Machine Learning Applications to Dust Storms: A Meta-Analysis

Reem K. Alshammari*[1,2], Omer Alrwais[1], Mehmet Sabih Aksoy[1]

[1]*King Saud University—Information Systems Department*
[2]*King Abdulaziz City for Science and Technology—Space and Aeronautics Research Institute*
Riyadh, Saudi Arabia

* Corresponding author. E-*mail*: ralshammri@kacst.edu.sa

Table S1. A Comparison of Applications of Machine Learning to Dust Storms

	Study citation	**Study target**	**Data source**	**Event location**	**Machine learning model**	**Prediction type**
1	(Lu et al., 2006)	Dust storm prediction	NCEP	Northwestern China Years: 1981–1997	SVM	Daily
2	(Rivas-perea et al., 2010)	Dust storm detection	MODIS L1	31 events (in southwestern USA and northwestern Mexico areas)	Maximum likelihood and probabilistic neural network (PNN)	Real time
3	(Chacon-murguía et al., 2011)	Dust storm detection	MODIS L1	Eight events (in north region of Chihuahua State in Mexico and the southwest area of the state of Texas in the USA)	ANN	Real time
4	(El-ossta et al., 2013)	Dust storm detection	MODIS L1	North Africa 41 dust events and 9 with no dust data acquired between 2001 and 2010.	ANN	Real time
5	(Shahrisvand and Akhoondzadeh, 2013)	Dust storm detection	MODIS L1	Three events in the Middle East	Decision tree, multi-layer perception (MLP) neural network, and SVM	Real time
6	(Ma et al., 2015)	Dust storm prediction	CALIOP L2	Northeast, Middle and East China dust storm: March 19–22, 2010. Aerosol: March 2, 2008 (28–34 N) thick cloud: March 19, 2008 (24–30 N)	TrAdaBoost based on transfer learning	Daily

				thin cloud: March 25, 2008 (42.5–44.8 N)		
7	(Rivas-perea et al., 2015)	Dust storm detection	MODIS L1	30 known dust events globally	support vector regression (SVR)	Near-real time
8	(Souri and Vajedian, 2015)	Dust storm detection	MODIS L1	Eight events in the Middle East	Random forests	Real time
9	(Xiao et al., 2015)	Dust storm detection	- MTSAT-2 - MODIS L2 “Land and Ocean” AOT products - NOAA HYSPLIT model	East Asia deserts including (the Gobi, the Taklimakan), and Arabian deserts	MLP	Near-real time and 72-h forward for dust transport pathways
10	(Xie et al., 2015)	Dust storm prediction	-NCEP grid data -Observation station	Yanchi and Tongxin district in Ningxia. Years: 1956–1965	SRU-AIBSMOTE based on SVM	Daily
11	(Zhang et al., 2015)	Dust storm prediction	Weather stations.	Weather stations in six northwestern provinces of China Years: 1961–2005	Combine SMOTE algorithm with AdaBoost algorithm and random forest algorithm	Frequency of dust storm occurrence monthly
12	(Iranmanesh et al., 2017)	Dust storm prediction	Weather station. Years: 2005 until 2009.	Khuzestan is located in southwestern of Iran	Local Linear Neuro Fuzzy Model	Daily
13	(Murayziq et al., 2017)	Dust storm prediction	Weather stations.	Riyadh 7,000 dust cases	hybrid technique that integrated case-based reasoning (CBR) and Bayesian networks	Daily
14	(Nabavi et al., 2018)	Dust storm prediction	-MODIS-AOD, -ECMWF ERA-Interim, -ESA-CCI GPCC -GIMMS	West Asia, Months (April to September) between 2003 and 2013	Random Forest and Multivariate Adaptive Regression Splines (MARS), SVM, ANN, and Multiple Linear Regression	Monthly
15	(Shaiba et al., 2018)	Dust storm prediction	Data from the Weather Underground website: https://www.wunderground.com	Riyadh, Jeddah, and Dammam	Logistic regression, naïve Bayes, and the classification and regression tree (CART) decision tree	Daily

16	(Ali et al., 2019)	Dust storm prediction	Weather stations	Cairo weather reports Years: 2010 - 2015	Decision tree, k-NN algorithm, and naïve Bayes	Daily
17	(Kh Zamim et al., 2019)	Dust storm prediction	Weather stations	Iraqi cities (Baghdad, Basrah, Samawa, and Nasiriya)	ANN	Daily
18	(Shi et al., 2020)	Dust storm detection	- MODIS L1B - MODIS L2 Aerosol optical depth (AOD) - OMI Ultraviolet aerosol index (UVAI)	Four cases in the Arabian, Gobi, and Taklimakan deserts	New SVM-based method	Near-real time
19	(Shi et al., 2019)	Dust storm detection	-MODIS L1 -CALIOP	Coast off North Africa on July 15,2007	logistic regression, Random Forest, SVM, ANN and one stacking classifiers.	Real time
20	(Tiancheng et al., 2019)	Dust storm detection	-Infrared satellite cloud image data set from National Meteorological Science Data Sharing Service Platform -Daily Dataset of China's Ground Climate Data -China Land Regional Cloud Map (IR1)" and China strong dust storm sequence and its supporting data set	Inner Mongolia-China Years: 2005–2009	naive Bayesian, CNN classification algorithm	Real time
21	(Qing-dao-er-ji et al., 2020)	Dust storm detection	-Infrared satellite cloud image data set from National Meteorological Science Data Sharing Service. -"China Land Regional Cloud Map (IR1)" and "China strong dust storm sequence and its supporting data set".	Inner Mongolia -China Years: 2005–2009	convolution neural network algorithm of transfer learning	Real time

22	(Hou et al., 2020)	Dust storm detection	-CALIPSO for label, -*VIIRS* -CALIOP	Global data, 10,000 samples in March 2012	Deep learning	Real time
23	(Lee et al., 2021)	Dust storm detection	-VIIRS -CALIOP	Global data, 2014	Logistic Regression, K Nearest Neighbor, Random Forest, Feed Forward Neural Network (FFNN), and CNN	Daytime
26	(Ebrahimi-khusfi et al., 2021a)	Dust storm prediction	- Meteorological data - Satellite data	The eastern half of Iran Years: 2000-2018	Granger-Ramanathan averaging (GRA) (MARS, LASSO, k-NN, genetic programming, SVR, Cubist, ANN, Extreme Gradient Boosting, Random forests, and GRA)	Daily
25	(Ebrahimi-khusfi et al., 2021b)	Dust storm prediction	-Meteorological data - Satellite data	semi-arid regions of Iran Years: 2000-2017	ANFIS model	Daily
26	(Berndt et al., 2021)	Dust storm detection at night	GOES-16	Southwest CONUS. 17 dust cases and 11 null cases from January 2018 to June 2020.	Random forests	Real time
27	(Dar et al., 2022)	Dust storm prediction	-25 weather stations - Grid data from Climatic Research Unit	Pakistan Years: 1986–2015	Linear regression	Frequency of dust storm occurrence seasonally
28	(Wang et al., 2022)	Dust storm detection	-MODIS -MODIS land cover type - The digital elevation model (DEM) data	Eight events over Arid Central Asia (the Aralkum and the Taklimakan)	SVM	Real time
29	(Aryal, 2022a)	Dust storm prediction (dust emission of PM2.5 and PM10)	-The Interagency Monitoring of Protected Visual Environments network - The North American Regional Reanalysis (NARR) - the National Oceanic and Atmospheric	Southwestern United States Years: 1988 to 2020	MLR, SVM, Random forests, Bayesian Regularized Neural Networks (BRNN), and Cubist	Frequency of dust storm occurrence monthly

			Administration (NOAA)			
30	(Aryal, 2022b)	Dust storm predication including fine dust (PM2.5) and coarse dust (PM10)	NARR dataset	Southwestern United States Years: 1990 to 2020	ANFIS model	Frequency of dust storm occurrence monthly and seasonally
31	(Jiang et al., 2022)	Dust storm detection	Advanced Geostationary Radiation Imager (AGRI)	48 dust storm events in the Tarim Basin Years :2018 to 2020	CNN	Real time

Table S2. Comparison of Performance of Different Machine Learning Methods Used for Dust Storm Detection and Prediction

	Machine learning method	Training set	Testing set	Accuracy	RMSE	Recall	Precision	Study citation
Methods to predict dust storms	SVM	1,627	400	N/A	N/A	N/A	N/A	(Lu et al., 2006)
	TrAdaBoost (transfer learning)	280 samples	40 samples	N/A	N/A	N/A	N/A	(Ma et al., 2015)
	SRU-AIBSMOTE based SVM	N/A	N/A	N/A	N/A	87.91%	86.52%	(Xie et al., 2015)
	Combined SMOTE algorithm with AdaBoost and random forest algorithms	500	500	N/A	N/A	100%	96.51%	(Zhang et al., 2015)
	Local linear neuro fuzzy model	70%	30%	N/A	N/A	N/A	N/A	(Iranmanesh et al., 2017)
	CBR with Bayesian networks	700 cases	N/A	80–90%	N/A	N/A	N/A	(Murayziq et al., 2017)
	Random Forest, Multivariate Adaptive, Regression Splines, SVM, ANN, and Multiple Linear Regression	partitions of 2003–2010	partitions of 2011–2013	N/A	a. 0.16 b. 0.315 c. 0.273 d. 0.059	N/A	N/A	(Nabavi et al., 2018)
	a. Logistic regression b. Naïve Bayes c. CART decision tree	165,929 cases	165,929 cases	N/A	N/A	N/A	N/A	(Shaiba et al., 2018)
	a. Decision tree b. k-NN c. Naïve Bayes	500	500	a. 97.45% b. 77.34% c. 97.45%	a.0.148 b.0.365 c.0.143	N/A	N/A	(Ali et al., 2019)
	ANN	70%	15% and validatio n 15%	90%	0.84	N/A	N/A	(Kh zamim et al., 2019)
	Granger-Ramanathan averaging (GRA) averaging method	10-fold cross-validation strategy with 10 repeats	N/A	N/A	N/A	0.72	N/A	(Ebrahimi-khusfi et al., 2021a)
	ANFIS model	70%	30%		86.25	N/A	N/A	(Ebrahimi-khusfi et al., 2021b)
	Linear regression	N/A	N/A	N/A	N/A	N/A	N/A	(Dar et al., 2022)
	a. MLR b. SVM c. Random forests	1988 to 2010	2011 to 2020	N/A	a. 0.46 b. 0.47 c. 0.40	N/A	N/A	(Aryal, 2022a)

	d. Bayesian Regularized Neural Networks (BRNN) e. Cubist.				d. 0.48 e. 0.53			
	a. ANFIS model	1988 to 2009	2010 to 2020	N/A	0.45	N/A	N/A	(Aryal, 2022b)
Methods to detect dust storms	a. Maximum likelihood b. PNN	552	552	a. 67.79% b. 84.12%	N/A	N/A	a. 52.55% b. 76.64%	(Rivas-perea et al., 2010)
	ANN	1,045	448	N/A	N/A	N/A	N/A	(Chacon-murguía et al., 2011)
	ANN	60%	40%	88%	N/A	N/A	N/A	(El-ossta et al., 2013)
	a. Decision tree b. SVM c. MLP	3 cases	N/A	a. N/A b. 84.04% c. 81.41%	N/A	N/A	N/A	(Shahrisvand and Akhoondzadeh, 2013)
	a. Detecting dust storm pixels in the MLP-BP b. AOT retrieval using the MLP-BP	8,143	3500	a. 96.6% b.89.39%.	N/A	N/A	N/A	(Xiao et al., 2015)
	SVR	42 million	N/A	92%	N/A	N/A	N/A	(Rivas-perea et al., 2015)
	Random forests	8 cases	N/A	N/A	N/A	N/A	N/A	(Souri and Vajedian, 2015)
	New SVM-based method	1 of 4 cases	N/A	N/A	N/A	77%	91%	(Shi et al., 2020)
	a. Random forests b. Logistic regression c. ANN d. SVM e. Stacking all classifiers (RF, LR, ANN, SVM)	70%	30%	a. 67.2% b. 82.0% c. 69.8% d. 59.6% e. 63.7%	N/A	N/A	N/A	(Shi et al., 2019)
	a. Naive Bayesian algorithm b. Convolutional neural network algorithm c. Improved algorithm			a. 76.9% b. 88.3% c.88.6%	N/A	N/A	N/A	(Tiancheng et al., 2019)
	Deep learning	10,000 samples	randomly chosen	71.1%	N/A	N/A	N/A	(Hou et al., 2020)
	convolution neural network algorithm of transfer learning	15000	5000	84.1%	N/A	N/A	N/A	(Qing-dao-er-ji et al., 2020)
	Random forests	60%	20%	85%	N/A	N/A	N/A	(Berndt et al., 2021)

a. Logistic Regression b. K Nearest Neighbor c. Random forests d. Artificial Neural Network e. Convolutional Neural Network (CNN)	89%	11%	a. 74.1% b. 78.58% c. 82.48% d. 84.99% e. 83.31%	N/A	N/A	N/A	(Lee et al.,2021)
CNN	70%	30%	N/A	N/A	N/A	N/A	(Jiang et al., 2022)
SVM	500 samples	500 samples	98%	N/A	N/A	N/A	(Wang et al., 2022)

Table S3. Performance of Different Machine Learning Methods Used for Dust storm Detection and Prediction

Study citation	Machine learning method	Performance metric	Accuracy
(Lu et al., 2006)	SVM	- Critical success index (CSI) - Running time	- 58.2% - 2.82 s
(Rivas-perea et al., 2010)	Maximum likelihood and PNN	(1) Precision (2) Accuracy (3) Running time	Maximum likelihood: (1) 52.55% (2) 67.79% (3) 0.1484 s PNN: (1) 76.64% (2) 84.12% (3) 2.5198 s
(Chacon-murguía et al., 2011)	ANN	Confusion matrix	97.45%
(El-ossta et al., 2013)	ANN	Accuracy	88%
(Shahrisvand and Akhoondzadeh, 2013)	Decision tree, MLP, and SVM	(1) Accuracy (2) Kappa coefficient	Decision tree: N/A SVM: (1) 84.04% (2) 0.8091 MLP: (1) 81.41% (2) 0.771
(Ma et al., 2015)	TrAdaBoost	N/A	N/A
(Rivas-perea et al., 2015)	SVR	Accuracy	92%
(Xiao et al., 2015)	MLP	Accuracy	Detecting dust storm pixels in the MLP-BP NN:96.6% AOT retrieval using the MLP-BP NN: 89.39%.
(Xie et al., 2015)	SRU-AIBSMOTE based on SVM	(1) Sensitivity (2) Precision (3) F-measure	(1) 87.91% (2) 86.52% (3) 87.21%

(Zhang et al., 2015)	Combine SMOTE with AdaBoost and random forest algorithms	(1) Precision (2) Miss reporting rate (3) Recall rate	(1) 96.51% (2) 0.28% (3) 100%
(Souri and vajedian, 2015)	Random forest	(1) POFD (2) POMD	(1) 7% (2) 6%
(Iranmanesh et al., 2017)		(1) Main Absolut Percentage Error (MAPE) (2) R^2	(1) 0.63 (2) 0.9602
(Murayziq et al., 2017)	CBR with Bayesian networks	Accuracy	Between 80 and 90%
(Nabavi et al., 2018)	a. Random forests b. Multivariate Adaptive Regression Splines (MARS) c. SVM d. ANN	root mean square error (RMSE)	a. 0.16 b. 0.315 c. 0.273 d. 0.059
(Shaiba et al., 2018)	Logistic regression, naïve Bayes, and CART decision tree.	Area under the ROC curve	-
(Shi et al., 2020)	New SVM-based method	(1) Precision (2) Recall (3) Harmonic mean value (F)	(1) 91% (2) 77% (3) 84%
(Shi et al., 2019)	a. Random forests b. Logistic regression c. ANN d. SVM e. Stacking classifiers (Random Forest, Logistic regression, ANN, SVM)	Accuracy	Random Forest (1)79.8% (2)0.765 (3)0.436 Logistic regression (1) 83.9% (2)0.864 (3)0.654 ANN (1) 64.7% (2)0.833 (3)0.429 SVM (1) 65.8% (2) 0.648 (3)0.376 Stacking classifiers (1) 75.6% (2)0.68 (3)0.370

(Ali et al., 2019)	Decision Tree, k-NN algorithm, and naïve Bayes	(1) Accuracy (2) RMSE (3) Correlation	Decision tree: (1) 97.45% (2) 0.148 (3) 0.972 k-NN: (1) 77.34% (2) 0.365 (3) 0.744 Naïve Bayes: (1) 97.45% (2) 0.143 (3) 0.972
(Kh zamim et al., 2019)	ANN	(1) MAPE (2) RMSE (3) Accuracy (4) coefficient of correlation (R)	(1)10% (2)84% (3)90% (4)92%
(Tiancheng et al., 2019)	a. Naive Bayesian algorithm b. CNN c. Improved algorithm	Accuracy	(1) 76.9% (2) 88.3% (3)88.6%
(Hou et al., 2020)	Deep learning	Accuracy	71.1%
(Qing-dao-er-ji et al., 2020)	CNN based on transfer learning	Accuracy	84.1%
(Lee et al.,2021)	a. Logistic Regression b. K Nearest Neighbor, c. Random Forest d. ANN e. CNN	Accuracy	a. 74.1% b. 78.58% c. 82.48% d. 84.99% e. 83.31%
(Ebrahimi-khusfi et al., 2021a)	Granger-Ramanathan averaging (GRA) averaging method	(1)R^2 (2) RMSE (3) MAE (4) CCC (5) % error	(1)0.69 (2) 0.72 (3)0.49 (4)0.70 (5)9.29%
(Ebrahimi-khusfi et al., 2021b)	ANFIS model	(1) RMSE (2) mean absolute error (MAE) (3) Lin's concordance correlation coefficient Lin's (CCC) (4) Pearson correlation coefficient (R)	(1)86.25 (2)50.62 (3)0.88 (4)0.91
(Berndt et al., 2021)	Random forests	-AUC -Accuracy	- 0.97 - 85%
(Aryal, 2022a)	a.MLR b.SVM c. Random forests (RF) d.BRNN e. Cubsit	(1) Correlation (2) RMSE	Fine Dust (PM2.5): *MLR*: (1)0.73 (2)0.46 *SVM*: (1)0.75 (2) 0.47 *RF*: (1)0.81 (2) 0.40 *BRNN*: (1)0.75 (2) 0.48 *Cubist*: (1)0.6 (2) 0.53 Coarse Dust (PM2.5–10): *MLR*:(1)0.71 (2) 2.12

			SVM:(1)0.67 (2) 2.27 *RF*: (1)0.71 (2)2.08 *BRNN*: (1)0.70 (2) 2.12 *Cubist*: (1)0.70 (2) 2.18
(Aryal, 2022b)	a. ANFIS model	(1) Correlation (2) RMSE (3) Bias	Fine Dust (PM2.5): ANFIS: (1)0.7 (2)0.45 (3) 40.64 % Coarse Dust (PM2.5–10): (2) 2.57 (3) 42.08%
(Dar et al., 2022)	Linear regression	N/A	N/A
(Jiang et al., 2022)	CNN	*F*-measure	99.5%
(Wang et al., 2022)	SVM	Accuracy	98%