



OPEN

# Novel deep learning hybrid models (CNN-GRU and DLDL-RF) for the susceptibility classification of dust sources in the Middle East: a global source

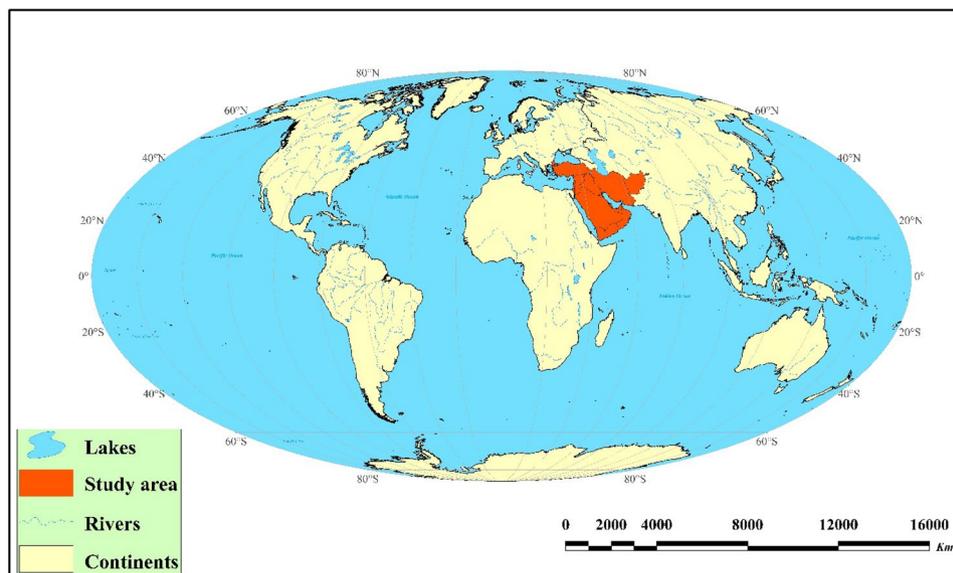
Hamid Gholami<sup>✉</sup> & Aliakbar Mohammadifar

Dust storms have many negative consequences, and affect all kinds of ecosystems, as well as climate and weather conditions. Therefore, classification of dust storm sources into different susceptibility categories can help us mitigate its negative effects. This study aimed to classify the susceptibility of dust sources in the Middle East (ME) by developing two novel deep learning (DL) hybrid models based on the convolutional neural network-gated recurrent unit (CNN-GRU) model, and the dense layer deep learning-random forest (DLDL-RF) model. The Dragonfly algorithm (DA) was used to identify the critical features controlling dust sources. Game theory was used for the interpretability of the DL model's output. Predictive DL models were constructed by dividing datasets randomly into train (70%) and test (30%) groups, six statistical indicators being then applied to assess the DL hybrid model performance for both datasets (train and test). Among 13 potential features (or variables) controlling dust sources, seven variables were selected as important and six as non-important by DA, respectively. Based on the DLDL-RF hybrid model – a model with higher accuracy in comparison with CNN-GRU–23.1, 22.8, and 22.2% of the study area were classified as being of very low, low and moderate susceptibility, whereas 20.2 and 11.7% of the area were classified as representing high and very high susceptibility classes, respectively. Among seven important features selected by DA, clay content, silt content, and precipitation were identified as the three most important by game theory through permutation values. Overall, DL hybrid models were found to be efficient methods for prediction purposes on large spatial scales with no or incomplete datasets from ground-based measurements.

Dust storms have significant effects on the terrestrial and marine environments, climate, biogeochemical cycle, socio-economic aspects of human societies, and their health<sup>1–5</sup>. Such storms are a serious threat in different countries and regions of Asia (e.g., west Asia<sup>6,7</sup>, China<sup>8</sup>, Iran<sup>9</sup>, and central Asia<sup>10</sup>), and especially in the ME<sup>11</sup>. Therefore, a susceptibility classification of dust sources is deemed necessary for the mitigation of the on-site (e.g., soil degradation, abrasion damage, and other damages consisting of loss of seeds, burial of plants, and etc.) and off-site adverse effects of dust storms.

Dust sources can be controlled by climatic factors, ground and land characteristics (e.g., vegetation cover, topographic variables, land surface roughness, and soil properties)<sup>12,13</sup>. Therefore, among variables suggested to represent dust source controls, identifying the influential or important variables is the key to accurate spatial modeling. Unfortunately, there is a general dearth of literature on the use of feature selection algorithms in the aeolian realm<sup>14,15</sup>. One study applied the leave-one-feature-out (LOFO) algorithm for identifying the influential variables controlling dust sources in central Asia<sup>10</sup>. In another work<sup>16</sup>, three variables comprising precipitation, digital elevation model, and soil organic carbon were identified as the key variables controlling dust emissions in southeastern Iran by the Boruta algorithm. Soil, geomorphology, and slope were identified as the three critical variables controlling dust sources in the Sistan watershed (Iran-Afghanistan borders) by two statistical-based predictive models, including frequency ratios (FR) and weights of evidence (WOE)<sup>15</sup>. In the present study, we

Department of Natural Resources Engineering, University of Hormozgan, Bandar-Abbas, , Hormozgan, Iran. ✉email: gholami@hormozgan.ac.ir



**Figure 1.** Location of the ME on the globe. This map was generated in ArcGIS 10.4.1 (<https://www.esri.com/en-us/about/about-esri/overview>).

applied the Dragonfly optimization (FDO) algorithm (a technique less applied to feature selection) to identify influential variables controlling dust provenance in the study area.

Deep learning (DL) models—a new generation of the machine learning (ML) models—are sophisticated tools and promising techniques for prediction purposes at different spatial and temporal scales<sup>17,18</sup>. Among the DL models, the most typical include SAE, DBN, CNN, and RNN models<sup>19</sup>. DL models are more efficient for prediction purposes and spatial modeling because they overcome many of the restrictions placed on the modeling process by traditional models<sup>14,17</sup>.

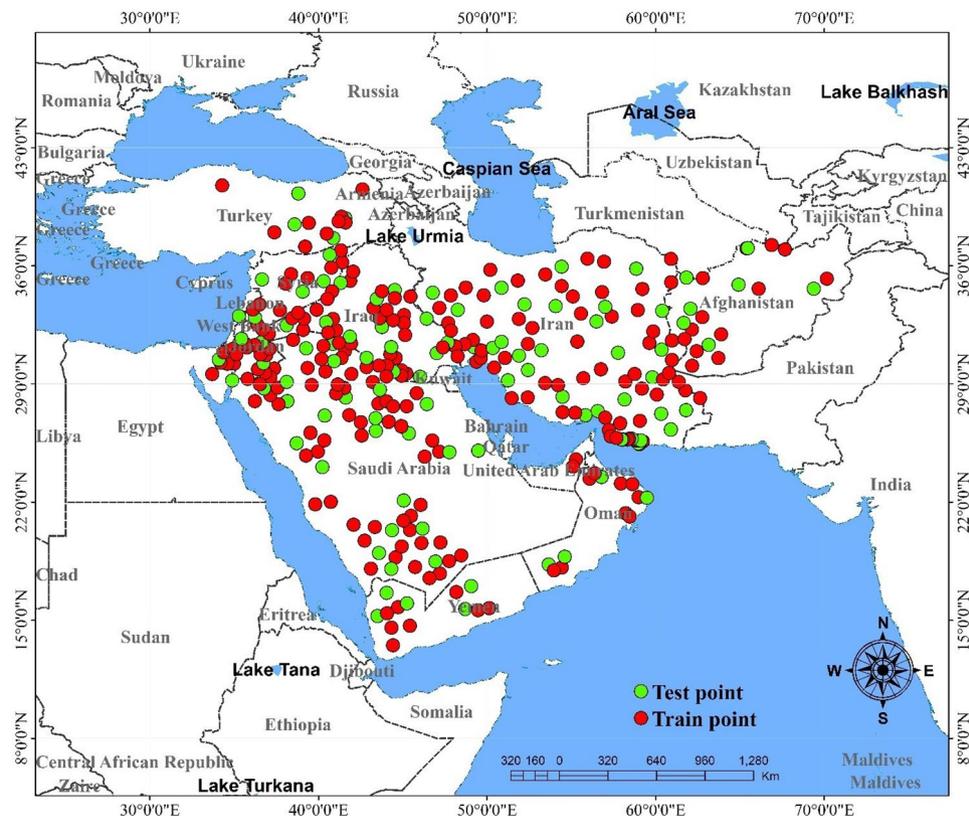
The application of the DL models (e.g., CNN, LSTM, GRU, etc.) has been reported frequently in different fields such as computer science, detecting anomaly, human interaction recognition<sup>20–24</sup>, etc., but its application in aeolian geomorphology is very new. For example, the successful application of two DL models consisting of RNN and RBM reported for spatial modeling of land susceptibility to dust emissions in the Kerman province, southeastern Iran<sup>16</sup>. In another work<sup>10</sup>, two DL models (e.g., gcForest and Bi-LSTM) and a copula-gcForest hybrid model were applied to mapping dust sources in central Asia.

To the best of our knowledge, application of individual CNNs or DL hybrid architectures for the classification of dust sources has not been reported in the literature before. Given the above context, our research applies, for the first time, the Dragonfly algorithm (DA) for discriminating the important features controlling dust source from the non-important features. In a second step, two DL hybrid models were developed based on the convolutional neural network and gated recurrent unit (CNN-GRU model) models, and a dense layer deep learning and random forest (DLDL-RF model) model to classify the spatial sources of dust in the ME. In a third step, six statistical indicators and game theory were used to assess the performance of the DL hybrid predictive model and its interpretability, respectively.

## Materials and methods

**Case study.** The Middle East includes Iran, Afghanistan, Turkey, Iraq, Syria, Kuwait, Bahrain, Saudi Arabia, Jordan, Oman, Yemen, United Arab Emirates, and Qatar. With an area of  $6.69 \times 10^6$  km<sup>2</sup>, the ME is one of the critical sources of dust storms in the global system (Fig. 1). Dust storms are a serious threat especially in the drylands of the ME<sup>25</sup>. The key dust-bearing winds in the study area involve the Shamal (north wind), the Sad-ou-bist bad (wind of 120 days), the Belat, the Simoom (poison wind), the Khamsin (fifty), and the Shlour. They affect the Persian Gulf, southeastern Iran (especially the Sistan plain), southeastern Arabia, Kuwait, Egypt, Syria, and the Lebanon, respectively<sup>26</sup>. The elevation in the study area ranges between  $-418$  and  $6774$  m. The mean annual precipitation ranges from 4 to 2101 mm.

**Mapping factors controlling dust sources and its inventory.** Almost none the ME countries have complete ground-based observational datasets. Remote sensing have high potential to monitor dust events<sup>7</sup>. In the present study we mapped 13 factors controlling dust sources, these comprising land-surface aspect, clay content, silt content, sand content, DEM, lithology, land use, organic carbon content, precipitation, saturated hydraulic conductivity, available water content, soil bulk density, and wind speed. All soil properties were extracted from soil grids, an open source global soil assessment system (<https://soilgrids.org>). A DEM with a resolution of 90 m was downloaded from the earth explorer. The long-term climatic variables were downloaded from the Worldclime website ([www.worldclime.org](http://www.worldclime.org)) and then mapped in ArcGIS. Land use was downloaded from the website <https://maps.elie.ucl.ac.be>. Overall, all input layers used for the modeling process were con-



**Figure 2.** The location of the train and test samples in the study area—the ME. This map was generated in ArcGIS 10.4.1 (<https://www.esri.com/en-us/about/about-esri/overview>).

verted to a spatial resolution of  $50 \times 50$  m. The spatial maps for potential variables controlling dust sources are presented in Supplementary Figs. (S1–S13).

For building the predictive DL models of the dust sources, an inventory map<sup>10</sup>—a map showing current sources of dust—is needed. To obtain it, we consulted the inventory of dust sources provided by Boroughani et al.<sup>15</sup> and the World Bank<sup>27</sup>. Predictive DL models were then constructed by dividing datasets randomly into train (70% or 222 samples) and test procedures (30% or 99 samples) (Fig. 2).

**Feature selection (FS) by the Dragonfly algorithm (DA).** FS is a main stage in the spatial modeling by data mining<sup>28</sup>. The goal of this step is to select the smallest subset of relevant variables (features) from a dataset that captures data characteristics for the generation of an adequate classification, which can be achieved by eliminating repeated, unnecessary, or noisy variables. This procedure speeds up data mining models and improves their performance.

DA—a swarm-based optimization technique—imitates the hunting and migration mechanisms of idealized dragonflies<sup>29</sup>. DA as a swarm intelligence, metaheuristic algorithm<sup>30</sup> include five key steps and all five principles require control of five variables, including separation ( $S_i$ ), alignment ( $A_i$ ), cohesion ( $C_i$ ), attraction towards the food source ( $F_i$ ), and distraction from the enemy ( $E_i$ ).

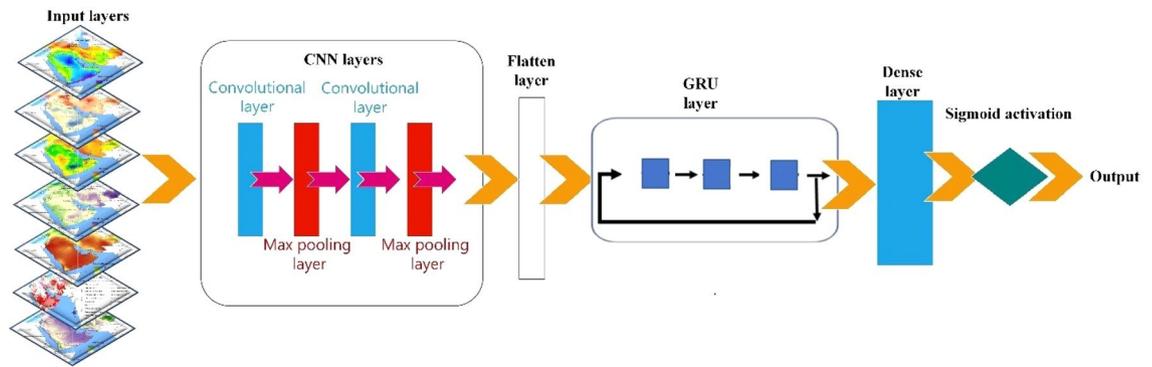
$$S_i = - \sum_{j=1}^N X - X_i \quad (1)$$

$$A_i = \frac{\sum_{j=1}^N V_j}{N} \quad (2)$$

$$C_i = \frac{\sum_{j=1}^N x_j}{N} - X \quad (3)$$

$$F_i = F_{loc} - X \quad (4)$$

$$E_i = E_{loc} + X \quad (5)$$



**Figure 3.** The structure of the CNN-GRU network for predicting dust source.

where  $V_j$ ,  $F_{loc}$ , and  $E_{loc}$  indicate the speed of the  $j$ th neighbor, the position of the food source, and the enemy's position, respectively; and  $X_i$  is the population ( $i = 1, 2, \dots, n$ ). We used the 'zoofs' library to run the DA method.

Successful applications to solve various optimization problems by DA have been reported by several authors<sup>31–33</sup>, but there are no reports to date on the application of this method for features controlling wind erosion and dust sources.

**DL hybrid models for prediction of dust sources.** *CNN-GRU hybrid model.* CNNs—the best architectures of neural networks that learn a hierarchy of complex features by sequential convolution, pooling, and activation function<sup>34</sup>—were initially designed for image classification and recognition, but are today also used in different fields of computer science such as image super-resolution, semantic segmentation, localization, object detection, etc.<sup>35,36</sup>. Due to the high ability of CNNs to learn problem-specific features from raw input data, this architecture mainly reduces the need for handcrafted features<sup>36</sup>.

GRU—an updated version of LSTM—consists of update ( $z_t$ ) and reset ( $r_t$ ) gates<sup>37–39</sup>. GRU optimizes the LSTM network while maintaining LSTM performance. The update and reset gates are used to control the extent to which the information of the previous moment is brought into the current moment and to control the degree of ignoring the information of the previous moment, respectively.  $r_t$  and  $z_t$  determine how the new input is combined with the last memory and how much of the previous memory is to be retained, respectively.  $r_t$ ,  $z_t$  are new memories ( $h_{ti}$ ) and hidden states ( $h_t$ ) that can be expressed as follows:

$$r_t = \sigma \left( W^{(r)}x_t + U^{(r)}h_{t-1} \right) \quad (6)$$

$$z_t = \sigma \left( W^{(z)}x_t + U^{(z)}h_{t-1} \right) \quad (7)$$

$$h_{ti} = \tanh \left( r_t \circ U h_{t-1} + W x_t \right) \quad (8)$$

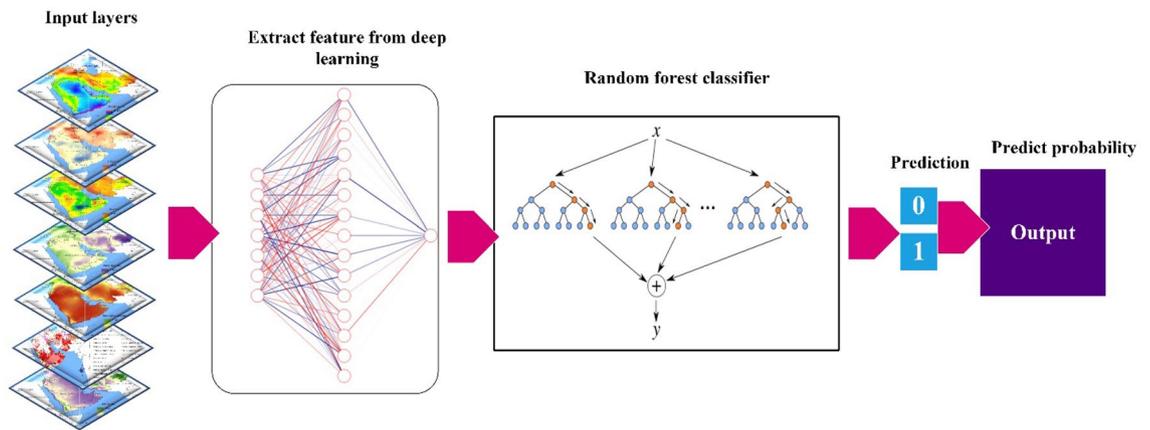
$$h_t = (1 - z_t) \circ h_{ti} + z_t \circ h_{t-1} \quad (9)$$

where  $\sigma(\dots)$  is a sigmoidal function, and  $W^{(r)}$  and  $W^{(z)}$  are weight matrices, while  $\circ$  is the element-wise product.

Here, we present our newly-designed CNN-GRU model (Fig. 3) for the classification of the dust sources in the ME and provide a detailed description of the hyper-parameters by which the model is tuned. Grid-search technique is used to find the optimal hyper-parameters of a model which results in the most 'accurate' predictions. At this method, the parameters of the estimator used to apply these methods are optimized by cross-validated grid-search over a parameter grid.

The network structure of the CNN-GRU model has nine layers: (1) one-dimensional convolution or convolution1D (with input dim = 7, number of neurons = 128, and with the rectifier linear unit (ReLU) as activation function), (2) max-pooling layer, (3) convolution1D (with neuron number = 64 and the activation function = ReLU), (4) max-pooling layer, (5) flatten layer, (6) GRU layer with the neuron number equal to 32, (7) dropout layer with a value of 0.25, (8) GRU layer (number of neurons = 16), and (9) fully connected layer with the neuron number = 1. At the compilation stage, the binary cross-entropy is selected as the loss function. The Adam optimization algorithm<sup>40</sup> was chosen as optimizer because the model's accuracy with this optimizer was higher than that with Adamax. The metrics function denotes the accuracy (metrics function = accuracy). At the fitting stage of the model, among different numbers of the epoch (200, 500, and 1000) and batch size (10, 15, and 20), the model with epoch = 1000, and batch size = 10 performed better than other possible combinations. The ReLU activation function ( $g(z)$ ) and its derivative ( $g(z)'$ ) can be expressed as:

$$g(z) = \max(0, z) \quad (10)$$



**Figure 4.** Conceptual diagram for the DLDL-RF model for prediction of dust sources.

$$g(z)' = \begin{cases} 0, & z < 0 \\ 1, & z > 1 \end{cases} \quad (11)$$

**Dense layer deep learning (DLDL)—random forest (RF) (DLDL-RF) hybrid model.** We introduce a new DL hybrid model (DLDL-RF) based on the dense layer deep learning (DLDL) or fully connected layer and random forest (RF)<sup>41</sup> models (Fig. 4) to classify dust sources in the ME.

New data were generated by DLDL, which were then entered into the RF model to classify the dust sources. Here, we describe the hyper-parameter tuning of the dense-RF model. The structure of the model's network includes three dense layers. In the first dense layer, the input dim is equal to seven, the number of neurons is 30 and ReLU is the activation function. In the second dense layer, the number of neurons is 15 and ReLU is selected as an activation function. The third dense layer has a neuron number of 1 and a sigmoidal activation function. At the compilation stage of the dense-RF model, the loss function, optimizer and metrics function are binary cross-entropy, Adam, and accuracy functions, respectively. At the model's fitting stage, the number of the epoch and batch size were selected as 3000 and 10, respectively, because the model with this epoch number and batch size performed better than other combinations. Finally, all parameters for both models were adjusted by the grid-search automat method.

**Assessment of the performance and interpretability of DL hybrid models.** The model's predictive performance was assessed by six statistical criteria consisting of accuracy, precision, recall, F1 score, Cohens kappa, and receiver operating characteristic—area under curve (ROC-AUC). The game theory through the permutation feature importance measure (PFIM) suggested by Breiman (2001) determines the importance of the influential variables for controlling dust source. A variable is “important” if permutation of its values increases the model error relative to the other features, and a feature is “unimportant” if permutation of its values keeps the model error relatively unchanged<sup>14</sup>.

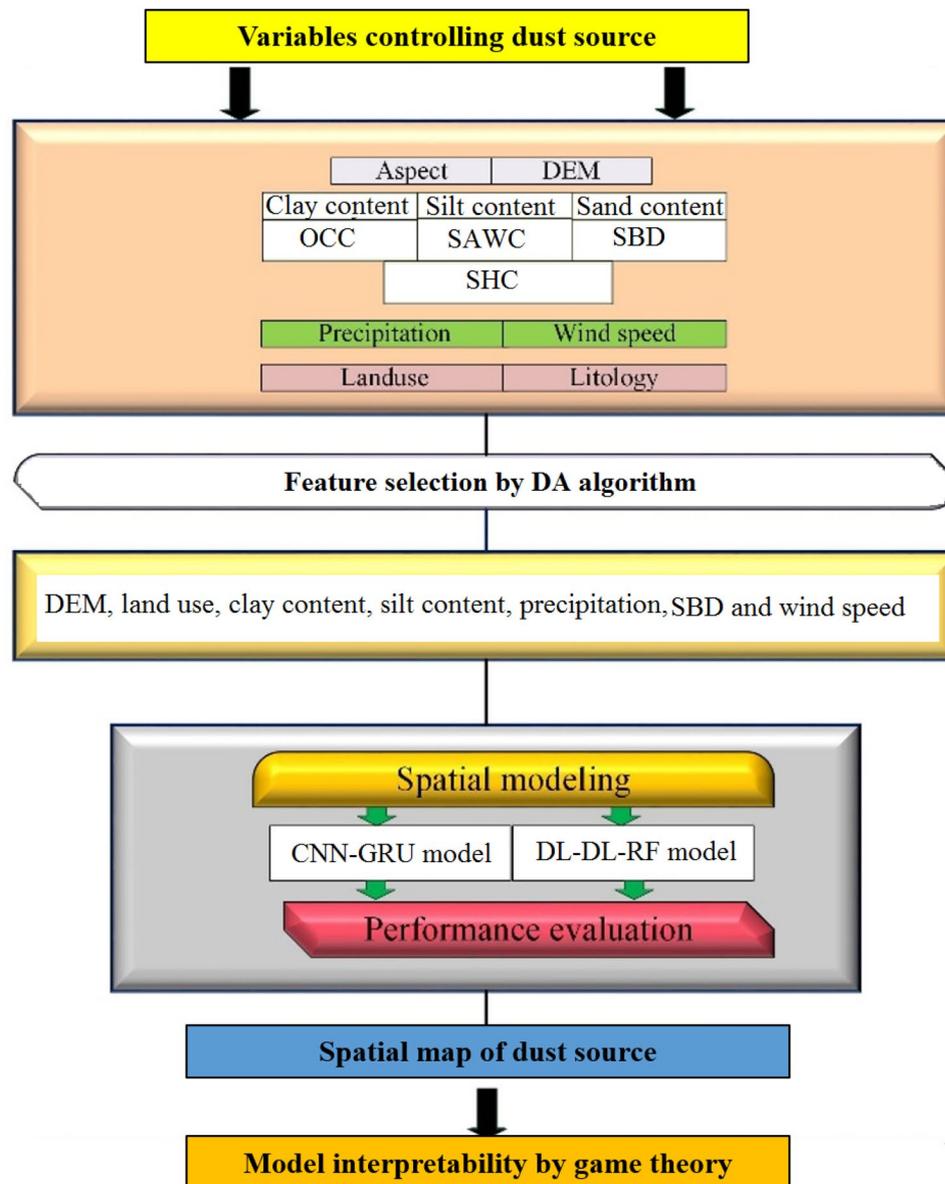
All of the key steps for the spatial modeling of dust sources by feature selection, DL models, and game theory are presented in Fig. 5.

## Results and discussion

**Important variables controlling dust source.** Many environmental, geographic, and bioclimatic variables consist of soil properties, digital elevation models (DEM), topographic characteristics, vegetation cover, land surface roughness, wind conditions, precipitation, and slope which together control dust sources or dust emissions from the land surfaces<sup>4,10,13,26</sup>. Therefore, discriminating important features from the non-important ones by feature selection is a key step for a successful modeling process that can serve to increase the accuracy of predictive models.

Features selected by DA are presented in Fig. 6. Among the 13 features explained as potential variables controlling dust sources, seven features consisting of DEM or elevation, land use, clay content, silt content, precipitation, soil bulk density and wind speed were selected as important variables by DA, the model having the lowest error when these features were included in the model. The objective function is optimized when the difference between observed and predicted values is the lowest, i.e. where the lowest value of the objective function is observed when the seven above mentioned variables are included in the model. According to Fig. 6, the lowest value of the objective function occurred after 18 iterations. Precipitation, soil bulk density and slope were selected as the three most important variables in Central Asia<sup>10</sup>. Three critical variables controlling dust source in the Sistan basin comprise wind speed, elevation and soil organic carbon, all three having been identified by a genetic algorithm<sup>41</sup>.

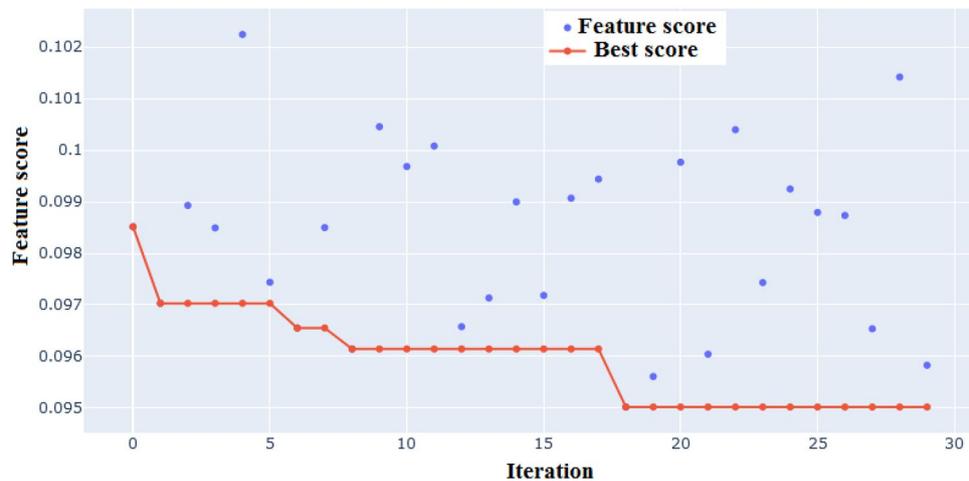
**Spatial maps of dust source.** The spatial maps of dust sources produced by CNN-GRU and DLDL-RF are presented in Figs. 7 and 8. The area of land susceptibility classes as dust sources predicted by the two DL hybrids



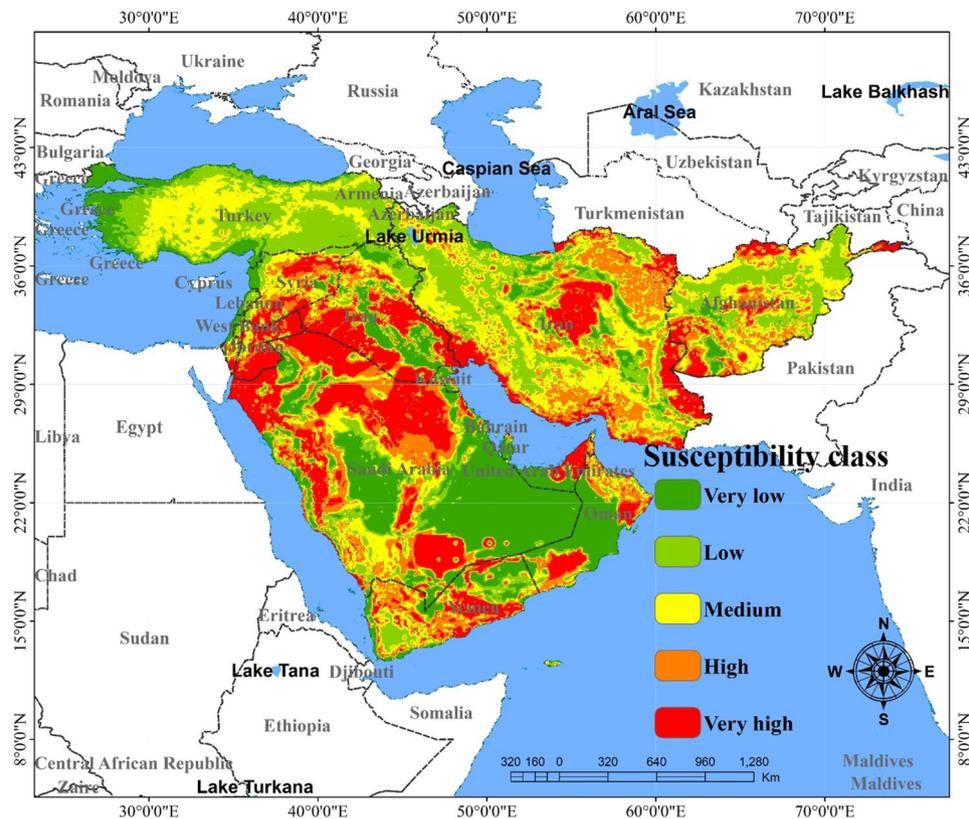
**Figure 5.** Flowchart of the main stages for mapping dust sources by DL modeling and game theory. OCC, SAWC, SBD and SHC indicate the organic carbon content, soil available water content, soil bulk density, and saturated hydraulic conductivity, respectively.

are shown in Table 1. The predictions related to the land susceptibility classes as dust sources can be divided into four classes<sup>10</sup>, or five classes. In the present study the land susceptibility was divided into five classes (very low, low, moderate, high and very high). Based on the results, the spatial map of dust sources generated by the CNN-GRU hybrid model suggested that 19.7, 21.7, and 22.9% of the study area were classified as having very low, low and moderate susceptibility, whereas 17.2 and 18.9% of the area were classified as having high and very high susceptibility, respectively (Fig. 7). The results of the DL-DL-RF model predicted that 23.1, 22.8, 22.2, 20.2, and 11.7% of the total area belonged to the low, moderate, high and very high susceptibility classes, respectively (Fig. 8). Based on the two models, more than 31% of the total area is classified as having high and very high susceptibility. Overall, the most important dust sources in the study area can be divided into 12 regions consisting of southeastern Iran and the border of Iran/Afghanistan (Sistan plain), central Iran (e.g., Yazd, Esfahan and Semnan provinces), southwestern Iran (especially the southern parts of Khuzestan province), a minor source in northern Qatar, a minor source in the middle of Kuwait, eastern Iraq and the border of Iran/Iraq (especially the desiccated bed of the Hawizeh Marshes), border of Iraq/Saudi Arabia (southern Iraq and northern Saudi Arabia), northern Syria, eastern Jordan (or border of Jordan/Saudi Arabia), central Yemen, northern United Arab Emirates, and southwestern Oman.

A large area of the ME consisting of arid and semi-arid climates, is impacted by dust storms<sup>25</sup>. This region—ME—has several dust sources such as Saudi Arabia, Syria, Iraq, and Iran<sup>42</sup>. According to the analysis of TOMS



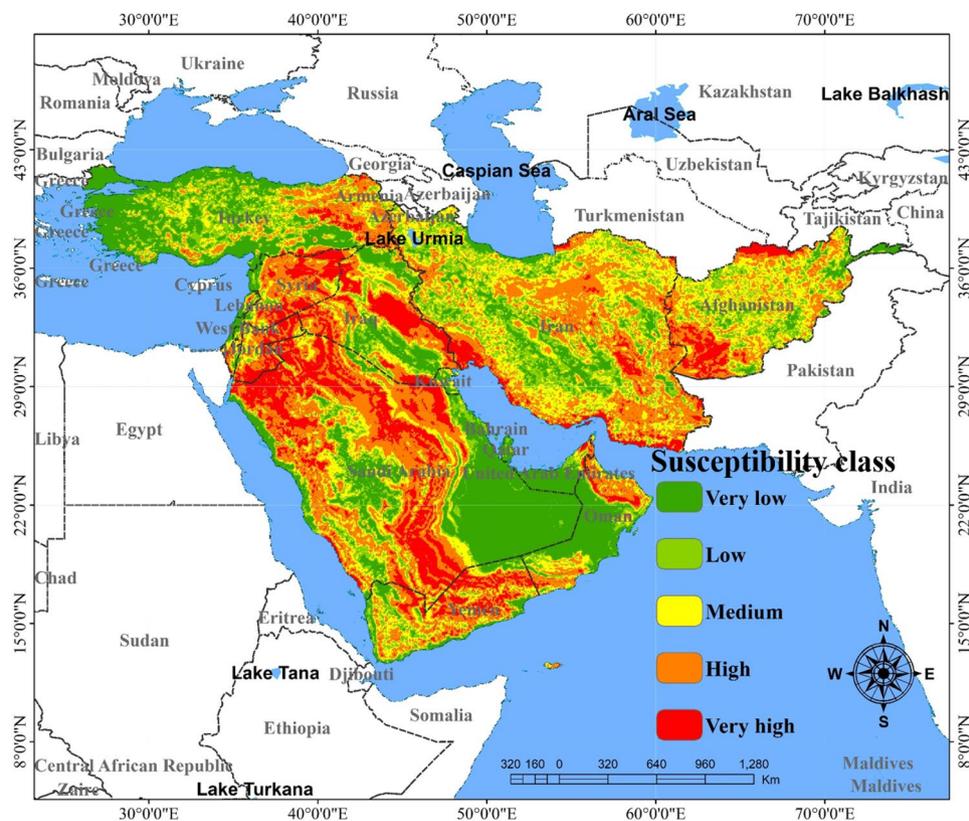
**Figure 6.** Feature selection by DA.



**Figure 7.** The spatial map of dust sources generated by the CNN-GRU hybrid model. The values for pixels was predicted in python, and then, values of pixels were mapped by ArcGIS 10.4.1 (<https://www.esri.com/en-us/about/about-esri/overview>).

data by<sup>26</sup>, the ME is one of the key sources for generating dust particles on earth. Based on the satellite remote sensing and visibility data taken from meteorological stations, Khuzestan province and the Sistan basin are the most dust-affected regions in Iran<sup>5,43</sup>, delivering consistent results in the present study. Recent droughts in external dust storm source areas (Iraq and Syria sources) have increased dust events in western Iran<sup>44</sup>.

Arabia is identified as one of the five key world regions where dust storms frequently occur<sup>45</sup>. A dust source with more significant extension and lesser intensity is located in northeastern Saudi Arabia to the north of the great sand sea—Rub Al Khali<sup>26</sup>, which is consistent with our results. The land along the eastern shore of the Red Sea or the western part of Saudi Arabia can be another primary source in this country. Dust storms frequently occur in the capital of this country—Riyadh—, where the visibility decreases markedly over many days<sup>26</sup>. Based



**Figure 8.** The spatial map of dust sources generated by the DLDL-RF hybrid model. The values for pixels was predicted in python, and then, values of pixels were mapped by ArcGIS 10.4.1 (<https://www.esri.com/en-us/about/about-esri/overview>).

Predictive DL hybrid model	Area	Susceptibility class				
		Very low	Low	Moderate	High	Very high
CNN-GRU	Km <sup>2</sup>	1,318,370	1,453,116	1,533,600	1,150,028	1,237,576
	%	19.7	21.7	22.9	17.2	18.5
DLDL-RF	Km <sup>2</sup>	1,545,082	1,527,326	1,484,457	1,350,140	785,685
	%	23.1	22.8	22.2	20.2	11.7

**Table 1.** The area of dust source susceptibility classes predicted by the CNN-GRU and DLDL-RF models.

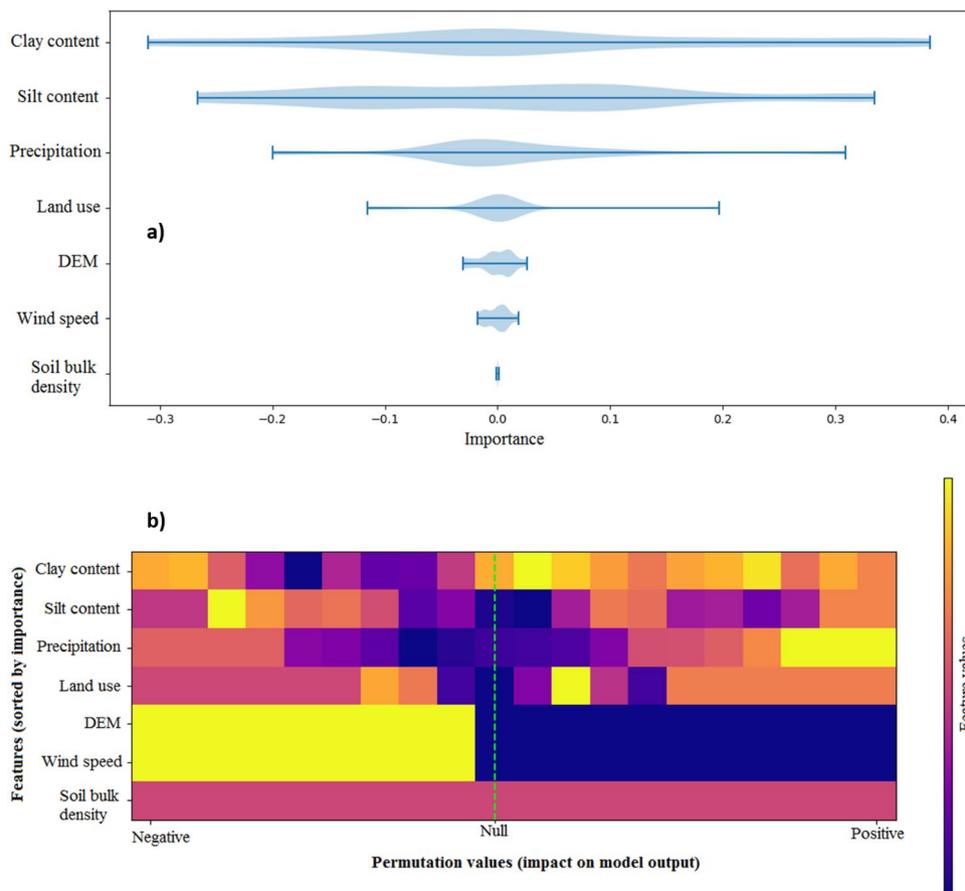
on the meteorological data analyzed by<sup>46,47</sup>, the highest frequency of dust events was observed in southern Iraq and Kuwait. The most intense dust storms in Qatar, Kuwait, and Iraq occur from April to August each year<sup>26</sup>.

**Assessment of the DL hybrid model's performance.** Table 2 presents the results of the predictive DL hybrid models performance assessment for both datasets (train and test) by six statistical indicators. Based on all six indicators used to assess the model's performance, the accuracy of the DLDL-RF hybrid model performed slightly better for both datasets than did the CNN-GRU model. A hybrid approach by synthesizing  $\geq$  two models into an overall model can improve the prediction accuracy and provide reasonably good results<sup>48</sup>. The deep learning architecture with a low false-positive diagnostic rate performed better than machine learning and shallow neural networks<sup>49</sup>. For example, Gholami et al.<sup>10</sup> reported that the copula-gcForest hybrid model (as a DL model) performed better than the individual DL models (e.g., gcForest and Bi-LSTM) for mapping different classes of dust sources in Central Asia. DL models (RNN and RBM) as novel techniques in aeolian geomorphology generate accurate spatial maps of dust sources to help target mitigation of detrimental dust effects on climate, ecosystems and human health<sup>16</sup>.

Until today, applications of CNN-GRU and DLDL-RF hybrid models for spatial modeling of dust sources have not been reported, but there are a limited number of studies that have applied the CNN to the temporal prediction of dust and PM<sub>10</sub>. For example, more recently, Sharma et al.<sup>50</sup> reported that their DL hybrid model or CNN-GRU model outperformed other models for forecasting PM<sub>10</sub>. A new DL hybrid model—CSVR—based

Criteria	Train dataset		Test dataset	
	CNN-GRU	DLDL-RF	CNN-GRU	DLDL-RF
Accuracy	90.8	99.8	87	96
Precision	92.7	100	90	97
Recall	79.7	99.5	84	94
F1 score	85.7	99.8	87	95.4
Cohens kappa	79	99.6	74	92.4
ROC-AUC	88	99.8	87	96.2

**Table 2.** The values of the six statistical indices for assessing the DL model's performance based on train and test datasets.



**Figure 9.** The importance and permutation values for the important features (variables) controlling dust sources calculated by game theory.

on CNN and support-vector-regression (SVR) proposed by Chimire et al.<sup>51</sup> with lower root mean square error (RMSE) and mean absolute error (MAE) performed better than DL and machine learning models (e.g., LSTM, DBN, MARS, RBF, etc.) for solar radiation prediction. Zhang et al.<sup>52</sup> introduced a CNN-GRU model based on multi-task deep learning for the prediction of air quality. They reported that their proposed model had good temporal stability and generalization ability. In comparison with different DL models (e.g., deep belief network (DBN), deep neural network, artificial neural network, etc.), a DL hybrid model consisting of CNN, an extreme gradient boosting with RF, and a Harris Hawks Optimization, was more efficient for predicting boosting solar radiation<sup>53</sup>. The spatial map of soil salinity generated by a one-dimensional convolution neural network—long short-term memory (1DCNN-LSTM) DL hybrid model was more accurate than the salinity map produced by deep Boltzmann machine (DBM) DL individual model<sup>54</sup>.

**Interpretability of DL models by game theory.** The relative importance of the important variables controlling dust sources and their contribution determined by game theory are presented in Fig. 9. Among

seven important features selected by DA, three features (e.g., clay content, silt content, and precipitation) have the highest importance scores and, based on the permutation values, have the highest contribution or impact on model output. Overall, pedo-climatic variables, land surface conditions and roughness are the important variables controlling wind erosion and dust emissions<sup>55,56</sup>.

## Conclusion

This study is the first attempt to develop two DL hybrid models (CNN-GRU and DLDL-RF) for the classification of dust sources in the ME. DA and game theory were applied to the feature selection and interpretability of the DL hybrid models, respectively. Both DL hybrid models performed well for the classification of dust sources, but the DLDL-RF hybrid model performed slightly better than the CNN-GRU hybrid model.

In comparison with existing techniques (e.g., sediment source fingerprinting, geochemical source tracing methods, field-based methods and physical models) for studying dust sources, our DL hybrid models are valuable tools for the classification of dust sources (very low, low, moderate, high, and very high) at large spatial scales (e.g., country, continent, and globe). They can also be applied to other regions experiencing frequent dust storms such as southwestern Asia, northern Africa, the arid and semi-arid regions of USA, Australia, and areas with poor ground-based observation datasets, particularly as the models are inexpensive and do not require any field sampling and laboratory measurements. The main limitations for research are related to the features used in the modeling process. For example, additional accuracy for models may be achieved with higher spatial resolution of the imagery, and more detailed soil maps. Overall, we recommend that future research apply CNNs with various sequential learning mechanisms to the spatial mapping of different environmental hazards.

## Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 6 June 2022; Accepted: 9 November 2022

Published online: 11 November 2022

## References

- Mahowald, N. M. *et al.* Atmospheric global dust cycle and iron inputs to the ocean. *Glob. Biogeochem. Cycles* <https://doi.org/10.1029/2004GB002402> (2005).
- Goudie, A. S. Desert dust and human health disorders. *Environ. Int.* **63**, 101–113 (2014).
- Middleton, N. J. Desert dust hazards: A global review. *Aeol. Res.* **24**, 53–63 (2017).
- Prospero, J. M., Ginoux, P., Torres, O., Nicholson, S. E. & Gill, T. E. Environmental characterization of global sources of atmospheric soil dust identified with the Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) absorbing aerosol product. *Rev. Geophys.* **40**(1), 2–1 (2002).
- Cao, H., Liu, J., Wang, G., Yang, G. & Luo, L. Identification of sand and dust storm source areas in Iran. *J. Arid. Land* **7**(5), 567–578 (2015).
- Kaskaoutis, D. G. *et al.* Assessment of changes in atmospheric dynamics and dust activity over southwest Asia using the Caspian Sea-Hindu Kush Index. *Int. J. Climatol.* **37**, 1013–1034 (2017).
- Nabavi, S. O., Haimberger, L. & Samimi, C. Climatology of dust distribution over West Asia from homogenized remote sensing data. *Aeol. Res.* **21**, 93–107 (2016).
- Feng, J. L., Hu, Z. G., Ju, J. T. & Lin, Y. C. The dust provenance and transport mechanism for the Chengdu Clay in the Sichuan Basin, China. *CATENA* **121**, 68–80 (2014).
- Goudarzi, G., Shirmardi, M., Naimabadi, A., Ghadiri, A. & Sajedifar, J. Chemical and organic characteristics of PM<sub>2.5</sub> particles and their in-vitro cytotoxic effects on lung cells: The Middle East dust storms in Ahvaz, Iran. *Sci. Total Environ.* **655**, 434–445 (2019).
- Gholami, H. *et al.* Integrated modelling for mapping spatial sources of dust in central Asia-An important dust source in the global atmospheric system. *Atmos. Pollut. Res.* **12**(9), 101173 (2021).
- Khaniabadi, Y. O. *et al.* Impact of Middle Eastern dust storms on human health. *Atmos. Pollut. Res.* **8**(4), 606–613 (2017).
- Parajuli, S. P., Stenchikov, G. L., Ukhov, A. & Kim, H. Dust emission modeling using a new high-resolution dust source function in WRF-Chem with implications for air quality. *J. Geophys. Res.: Atmos.* **124**(17–18), 10109–10133 (2019).
- Rashki, A., Arjmand, M. & Kaskaoutis, D. G. Assessment of dust activity and dust-plume pathways over Jazmuriyan Basin, southeast Iran. *Aeol. Res.* **24**, 145–160 (2017).
- Mohammadifar, A., Gholami, H., Comino, J. R. & Collins, A. L. Assessment of the interpretability of data mining for the spatial modelling of water erosion using game theory. *CATENA* **200**, 105178 (2021).
- Boroughani, M., Pourhashemi, S., Gholami, H. & Kaskaoutis, D. G. Predicting of dust storm source by combining remote sensing, statistic-based predictive models and game theory in the Sistan watershed, southwestern Asia. *J. Arid. Land* **13**(11), 1103–1121 (2021).
- Gholami, H., Mohammadifar, A., Golzari, S., Kaskaoutis, D. G. & Collins, A. L. Using the Boruta algorithm and deep learning models for mapping land susceptibility to atmospheric dust emissions in Iran. *Aeol. Res.* **50**, 100682 (2021).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**(7553), 436–444 (2015).
- Shao, Z., Cai, J., Fu, P., Hu, L. & Liu, T. Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. *Remote Sens. Environ.* **235**, 111425 (2019).
- Zhang, Q., Yang, L. T., Chen, Z. & Li, P. A survey on deep learning for big data. *Inform. Fusion* **42**, 146–157 (2018).
- Haroon, U. *et al.* A multi-stream sequence learning framework for human interaction recognition. *IEEE Trans. Hum.-Mach. Syst.* **52**(3), 435–444 (2022).
- Ullah, W., Hussain, T., Khan, Z. A., Haroon, U. & Baik, S. W. Intelligent dual stream CNN and echo state network for anomaly detection. *Knowl.-Based Syst.* **253**, 109456 (2022).
- Ullah, W. *et al.* Artificial Intelligence of Things-assisted two-stream neural network for anomaly detection in surveillance Big Video Data. *Futur. Gener. Comput. Syst.* **129**, 286–297 (2022).
- Hussain, A., Hussain, T., Ullah, W. & Baik, S. W. Vision transformer and deep sequence learning for human activity recognition in surveillance videos. *Comput. Intell. Neurosci* <https://doi.org/10.1155/2022/3454167> (2022).
- Ullah, W., Ullah, A., Hussain, T., Khan, Z. A. & Baik, S. W. An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos. *Sensors* **21**(8), 2811 (2021).

25. Middleton, N. J. Dust storms in the Middle East. *J. Arid Environ.* **10**(2), 83–96 (1986).
26. Goudie, A. S., & Middleton, N. J. (2006). *Desert dust in the global system*. Springer Science & Business Media]
27. World Bank. (2019). Sand and dust storms in the Middle East and North Africa Region: Sources, Costs, and Solutions]
28. Guyon, I. & Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182 (2003).
29. Hammouri, A. I., Mafarja, M., Al-Betar, M. A., Awadallah, M. A. & Abu-Doush, I. An improved dragonfly algorithm for feature selection. *Knowl.-Based Syst.* **203**, 106131 (2020).
30. Mirjalili, S. Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural Comput. Appl.* **27**(4), 1053–1073 (2016).
31. Guha, D., Roy, P. K. & Banerjee, S. Optimal tuning of 3 degree-of-freedom proportional-integral-derivative controller for hybrid distributed power system using dragonfly algorithm. *Comput. Electr. Eng.* **72**, 137–153 (2018).
32. Sureshkumar, K. & Ponnusamy, V. Power flow management in micro grid through renewable energy sources using a hybrid modified dragonfly algorithm with bat search algorithm. *Energy* **181**, 1166–1178 (2019).
33. Yuan, Y., Lv, L., Wang, X. & Song, X. Optimization of a frame structure using the Coulomb force search strategy-based dragonfly algorithm. *Eng. Optim.* **52**(6), 915–931 (2020).
34. Zareapoor, M., Shamsolmoali, P. & Yang, J. Learning depth super-resolution by using multi-scale convolutional neural network. *J. Intell. Fuzzy Syst.* **36**(2), 1773–1783 (2019).
35. Shamsolmoali, P., Li, X. & Wang, R. Single image resolution enhancement by efficient dilated densely connected residual network. *Sig. Process.: Image Commun.* **79**, 13–23 (2019).
36. Basha, S. S., Dubey, S. R., Pulabaigari, V. & Mukherjee, S. Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing* **378**, 112–119 (2020).
37. Athira, V., Geetha, P., Vinayakumar, R. & Soman, K. P. Deepairnet: Applying recurrent networks for air quality prediction. *Proc. Comput. Sci.* **132**, 1394–1403 (2018).
38. Li, W. *et al.* Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). *Inform. Process. Agric.* **8**(1), 185–193 (2021).
39. Pan, E., Mei, X., Wang, Q., Ma, Y. & Ma, J. Spectral-spatial classification for hyperspectral image based on a single GRU. *Neurocomputing* **387**, 150–160 (2020).
40. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
41. Gholami, H., Mohammadifar, A., Pourghasemi, H. R. & Collins, A. L. A new integrated data mining model to map spatial variation in the susceptibility of land to act as a source of aeolian dust. *Environ. Sci. Pollut. Res.* **27**, 1–18 (2020).
42. Shao, Y. *et al.* Dust cycle: An emerging core theme in Earth system science. *Aeol. Res.* **2**(4), 181–204 (2011).
43. Rezaadeh, M., Irannejad, P. & Shao, Y. Climatology of the Middle East dust events. *Aeol. Res.* **10**, 103–109 (2013).
44. Zoljoodi, M., Didevarasl, A., & Saadatabadi, A. R. (2013). Dust events in the western parts of Iran and the relationship with drought expansion over the dust-source areas in Iraq and Syria]
45. Idso, S. B. Dust storms. *Sci. Am.* **235**, 108–114 (1976).
46. Abdulaziz, A. O. A study of three types of wind-blown dust in Kuwait. *J. Meteorol.-Trowbridge Bradford Avon* **19**, 19–19 (1994).
47. Al-Najim, F. A. Dust storms in Iraq. *Bull. Coll. Sci.* **16**, 437–451 (1975).
48. Neshat, A. & Pradhan, B. An integrated DRASTIC model using frequency ratio and two new hybrid methods for groundwater vulnerability assessment. *Nat. Hazards* **76**(1), 543–563 (2015).
49. Kasim, Ö. A Robust DNS flood attack detection with a hybrid deeper learning model. *Comput. Electr. Eng.* **100**, 107883 (2022).
50. Sharma, E. *et al.* Novel hybrid deep learning model for satellite based PM10 forecasting in the most polluted Australian hotspots. *Atmos. Environ.* **279**, 119111 (2022).
51. Ghimire, S., Bhandari, B., Casillas-Pérez, D., Deo, R. C. & Salcedo-Sanz, S. Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in Queensland, Australia. *Eng. Appl. Artif. Intell.* **112**, 104860 (2022).
52. Zhang, Q., Wu, S., Wang, X., Sun, B. & Liu, H. A PM2.5 concentration prediction model based on multi-task deep learning for intensive air quality monitoring stations. *J. Clean. Prod.* **275**, 122722 (2020).
53. Ghimire, S., Deo, R. C., Casillas-Pérez, D. & Salcedo-Sanz, S. Boosting solar radiation predictions with global climate models, observational predictors and hybrid deep-machine learning algorithms. *Appl. Energy* **316**, 119063 (2022).
54. Mohammadifar, A., Gholami, H. & Golzari, S. H. Assessment of the uncertainty and interpretability of deep learning models for mapping soil salinity using DeepQuantreg and game theory. *Sci. Rep.* **12**, 15167 (2022).
55. Gholami, H., Mohammadifar, A., Sorooshian, A. & Jansen, J. D. Machine-learning algorithms for predicting land susceptibility to dust emissions: The case of the Jazmurian Basin, Iran. *Atmos. Pollut. Res.* **11**(8), 1303–1315 (2020).
56. Shao, Y. (2008). *Physics and modelling of wind erosion* (Vol. 37). Springer Science & Business Media.

## Acknowledgements

The authors would like to thank the Faculty of Agriculture and Natural Resources, University of Hormozgan, Iran for supporting this joint research project. We thank Professor Burghard W. Flemming for his assistance to prepare this paper.

## Author contributions

H.G: Formal analysis, Investigation, Visualization, Writing original draft, Supervision, Project administration, Review & editing. A.M: Software, Formal analysis, Research.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-24036-5>.

**Correspondence** and requests for materials should be addressed to H.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022