University of Tehran Press

Online ISSN: 2345-475X

DESERT

Home page: https://jdesert.ut.ac.ir/

Predicting Dust Sources Susceptibility Using Machine Learning Techniques in the Future (Case Study: South Khorasan, Iran)

Amir Ebrahimi[1*], Saeedreza Ahmadizadeh[1]

1. Department of Environment, Faculty of Natural Resources and Environment, University of Birjand, Birjand, Iran.
E-mail: am.ebrahimi@birjand.ac.ir

Article Info.

Article type:
Research Article

Article history:
Received: 16 Aug. 2025
Received in revised from: 13 Sep. 2025
Accepted: 17 Sep. 2025
Published online: 18 Sep. 2025

Keywords:
Asia,
Climate Change,
Dust Source Susceptibility,
Land Use Change,
Machine Learning,
South Khorasan.

ABSTRACT

Mineral dust significantly affects air quality, visibility, and Earth's radiation balance. Dust storms frequently occur in arid, semi-arid lands, flat regions with erodible soils, where drought and land-use changes have increased their occurrence, harming agriculture and communities. Central Eurasia, particularly the Middle East, is a major dust source region. This study employed machine learning to evaluate dust emission susceptibility in South Khorasan, Iran, by analyzing environmental factors and enhancing existing dust prediction models. Researchers used land use/land cover (2004 and 2019) maps, lithology, elevation, and climate variables from ACCESS-CM2 and CANESM5 models under IPCC6's SSP5-8.5 scenario to predict dust source susceptibility. Among SVM, CART, and Linear Regression algorithms, Random Forest performed best for LULC classification and wind speed prediction. The study combined CA-Markov for LULC prediction with Maximum Entropy modeling to calculate the Dust Source Susceptibility Index (DSSI). Results showed CANESM5 projected higher dust susceptibility than ACCESS-CM2, with over 10,340 km² falling into the highest-risk DSSI category. Wind plays a determining role in starting dust storms. The research demonstrates that integrating multiple modeling approaches and validation metrics (Kappa, AUC, R^2) provides an effective framework for investigating dust source susceptibility, offering improved predictive capability for dust storm management and mitigation strategies.

Cite this article: Ebrahimi, A., Ahmadizadeh, S.R. (2025). Predicting dust sources susceptibility using Machine Learning techniques in the future (case study: South Khorasan, Iran). DESERT, 30 (1), DOI: 10.22059/jdesert.2025.104762

DOI: 10.22059/jdesert.2025.104762

Publisher: University of Tehran Press

1. Introduction

Mineral dust represents the most extensive terrestrial source of particulate matter in the atmosphere. Once in the atmosphere, dust degrades the air quality and visibility, supplies a source of nutrients to remote oceans and ecosystems (Mahowald *et al*., 2005), provides a surface for the heterogeneous reaction of trace gases, and impacts the Earth's radiative budget (Bauer *et al*., 2004). The formal definition of dust storms results from turbulent winds raising large quantities of dust into the air and reducing visibility to less than 1000 meters (Jarraud, 2008). The main dust areas of the world are flat, physiographically dry areas with erosion-sensitive soils and sparse vegetation that is easily eroded by wind (Boroughani *et al*., 2020). In recent years, due to drought and Land use/Land cover Change, the frequency of days with dust storms has increased significantly, which causes adverse biological effects and extensive damage to agriculture, industry, and society. The combined development of this event along with the accelerated trend of development, industrialization, and population growth in urban areas has doubled the environmental tensions (Shaheen *et al*., 2020; Boloorani *et al*., 2020).

Central Eurasia included many dust hotspots and has a key role in particle dispersion. The Middle East is a transcontinental region, primarily situated in Western Asia but extending into northeastern Africa. The frequent dust storm events in most parts of the east and southeast of Iran have devastating effects on people's lives and resulted in heavy financial losses and casualties (Ahmadi *et al*. 2015; Karami *et al*., 2019). The major sources of dust emissions in southwest Asia are the Karakum Desert which located in in Central Asia, Dasht-e-Lut, the Margo and Registan deserts, the Sistan Basin, and the Hamun-e-Jaz Murian-a seasonal lake in the Jazmurian Basin (Kaskaoutis *et al*., 2018; Crosbie *et al*., 2014; Namdari *et al*., 2018). According to previous studies on the dust sources in the world, it can be seen that there are two critical sources in the top and bottom of the Khorasan region. This region is bounded by the desert of Kazakhstan in the north and the Sistan and Baluchestan region in the south (Baghi *et al*., 2020).

Ecological models of landscape change have potential to increase the precision of dust emission assessments across land cover types and the evaluation of accelerated and anthropogenic wind erosion (Webb and Pierre, 2018). Machine learning (ML) is a part of data science and a branch of artificial intelligence that can help solve a specific problem by examining the pattern of changes in big data and its prediction (L'heureux *et al*., 2017; Qiu *et al*., 2016). This method spans a broad set of models that are used to discover patterns in data and to make predictions, whereby the process of model 'training' is synonymous with a type of 'learning' (Witten and Frank, 2002). In recent years, the use of MLAs for predicting various variables has significantly grown, especially in The Earth science (Khaledian and Miller, 2020; Heuvelink *et al*., 2021; Mandal & Vipparthi, 2021).

Previously, some studies were conducted in relation to dust emission susceptibility, in which mathematical models were used in the spatial modeling of dust emission susceptibility (Boroughani *et al*., 2020; Gholami *et al*., 2020; Rahmati *et al*., 2020). In this research, through the use of machine learning for spatial and numerical data and other mathematical-based models, the dust emission susceptibility of South Khorasan in 2040 was calculated using environmental factors.

2. Materials and methods

2.1. Study Area

Central Eurasia (a) is a dust-prone region, with frequent storms occurring in Kazakhstan, Turkmenistan, and Iran (Nobakht *et al*., 2021; OĞUZ, 2020; Rashki *et al*., 2021). This study

focuses on South Khorasan, Iran (b), located on the northeastern margin of the Dasht-e-Lut desert, which connects Turkmenistan to the Sistan plain (Fig. 1). The region's diverse climate and landforms make it a critical area for dust research. Six long-term synoptic stations in South Khorasan provide essential climatic data, including visibility and dust event records. Governmental reports indicate that the local population is increasingly harmed by these dust events. Despite its significance, South Khorasan remains understudied, representing a missing piece in Central Eurasian dust research. Previous studies have identified dominant dust transport pathways. Ahmadi *et al*. (2015) found that winds carry dust from the eastern Caspian Sea and Turkmenistan deserts through eastern Iran's drylands to the southeast. Supporting this, HYSPLIT particle tracking from June-September 2017 confirmed that dust in northeastern Iran originates mainly from Turkmenistan, with a minor contribution from Uzbekistan (Baghi *et al*., 2020; Rashki *et al*., 2021).

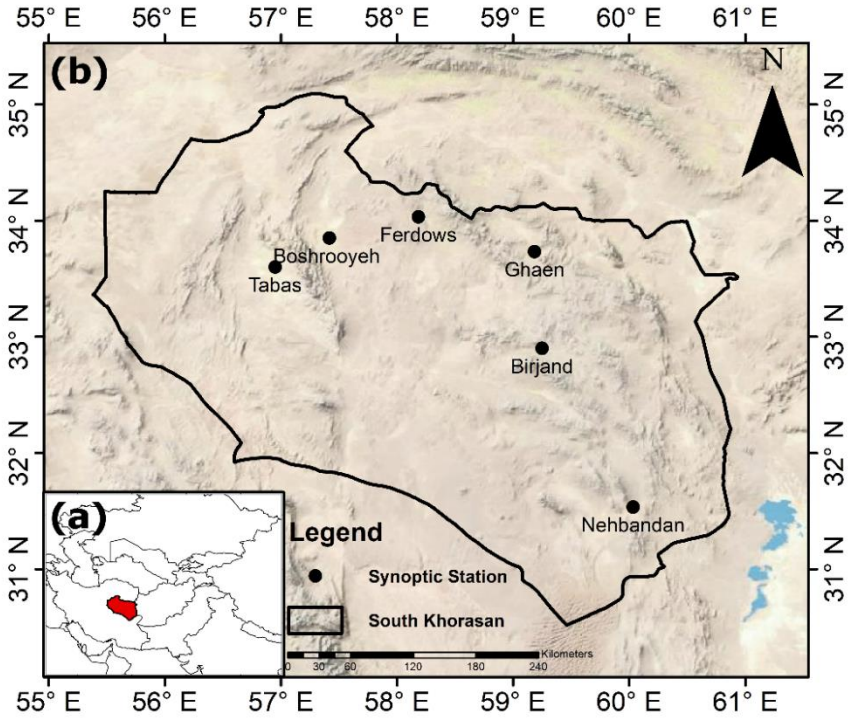

Fig 1. Study area map showing the location of the Synoptic Stations

2.2. Dataset

2.2.1. Climate Variables

2.2.1.1. Observation Data

Wind speed was obtained from South Khorasan Meteorological Organization. The dataset corresponds to daily observations from Birjand, Boshrooyeh, Ferdows, Ghaen, Nehbandan and Tabas synoptic stations.

2.2.1.2. Predicted Data

CMIP[1] Phase 6 (CMIP6) is currently in progress and a series of new versions GCMs have been released. The climate model outputs from CMIP6 provide a foundational baseline for the next

[1] Coupled Model Intercomparison Project

sixth assessment report of the Intergovernmental Panel on Climate Change (IPCC6) (Eyring *et al*., 2016). The CMIP6 model algorithm was designed to overcome the drawbacks that existed in CMIP5 like the overestimation of annual and seasonal precipitation and the physical algorithm has been improved (Rivera and Arnould, 2020). Changes to the GCM [1]algorithm may affect the precipitation simulation. To increase the confidence in future climate projections and the fidelity of the IPCC[2]6, the performance of GCMs from CMIP6 requires rigorous assessments (Tian and Dong, 2020). Despite such much evaluation works for CMIP5, the assessments of CMIP6 are still ongoing in various regions.

ACCESS-CM2[3], as used for CMIP6, comprises the following components: UM10.6 GA7.1 configuration for the atmosphere; the Community Atmosphere Biosphere Land Exchange (CABLE) model version 2.5 (coupled directly to the UM) for the land surface; CICE5.1.2 for the sea ice; MOM5 for the ocean; OASIS3-MCT for the numerical coupler; and the Rose/Cylc (Oliver *et al*. 2019) framework for experiment management such as model configuration and simulation control. Since all the submodels have been well documented by their individual developers, we provide brief descriptions, mainly documenting the major changes occurring in the scientific configurations of the model components since the implementation of our CMIP5 model ACCESS1.3 (Bi *et al*. 2020).

The Canadian Earth System Model version 5 (CanESM5) is the latest Canadian Centre for Climate Modelling and Analysis (CCCma) global climate model. CanESM5 has the capability to incorporate an interactive carbon cycle and was developed to simulate historical climate change and variability, to make centennial-scale projections of future climate, and to produce initialized climate predictions on seasonal to decadal timescales (Swart *et al*., 2019b, this issue, hereafter S19). S19 summarizes CanESM5 components and their coupling, together with the model's ability to reproduce large-scale features of the historical climate and its response to external forcing. This paper examines the predictive ability of CanESM5 on decadal timescales. CanESM5 decadal climate predictions are CCCma contribution to Component A of the Decadal Climate Prediction Project (Boer *et al*., 2016) endorsed by phase 6 of the Coupled Model Intercomparison Project (CMIP6).

In this study, Mean Temperature, Total Precipitation were received from WorldClim based on ACCESS-CM2 and CANESM5 from IPCC6 thorough SSP 5-8.5 scenario which is high emission scenario (Nazarenko *et al*., 2022).

2.2.2. LULC

Land Use and Land Cover (LULC) are two alterable terms that have different connotations in modern science (Lo, 1986). Monitoring of land use and land cover (LULC) change is fundamental aspect of the landscape dynamics or environmental health evaluation at different spatio-temporal scales (Chamling & Bera, 2020). LULC changes are fundamental processes on the earth's surface and have significant impacts on human society, climate, biodiversity, hydrological cycles, ecosystems, biogeochemical cycles and many other processes (Were *et al*., 2014; Lin *et al*., 2018).

The CA-Markov model is considered a robust approach because of the quantitative estimation and the spatial and temporal dynamic it has for modeling the LULC dynamic (Mishra *et al*., 2014). The CA model, together with the Markov transition probability, shows the ability to predict the trend for landscape change. This spatio-temporal model provides not only a

[1] General Circulation Models
[2] Intergovernmental Panel on Climate Change
[3] Australian Community Climate and Earth System Simulator-Coupled Model-v2

quantitative description of change in the past, but also the direction and amount of change in the future (Surabuddin Mondal *et al*., 2019). In this study, CA-Markov was used to predict land cover in 2040.

LULC were obtained from Landsat 8 images. First, 11 images of the study area were stitched together on the Google Earth Engine platform through mosaicking. Then, by visiting and examining control points, the land uses in the integrated image were classified using three methods (RF[1], SVM[2], CART[3]), for the years 2004 and 2019. Finally, using a Markov chain, land use in 2040 was predicted.

2.2.3. Lithology

High resolution global lithological map (GLiM) was assembled from existing regional geological maps translated into lithological information with the help of regional literature. In GLIM data, the area values represent the mapped surface coverage of each rock unit, which is classified based on its lithology (rock type) and origin, not its intrinsic quality or strength. Statistical analysis of the geological layer areas reveals that the region is predominantly covered by **superficial deposits**, which dominate with a substantial 55.2% of the total area, indicating the prevalence of young geology and unconsolidated sediments. This is followed by **soil complex** at 18.5% and **intrusive igneous rock** at 10.9%, whereas hydrogeologically significant layers such as pre-Quaternary aquifers constitute only 3.4% and aquitards account for less than 0.5%. This uneven distribution, coupled with the minimal presence of metamorphic rock and weathered layers - each less than 0.5% - underscores a geological framework dominated by sedimentary and igneous formations, with limited development of hard and ancient rock layers in the area (fig 2).

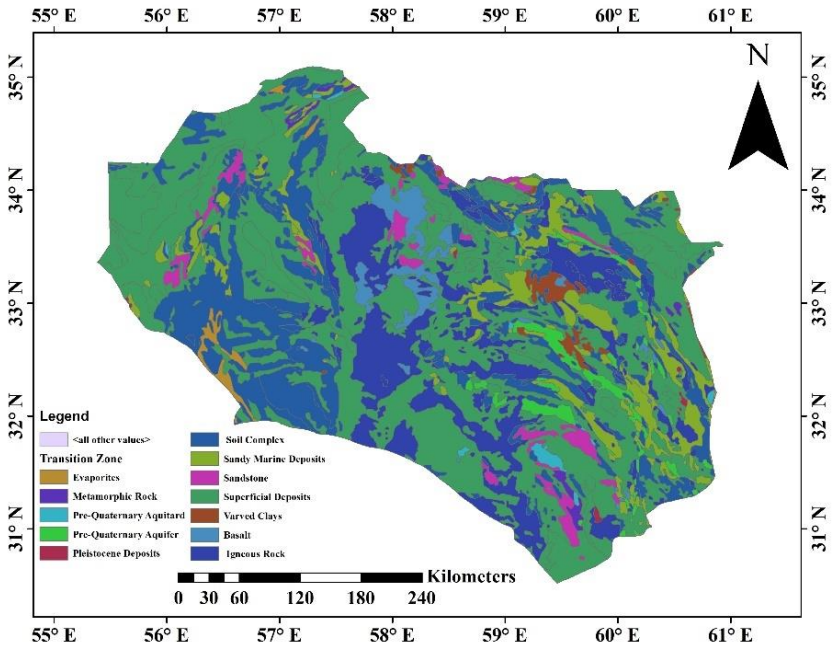

Fig 2. Lithology map of the study area

[1] Random Forest
[2] Support Vector Machine
[3] Classification and Regression Tree

2.3. Methodology

2.3.1. Interpolation

There are many ways to convert numerical climate data to climate maps. Interpolation is a method to find the value between the known data points (Cressie, 2015). Although choosing the best interpolation method is an entirely relative choice, considering the number of stations, Inverse distance weighting (IDW) was the most appropriate interpolation method for this study (Shepard, 1968). For this purpose, six Synoptic stations were used for Wind Speed interpolation.

2.3.2. Dust Source Susceptibility

Recent studies on dust source susceptibility have used a wide range of methodologies, evolving from traditional statistical approaches to advanced machine learning techniques. Key methods include multi-criteria decision analysis (MCDA) with the Analytical Hierarchy Process (AHP) as in Al-Hemoud *et al*. (2024) and Fuzzy-AHP hybrids in Gholami *et al*. (2020), as well as geostatistical models integrating remote sensing data, demonstrated by Darvishi Boloorani *et al*. (2022). Machine learning approaches have become more prominent, with studies employing Random Forest algorithms (Jafari *et al*., 2022), Logistic Model Trees (Jafari & Malekian, 2019), and Frequency Ratio models (Boroughani *et al*., 2020). Other methods include multivariate logistic regression (Middleton, 2017), Principal Component Analysis with clustering (Zoljoodi *et al*., 2021), and physical-based indices such as Dust Uplift Potential (Yu *et al*., 2018). These approaches typically integrate multiple environmental factors – including soil properties, vegetation indices, land use, wind patterns, and topographic features – validated against ground observations and satellite imagery to produce accurate susceptibility maps.

There are several methods to evaluate the susceptibility of lands. Elevation, Lithology, LULC, Mean Temperature, Total Precipitation and Wind Speed were used to measure the susceptibility to dust sources. Then, Nominal maps were divided into different classes based on the defined classes.

Maximum entropy (MaxEnt) is increasingly being considered in studying various earth system processes (Dyke and Kleidon, 2010). This Model compares the conditional density function of covariates (predictor variables) at presence sites to the marginal (background) density of covariates in the study area in order to derive the conditional occurrence probability (Elith *et al*., 2011). Maximum entropy models derive from information theory (as opposed to thermodynamic entropy models) and have shown promise in various earth science applications (Ruddell *et al*., 2013). Before this research, Maxent was used to determine different areas' suitability to create a dust source (Ansari *et al*., 2017; Lababpour, 2020). Finally, based on the prediction rate, Dust Source Susceptibility Index, or DSSI can be calculated (Pourhashemi *et al*., 2019).

$$DSSI = \sum (FR)_i \qquad (i = 1,2,3, \dots, n)$$

Which FR is Frequency Ratio.

The flowchart of methodology for predicting DSSI in 2040 is given in Fig 3.

2.3.3. Methods Validation

A wide range of techniques has been used to evaluate the performance and efficiency of ML algorithms; for example, Receiver Operating Characteristic Area Under Curve (ROC-AUC) (Pham *et al*., 2018; Felicísimo *et al*., 2013), root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), determination coefficient (R^2) (Khosravi *et al*., 2019; Fan *et al*., 2018). R^2, MSE, RMSE, and MAE can be expressed as:

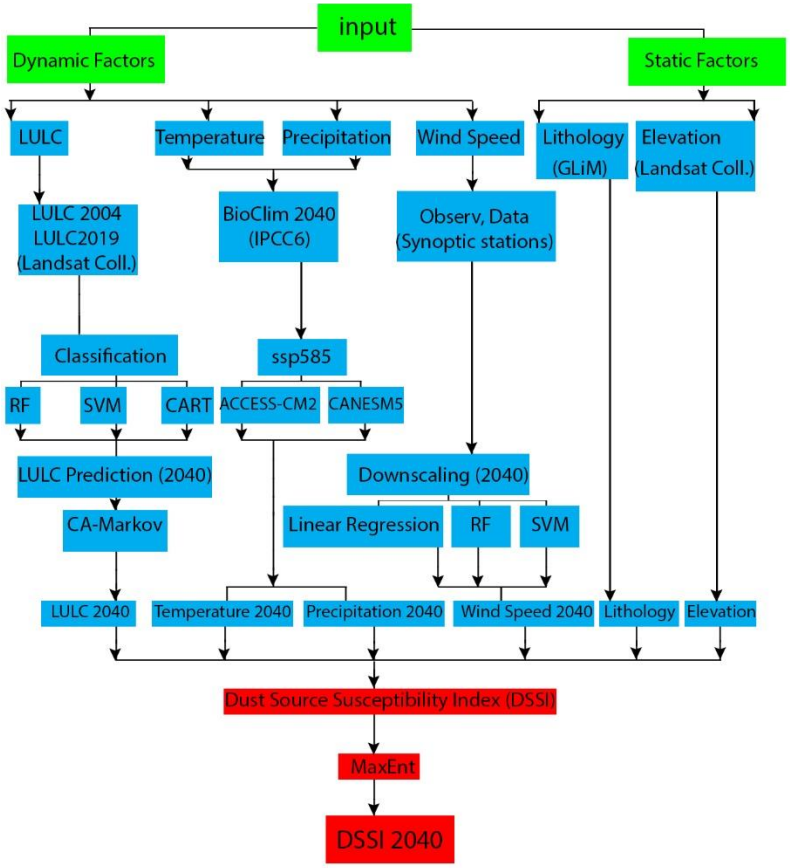

Fig 3. Descriptive flowchart outlining our approach to predicting dust susceptibility in 2040

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}}$$

$$MSE = \frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}$$

$$MAE = \frac{1}{\mathrm{n}} \times \sum_{i=1}^{n} |O_i - P_i|$$

$$r = \frac{n\left(\sum X_{obs,i}X_{model,i}\right) - \left(\sum X_{obs,i}\right)\left(\sum X_{model,i}\right)}{\sqrt{[n\sum(X_{obs,i})^2 - (\sum X_{obs,i})^2][n\sum(X_{model,i})^2 - (\sum X_{model,i})^2]}}$$

where n is the number of observations, and X_{obs}, X_{model}, and X_{AK} indicate the measured, estimated, and the mean of measured values, respectively.

Matrix by Kappa coefficient is a common and typical method (Congalton & Green, 2019). KAPPA analysis is a discrete multivariate technique used in accuracy assessments (Jensen *et al*., 1996). KAPPA analysis yields a Khat statistic (an estimate of KAPPA) that is a measure of agreement or accuracy (Congalton, 1991).

$$K = \frac{\mathrm{N}\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(\mathrm{x}_i + \mathrm{X}x_{+1})}{N^2 - \sum_{i=1}^{r}(\mathrm{x}_{ii}\mathrm{X}x_{+1})}$$

where;

r = number of rows and columns in error matrix, N = total number of observations (pixels)
X_{ii} = observation in row i and column i,
X_i+ = marginal total of row i, and X+i = marginal total of column i

The overall classification accuracy is the percentage of correctly classified samples of an error matrix. It is computed by dividing the total number of correctly classified samples by the total number of reference samples. It can be expressed by the following equation:

$$Overall\ Accuracy = \frac{1}{\mathrm{N}}\sum_{k-1}^{n} a_{kk}$$

Producer's accuracy (PA) informs the image analyst of the number of pixels correctly classified in a particular category as a percentage of the total number of pixels actually belonging to that category in the image. Producer's accuracy measures errors of omission (Anand, 2017).

The consumer's accuracy (CA) is computed using the number of correctly classified pixels to the total number of pixels assigned to a particular category. It takes errors of commission into account by telling the consumer that, for all areas identified as category X, a certain percentage are actually correct (Anand, 2017).

Final dust sources susceptibility maps were evaluated using the Area Under Curve (AUC). The area under the ROC curve, called AUC, indicates the amount of system prediction through its ability to accurately estimate its occurrence (presence of dust sources) and non-occurrence (absence of dust sources). It varies from 0.5 to 1 (Zhu and Wang, 2009). Finally, the range of values of dust sources susceptibility maps was reclassified between 0 and 1.

To evaluate LULC classification Kappa, Overall Accuracy, Producer's Accuracy and Consumer's Accuracy, and to evaluate wind speed, RMSE, MSE, MAE and R^2, and to evaluate DSSI, AUC were used.

3. Results and discussion

3.1. LULC Classification

The following is a land use map of the study area in 2004 and 2019 (Fig 4).

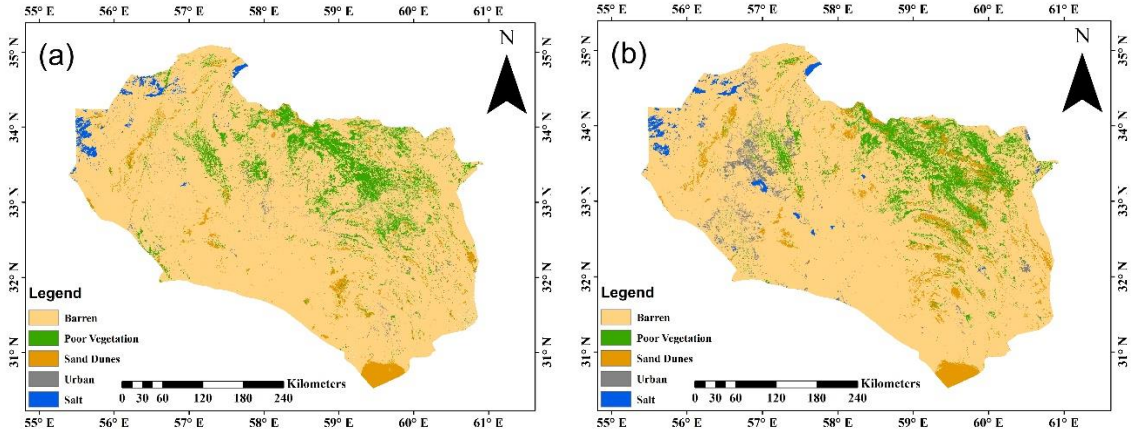

Fig 4. Classified map of 2004 (a) and 2019 (b)

Analysis of the LULC area changes from 2004 to 2019 reveals significant and concerning environmental trends. The most striking changes are the dramatic expansion of urban areas, which more than doubled from 1,570.74 km² to 3,542.73 km², and the substantial spread of sand dunes, which increased from 5,237.80 km² to 7,796.22 km², indicating severe

desertification. Concurrently, salt-affected lands grew from 1,524.65 km² to 1,753.64 km², signaling increasing soil salinity, while areas of poor vegetation saw a minor decline. These concurrent trends point to mounting pressure on the natural environment from both human development and ecological degradation (Table 1).

Table 1. Area of LULC classes in 2004 and 2019

	Area (km^2)	
Class	2004	2019
Barren	131741.48	127447.49
Poor Vegetation	14335.28	14189.29
Sand Dunes	5237.80	7796.22
Urban	1570.74	3542.73
Salt	1524.65	1753.64

Based on the results, RF is the best model for LULC classification in the study area. In addition to having the highest Kappa value, this model also has the best Validation Overall Accuracy. On the other hand, SVM is the weakest model for LULC classification, and its kappa value does not reach 0.5 (Table 2).

Table 2. Values of overall accuracy metrics for LULC classification

		Training Overall Accuracy	Validation Overall Accuracy	Kappa
LULC (2004)	SVM	0.685	0.655	0.477
	CART	1.000	0.782	0.687
	RF	0.956	0.832	0.757
LULC (2019)	SVM	0.662	0.663	0.465
	CART	1.000	0.696	0.789
	RF	0.962	0.832	0.755

Consumer and producer accuracy in 2004 show that the SVM model does not have the necessary ability to distinguish “urban” and “sand dunes” classes. These two classes are also weak in CART classification in 2004, but their value is high in 2019. Overall, the difference in value between CART and RF is small in 2019, but RF model performed slightly better. This difference is more in the classification of 2004 classification (Table 3).

After predicting the land cover in 2040, the results showed that 85% of the barren areas remain in their current form. On the other hand, 67% of areas with poor vegetation will become areas without vegetation in 2040, which shows the trend of desertification in the future (Fig 5).

It is also observed that the Sand Dunes have increased significantly. The area of salt flats has also increased. However, it is not as much as the increase in the area of Sand Dunes (Table 4).

The study's results, showing RF achieving the highest Kappa value and overall accuracy, are consistent with previous research where RF consistently ranks high in classification tasks involving remote sensing data. In a study by Krivoguz *et al*. (2023), various machine learning models were evaluated for LULC classification using Landsat-5 data. The RF model demonstrated high accuracy, similar to the current study's findings, where RF was favored over SVM and CART models. This reinforces the notion that RF is particularly effective in distinguishing between land cover types, especially in regions with complex landscapes or overlapping spectral signatures (Basheer *et al*., 2022).

Table 3. Values of consumer and producer accuracy in for LULC classification

Consumer's Accuracy						
Class	2004			2019		
	SVM	CART	RF	SVM	CART	RF
Barren	0.550	0.787	0.810	0.561	0.779	0.756
Poor vegetation	0.950	0.789	0.821	0.888	0.682	0.776
Sand dunes	0.000	0.200	0.833	0.000	0.857	0.750
Urban	0.000	0.684	0.800	0.925	0.779	0.906
Salt	0.764	0.809	0.891	0.750	0.861	0.861
Producer's Accuracy						
Class	2004			2019		
	SVM	CART	RF	SVM	CART	RF
Barren	0.894	0.824	0.900	0.908	0.805	0.885
Poor vegetation	0.483	0.875	0.883	0.210	0.763	0.776
Sand dunes	0.000	0.166	0.833	0.000	0.750	0.750
Urban	0.000	0.393	0.363	0.409	.754	0.639
Salt	0.702	0.765	0.810	0.700	.775	0.775

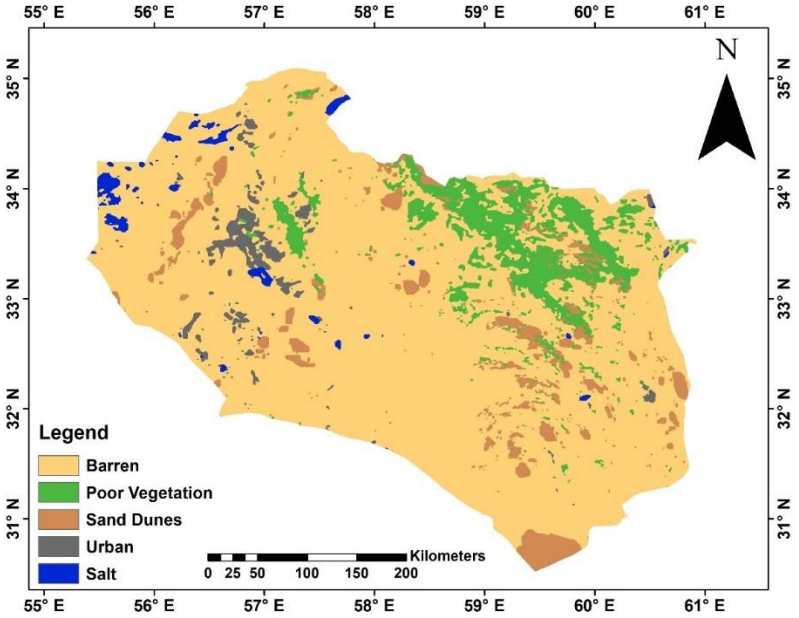

Fig 5. LULC predicted map using CA-Markov in 2040

Table 4. Areas of classified map in 2040

Class	Area (Km^2)
Barren	120872.93
Poor Vegetation	13384.62
Sand Dunes	9516.31
Urban	3411.62
Salt	1880.75

3.2. Climatic Variables Downscaling

Downscaling of wind speed in different stations shows that, like mean temperature and precipitation, for downscaling of wind speed, RF has the best performance compared to other regressors. The value of R2 gauge in Ghaen station is estimated to be higher than other stations. The value of RMSE in this station is also lower than other stations and considering the measurements of other regressors, it can be seen that the downscaling of wind speed in Ghaen station is done with higher quality than other stations. Also, LR has a good performance (Table 5).

Table 5. Calibration and Validation metrics of Wind Speed in all synoptic stations.

		LR		RF		SVM (rbf[1])	
		cal	val	cal	val	cal	val
Birjand	R^2	0.190	0.236	0.841	0.849	0.227	0.222
	RMSE	2.353	2.203	1.040	0.979	2.298	2.224
	MSE	5.539	4.854	1.082	0.959	5.282	4.947
	MAE	1.826	1.673	0.752	0.714	1.718	1.669
Boshrooyeh	R^2	0.182	-0.870	0.556	0.779	0.229	-0.794
	RMSE	1.186	4.276	0.873	1.467	1.150	4.186
	MSE	1.406	18.290	0.763	2.153	1.324	17.529
	MAE	0.865	3.376	0.637	1.095	0.828	3.290
Ferdows	R^2	0.208	0.037	0.843	0.808	0.237	0.094
	RMSE	1.795	1.710	0.797	0.761	1.762	1.658
	MSE	3.224	2.924	0.635	0.580	3.106	2.750
	MAE	1.425	1.311	0.590	0.556	1.394	1.261
Ghaen	R^2	0.699	0.216	0.924	0.894	0.750	0.360
	RMSE	1.094	2.228	0.550	0.815	0.997	2.012
	MSE	1.197	4.964	0.302	0.665	0.994	4.051
	MAE	0.861	1.679	0.399	0.581	0.776	1.496
Nehbandan	R^2	0.462	0.287	0.896	0.875	0.469	0.435
	RMSE	2.578	2.510	1.133	1.050	2.562	2.235
	MSE	6.650	6.302	1.284	1.102	6.564	4.996
	MAE	1.975	1.965	0.812	0.768	1.907	1.730
Tabas	R^2	0.055	-0.045	0.795	0.809	0.060	0.0302
	RMSE	2.492	2.694	1.161	1.151	2.486	2.594
	MSE	6.210	7.260	1.348	1.324	6.183	6.732
	MAE	1.901	2.064	0.833	0.836	1.894	1.967

[1] Radial Basis Function

3.3. DSSI

In Obisesan *et al* (2024)'s study, various machine learning models were trained to predict meteorological variables in a tropical location. The results demonstrated that the Random Forest model had the highest performance metrics, including an R^2 value of 0.93 for temperature prediction and 0.79 for wind speed. showed that Random Forest achieved an accuracy of 95.64%, surpassing the Decision Tree's accuracy of 94.85%. This reinforces the idea that RF is more effective for predicting rainfall due to its ensemble nature, which enhances predictive performance. Furthermore, Mecikalski *et al* (2021) utilized Random Forest to assess predictor importance related to severe storm warnings. The study highlighted how incorporating satellite data into RF models significantly improved prediction accuracy for severe weather events, showcasing RF's versatility in different meteorological applications.

The results showed that CANESM5 was implemented with higher quality according to the AUC assessment (Table 6). Also, the Jackknife of AUC presented for the two models indicated that the common point of the two models is the high importance of wind speed, the high value of which indicates its importance in the suitability of the habitat for the formation of dust sources (Fig 6).

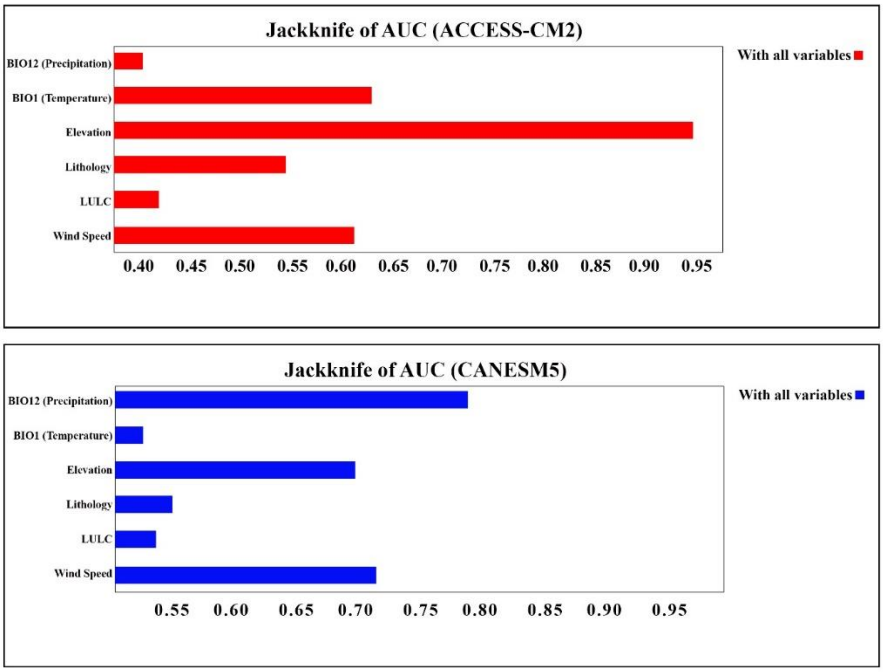

Fig 6. Jackknife of AUC for ACCESS-CM2 and CANESM5

Table 6. Values of AUC

	AUC	
Model	Training data	Test data
ACCESS-CM2	0.882	0.876
CANESM5	0.896	0.966

The results show that in CANESM5, the areas with the highest potential to form dust sources are 2000 km^2 more than ACCESS-CM2 (Table 7). Also, the area of the good floor is more in CANESM5 than in model 1, which indicates the higher area of areas prone to dust formation in CANESM5 (Fig 7).

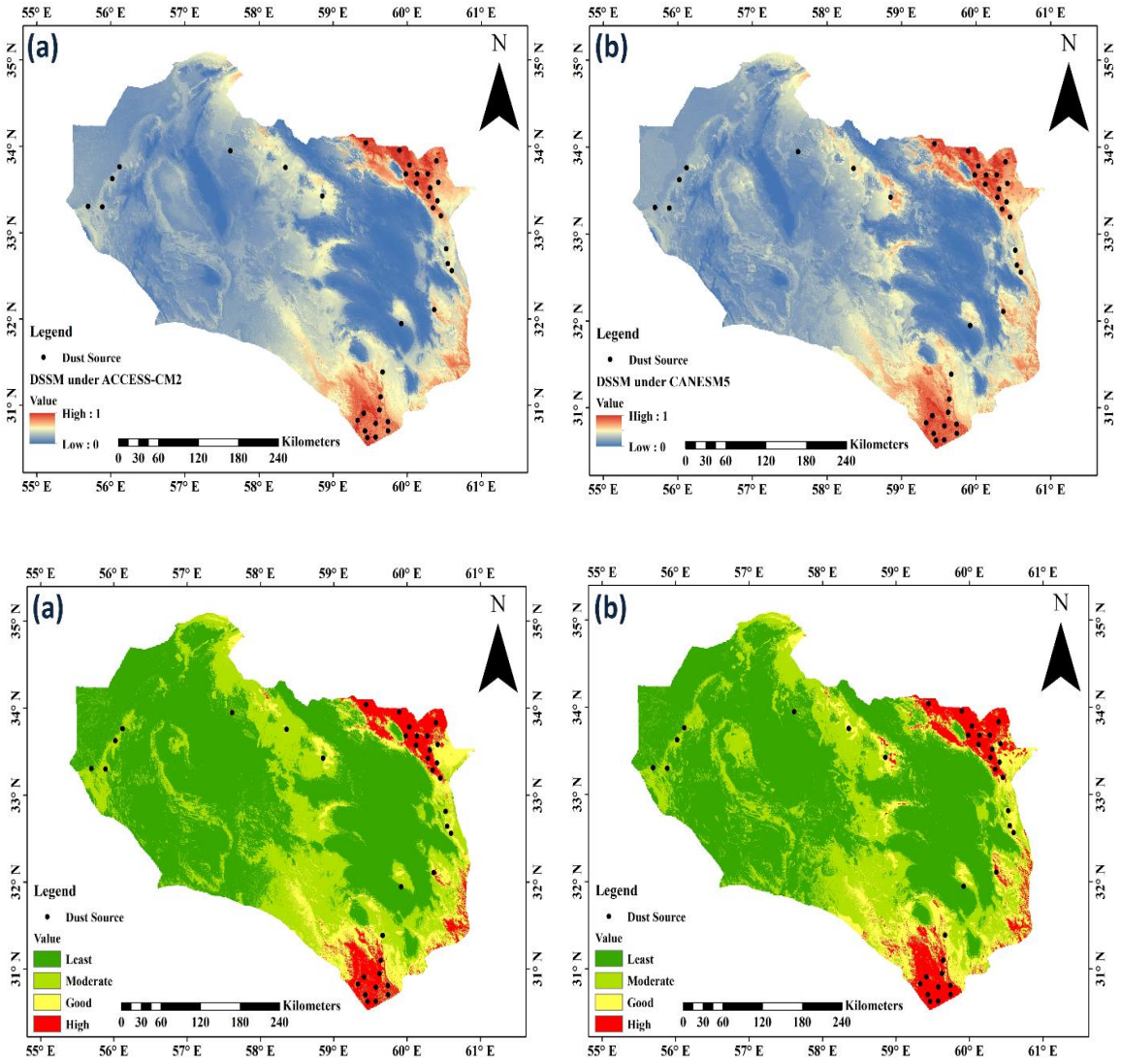

Fig 7. Continuous and classified map of DSSI in ACCESS-CM2 and CANESM5 under SSP 5-8.5 scenario

In similar research, Boroughani *et al* (2022) focused on mapping dust sources using satellite imagery combined with machine learning techniques. The Random Forest model achieved an accuracy of 63.5%, making it the most accurate among the algorithms tested. Utilizing the Google Earth Engine, Wang *et al* (2023) employed four machine learning methods to predict sources of sand and dust storms in arid areas. The results indicated that machine learning approaches effectively identified potential dust sources, providing valuable insights for environmental management and planning. Aryal (2022) compared various machine learning models, including Random Forest (RF), Support Vector Machine (SVM), and others, for predicting fine and coarse dust concentrations. The findings revealed that non-linear machine

learning models outperformed linear regression, particularly in predicting fine dust. Also, they found Air temperature was identified as the most significant predictor.

There are no other similar studies in this field and with these methods, and this study provides a new and comprehensive method by incorporating climate and land use, making it possible to predict dust sources.

Table 7. Areas of DSSI classified map in ACCESS-CM2 and CANESM5 under SSP 5-8.5 scenario

DSSI class	Area (Km^2)	
	ACCESS-CM2	CANESM5
Least	86476.81	83213.75
Moderate	39168.00	38594.06
Good	14038.50	16329.06
High	8801.93	10348.37

4. Conclusion

In this study, three approaches were used to classify land use: Random Forest, CART and SVM. Finally, Random Forest was selected from the two methods CART and Random Forest, which differed only slightly. After the model was modeled using two classified land use maps from 2004 and 2019, a LULC map for 2040 was created. After downloading the data of two temperature variables from IPCC6 models ACCESS-CM2 and CANESM5, the observed wind speed data for 2040 was predicted using three linear regression models, Random Forest and SVM. The R2 showed that the Random Forest model performed better.

Finally, by combining static factors such as lithology and elevation and dynamic factors including two categories of received data, including temperature and precipitation, and predicted data, including wind speed and land cover, the dust source sensitivity model was implemented. For this purpose, the maximum entropy model was used and two different models were created. One model used the temperature and precipitation data of ACCESS-CM2 and CANESM5 used the temperature and precipitation data of CANESM5. In ACCESS-CM2, temperature, wind speed and elevation data played the most important role in the implementation of the model, while in CANESM5, precipitation, elevation and wind speed data played the most important role. Therefore, it can be said that elevation and wind speed are among the most important factors in DSSI. Also, in CANESM5, the area of high-sensitivity regions is 2000 square kilometers higher than in ACCESS-CM2, which indicates the greater vulnerability of the region to the characteristics of this model.

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 all participants of the present study.

Ethical considerations

The authors avoided from data fabrication and falsification.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest
The authors declare no conflict of interest.

References

Ahmadi, Z., Doostan, R., & Mofidi, A. (2015). Synoptic analysis of dust from the warm half of the year in southern khorasan province.

Al-Hemoud, A., Naghibi, A., Hashemi, H., Petrov, P., Kamal, H., Al-Senafi, A., ... & Boloorani, A. D. (2024). Dust source susceptibility in the lower Mesopotamian floodplain of Iraq. *Remote Sensing Applications: Society and Environment*, *36*, 101355.

Anand, A. (2017). Unit-14 Accuracy Assessment. *IGNOU: New Delhi, India*.

Ansari, A. (2017). Determination of dust emissions concentration in desert wetlands (Case study: Meighan wetland, Iran). *Journal of Biodiversity and Environmental Sciences (JBES)*, *10*, 89-97.

Aryal, Y. (2022). Evaluation of machine-learning models for predicting aeolian dust: A case study over the southwestern USA. *Climate*, *10*(6), 78.

Baghi, M., Rashki, A., & Mahmudy Gharaie, M. H. (2020). Investigation of Chemical and Mineralogical Properties of Dust Entering Northeastern Iran and its Pathogenic Potential. *Journal of Geography and Environmental Hazards*, *9*(1), 139-153.

Basheer, S., Wang, X., Farooque, A. A., Nawaz, R. A., Liu, K., Adekanmbi, T., & Liu, S. (2022). Comparison of land use land cover classifiers using different satellite imagery and machine learning techniques. *Remote Sensing*, *14*(19), 4978.

Bauer, S. E., Balkanski, Y., Schulz, M., Hauglustaine, D. A., & Dentener, F. (2004). Global modeling of heterogeneous chemistry on mineral aerosol surfaces: Influence on tropospheric ozone chemistry and comparison to observations. *Journal of Geophysical Research: Atmospheres*, *109*(D2).

Bi, D., Dix, M., Marsland, S., O'farrell, S., Sullivan, A., Bodman, R., ... & Heerdegen, A. (2020). Configuration and spin-up of ACCESS-CM2, the new generation Australian community climate and earth system simulator coupled model. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1), 225-251.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., ... & Eade, R. (2016). The decadal climate prediction project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, *9*(10), 3751-3777.

Boloorani, A. D., Kazemi, Y., Sadeghi, A., Shorabeh, S. N., & Argany, M. (2020). Identification of dust sources using long term satellite and climatic data: A case study of Tigris and Euphrates basin. *Atmospheric Environment*, *224*, 117299.

Boloorani, A. D., Samany, N. N., Papi, R., & Soleimani, M. (2022). Dust source susceptibility mapping in Tigris and Euphrates basin using remotely sensed imagery. *Catena*, *209*, 105795.

Boroughani, M., Pourhashemi, S., Hashemi, H., Salehi, M., Amirahmadi, A., Asadi, M. A. Z., & Berndtsson, R. (2020). Application of remote sensing techniques and machine learning algorithms in dust source detection and dust source susceptibility mapping. *Ecological Informatics*, *56*, 101059.

Boroughani, M., Pourhashemi, S., Hashemi, H., Salehi, M., Amirahmadi, A., Asadi, M. A. Z., & Berndtsson, R. (2020). Application of remote sensing techniques and machine learning algorithms in dust source detection and dust source susceptibility mapping. *Ecological Informatics*, *56*, 101059.

Boroughani, M., Mirchooli, F., & Mohammadi, M. (2022). Dust source mapping using satellite imagery and machine learning models. *Journal of Arid Regions Geographic Studies*, *13*(47), 1-13.

Chamling, M., & Bera, B. (2020). Spatio-temporal patterns of land use/land cover change in the Bhutan–Bengal foothill region between 1987 and 2019: study towards geospatial applications and policy making. *Earth Systems and Environment*, *4*(1), 117-130.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, *37*(1), 35-46.

Congalton, R. G., & Green, K. (2019). *Assessing the Accuracy of Remotely Sensed Data: principles and practices*. CRC press.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Crosbie, E., Sorooshian, A., Monfared, N. A., Shingler, T., & Esmaili, O. (2014). A multi-year aerosol characterization for the greater Tehran area using satellite, surface, and modeling data. *Atmosphere*, *5*(2), 178-197.

Dyke, J., & Kleidon, A. (2010). The maximum entropy production principle: Its theoretical foundations and applications to the earth system. *Entropy*, *12*(3), 613-630.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, *17*(1), 43-57.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937-1958.

Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, 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*, *164*, 102-111.

Felicísimo, Á. M., Cuartero, A., Remondo, J., & Quirós, E. (2013). Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. *Landslides*, *10*(2), 175-189.

Gholami, H., Mohamadifar, A., Sorooshian, A., & Jansen, J. D. (2020). Machine-learning algorithms for predicting land susceptibility to dust emissions: The case of the Jazmurian Basin, Iran. *Atmospheric Pollution Research*, *11*(8), 1303-1315.

Heuvelink, G. B., Angelini, M. E., Poggio, L., Bai, Z., Batjes, N. H., van den Bosch, R., ... & Sanderman, J. (2021). Machine learning in space and time for modelling soil organic carbon change. *European Journal of Soil Science*, *72*(4), 1607-1623.

Jafari, R., Amiri, M., Asgari, F., & Tarkesh, M. (2022). Dust source susceptibility mapping based on remote sensing and machine learning techniques. *Ecological Informatics*, *72*, 101872.

Jafari, M., & Malekian, M. (2019). Comparison and evaluation of dust source susceptibility mapping using logistic regression and logistic model tree models. *Environmental Earth Sciences*, *78*(14), 416.

Jarraud, M. (2008). Guide to meteorological instruments and methods of observation (WMO-No. 8). *World Meteorological Organisation: Geneva, Switzerland*, *29*.

Jensen, J. R. (1996). Introductory digital image processing: a remote sensing perspective (No. Ed. 2). Prentice-Hall Inc..

Karami, S., Hamzeh, N. H., Noori, F., & Ranjbar, A. (2019). Investigation of dust storms in Ilam and the performance analysis of simulation of 6 numerical prediction models at a severe dust storm in west of Iran. *Journal of Air Pollution and Health*.

Kaskaoutis, D. G., Houssos, E. E., Minvielle, F., Rashki, A., Chiapello, I., Dumka, U. C., & Legrand, M. (2018). Long-term variability and trends in the Caspian Sea–Hindu Kush Index: Influence on atmospheric circulation patterns, temperature and rainfall over the Middle East and Southwest Asia. *Global and Planetary Change*, *169*, 16-33.

Khaledian, Y., & Miller, B. A. (2020). Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, *81*, 401-418.

Khosravi, K., Daggupati, P., Alami, M. T., Awadh, S. M., Ghareb, M. I., Panahi, M., ... & Yaseen, Z. M. (2019). Meteorological data mining and hybrid data-intelligence models for reference evaporation simulation: A case study in Iraq. *Computers and Electronics in Agriculture*, *167*, 105041.

Krivoguz, D., Chernyi, S. G., Zinchenko, E., Silkin, A., & Zinchenko, A. (2023). Using Landsat-5 for accurate historical LULC classification: A comparison of machine learning models. *Data*, *8*(9), 138.

Lababpour, A. (2020). The response of dust emission sources to climate change: Current and future simulation for southwest of Iran. *Science of The Total Environment*, *714*, 136821.

L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. *Ieee Access*, *5*, 7776-7797.

Lin, X., Xu, M., Cao, C., P. Singh, R., Chen, W., & Ju, H. (2018). Land-use/land-cover changes and their influence on the ecosystem in Chengdu City, China during the period of 1992–2018. *Sustainability*, *10*(10), 3580.

Lo, C. P. (1986). Applied remote sensing.

Mahowald, N. M., Baker, A. R., Bergametti, G., Brooks, N., Duce, R. A., Jickells, T. D., ... & Tegen, I. (2005). Atmospheric global dust cycle and iron inputs to the ocean. *Global Biogeochemical Cycles*, *19*(4).

Mandal, M., & Vipparthi, S. K. (2021). An empirical review of deep learning frameworks for change detection: Model design, experimental frameworks, challenges and research needs. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 6101-6122.

Mecikalski, J. R., Sandmæl, T. N., Murillo, E. M., Homeyer, C. R., Bedka, K. M., Apke, J. M., & Jewett, C. P. (2021). A random-forest model to assess predictor importance and nowcast severe storms using high-resolution radar–GOES satellite–lightning observations. *Monthly Weather Review*, *149*(6), 1725-1746.

Middleton, N. J. (2017). Desert dust hazards: A global review. *Aeolian Research*, *24*, 53-63.

Mishra, V. N., Rai, P. K., & Mohan, K. (2014). Prediction of land use changes based on land change modeler (LCM) using remote sensing: A case study of Muzaffarpur (Bihar), India. *Journal of the Geographical Institute" Jovan Cvijic", SASA*, *64*(1), 111-127.

Namdari, S., Karimi, N., Sorooshian, A., Mohammadi, G., & Sehatkashani, S. (2018). Impacts of climate and synoptic fluctuations on dust storm activity over the Middle East. *Atmospheric Environment*, *173*, 265-276.

Nazarenko, L. S., Tausnev, N., Russell, G. L., Rind, D., Miller, R. L., Schmidt, G. A., ... & Yao, M. S. (2022). Future climate change under SSP emission scenarios with GISS-E2. 1. *Journal of Advances in Modeling Earth Systems*, *14*(7), e2021MS002871.

Nobakht, M., Shahgedanova, M., & White, K. (2021). New inventory of dust emission sources in Central Asia and northwestern China derived from MODIS imagery using dust enhancement technique. *Journal of Geophysical Research: Atmospheres*, *126*(4), e2020JD033382.

Obisesan, O. E. (2024). Machine learning models for prediction of meteorological variables for weather forecasting. *Int J Environ Clim Change*, *14*(1), 234-252.

Oğuz, K. (2020). Analysis of dust event in Turkmenistan and its source regions. *Eskişehir Teknik Üniversitesi Bilim ve Teknoloji Dergisi B-Teorik Bilimler*, *8*(1), 61-72.

Oliver, H., Shin, M., Matthews, D., Sanders, O., Bartholomew, S., Clark, A., ... & Drost, N. (2019). Workflow automation for cycling systems. *Computing in Science & Engineering*, *21*(4), 7-21.

Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, *2016*(1), 67.

Pham, B. T., Prakash, I., & Bui, D. T. (2018). Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology*, *303*, 256-270.

Pourhashemi, S., Boroghani, M., Amirahmadi, A., Zanganeh Asadi, M., & Salhi, M. (2019). Dust source prioritization with using statistical models (Case study: Khorasan Razavi provience). *Journal of Range and Watershed Managment*, *72*(2), 343-358.

Rahmati, O., Panahi, M., Ghiasi, S. S., Deo, R. C., Tiefenbacher, J. P., Pradhan, B., ... & Bui, D. T. (2020). Hybridized neural fuzzy ensembles for dust source modeling and prediction. *Atmospheric Environment*, *224*, 117320.

Rashki, A., Middleton, N. J., & Goudie, A. S. (2021). Dust storms in Iran–Distribution, causes, frequencies and impacts. *Aeolian Research*, *48*, 100655.

Rivera, J. A., & Arnould, G. (2020). Evaluation of the ability of CMIP6 models to simulate precipitation over Southwestern South America: Climatic features and long-term trends (1901–2014). *Atmospheric Research*, *241*, 104953.

Ruddell, B. L., Brunsell, N. A., & Stoy, P. (2013). Applying information theory in the geosciences to quantify process uncertainty, feedback, scale. *Eos, Transactions American Geophysical Union*, *94*(5), 56-56.

Shaheen, A., Wu, R., & Aldabash, M. (2020). Long-term AOD trend assessment over the Eastern Mediterranean region: A comparative study including a new merged aerosol product. *Atmospheric Environment*, *238*, 117736.

Shepard, D. (1968, January). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM National Conference* (pp. 517-524).

Surabuddin Mondal, M., Sharma, N., Kappas, M., & Garg, P. K. (2019). Ca Markov modeling of land use land cover dynamics and sensitivity analysis to identify sensitive parameter (S). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 723-729.

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., ... & Winter, B. (2019). The Canadian earth system model version 5 (CanESM5. 0.3). *Geoscientific Model Development*, *12*(11), 4823-4873.

Tian, B., & Dong, X. (2020). The double-ITCZ bias in CMIP3, CMIP5, and CMIP6 models based on annual mean precipitation. *Geophysical Research Letters*, *47*(8), e2020GL087232.

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.

Webb, N. P., & Pierre, C. (2018). Quantifying anthropogenic dust emissions. *Earth's Future*, *6*(2), 286-295.

Were, K., Dick, Ø. B., & Singh, B. R. (2014). Exploring the geophysical and socio-economic determinants of land cover changes in Eastern Mau forest reserve and Lake Nakuru drainage basin, Kenya. *GeoJournal*, *79*(6), 775-790.

Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record*, *31*(1), 76-77.

Yu, Y., Notaro, M., & Kalashnikova, O. V. (2018). Assessing the contribution of the Middle East to East Asian dust emissions. *Geophysical Research Letters, 45*(15), 7848-7857.

Zhu, H., & Wang, J. (2009). Chunk-based resource allocation in OFDMA systems-part I: chunk allocation. *IEEE Transactions on Communications*, *57*(9), 2734-2744.

Zoljoodi, M., Ranjbar, A., & Mousavi, S. H. (2021). Identification and classification of dust sources in the Middle East using PCA and hierarchical clustering. *Journal of Arid Environments*, *184*, 104312.