخاب‌شده
SVM-SFS	۹-۲



شکل ۵. نتایج به‌دست‌آمده از روش SVM-PSO

در نهایت، نتایج روش SVM-Adaboost در شکل ۷ نشان داده شده است.



شکل ۷. نتایج به‌دست‌آمده از روش SVM-Adaboost

ویژگی‌های منتخب با این روش در جدول ۳ بیان شده‌اند.

جدول ۳. ویژگی‌های انتخاب‌شده با روش SVM-PSO

نام روش	ویژگی‌های انتخاب‌شده
SVM-PSO	۱-۵-۷-۱۰-۱۱-۱۲-۱۳-۱۷-۱۸-۲۰-۲۱-۲۴
	۲۵-۲۶-۳۰-۳۱-۳۴-۳۶-۳۹-۴۱-۴۲

ویژگی‌های انتخاب‌شده براساس این روش در جدول ۵ آمده است.

جدول ۵. ویژگی‌های انتخاب‌شده با روش SVM-Adaboost

نام روش	ویژگی‌های انتخاب‌شده
SVM-Adaboost	۶-۹-۱۱-۱۸-۲۷-۲۸-۳۷-۳۸-۳۹-۴۱

در ادامه، شکل ۶ بیانگر نتایج به‌دست‌آمده از روش SVM-SFS است.

1. Confusion matrix

پس از آنکه مدل‌سازی‌ها انجام شد و کلاس‌های متناظر با ویژگی‌های انتخابی حاصل از هر سه روش به‌دست آمد، معیارهای ارزیابی آنها در قالب یک جدول، با یکدیگر مقایسه و ارزیابی شد. دلیل این کار معرفی بهترین روش دارای بهترین کلاس انتخاب ویژگی‌ها بود. مقایسهٔ مقادیر به‌دست‌آمده برای معیارهای ارزیابی حاصل از هر سه روش در جدول ۶ آمده است.

با توجه به نتایج ثبت‌شده در جدول ۶ و پس از محاسبهٔ مقدار ماکزیمم برای هر یک از معیارهای ارزیابی به‌دست‌آمده از هر سه روش، مشخص شد روش SVM-PSO دقت بیشتری از دیگر روش‌ها در ارائهٔ کلاس مناسب درمورد ویژگی‌های انتخابی دارد؛ به همین دلیل، به‌منزلهٔ بهترین روش برای انتخاب ویژگی‌های مناسب، به‌منظور استفاده در مراحل بعدی مدل‌سازی، معرفی شد.

۴-۳-۲- مدل‌سازی با استفاده از روش‌های یادگیری ماشین در مرحلهٔ قبل، ویژگی‌های مهم تأثیرگذار که در رسیدن به هدف اصلی این پژوهش نقش مهمی داشتند، پس از انتخاب اولیه وارد مدل‌های انتخاب

ویژگی شدند؛ سپس با استفاده از مدل SVM کلاس‌بندی مناسبی برای آنها در نظر گرفته شد تا بهترین روش حاصل شود. مطابق با نتایج نیز، روش SVM-PSO به‌منزلهٔ بهترین روش در این بخش انتخاب شد. در مرحلهٔ بعد، ویژگی‌های انتخاب‌شده از طریق SVM-PSO به‌منزلهٔ ورودی به مدل‌های منتخب طبقه‌بندی وارد شدند. بدین‌منظور، با بررسی مطالعات گذشته و تحلیل روش‌های پرکاربرد در زمینهٔ پیش‌بینی و خوشه‌بندی، روش‌های RF، GMDH، MLP، MLR و SVM به‌منزلهٔ روش‌های منتخب در شبیه‌سازی و مقایسه، انتخاب شده‌اند. بنابراین در پژوهش حاضر نیز، همین روش‌ها به‌منزلهٔ روش‌های اصلی تحلیل و ارزیابی و مقایسه با رویکرد اصلی که همان شبکهٔ عصبی GMDH توسعه‌یافته است، در طبقه‌بندی به‌کار رفت و ماتریس درهم‌ریختگی برای آنها تشکیل شد تا میزان دقت نتایج حاصل از آنها سنجیده شود. برای آنکه مشخص شود کدام روش بهترین نتیجه را دربرداشته است، نتایج معیارهای ارزیابی روش‌های یادشده با یکدیگر، در قالب جدول ۷، مقایسه شدند.

جدول ۶. گزینش بهترین روش انتخاب ویژگی

معیارهای ارزیابی					روش‌ها
F-score	Specificity	Sensitivity	Precision	Accuracy	
۰/۷۳۸۴	۰/۹۰۱۹	۰/۷۳۹۸	۰/۷۳۶۹	۰/۷۴۶۹	SVM-PSO
۰/۵۷۴۶	۰/۸۰۸۸	۰/۵۷۸۷	۰/۵۷۰۵	۰/۵۸۲۰	SVM-SFS
۰/۵۳۹۲	۰/۷۸۱۱	۰/۵۳۹۸	۰/۵۳۸۵	۰/۵۴۱۰	SVM-Adaboost

جدول ۷. انتخاب بهترین روش طبقه‌بندی

معیارهای ارزیابی					روش‌ها
F-score	Specificity	Sensitivity	Precision	Accuracy	
۰/۵۴۸۳	۰/۷۶۷۸	۰/۵۷۸۷	۰/۵۲۱۰	۰/۵۲۴۷	GMDH
۰/۷۵۱۴	۰/۸۹۸۷	۰/۷۵۱۴	۰/۷۵۱۳	۰/۷۴۶۹	MLP
۰/۵۲۱۴	۰/۷۴۰۲	۰/۵۵۶۱	۰/۴۹۰۸	۰/۴۷۵۳	MLp
۰/۹۳۲۹	۰/۹۷۶۴	۰/۹۳۲۴	۰/۹۳۳۵	۰/۹۳۲۱	RF
۰/۷۳۸۴	۰/۹۰۱۹	۰/۷۳۹۸	۰/۷۳۶۹	۰/۷۴۶۹	SVM

از همه روش‌ها به‌همراه بهترین مدل فاز قبل (مدل RF)، در جدول ۸ با یکدیگر مقایسه شدند. با توجه به مقادیر موجود، بیشترین مقدار در روش‌های مدل‌سازی به‌دست آمد و مشخص شد که نتایج را در مقیاس با سایر روش‌ها دارند. Ensemble SVM و Ensemble RF بهترین نتایج را در دو شاخص Precision و F-score بهتر عمل کرده و در مقابل، Ensemble RF در دو شاخص Sensitivity و Specificity بهتر بوده است. این نتیجه نشان می‌دهد، با ترکیب خروجی‌های مدل‌ها از طریق رویکرد گروهی، دقت مناسب‌تری در مقایسه با استفاده مستقل از هر یک از آنها حاصل می‌شود. به عبارت دیگر، می‌توان مشاهده کرد که با ترکیب توانایی‌های مدل‌های طبقه‌بندی رایج در قالب رویکرد گروهی‌کننده غیرخطی، می‌توان به دقت مناسب‌تری در تشخیص کلاس دید افقی دست یافت.

۳-۳-۴- مدل‌سازی با روش ML-Based GMDH

تنظیم‌شده با الگوریتم TLBO

پس از مشاهده نتایج رویکرد اول، در اقدامی مستقل و با استفاده روش توسعه‌یافته ML-Based GMDH از سوی محققان، غلظت گردوغبار تقریب زده شد. همان‌طور که بیان شد، برای دستیابی به بهترین نتیجه، پارامترهای مدل توسعه‌یافته با الگوریتم بهینه‌سازی TLBO به‌صورت بهینه تنظیم شدند. شکل ۸ نحوه همگرایی میزان دقت را در تنظیم پارامتر از طریق TLBO، نشان داده است.

مطابق نتایج ثبت‌شده در جدول ۷، روش جنگل تصادفی (RF) در تمامی معیارهای ارزیابی، پاسخی مطلوب‌تر از دیگر روش‌ها در برداشت و از این‌رو به‌منزله بهترین روش در این بخش انتخاب می‌شود. پس از این مرحله، خروجی حاصل از هر یک از روش‌های مدل‌سازی برای طبقه‌بندی ویژگی‌ها، با رویکرد گروهی، دوباره استفاده شد تا توانایی‌های هر یک از پنج مدل به‌کاررفته در قالب رویکردی ترکیب‌کننده تجمیع شود. برخلاف رویکردهای رایج و استفاده از روش‌های مبتنی بر رأی‌گیری^۱، در این پژوهش از هر یک از مدل‌های طبقه‌بندی گام پیشین، بار دیگر، در قامت مدلی گروهی‌کننده استفاده شده است. به این ترتیب مدل‌های Ensemble MLP، Ensemble GMDH، Ensemble SVM و Ensemble MLR، Ensemble RF ارزیابی شدند که ورودی آنها کلاس‌های تقریب‌شده از طریق پنج مدل GMDH، RF، MLP، MLR و SVM بود. نتایج به‌دست‌آمده از این مدل‌ها نیز در جدول ۸ مطرح شده است. با توجه به معیارهای ارزیابی، می‌توان نتیجه گرفت روش Ensemble RF، در پیش‌بینی کلاس‌بندی داده‌ها، دقت بسیار مناسبی دارد. تمامی معیارهای ارزیابی دارای مقادیری حدود ۰/۹۵ هستند (جز Specificity که مقدار ۰/۹۸ دارد)؛ این نکته حاکی از دقت بسیار بالای این روش است. دقت این روش به روش Ensemble SVM نزدیک است. پس از آنکه تمامی مدل‌ها، در این بخش، ارزیابی و سنجش شدند و معیارهای ارزیابی برای آنها به‌دست آمد، برای مشخص نمودن بهترین روش، معیارهای ارزیابی حاصل

جدول ۸. مقایسه معیارهای ارزیابی به‌دست‌آمده از روش‌های Ensemble

معیارهای ارزیابی				روش‌ها	
F-score	Specificity	Sensitivity	Precision	Accuracy	
۰/۸۰۶۷	۰/۹۲۲۱	۰/۸۱۸۳	۰/۷۹۵۴	۰/۷۹۰۱	Ensemble GMDH
۰/۹۳۸۹	۰/۹۷۹۵	۰/۹۴۳۸	۰/۹۳۳۹	۰/۹۳۸۳	Ensemble MLP
۰/۸۶۸۱	۰/۹۵۱۲	۰/۸۷۰۳	۰/۸۶۵۹	۰/۸۶۴۲	Ensemble MLR
۰/۹۴۷۰	۰/۹۸۱۴	۰/۹۵۰۸	۰/۹۴۳۲	۰/۹۴۴۴	Ensemble RF
۰/۹۴۷۱	۰/۹۸۰۶	۰/۹۴۵۹	۰/۹۴۸۲	۰/۹۴۴۴	Ensemble SVM
۰/۹۳۲۹	۰/۹۷۶۴	۰/۹۳۲۴	۰/۹۳۳۵	۰/۹۳۲۱	RF

1. Voting

یادگیری ماشین و یادگیری عمیق، تنظیم مناسب پارامترهای تنظیمی، در دستیابی به پاسخ مطلوب اهمیت بسیاری دارد. برای ایجاد توانایی در مدل به‌منظور تعیین خودکار پارامترهای تنظیمی، برای تنظیم پارامتر مدل GMDH توسعه‌یافته، از الگوریتم بهینه‌سازی مبتنی بر یادگیری و آموزش (TLBO) استفاده شد؛ به‌گونه‌ای که با استفاده از الگوریتم TLBO نوع مدل‌های جزئی، بیشینه تعداد نورون‌های هر لایه، درصد داده‌های آموزش و اهرم انتخاب α تعیین می‌شوند. در شکل ۱۰، شیوه اعمال الگوریتم TLBO برای تنظیم پارامترهای شبکه ML-Based GMDH توسعه‌یافته نمایش داده شده است.

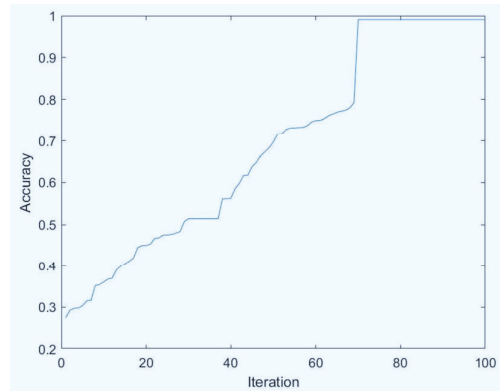
مقادیر به‌دست‌آمده در قطر اصلی ماتریس درهم‌ریختگی شکل ۱۰ نیز گواه همین نکته است. در ادامه، مقادیر به‌دست‌آمده برای معیارهای ارزیابی حاصل از این روش نیز در جدول ۹ نشان داده شده است.

جدول ۹. مقادیر به‌دست‌آمده برای معیارهای ارزیابی حاصل از

روش ML-Based GMDH

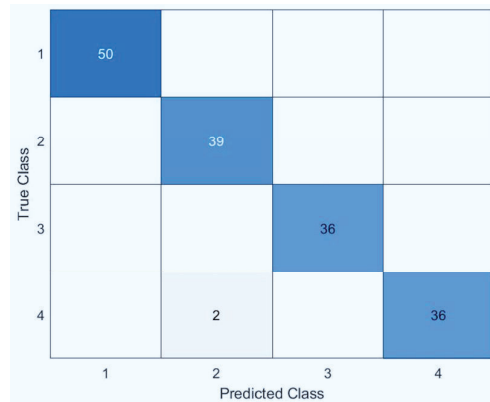
مقدار به‌دست‌آمده	معیار
۰/۹۸۷۷	Accuracy
۰/۹۸۶۸	Precision
۰/۹۸۷۸	Sensitivity
۰/۹۹۶۰	Specificity
۰/۹۸۷۳	F-score

با توجه به معیارهای ارزیابی به‌دست‌آمده درمورد روش ML-Based GMDH، می‌توان نتیجه گرفت که این روش دقت بسیار مناسبی در پیش‌بینی کلاس‌بندی داده‌ها دارد. تمامی معیارهای ارزیابی مقادیری حدود ۰/۹۹ دارند که این حاکی از دقت بسیار بالای این روش است. توجه در اعداد به‌دست‌آمده برای معیارهای ارزیابی، این نتیجه را دربردارد که دقت این روش نزدیک از تمامی روش‌های یادشده بهتر است. شکل ۱۱ تصویر طبقه‌بندی‌شده دو تصویر متعلق به پنجم و ششم ژانویه ۲۰۱۸ را نشان می‌دهد.



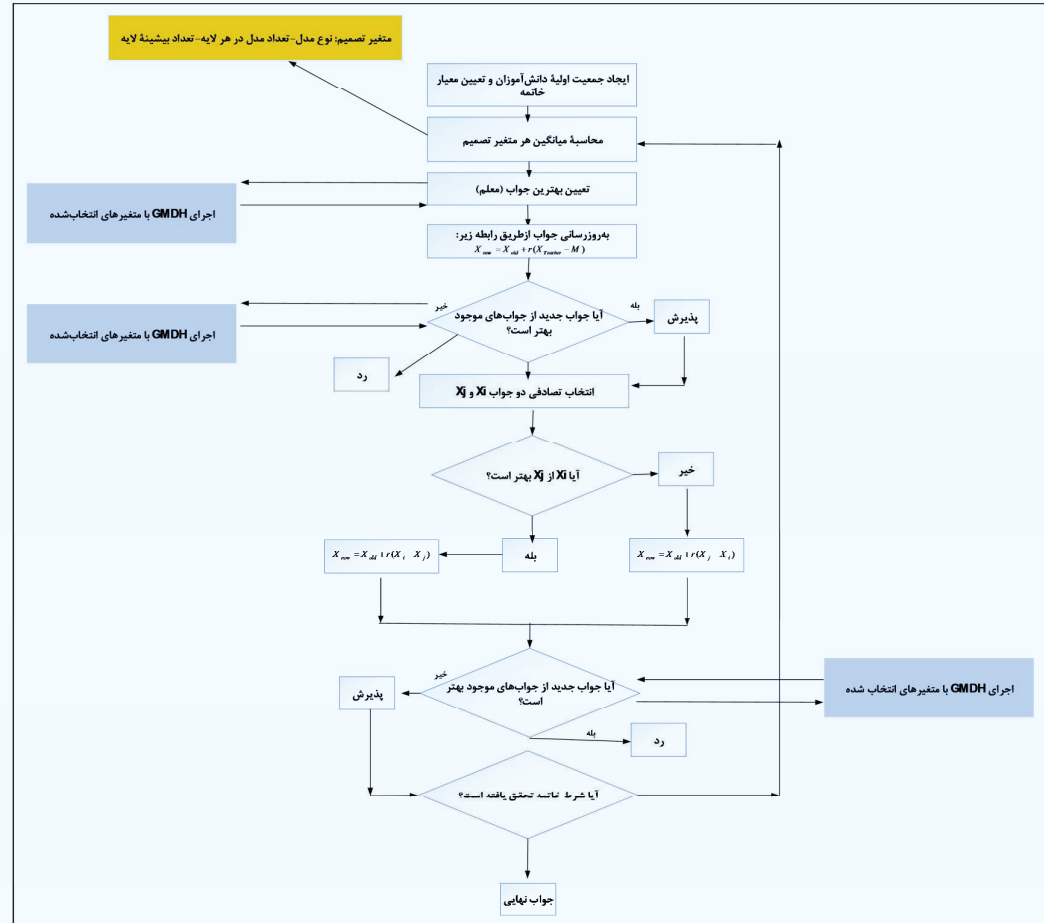
شکل ۸. نمودار همگرایی میزان دقت در تنظیم پارامترهای ML-Based GMDH با استفاده از TLBO

پس از تنظیم پارامترها، مدل ML-Based GMDH برای تقریب غلظت گردوغبار به کار رفت. شکل ۹ نتایج به‌دست‌آمده برای ماتریس درهم‌ریختگی حاصل از روش مورد استفاده را نشان می‌دهد. با توجه به شکل ۹، مشاهده می‌شود پیش‌بینی با روش ML-Based GMDH از دقت بسیار مطلوبی برخوردار بوده است.

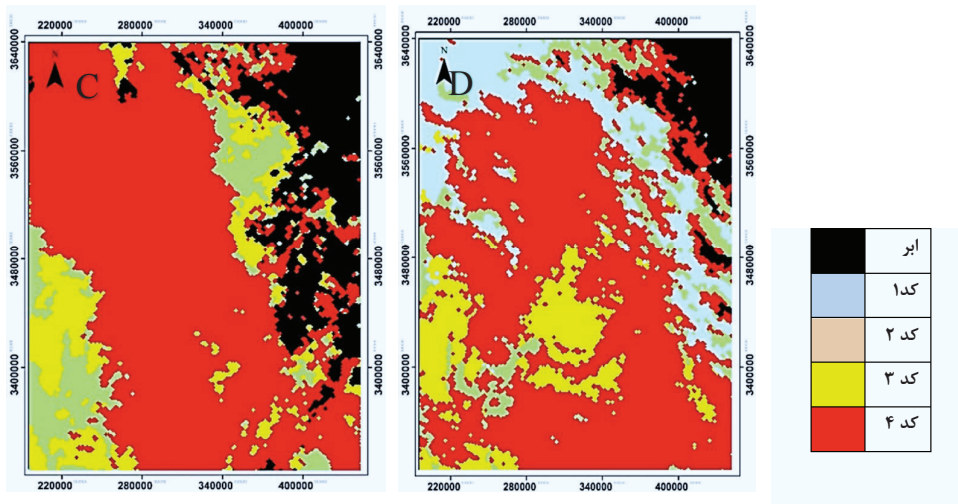


شکل ۹. نتایج حاصل از روش ML-Based GMDH

مدل ML-Based GMDH، با توجه به اینکه از نورون‌هایی از جنس مدل‌های یادگیری ماشین تشکیل می‌شود، توانایی بهتری در شناسایی الگوهای پنهان خواهد داشت. بنابراین مدل ML-Based GMDH به‌صورت مستقیم روی ویژگی‌های استخراج‌شده از تصاویر آموزش داده شد. همانند تمامی مدل‌های حوزه



شکل ۱۰. تنظیم پارامترهای شبکه ML-Based GMDH توسعه‌یافته با الگوریتم TLBO



شکل ۱۱. دید افقی تخمین‌زده‌شده در پنجم و ششم ژانویه ۲۰۱۸

- Neural Network**, Iranian Journal of Remote Sensing and GIS, 12(1), PP. 37-54
- Chen, Z., Chen, D., Xie, X., Cai, J., Zhuang, Y., Cheng, N., He, B. & Gao, B., 2019, **Spatial Self-Aggregation Effects and National Division of City-Level PM2.5 Concentrations in China Based on Spatio-Temporal Clustering**, Journal of Cleaner Production, 207, PP. 875-881.
- Fan, W., Qin, K., Cui, Y., Li, D. & Bilal, M., 2020, **Estimation of Hourly Ground-Level PM2.5 Concentration Based on Himawari-8 Apparent Reflectance**, IEEE Transactions on Geoscience and Remote Sensing, 59(1), PP. 76-85.
- Feng, L., Li, Y., Wang, Y. & Du, Q., 2020, **Estimating Hourly and Continuous Ground-Level PM2.5 Concentrations Using an Ensemble Learning Algorithm: The ST-Stacking Model**, Atmospheric Environment, 223, P. 117242.
- Guo, J.-P., Zhang, X.-Y., CHE, H., Gong, S.-L., An, X., Cao, Ch., Guang, J., Zhang, H., Wang, Y., Zhang, X.-CH., Xue, M. & Li, X.-W., 2009, **Correlation between PM Concentrations and Aerosol Optical Depth in Eastern China**, Atmospheric Environment, 2009, 43(37), PP. 5876-5886.
- Harba, H.S., Harba, E. & Farttoos, M., 2020, **Prediction of Dust Storm Direction from Satellite Images by Utilized Deep Learning Neural Network**, 2020 6th International Engineering Conference "Sustainable Technology and Development"(IEC), IEEE, Erbil, Iraq (23 June 2020), DOI.org/10.1038/s41612-023-00348-9.
- Ji, L. & Fan, K., 2019, **Climate Prediction of Dust Weather Frequency over Northern China Based on Sea-Ice Cover and Vegetation Variability**, Climate Dynamics, 53(1-2), PP. 687-705.
- Li, X. & Zhang, X., 2019, **Predicting Ground-Level PM2.5 Concentrations in the Beijing-Tianjin-Hebei Region: A Hybrid Remote Sensing and Machine Learning Approach**, Environmental Pollution, 249, PP. 735-749.

۵- نتیجه‌گیری

در این پژوهش، نتایج حاصل از روش‌ها به‌طور کامل نشان داده شده است و تلاش شد، در زمینه مقادیر به‌دست‌آمده از هر روش برای معیارهای ارزیابی انتخاب‌شده، بحث و بررسی انجام شود. همچنین روش‌های برتر در مراحل گوناگون تعیین شدند؛ به‌این‌ترتیب که روش SVM-PSO به‌منزله روش برتر در مرحله انتخاب ویژگی، روش RF به‌منزله روش برتر در میان روش‌های پایه دسته‌بندی و روش‌های Ensemble RF و Ensemble SVM به‌منزله روش‌های برتر در مرحله گروهی و دسته‌بندی انتخاب شدند. به‌علاوه، مشاهده شد که با استفاده از رویکرد گروهی، بهبود مناسبی در تشخیص دسته دید افقی حاصل شد. در رویکرد دوم، روشی با عنوان ML-Based GMDH استفاده شد که حاصل بهبود شبکه عصبی GMDH از طریق تغییر توابع جزئی با مدل‌های یادگیری ماشین است و برای تقریب غلظت گردوغبار به کار می‌رود. همچنین، برای دستیابی به دقت مناسب، ابرپارامترهای این مدل با استفاده از الگوریتم بهینه‌سازی TLBO با دقت بسیار بالا تنظیم شدند. نتایج نشان دادند که این روش با ایجاد بهبود، درقیاس با بهترین روش‌های انتخابی از رویکرد اول، دقت مطلوب‌تری را در تقریب غلظت گردوغبار و عمق دید افقی فراهم کرده است.

۶- منابع

- Amiri, M. & Soleimani, S.A., 2021, **ML-Based Group Method of Data Handling: An Improvement on the Conventional GMDH**, Complex Intell. Syst., 7, PP. 2949-2960.
- Amiri, M. & Soleimani, S.A., 2022, **Hybrid Atmospheric Satellite Image-Processing Method for Dust and Horizontal Visibility Detection through Feature Extraction and Machine Learning Techniques**, J. Indian Soc. Remote Sens., 50, PP. 523-532.
- Amiri, M., Soleimani, S.A. & Soltani Tafreshi, F., 2020, **Dust and Sand Extraction from MODIS Satellite Imagery Using Artificial**

- Spatiotemporal Kriging**, Science of The Total Environment, 740, P. 139761.
- Taghavi, F., Olad, A. & Safarrad, T., 2012, **Highlighting the Dust Storms of Western Iran Using MODIS Spectral Features**, Proceedings of the 15th Iran Geophysics Conference, Atmospheric Sciences Section, Ordibehesht, P. 109.
- Wang, X. & Sun, W., 2019, **Meteorological Parameters and Gaseous Pollutant Concentrations as Predictors of Daily Continuous PM2.5 Concentrations Using Deep Neural Network in Beijing–Tianjin–Hebei, China**, Atmospheric Environment, 211, PP. 128-137.
- Wang, W., Zhao, S., Jiao, L., Taylor, M., Zhang, B., Xu, G. & Hou, H., 2019, **Estimation of PM2.5 Concentrations in China Using a Spatial Back Propagation Neural Network**, Scientific Reports, 9(1), PP. 1-10.
- Yoon, J.H., Li, Y., Lee, M.S. & Jo, M., 2019, **Deep Learning Drone Flying Height Prediction for Efficient Fine Dust Concentration Measurement**, Proceedings of the 13th International Conference on Ubiquitous Information Management and Communication (IMCOM) 2019 (pp.1112-1119)a, International Conference on Ubiquitous Information Management and Communication, Springer.
- Zamim, S.K., Faraj, N.S., Aidan, I.A., Al-Zwainy, F.M., AbdulQader, M.A. & Mohammed, I.A., 2019, **Prediction of Dust Storms in Construction Projects Using Intelligent Artificial Neural Network Technology**, Periodicals of Engineering and Natural Sciences, 7(4), PP. 1659-1666.
- Zhao, T.X.P., 2012, **Asian Dust Detection from the Satellite Observations of Moderate Resolution Imaging Spectroradiometer (MODIS)**, Aerosol and Air Quality Research, 12(6), PP. 1073-1080.
- Zhao, T.X.P., Ackerman, S. & Guo, W., 2010, **Dust and Smoke Detection for Multi Channel Imagers**, International Journal of remote sensing, 2(10), PP. 2347-2367.
- Li, S., Zou, B., Fang, X. & Lin, Y., 2020, **Time Series Modeling of PM2.5 Concentrations with Residual Variance Constraint in Eastern Mainland China during 2013-2017**, Science of the Total Environment, 710, P. 135755.
- Liu, G. & Park, S.U., 2007, **The Logarithm-Linear Relationship of the Occurrence Frequency to the Duration of Sand-Dust Storms: Evidencee from Observational Data in China**, J. Arid. Environments, 71(2), PP. 243-249.
- Park, S., Shin, M., Im, J., Song, C.-K., Choi, M., Kim, J., Lee, S., Park, R., Kim, J. & Lee, D.-W., 2019, **Estimation of Ground-Level Particulate Matter Concentrations through the Synergistic Use of Satellite Observations and Process-Based Models over South Korea**, Atmospheric Chemistry and Physics, 19(2), PP. 1097-1113.
- Rahmati, O., Panahi, M., Ghiasi, S.S., Deo, R.C., Tiefenbacher, J.P., Pradhan, B., Jahani, A., Goshtasb, H., Kornejady, A. & Shahabi, H., 2020, **Hybridized Neural Fuzzy Ensembles for Dust Source Modeling and Prediction**, Atmospheric Environment, 224, P. 117320.
- Rezapour, K., Tavosi, T. & Khsrovi, M., 2010, **A Study of the Causes of Formation of Arabian Dust Storms and Its Expansion over Iran**, The Fourth International Congress of Geographers of the Islamic World, Zahedan.
- Sahu, S.K., Sharma, S., Zhang, H., Chejarla, V., Guo, H., Hu, J., Ying, Q., Xing, J. & Kota, S.H., 2020, **Estimating Ground Level PM2.5 Concentrations and Associated Health Risk in India Using Satellite Based AOD and WRF Predicted Meteorological Parameters**, Chemosphere, 255, P. 126969.
- Samadi, M. & Darvishi Bolorani, A., 2014, **Global Dust Detection Index (GDDI); A New Remotely Sensed Methodology for Dust Storms Detection**, Journal of Environmental Health Science & Engineering, 12, P. 20.
- Shao, Y., Ma, Z., Wang, J. & Bi, J., 2020, **Estimating Daily Ground-Level PM2.5 in China with Random-Forest-Based**