A data-mining-based approach for aeolian desertification susceptibility assessment: A case study from Northern China

Yue Yaojie1,2 | Li Min2 | Wang Lin2 | Zhu A-Xing3,4

1 Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education, Beijing Normal University, Beijing 100875, PR China
2 School of Geography, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, PR China
3 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application and School of Geography, Nanjing Normal University, Nanjing 210023, PR China
4 Department of Geography, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

Correspondence
Zhu A-Xing, Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application and School of Geography, Nanjing Normal University, Nanjing, PR China.
Email: azhu@wisc.edu

Funding Information
National Key Research and Development Program of China, Grant/Award Number: 2016YFA0602402; National Basic Research Program of China, Grant/Award Number: 2015CB954102; Foundation for Innovative Research Groups of the National Natural Science Foundation of China, Grant/Award Number: 41271286; Outstanding Innovation Team in Colleges and Universities in Jiangsu Province of China

Abstract
Desertification is a grave threat to the environment and livelihoods. Desertification susceptibility assessment (DSA) plays a critical role in reasonable desertification prevention planning by mapping the extent, intensity, and classification of desertification. Numerous desertification maps have been produced using various DSA methods. However, the method of rapid desertification mapping by objectively discovering valuable DSA knowledge from experienced experts stored in such maps has rarely been explored. We propose a data-mining-based approach to mapping aeolian desertification that applies the decision tree (DT) C5.0 (C5) algorithm as a knowledge discovery tool to the reference map and corresponding environmental variables. The results of our case study in Northern China show that the overall accuracy of aeolian desertification classification based on C5 is 86.69%, and the predicted map is highly consistent with the reference map. The DT algorithm outperforms the artificial neural network and naive Bayes approaches. Our results highlight the importance of selecting more representative training samples across where interleaved distributions of multiple aeolian desertification land exist when applying the DT algorithm. The findings of the present study are valuable for highlighting the significance of the data mining approach in DSA, with the growth of desertification maps. Given that aeolian desertification is a complex process coupling natural and human factors, and there are significant regional and scale differences in Northern China, further studies at a fine-scale regarding human factors deserve more attention.

KEYWORDS
aeolian desertification susceptibility, data mining, DT algorithm, environmental indicators, Northern China

1 | INTRODUCTION

Desertification may rank among today's greatest environmental challenges. It has serious local and global impacts and affects the livelihoods of millions of people (MEA, 2005). Furthermore, studies have estimated that desertification is likely to expand substantially because of population growth (MEA, 2005; Reynolds et al., 2007) and climate change resulting in precipitation variability, increased drought frequency, and persistence of dry conditions (D’Odorico, Bhattachan, Davis, Ravi, & Runyan, 2013; Huang, Yu, Guan, Wang, & Guo, 2016; IPCC, 2007). However, the geographic extent of desertification, which is of vital importance in desertification...
prevention planning (Verstraete, Scholes, & Smith, 2009; Yue et al., 2016), remains poorly understood (Daily, 1995; D’Odorico et al., 2013) given the importance of the problem and its recognition as a global issue. Desertification susceptibility assessment (DSA), also known as desertification sensitivity assessment or desertification assessment, is the primary method used to map the locations, extensions, and classifications of desertification (Lavado Contador, Schnabel, Gutiérrez, & Fernández, 2009; Sepehr, Hassanli, Ekhtesasi, & Jamali, 2007). Such assessment is not only of vital practical importance but also an urgent scientific issue in desertification prevention planning at local, regional, and global scales.

One method that has been widely used, among a variety of others, is a typical semiquantitative desertification assessment proposed by Food and Agriculture Organization/United Nations Environment Programme [FAO/UNEP] (1984). This approach is based on 22 indicators covering five aspects of desertification, namely, the current situation, expansion rate, inherent risk, population pressure, and livestock pressure (FAO/UNEP, 1984; Mabbutt, 1986). The relationship between these indicators and desertification level can be expressed as a matrix, in which the rows refer to qualitative or quantitative indicators such as soil and vegetation, and the columns refer to desertification magnitude (FAO/UNEP, 1984; Verón, Paruelo, & Oesterheld, 2006). The elements of the matrix are then integrated in a single index that summarizes the desertification status in one of four classes, that is, slight, moderate, severe, and very severe (FAO/UNEP, 1984; Sepehr et al., 2007; Verón et al., 2006). The first edition of the World Atlas of Desertification was generated using this approach (Oldeman, Hakkeling, & Sombroek, 1990; UNEP, 1992). However, the indicators used are complicated and the assessment process is labour-intensive and time-consuming (Verón et al., 2006).

The MEDALUS model (Kosmas, Kirkby, & Geeson, 1999), proposed by the Mediterranean Desertification and Land Use Project and funded by the European Commission, has also been widely used (Basso et al., 2000; Ferrara, Salvati, Sateriano, & Nole, 2012; Izzo, Araujo, Aucelli, Maratea, & Sánchez, 2013; Ladisa, Todorovic, & Liuzzi, 2012; Lavado Contador et al., 2009; Sepehr et al., 2007; Symeonakis, Karathanasis, Koukoulas, & Panagopoulos, 2014). This model can be regarded as a further development of the FAO/UNEP method. It selects variables such as landform, soil, geology, vegetation, climate, and human activity based on four main indicators, namely, soil, climate, vegetation, and management qualities. Each of these variables is grouped into various uniform classes, and a score is assigned to each class. Then, the desertification level is determined by combining and weighting the four quality layers using a geometric mean method (Ferrara et al., 2012; Ladisa et al., 2012). MEDALUS has the advantage of being relatively easy to implement in conjunction with a geographic information system (GIS) and being relatively flexible in indicator selection and weight setting. However, the approach assumes that climate, soil, vegetation, and management all contribute equally to desertification, neglecting differences between their contributions to desertification development (Salvati & Zitti, 2009; Symeonakis et al., 2014).

In summary, both the FAO/UNEP and MEDALUS techniques confirm that desertification is closely related to the state of impact factors, and specific threshold values of each factor correspond to specific desertification levels (Sommer et al., 2011). However, these two approaches simplified interactions between indicators and desertification as a linear relationship (Salvati & Zitti, 2009). That is, the complex interaction mechanism between indicators and desertification is somewhat underestimated (Salvati & Zitti, 2009). Therefore, how to express the complex quantitative relationships between environmental indicators and desertification susceptibility remains an urgent core scientific issue in desertification assessment research.

More recently, remote sensing has become an important method for monitoring and assessment of desertification susceptibility at global and regional scales, with a large number of up-to-date and high-accuracy desertification detection results (De Jong, de Bruin, Schaepman, & Dent, 2011; Metternicht, Zinck, Blanco, & Del Valle, 2010; Prince, Becker-Reshef, & Rishmawi, 2009; Vorovencii, 2015; Yue, Li, et al., 2016, and references therein). However, remote sensing is unable to reveal the impact mechanism of desertification impact factors (e.g., soil, vegetation, and climate) on desertification. Further, it can hardly project the future trend of desertification caused by natural factors, for example, climate change.

Aeolian desertification, also known as wind-induced or sandy desertification, is a type of land degradation caused by wind erosion in drylands (Wang, 2004; Wang, Chen, Hasi, & Li, 2008; Wang, Eerdun, Zhou, & Liu, 2007; Wang, Xue, Zhou, & Guo, 2015; Wang, Yang, Dong, & Zhang, 2009; Yue, Ye, Zou, Wang, & Gao, 2016). It accounts for ~46.4% of land degradation and is the most important land-degradation type in China (Du, Xu, Zhang, Wu, & Liu, 2016; Guo, Huang, Dong, Pelt, & Zobeck, 2014; Wang, 2004; Wang et al., 2008, 2007, 2009), affecting ~17.93% of the Country’s territory (State Forestry Administration of China, 2015). Aeolian desertification threatens not only China’s environmental sustainability but also that of the world (Liu & Diamond, 2005). As a populous developing country with vast arid areas, China plays a vital role in global aeolian desertification prevention and control, toward achieving sustainable land management (Yue, Ye, et al., 2016). Aeolian desertification assessment in Northern China began in the mid-1980s. Zhu and Liu (1984) first proposed a series of desertification assessment indicators. Later, Dong (1996) assessed aeolian desertification using the FAO/UNEP method. Recently, Shen, Qiu, and Yue (2017) assessed the sandy desertification hazard in Northern China using MEDALUS. However, the above studies still heavily relied on experts’ opinions, and methods of aeolian desertification assessment based on objective knowledge have not been established.

Motivated by the growing demand for information on desertification type and extent for formulating desertification prevention plans under a changing climate, this paper proposes an aeolian desertification susceptibility assessment (ADSA) method that meets the following scientific objectives: (a) objectively discovering expert knowledge of desertification assessment, that is, complex nonlinear relationships between environmental indicators and aeolian
Predicting the type and extent of desertification by studying the characteristics of driving factors is the main task of DSA (Bouabid, Rouchdi, Badraoui, Diab, & Louafi, 2010; Dasgupta, Sastry, Dhinwa, Rathore, & Nathawat, 2013; Dong, 1996; Symeonakis et al., 2014). Specifically, in this paper, we define the ADSA as a classification of lands affected by aeolian desertification (Dong, 1996). Similarity in geographic environment leads to similarity of geographic features (Hudson, 1992; Qi & Zhu, 2003). In other words, regions with the same environmental conditions are likely to experience similar types of aeolian desertification (Zhu, Lu, Liu, Qin, & Zhou, 2018). Accordingly, it is possible to predict the type of such desertification in areas that meet specific conditions based on environmental variables (Yeong, Han, & Ryu, 2010).

To predict aeolian desertification, new experts must develop their own knowledge from scratch. However, if the knowledge of experienced desertification experts can be retrieved and presented in proper form, new scientists could then build upon it (Qi & Zhu, 2003). The desired products of this application of expert knowledge are maps, which constitute an effective medium for presenting spatial information and geographic relationships. For regions where there is no experienced human expertise available, desertification maps produced from previous surveys are one potential source of knowledge (Qi & Zhu, 2003). We argue that existing aeolian desertification maps (data) are vast deposits of knowledge, specifically that regarding the relationships between aeolian desertification types and their indicators. This knowledge can be used to quickly and automatically predict those types in other regions or periods.

If there is a method by which it is possible to quickly and accurately obtain knowledge about the relationships between aeolian desertification susceptibility and its indicators hidden in the maps, then that knowledge can be used to automatically predict such susceptibility in other regions or periods. The development of data mining technology makes possible the visualization of knowledge hidden in maps, which can accelerate map updating. The data mining, which can automatically extract potentially valuable knowledge from large or incomplete datasets, has proven to be a powerful tool for discovering underlying relationships and patterns hidden among dataset variables (Braun, Fernandez-Steeeger, Havenith, & Torgoev, 2015; Delen, Walker, & Kadam, 2005; Fayyad, Piatesky-Shapiro, & Smyth, 1996; Moran & Bui, 2002; Qi & Zhu, 2003; Quinlan, 1993, 2001; Wu et al., 2008). As a step in the overall process of knowledge discovery in a database, data mining is the application of specific algorithms to extracting patterns from data (Fayyad et al., 1996).

On the above basis, we performed the following work to achieve our research goals. Northern China was chosen as a case for ADSA (refer to Section 2.1 for details). The algorithms decision tree (DT) C5, naïve Bayes (NB), and artificial neural network (ANN) were selected to execute the aeolian desertification knowledge discovery (Section 2.2 has details). Five environmental factors covering three aspects of desertification were chosen as indicators for predicting aeolian desertification classifications. These factors are precipitation, aridity index (AI), wind speed, vegetation index, and soil erodibility, and the aspects they cover are climate, vegetation, and soil (Section 2.3 has details). The map of the desert and aeolian desertification in China (MDADC; Wang, Xue, & Chen, 2005) was established as the reference map from which to mine expert knowledge (Section 2.4 has details). Maps of aeolian desertification in Northern China were generated and validated (Section 3 has details), and the feasibility of our proposed method is discussed in Section 4.

2 | METHODOLOGY

2.1 | Study area

According to the FAO (1989), the term ‘arid zone’ collectively represents hyperarid (AI ≤0.03), arid (AI 0.03–0.20), semiarid (AI 0.20–0.50), and subhumid zones (AI 0.50–0.75). China’s hyperarid zone is mostly desert and mobile sand, and aeolian desertification mainly occurs in arid, semiarid, and some semihumid regions of Northern China north of 35°N, where the sandy and desertified lands of the Gobi, China’s major desert, are distributed (Chen & Tang, 2005; State Forestry Administration of China, 2015; Wang et al., 2005; Wang et al., 2008; Wang et al., 2011; Wang, Wu, Xue, Sun, & Chen, 2004; Figure 1). Administratively, this region includes Inner Mongolia, Ningxia, Gansu, Qinghai, Xinjiang, Beijing, Tianjin, Shanxi, Shaanxi, Hebei, Jilin, Liaoning, and Heilong Jiang.

The physiographic character of Northern China provides the background to aeolian desertification. With the exception of eastern and coastal provinces, annual precipitation is <450 mm and spring precipitation <90 mm in most of Northern China (Wang et al., 2008). In accord with this variation, vegetation cover increases from <10% in the west to ~40% in the east (Wang et al., 2007). Furthermore, there are 30–210 days annually in Northern China when the wind speed exceeds the threshold required to transport particles of sand and silt (Chen & Tang, 2005). Also, underlying sediments in most of the arid and semiarid parts of Northern China are extensive fluvial, lacustrine, residual, alluvial, and diluvial deposits, which are particularly vulnerable to erosion (Chen & Tang, 2005; Wang et al., 2007). An arid climate, abundance of sandy soils, strong spring winds, and sparse vegetation make this region very susceptible to aeolian desertification. It is estimated that the area of aeolian desertified lands as of the year 2000 was 385,700 km² within a monitoring area of 2,560,000 km² (Wang et al., 2005). As a result, China has long been deemed a global hot spot for aeolian desertification study.

2.2 | Selection of data mining algorithms and experimental framework

The DT is one of the most popular classification algorithms used in data mining. It is a method that aims to discover the mapping relationships between factor values and categories by learning from a
group of sequence-free and irregular cases (Braun et al., 2015; Quinlan, 1993). It makes no statistical assumptions and can process data represented on different measurement scales, as well as noisy or incomplete data, at a high processing speed (Braun et al., 2015; Yeon et al., 2010). In addition, the DT is suitable for exploratory knowledge discovery (Braun et al., 2015; Chen et al., 2014). Currently, DT algorithms have been successfully used in soil mapping (Braun et al., 2015; Yeon et al., 2010), mineral prospecting mapping (Chen et al., 2014), landslide susceptibility mapping (Braun et al., 2015; Yeon et al., 2010), land cover classification (Colstoun & Walthall, 2006), and desertification assessment (Afrasinei et al., 2017). In the present study, the C5 algorithm (Quinlan, 1993, 2001; Wu et al., 2008) was used to derive DTs from training data, because it proved to be both fast and highly accurate (e.g., Braun et al., 2015; Delen et al., 2005; Pal & Mather, 2003; Qi & Zhu, 2003; Vorovencii, 2015; Wu et al., 2008) and allowed for viewing established rules. This makes it widely applicable in various areas.

To examine the algorithm feasibility in terms of aeolian desertification classification accuracy (CA), we also experimented with two other algorithms, NB (Domingos & Paazzani, 1997) and ANN (Mitchell, 1997). The proposed approach consists of three general steps (Figure 2): (a) selection of aeolian desertification susceptibility indicators, (b) data preparation and ADSA model building based on data mining, and (c) application and validation of the ADSA model.

2.3 | Selection of aeolian desertification indicators

Factors affecting aeolian desertification susceptibility are diverse. We argue that suitable desertification factors must be able to reflect degradation processes. Aeolian desertification is a hazard formed when vulnerable and erosion-prone land surfaces within fragile arid and semiarid environments are exposed to strong winds. This process is impacted by both climate change and human activity (Shen et al., 2017). A lack of rain is the basic environmental condition for aeolian desertification (Becerril-Pina, Mastachi-Loza, Gonzalez-Sosa, Diaz-Delgado, & Ba, 2015; Kassas, 1995; Du et al., 2016; Ferrara et al., 2012). Wind is the primary driving force behind soil erosion that generates aeolian desertification (Du et al., 2016; Wang et al., 2008, 2007, 2009; Yang, He, Li, Huo, & Ding, 2012; Zhu & Liu, 1984). Vegetation coverage determines the degree to which soil is exposed to the wind (D’Odorico et al., 2013; Du et al., 2016). Silt and clay content can decrease soil transport rates (Du et al., 2016), but when wind power exceeds soil resistance (i.e., soil erodibility), there is wind erosion. This erosion ultimately results in vegetation degradation and a decline in land productivity (Kassas, 1995; Wang & Zhu, 2003). Therefore, variables such as soil type, vegetation cover, climate aridity, precipitation, and wind speed are the major determinants of aeolian desertification, and the interaction of those determinants produces complex, nonlinear relationships between those variables and their effects (Du et al., 2016).

In accord with the principle of dominance, we suggest selecting key factors that are closely related to aeolian desertification as susceptibility indicators. These indicators cover three aspects (Table 1), that is, climate, vegetation, and soil, from which five indicators are eventually selected: mean annual precipitation, mean annual wind speed, mean annual AI, vegetation index, and soil erodibility. This selection is informed by expert knowledge from preceding studies (e.g., Du et al., 2016; FAO/UNEP, 1984; Kosmas et al., 1999; Shen et al., 2017; UNEP, 1992; Wessels, Prince, Frost, & van Zyl, 2004; Yue, Li, et al., 2016, and references therein). Among these, the precipitation, AI, and vegetation index represent environmental characteristics relevant to aeolian...
desertification. Soil erodibility depicts the vulnerability of soil to wind erosion, and wind speed characterizes the criticality of that erosion.

The proposed indicators are mainly natural factors affecting aeolian desertification. However, we argue that further study should focus on constructing a comprehensive indicator system that integrates both natural and human factors. We discuss this in Section 4.3.

### 2.4 Data preparation and model building

Data preparation is one of the most important steps in data mining (Delen et al., 2005). The main tasks in this stage include the processing of data resources and extraction of training samples. The former refers to processing existing maps and variables relevant to aeolian desertification into datasets with consistent data formats, projections and coordinate systems.

#### 2.4.1 Data preparation

Datasets used in our study are shown in Table 2.

In this paper, mean annual precipitation, wind speed, and AI data cover the periods 1951–2000, 1951–2000, and 1950–2000, respectively. The precipitation are meteorological station data, and we used ordinary kriging to interpolate them onto a 1-km grid, because the precision of that kriging has been proven the best among seven types of interpolation when applied to semiarid China (Wei, Li, & Liang, 2005).

The AI data are provided by the CGIAR Consortium for Spatial Information (CGIAR-CSI). The AI is usually used to quantify...
where \( Sa \) is sand content (%), \( Si \) is silt content (%), \( Cl \) is clay content (%), OM is organic matter content (%), and \( CaCO_3 \) is calcium carbonate content (%). All these soil parameters are derived from the Harmonized World Soil Database (Fischer et al., 2008). Finally, the soil erodibility index was converted to a 1-km grid.

The MDADC (1:4,000,000) is the most important research result of the State Key Research Development Program project “Study on the Process of Desertification and its Control in Northern China” (G20000487; Wang et al., 2005). It was edited by the project team on the basis of Thematic Mapper images of 2000, and the type and distribution of aeolian desertification have been extensively validated through field surveys by a large number of experienced professionals (e.g., Wang, 2004; Wang et al., 2011, 2004, 2005, and references therein). The MDADC has been widely recognized as a comprehensive and long-term research result for aeolian desertification in China, and the data have been used in many research studies (e.g., Wang, 2014; Wang et al., 2007; Wang et al., 2008; Wang et al., 2011; Wang et al., 2015, and references therein). Therefore, we consider data of this map to be reliable.

The MDADC was first digitized and then converted to a 1-km raster to keep it consistent with the other data. The map has nine land types including mobile sand, semifixed dune, fixed dune, Yardang, Gobi, oasis, slight, moderate, and severe desertification. Among these, we argue that the oasis is not a desertification type. The fixed and semifixed dunes are mainly in the semiarid and semi-humid regions of northeastern China and are customarily called desertified land (Wang et al., 2005). Based on this, we combined the desertification land types of slight, moderate and severe

**TABLE 2** Data sources

| Data name | Contents | Data sources |
|-----------|----------|--------------|
| Map of desert and aeolian desertification in China | 1:4,000,000 vector map Land types: mobile sand, semi-fixed dune, fixed dune, Yardang, Gobi, oasis, slight desertification, moderate desertification, severe desertification | Wang et al. (2005) |
| Precipitation | Average annual precipitation with range 13.55–1075.44 mm over period 1951–2010 of 757 stations in Northern China | China Meteorological Data Sharing Service (http://data.cma.gov.cn/) |
| Aridity index | Global 30 arc-second (~1 km) annual average aridity index with range 0.0084–0.9696 over period 1950–2000 | Bossio, Trabucco, Bossio, van Straaten, and Verchot (2008). Data available from CGIAR-CSI GeoPortal at http://www.csi.cgiar.org. |
| Wind speed | 1-km × 1-km grid, annual average speed with range 0.5–6.2 m/s over period 1951–2000 in Northern China | Institute of Geographical Sciences and Natural Resource Research |
| Normalized difference vegetation index (NDVI) | SPOT_VEGETATION, 1-km × 1-km grid, annual average in 2000 | Cold and Arid Regions Sciences Data Center at Lanzhou (http://westdc.westgis.ac.cn) |
| Soil | China soil map based on Harmonized World Soil Database (v1.1), 1:1,000,000 vector map, soil types and attributes | Fischer et al. (2008). Cold and Arid Regions Sciences Data Center at Lanzhou (http://westdc.westgis.ac.cn) |

precipitation availability over atmospheric water demand. According to CGIAR-CSI (2012) and Zomer et al. (2008), AI is calculated using

\[ AI = \frac{MAP}{MAE} \quad (1) \]

where MAP is mean annual precipitation and MAE is mean annual potential evapotranspiration.

We used the annual average SPOT_VEG NDVI of 2000, because NDVI is typically used to indicate the ultimate state of vegetation productivity after long-term desertification.

The soil erodibility index represents the susceptibility of soil to being lost to wind erosion. It is calculated according the equation proposed by Fryrear, Krammes, Williamson, and Zobeck (1994):

\[ SEI = \frac{29.09 + 0.315a + 0.175i + 0.335a/Cl - 4.66OM - 0.95CaCO_3}{100} \quad (2) \]

Note. After Wang (2004), Wang et al. (2008, 2007, 2004, 2005).
desertification, semifixed, and fixed dune into a single ‘desertified land’ category (Table 3). The entire dataset was configured using geographic coordinates and the WGS84 spheroid. All processes used ArcGIS software (version 10.02).

Table 3. Characteristics of aeolian desertification types in Northern China (after Wang et al., 2008, 2007, 2004, 2005)

| Type                | Description                          | Code  |
|---------------------|--------------------------------------|-------|
| Desertified land    | Desertified land                     | i     |
| Gobi                | Gobi                                 | ii    |
| Mobile sand         | Mobile sand                          | iii   |
| Yardang             | Yardang                              | iv    |

2.4.2 Sampling and model building

Because the areas of various aeolian desertification types in Northern China vary greatly, to obtain the optimal classification model for this desertification, a plurality of training sample sets were constructed based on area-weighted sampling principles (Liu, Zhu, Yang, Miao, & Zeng, 2017; Moran & Bui, 2002; Pal & Mather, 2003; Qi & Zhu, 2003; Zhu et al., 2015). The samples were spatially extracted using the “Create Fishnet” tool within ArcGIS. Variable sample numbers can be obtained by adjusting the fishnet size. Each sample point obtained by this method contains the aeolian desertification type and corresponding indicators values at the same location.

In order to achieve optimal aeolian desertification CA, it is necessary to select a number of training sets of different sizes and train the data mining algorithm repeatedly. Then, through analyzing the relationship between CA and sample set size, the best training sample size can be selected. On this basis, the training samples initially acquired were subjected to preprocessing steps such as removal of null sample points, data format conversion, and attribute coding, finally forming a training sample set suitable for data mining. All training samples were processed using the DT C5 algorithm built using IBM SPSS Modeler 14.1.

The aeolian desertification classifier model is presented in tree form and cannot be applied directly to subsequent aeolian desertification classification. Therefore, a rule transformation of the tree model is needed. Specifically, the tree model was expressed as an IF/THEN statement from the root node to each leaf node, wherein the IF part fully inspects of one path and corresponds to the value of each indicator. The THEN part corresponds to the most likely classification result, that is, the degree of aeolian desertification susceptibility. Each IF/THEN statement induced and formed by this method is mutually exclusive and complete, so there is no sequential problem in application. The rule set formed by all IF/THEN statements can express knowledge of the relationships between aeolian desertification susceptibility and its impact factors, and the ADSA model is thereby established.

2.5 Model accuracy verification

We suggest that the ADSA results be checked for CA and spatial distribution consistency (SDC) at the same time. This is because it may be that the extent of a desertification type is predicted accurately, but the locations at which that type are predicted differ from their actual locations.

The procedure is as follows. First, we overlay the predicted map of aeolian desertification with the reference map, that is, the MDADC. Second, an error matrix is obtained by random selected validation samples (training samples not included). This matrix enables us to check prediction accuracy and spatial consistency by comparing the reference and predicted maps grid-by-grid. Third, the CA, Kappa coefficient, and SDC are used to examine the prediction accuracy of the ADSA model.

The CA expresses the ability of an established model to correctly predict aeolian desertification susceptibility categories. In general, the greater the accuracy, the stronger the model’s prediction capability. The CA for aeolian desertification type \( i \) is expressed as

\[
CA_i = \frac{P_i}{N_i},
\]

where \( P_i \) refers to the number of samples that actually belong to type \( i \) and are also predicted as \( i \) (i.e., desertified land, Gobi, mobile sand, and Yardang), that is, the number of samples along the main diagonal of the error matrix; \( N_i \) refers to the total number of samples of \( i \).

The Kappa coefficient is used to assess consistency between the predicted aeolian desertification type and the original, the value of which is taken between 0 and 1 (Grinand, Arrouays, Martin, & Laroche, 2008). The closer to 1, the greater the consistency and the better the model. The calculation is

\[
\text{Kappa} = \frac{P_o - P_c}{1 - P_c},
\]

Here, \( P_o \) refers to the actual consistency rate and \( P_c \) to the theoretical one. TP means truly predicted, TP\(_i\) is the total number of samples of type \( i \) correctly predicted as that type, and \( ii \) means that the predicted \( i \) is truly \( i \) (i.e., desertified land, Gobi, mobile sand, and Yardang; that is, four aeolian desertification types). \( P_r \) refers to the total number of grids in the column containing type-i land, \( P_s \) to the total number of grids in the row containing type-i land, and \( N \) to the total number of samples.

The SDC shows how many grids are correctly predicted in terms of aeolian desertification type. It is measured by dividing the number of correctly predicted grids by the total number of grids in the study area. In this case, SDC equals \( P_o \).

Besides the above indicators, field investigation is key to desertification classification. By examining details of incorrectly predicted aeolian desertification types in an inconsistency region, we can identify possible reasons for classification errors. However, because the study area is so large, it is almost impossible to validate the aeolian desertification classification results by field survey. Additionally, the spatial resolution of the results is still relatively coarse, that is, 1:4 million considering the large spatial scope, so the validation samples collected at field scale in a small region are not comparable with our results. Therefore, we did not collect accuracy test samples onsite. However, we checked land-surface details of the predicted aeolian desertification map using remote sensing images and the assistance of Google Earth. The spatial resolution of satellite imagery is much higher than that of
our results, so satellite images can provide sufficient information to test and analyze the reliability of the predicted map, especially in a region as large as Northern China.

3 | RESULTS

3.1 | Aeolian DSA model programming

Taking the MDADC as reference map, there were over two million total samples of land affected by aeolian desertification in Northern China. To avoid over-training the DT algorithm, the influence of sample size on aeolian desertification CA was analyzed first. The training sample size ranges from 1.0% to 30.0% of the samples (Grinand et al., 2008). Data mining and aeolian desertification classification using the DT C5 algorithm were repeated 25 times. The experimental results of the classification method were validated using all grids except training samples from the reference map.

We found that the aeolian desertification CA was very stable, which means the C5 algorithm performed stably in building a DT. In addition, this indicates that CA improves with the number of training samples (Figure 3). However, the magnitude of accuracy improvement tends to increase very slowly when the sample size is >170,000. We further found that the DT algorithm performs best when the sample size is between 200,000 and 250,000 (Figure 3). More specifically, a total of 210,000 grids, which cover about 9.15% of Northern China, achieved the highest overall accuracy score. Therefore, we conclude that satisfactory accuracy in aeolian desertification classification can be obtained without increasing the number of training samples endlessly. Thus, approximately 210,000 grids were selected as the training sample.

An aeolian desertification classifier model with 268 total leaf nodes, each of which falls within one of the four desertification land types, was established. The numbers of leaf nodes corresponding to desertified land, mobile sand, Gobi, and Yardang were 73, 72, 112, and 11, respectively. Because the study area is so large and the land surface so diverse, the DT diagram with values of the five indicators corresponding to each desertification land type is very complex. Therefore, we show an example of the DT diagram (Figure S1 in the Supporting Information) and rule sets of the DT classifier model to reveal how a type of aeolian desertification land is classified.

As shown in Figure S1, the ‘IF A and B and C and ... THEN class X’ statement corresponding to the first leaf node is as follows: IF A1 ≤ 0.0587 and soil erodibility index ≤ 0.5021 and vegetation index ≤ 56, and wind speed < 1.1 then Class == Gobi. The entire set of these ‘IF...THEN...’ statements constitutes the ADSA model. These make up only a very small part of the diagram and model because they are too complex to show in their entirety.

3.2 | Validation of predicted aeolian desertification map

The predicted aeolian desertification map for Northern China in 2000 is shown in Figure 4, and comparisons between the reference maps, predicted maps, and remote sensing images of major deserts and sandy regions in the region are shown in Figure 5. The total land area of each of the four types of aeolian desertification is shown in Figure 4a as follows: Desertified land, 403,780 km²; mobile sand, 446,888 km²; Gobi, 599,876 km²; and Yardang, 29,367 km². These constitute 27.28%, 30.20%, 40.53%, and 1.98% of the total area, respectively. Compared with the reference map, the predicted area of Gobi is 1.81% larger, and those of mobile sand, desertified land, and Yardang are 0.15%, 1.64%, and 0.02% smaller. Therefore, the areas of predicted aeolian desertification lands are very similar to those shown on the reference map, indicating that the model performed well in predicting the extent of desertification types.

We constructed the error matrix for aeolian desertification susceptibility classification (Table 4), and the Kappa coefficient was calculated to be 0.83. The prediction accuracy for each of the four types of aeolian desertification are from high to low as follows: Gobi (91.76%), mobile sand (88.33%), desertified land (84.63%), and Yardang (81.85%). Furthermore, the spatial consistency reached 86.69%, indicating that the locations of aeolian desertification lands on the predicted map are highly consistent with those on the reference map.

As shown in Figure 4b, areas where consistency is relatively strong are characterized by single and monolithic aeolian desertification types, such as the Taklimakan desert, whereas regions with poor consistency tended to have one of the following characteristics. The first is that the region is dominated by a certain type of aeolian desertification, but the periphery contains a variety of other types in a staggered pattern. For example, the Taklimakan Desert region is dominated by mobile sand but surrounded by a sporadic distribution of desertified land, Gobi, and Yardang. Although the dominant desertification type was correctly predicted, other minor surrounding types were predicted less accurately. Second, the region has two desertification land types and is dominated by one type in the center, whereas the two types are staggered and roughly distributed equally at the periphery.
For example, desertified land and mobile dunes are mixed at the peripheries of both the Mu Us and Hobq Sandy Lands, which leads to one type being falsely identified as the other. The reasons for the poor prediction ability of the model in regions characterized by an interlaced distribution of multiple desertification land types are detailed in Section 4.4.

In summary, when compared with the MDADC (Figures 4 and 5), the predicted map obtained via the proposed model shows only small differences in total area and strong spatial consistency for each of the four aeolian desertification types.

4 | DISCUSSION

4.1 | Comparison of DT, ANN, and NB Algorithms

ADSA models based on the ANN and NB algorithms were constructed based on the same group of training sample sets used by the DT. Predicted aeolian desertification maps of Northern China based on the ANN and NB were generated, as well as spatial distribution consistencies for each of the two maps were compared (Figure S2).

The overall prediction accuracies of ANN and NB are 68.88% and 53.93%, respectively. Accordingly, the Kappa coefficients are 0.53 and 0.27, which are both much smaller than those of DT. That is, the accuracy of aeolian desertification classification using the DT algorithm is much higher than that from the other two algorithms. Our results show that classification accuracies for the four aeolian desertification land types vary substantially between the models generated by the three algorithms. The DT had the greatest prediction accuracy, >80% for all four desertification land types. The ANN had the greatest prediction accuracy, 72.17% for Gobi, followed by 71.64% for mobile sand and 66.12% for desertified land. However, the recognition rates for Yardang were as small as 0%. NB has the greatest prediction accuracy, 100% for Gobi, followed by 52.16% for desertified land, and 0% for the other two land types.

Using comparisons, we concluded that the DT has a relatively stable and balanced prediction accuracy for each desertification land type and gives the highest accuracy overall. Our finding agrees with previous works claiming that DT performance is acceptable for desertification assessment (Xu et al., 2009) and landslide susceptibility mapping (Yeon et al., 2010). More specifically, the C5 algorithm performs best in landslide susceptibility mapping (Braun et al., 2015), breast cancer survivability prediction (Delen et al., 2005), land cover classification (Pal & Mather, 2003; Vorovencii, 2015), and soil map production (Qi & Zhu, 2003) relative to other algorithms. This is partly because C5 adds a powerful boosting algorithm to improve...
CA by establishing the DT, which focuses on data that are wrongly classified. It is very suitable for mining knowledge hidden in desertification maps. Moreover, DT is very capable of solving nonlinear relationships and the thresholds of each indicator are very flexible, depending on the specific environmental condition of an area. For example, for a desertification type, its classification thresholds for each of the five indicators may be $A'\text{'}$ in region A, but in region B, these thresholds might be $B'\text{'}$, where $A'\text{'}$ and $B'\text{'}$ are not necessarily

**FIGURE 5** Visual comparisons of reference maps, decision tree predicted maps, and remote sensing images for major deserts and sand lands in Northern China: (a) Taklimakan Desert, (b) Gurbantunggut Desert, (c) Badain Jaran Desert, (d) Ordos Plateau including Mu Us Sandy Land and Hobq Desert, (e) Horqin Sandy Land, and (f) Hulunbeir Sandy Land [Colour figure can be viewed at wileyonlinelibrary.com]
identical. Therefore, the classification of desertification types by DT accurately reflects the characteristics of different climatic and ecological regions.

In addition, the rules (knowledge) discovered and generated by the DT are easier to understand than results from the other two algorithms, thus helping researchers interpret and evaluate models established according to those rules. Qi and Zhu (2003) and Braun et al. (2015) pointed out that compared with ANN and NB algorithms, the DT has the advantages of easily comprehensible mining results and high accuracy. Furthermore, the ANN algorithm requires a specialist to determine network structure and set various parameters, which affects the efficacy of models to some extent (Kim et al., 2011; Pal & Mather, 2003). Also, Pal and Mather (2003) indicated that the DT is more accurate than ANN in land-cover classification. In conclusion, the DT algorithm offers meaningful advantages in comparison to ANN and NB when applied to the task of discovering knowledge contained in existing maps and establishing the ADSA model.

### 4.2 Data sources and data homogenization

We used data from different periods, scales, and spatial resolutions. This created a challenge for data homogenization, even though the DT algorithm is capable of partitioning the input data into many homogeneous subsets by producing optimal rules that minimize error rates in the branches of the tree (Colstoun & Walthall, 2006).

With regard to period, we suggest that the selection of data sources should consider the evolution speed of aeolian desertification indicators. Though some factors might have strong annual volatility, aeolian desertification may not. In other words, the classification of aeolian desertification type may be more prone to error when using variable data from the same year, for example, data from 2000 to classify aeolian desertification types in that same year. Therefore, we suggest that if one wants to make a map of aeolian desertification for a certain year, one would be better served by using the annual average of an indicator for a relatively long period before that year. For factors that change very slowly, for example, soil erodibility, the data can come from the same year for which the aeolian desertification map is produced, or from proximate prior years. Therefore, we used mean annual precipitation and wind speed data plus the mean annual Al over the periods 1951–2000, 1951–2000, and 1950–2000, respectively, together with the 2000 annual average SPOT_VEG NDVI and soil erodibility based on the Harmonized World Soil Database (Fischer et al., 2008).

With regard to spatial resolution, the Al, wind speed, and soil erodibility are 1-km spatial resolution grid data. The mean annual precipitation data are from meteorological stations, and ordinary kriging was used to interpolate them onto a 1-km grid (Wei et al., 2005). The MDADC (Wang et al., 2005) was also converted to a 1-km grid to keep it consistent with precipitation, the Al, NDVI, wind speed, and soil erodibility. This does not mean that the spatial resolution of this map was interpolated to 1 km; it is still actually a map with scale 1:4,000,000. However, such a conversion is conducive to spatial linkage between the reference map and indicator data at grid scale, which is ultimately useful for selecting training points and data mining.

A reliable reference map plays a key role in extracting knowledge of aeolian desertification classification and validating CA. In our study, the learned DT was tested using independent samples from the same map to evaluate learning accuracy. To some extent, it is inappropriate to use a derived map as a source of validation points. This is because the reference map is derived, subject to a certain accuracy. The propagation of error might be amplified if desertified land types in the reference map are not accurate. However, we argue that it is not necessary to pursue ‘absolute’ correctness when using this map as a database for expert knowledge mining. In fact, any map is constructed with certain accuracy requirements. As for the MDADC, it is the crystallization of many experts’ collective wisdom and has been verified through extensive field survey. Limited by data availability, the MDADC still has the most reliable data available to depict desertification in Northern China. Nevertheless, our proposed approach can be further improved and validated in a smaller study area, using a high-precision desertification map.

Aeolian desertification is the result of the complex and nonlinear effects of many factors (Du et al., 2016). Locations where environmental factors have some differences will not necessarily form different types of aeolian desertification. This implies that such desertification is more ‘coarse’ regarding spatial differences than its impact factors. From this point of view, the role of the reference map is to provide knowledge of the factors influencing a certain aeolian desertification type, which usually covers a large area. We therefore argue that the reference map as a ‘knowledge background’ does not need to maintain
a spatial resolution as high as the impact indicators. In other words, the data resolution of the indicators is key to determining how fine the aeolian desertification classification can be.

One deficiency of our work is that the validation points were all taken from the reference map instead of field investigation. However, it is almost impossible to validate the aeolian desertification classification results through field survey because the study area is so large. Therefore, we did not collect accuracy test samples onsite. Instead, we used the CA, Kappa coefficient, and SDC to examine prediction accuracy of the ADSA model, by comparing each grid of the reference map to the corresponding grid from the predicted map. We also checked details about the predicted aeolian desertification types by examining the land-surface details with the aid of Google Earth (Figure 5). Considering the slow process of desertification, present satellite images can still provide enough information to test and analyze the reliability of desertification classification, especially in such a large region as Northern China. Nevertheless, field investigation should attach great importance to our proposed approach if it is applied to a smaller study area using data of finer resolution.

Another deficiency is that all fixed dunes in the reference map were combined into desertified land, which might be inappropriate. For example, some new mobile sand should be included in desertified lands, whereas many long-term fixed dunes should not be included in those lands. However, desertification is a gradual process that may last a long time (Wang, 2004; Wang et al., 2004). Operationally, from the MDADC, it is difficult to identify which mobile/fixed dunes are new and which are long term. Usually, the type of desertification is recognized according to the strength of dune activity or wind erosion (Wang, 2004; Wang et al., 2008, 2007, 2015, 2009). For example, sand land in a semifluid state is regarded as severely desertified land (Wang et al., 2005). From this aspect, it is appropriate to combine fixed dunes with desertified lands. Furthermore, other indicators such as the dune activity index (Wang et al., 2007, 2009) may help to classify the types of desertification lands more appropriately.

4.3 | Aeolian desertification assessment indicators

Desertification is a dynamic phenomenon in time and space and cannot be treated as a single-step problem, because it is influenced by several factors from both natural and socioeconomic perspectives (D’Odorico et al., 2013; MEA, 2005; Salvati & Zitti, 2009). Thus, the selection of suitable indicators is key in ADSA modeling. However, lack of agreement on the choice and application of indicators has been a major handicap in attempts to assess the status and trends of desertification (Mabbutt, 1986).

We mainly used natural factors, for example, precipitation, wind speed, AI, vegetation, and soil in this study, among which climatic factors dominated the process of aeolian desertification. Thus, trends and future projections of desertification under climate change have attracted much attention. Although several studies tried to reveal the future trend of aeolian desertification (Wang et al., 2009), global and regional projections of desertification linked to climate change remain unclear. Our proposed method provides an opportunity to look at future trends of aeolian desertification by coupling projected climatic variables. Nevertheless, human activities such as deforestation, irrational cultivation, and overgrazing are the most likely main causes of desertification besides climate change (e.g., Chen & Tang, 2005; D’Odorico et al., 2013; Wang et al., 2011, and references therein). However, there is much debate regarding the leading role of natural and human factors in aeolian desertification (Wang et al., 2008; Wang, Chen, & Dong, 2006). Indirect indicators such as human and socioeconomic phenomena have the potential advantage of integrative monitoring of a range of processes over broad areas, but with the risk of sensitivity to processes other than desertification (Mabbutt, 1986). Therefore, human factors have not been treated in the present study. Nevertheless, human activities such as ‘grain to green’ and ecological reconstruction might have strong effects toward the reversal of land degradation (Yue, Li, et al., 2016; Yue, Ye, et al., 2016). Because of this, we argue that human factors should be added to the indicators for a comprehensive ADSA.

Among the proposed indicators, the AI is closely related to precipitation. At one time, mean annual precipitation was used as a direct substitute index for aridity, but precipitation cannot reflect the magnitude of regional dryness or humidity comprehensively because it does not consider the consumption of water. On the other hand, the AI is more suitable for characterizing regional aridity (or on the contrary, humidity) than precipitation, because it considers both water supply and consumption. However, because the study area is very large and the land surface very complex, we argue that taking either aridity or precipitation as the sole indicator cannot fully reflect the impact of water status on aeolian desertification. This may lead to large error in aeolian desertification classification. To confirm this, we did additional experiments by removing either precipitation or the AI from the indicators. The corresponding overall accuracies in terms of aeolian desertification classification were 84.8% and 84.5%, respectively, both of which are substantially smaller than the accuracy attained using both indicators. Therefore, we recommend that both the AI and precipitation should be included in the index system. This has been done in a large number of studies (e.g., Basso et al., 2000; Bouabid et al., 2010; Ferrara et al., 2012; Izzo et al., 2013; Kosmas et al., 1999; Salvati & Zitti, 2009; Sepehr et al., 2007; Sommer et al., 2011).

The NDVI has been proven capable of assessing vegetation dynamics and relationships with land degradation (De Jong et al., 2011). Even so, with respect to the discrimination of sand bodies, most authors have relied on analyses of visible reflectance (0.4–0.7 μm; Collado, Chuvieco, & Camarasa, 2002), because bare soil shows high reflectance in the visible bands and green vegetation has strong absorption. However, visual interpretation of remote sensing images is time-consuming and laborious. In principle, the desertification indicators should be reliable, regularly updatable, available at low cost, and cover large areas with high spatial resolution (Ren, Dong, Hu, Zhang, & Li, 2016). Because of this, vegetation indices have been widely used to improve desertification monitoring efficiency at large scales (Yue, Li, et al., 2016). Among these, the NDVI has demonstrated its reliability in assessing or monitoring land desertification (Collado
et al., 2002; Runnstrom, 2003; Wessels et al., 2004; De Jong et al., 2011; Becerril-Pina et al., 2015; Ren et al., 2016, and references therein). Compared with the other vegetation indices, the NDVI has been demonstrated more reliable for detecting aeolian desertification in arid and semiarid areas (De Jong et al., 2011; Ren et al., 2016). Therefore, considering the huge arid area in Northern China, that index is well suited to our research needs. Some indices such as the moving standard deviation index and albedo have been proposed as powerful adjuncts to the NDVI for monitoring desertification patterns (Tanser & Palmer, 1999; Xu et al., 2009; Yue, Li, et al., 2016). Therefore, we argue that indicators such as albedo as an adjunct to NDVI may create a more accurate DT for classifying aeolian desertification.

### 4.4 Selection and size of training samples

The amount, spatial rationality, and representativeness of training samples in the study area have a critical impact on full expression of the relationship between aeolian desertification and environmental factors, which in turn influences the CA of that desertification. Undoubtedly, there is a need to set the number of training samples and their distribution based on characteristics of the entire study area (Grinand et al., 2008; Moran & Bui, 2002).

Currently, there are three main methods for selecting samples, random, equal-number, and area-weighted sampling. Foody (2002) argued that when the training sample set is large enough to ensure that each category has sufficient representative samples, random sampling can be used. However, studies of training sample size by Pal and Mather (2003) and Grinand et al. (2008) show that although the accuracy of the DT classification model increases with the number of training sample sets, this increase will reach a ceiling, as shown by Figure 3. Therefore, there are limitations on how much model performance can be improved simply by increasing the size of the training sample. Compared with random sampling, the area-weighted sampling method ensures that each category has a certain number of samples when the training sample set is not sufficiently large, so categories with smaller area proportion will not be neglected during sampling (Grinand et al., 2008; Moran & Bui, 2002). The model established using area-weighted sampling also proved to be more accurate than that using equal-number sampling in soil mapping (Moran & Bui, 2002). Literature also indicates that training sample selection based on the grading of soil types by area is an effective sampling technique (Moran & Bui, 2002). This study classified aeolian desertification types using five environmental factors (precipitation, wind speed, Al, soil erodibility, and vegetation index) as indicators, by applying the DT C5 algorithm as a knowledge discovery tool to the reference map and corresponding environmental variables. A case-study of Northern China shows that the algorithm offers substantial advantages for knowledge discovery using existing maps and ADSA modeling. The predicted aeolian desertification map is very consistent with that of the reference map, that is, an 86.69% match, with a Kappa coefficient of 0.83, showing that DT is very capable of classifying desertification type depending on the specific environmental condition of an area. The DT C5 algorithm was shown to have the greatest overall accuracy in classifying aeolian desertification land as compared with the ANN and NB algorithms.

We found that the DT C5 algorithm is less accurate in regions with interleaved distributions of multiple aeolian desertification land types. Indicating that the spatial distribution of desertification type might have a considerable impact on the effectiveness of the C5 algorithm when the area-weighted sampling method is used to obtain training samples. Although aeolian desertification CA does not always increase endlessly with the increasing overall training sample size, it is still recommended that more representative samples be established in regions with multiple mixed desertification types when acquiring samples, especially that desertification land with a small area, where smaller sample size may be not sufficient for forming an effective desertification classification mode.

This study advances the concept that for DSA, the same environmental conditions are likely to experience similar desertification types. That is, the desertification type is predictable in other regions or periods if the interaction relationships, that is, knowledge of the type of such desertification in a specific area and corresponding environmental variables, can be discovered. We argue that existing desertification maps are knowledge deposits from many experts that describe relationships between desertification types and environmental variables. Such knowledge can be, and should be, discovered and applied for rapid desertification mapping.
Our findings reveal that the DT algorithm is capable of discovering expert knowledge objectively from desertification maps, which suggests that data mining is promising as a powerful DSA approach, considering that the number of desertification maps is growing quickly. These findings highlight that our proposed method is very useful for promoting sustainable land planning strategies, not only for decision-makers but also for stakeholders, by quickly providing information about desertification types and their geographic extents. Further studies regarding human factors at fine scale are needed, with the aid of intensive field investigation.

ACKNOWLEDGMENTS

This research was financially supported by the National Key Research and Development Program (2016YFA0602402), National Basic Research Program (2015CB954102), Foundation for Innovative Research Groups of National Natural Science Foundation (41271286), and Outstanding Innovation Team in Colleges and Universities in Jiangsu Province of China.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Yue Yaojie https://orcid.org/0000-0001-5198-1281

REFERENCES

Afrasinei, G. M., Melis, M. T., Buttacu, C., Arras, C., Pistis, M., Zerrim, A., ... Ghiglieri, G. (2017). Euro-Mediterranean Journal for Environmental Integration, 2(1), 14. https://doi.org/10.1007/s41207-017-0021-1

Basso, F., Bove, E., Dumontet, S., Ferrara, A., Pisante, M., & Quaranta, G. (2000). Evaluating environmental sensitivity at the basin scale through the use of geographic information systems and remotely sensed data: an example covering the Agri basin (southern Italy). Catena, 40(1), 19-35. https://doi.org/10.1016/S0341-8162(99)00062-4

Becerril-Pina, R., Mastachi-Loza, C. A., Gonzalez-Sosa, E., Diaz-Delgado, C., & Ba, K. M. (2015). Assessing desertification risk in the semi-arid highlands of central Mexico. Journal of Arid Environments, 120, 4-13. https://doi.org/10.1016/j.jaridenv.2015.04.006

Bouabid, R. M., Rouchdi, M., Badraoui, A., Diab, S., & Louafi, S. (2010). Assessment of land desertification based on the MEDALUS approach and elaboration of an action plan: The case study of the Sous River Basin, Morocco. In P. Zdruli, M. Pagliai, S. Kapur, & A. Faz Cano (Eds.), Land degradation and desertification: Assessment, mitigation and remediation. Dordrecht: Springer. https://doi.org/10.1007/978-90-481-8657-0_10

Braun, A., Fernandez-Steeger, T., Havenith, H. B., & Torgoev, A. (2015). Landslide susceptibility mapping with data mining methods—A case study from Maily-Say, Kyrgyzstan. In G. Lolino, D. Giordan, G. B. Crosta, J. Corominas, R. Azzam, J. Wasowski, & N. Sica (Eds.), Engineering geology for society and territory (Vol. 2). Cham: Springer. https://doi.org/10.1007/978-3-319-09057-3_174

CGIAR-CSI. (2012). Global aridity and global PET—Methodology. http://www.cgiar-csi.org/wp-content/uploads/2012/11/GLOBAL-ARIDITY-AND-GLOBAL-PET-Methodology.pdf.

Chen, C., He, B., & Zeng, Z. (2014). A method for mineral prospectivity mapping integrating c4.5 decision tree, weights-of-evidence and m-branch smoothing techniques: a case study in the eastern Kunlun Mountains, China. Earth Science Informatics, 7(1), 13–24. https://doi.org/10.1007/s12145-013-0128-0

Chen, Y., & Tang, H. (2005). Desertification in north China: Background, anthropogenic impacts and failures in combating it. Land Degradation & Development, 16(4), 367–376. https://doi.org/10.1002/ldr.667

Collado, A. D., Chuvieco, E., & Camarasa, A. (2002). Satellite remote sensing analysis to monitor desertification processes in the cropp-rangeland boundary of Argentina. Journal of Arid Environments, 52(1), 121–133. https://doi.org/10.1016/j.jare.2001.0980

Colstoun, E. C. B. D., & Walthall, C. L. (2006). Improving global scale land cover classifications with multi-directional polder data and a decision tree classifier. Remote Sensing of Environment, 100(4), 474–485.

Daily, G. C. (1995). Restoring value to the world’s degraded lands. Science, 269(5222), 350–354. https://doi.org/10.1126/science.269.5222.350

Dasgupta, A., Sastry, K. L. N., Dhinwa, P. S., Rathore, V. S., & Nathawat, M. S. (2013). Identifying desertification risk areas using fuzzy membership and geospatial technique—A case study, Kota District, Rajasthan. Journal of Earth System Science, 122(4), 1107–1124. https://doi.org/10.1007/s12040-013-0331-x

De Jong, R., de Bruin, S., Schepman, M., & Dent, D. (2011). Quantitative mapping of global land degradation using Earth observations. International Journal of Remote Sensing, 32(21), 6823–6853. https://doi.org/10.1080/01431161.2010.512946

Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. Artificial Intelligence in Medicine, 34(2), 113–127. https://doi.org/10.1016/j.artmed.2004.07.002

D’Orsorio, P., Bhattachan, A., Davis, K. F., Ravi, S., & Runyan, C. W. (2013). Global desertification: Drivers and feedbacks. Advances in Water Resources, 51(1), 326–344. https://doi.org/10.1016/j.adwmat.2012.01.013

Domingos, P., & Pazzani, M. (1997). On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Machine Learning, 29(2-3), 103–130. https://doi.org/10.1023/A:1007413511361

Dong, Y. X. (1996). Study on the assessment of hazard degree and analysis of development trend of sandy desertification in China. Journal of Desert Research, 11(2), 127–131. (In Chinese with English abstract)

Du, Z., Xu, X., Zhang, H., Wu, Z., & Liu, Y. (2016). Geographical detector-based identification of the impact of major determinants on aeolian desertification risk. PLoS ONE, 11(3), e0151331. https://doi.org/10.1371/journal.pone.0151331

FAO. (1989). Arid Zone Forestry. A Guide for Field Technicians. Rome: Food and Agriculture Organization.

FAO/UNEP (1984). Provisional methodology for assessment and mapping of desertification. Rome: FAO, UNEP.

Fayyad, U., Platts-Kshapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI Magazine, 17(3), 37–54. https://doi.org/10.1609/aimag.v17i3.1230

Ferrara, A., Salvati, L., Sateriano, A., & Nole, A. (2012). Performance evaluation and cost assessment of a key indicator system to monitor desertification vulnerability. Ecological Indicators, 23, 123–129. https://doi.org/10.1016/j.ecolind.2012.03.015

Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H. T., Verelst, L., & Wiberg, D. (2008). Global agro-ecological zones assessment for agriculture (GAEZ 2008). Laxenburg: IIASA, Rome: FAO.
Wang, T. (2004). Progress in sandy desertification research of China. "Journal of Arid Environments," 43(4), 477–484. https://doi.org/10.1016/j.jar.1999.0568

UNEP (1992). World atlas of desertification. London: Edward Arnold.

Verón, S. R., Paruelo, J. M., & Oesterheld, M. (2006). Assessing desertification. "Journal of Arid Environments," 66(4), 751–763. https://doi.org/10.1016/j.jar.2006.01.021

Verstraete, M. M., Scholes, R. J., & Smith, M. S. (2009). Climate and desertification: Looking at an old problem through new lenses. "Frontiers in Ecology & the Environment," 7(8), 421–428. https://doi.org/10.1890/080119

Vorovencii, I. (2015). Assessing and monitoring the risk of desertification in Dobrogea, Romania, using Landsat data and decision tree classifier. "Environmental Monitoring & Assessment," 187(4), 204. https://doi.org/10.1007/s10661-015-4428-3

Wang, T. (2004). Progress in sandy desertification research of China. "Journal of Geographical Sciences," 14(4), 387–400. https://doi.org/10.1007/BF02837482

Wang, T. (2014). Atlas of sandy desert and aeolian desertification in Northern China. Beijing: Science Press.

Wang, X., Yang, Y., Dong, Z., & Zhang, C. (2009). Responses of dune activity and desertification in arid and semiarid Northern China. "Geomorphology," 86(3–4), 252–266. https://doi.org/10.1016/j.geomorph.2006.09.003

Wang, X., Yang, Y., Dong, Z., & Zhang, C. (2009). Responses of dune activity and desertification in China to global warming in the twenty-first century. "Global & Planetary Change," 67(3), 167–185. https://doi.org/10.1016/j.gloplacha.2009.02.004

Wang, T., & Zhu, Z. D. (2003). Study on Sandy Desertification in China—1. definition of sandy desertification and its connotation. "Journal of Desert Research," 23(3), 209–214. (In Chinese)

Wei, H., Li, J. L., & Liang, T. G. (2005). Study on the estimation of precipitation resources for rainwater harvesting agriculture in semi-arid land of China. "Agricultural Water Management," 71(1), 33–45. https://doi.org/10.1016/j.agwat.2004.07.002

Wessels, K. J., Prince, S. D., Frost, P. E., & van Zyl, D. (2004). Assessing the effects of human-induced land degradation in the former homelands of Northern South Africa with a 1 km AVHRR NDVI time-series. "Remote Sensing of Environment," 91(1), 47–67. https://doi.org/10.1016/j.rse.2004.02.005

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., & Zhou, Z. H. (2008). Top 10 algorithms in data mining. "Knowledge and Information Systems," 14(1), 1–37. https://doi.org/10.1007/s10115-007-0114-2

Xu, D., Kang, X., Qiu, D., Zhuang, D., & Pan, J. (2009). Quantitative assessment of desertification using Landsat data on a regional scale—A case study in the Ordos plateau, China. Sensors, 9(3), 1738–1753. https://doi.org/10.3390/s90301738

Yang, X. H., He, Q., Li, H. J., Hua, W., & Ding, G. F. (2012). Calculation and analysis on wind erosion climatic erosivity in the Tarim Basin. "Journal of Desert Research," 32(4), 990–995. (In Chinese with English abstract)

Yeon, Y. K., Han, J. G., & Ryu, K. H. (2010). Landslide susceptibility mapping in Injae, Korea, using a decision tree. "Engineering Geology," 116(3–4), 274–283. https://doi.org/10.1016/j.enggeo.2010.09.009

Yue, Y. J., Li, M., Zhu, A. X., Ye, X. Y., Mao, R., Wan, J. H., & Dong, J. (2016). Land degradation monitoring in the Ordos Plateau of China using an expert knowledge and BP-ANN-based approach. Sustainability, 8(11), 1174. https://doi.org/10.3390/su8111174

Yue, Y. J., Ye, X. Y., Zou, X. Y., Wang, J. A., & Gao, L. (2016). Research on land use optimization for reducing wind erosion in sandy desertified area: A case study of Yuyang County in Mu Us desert, China. "Stochastic Environmental Research & Risk Assessment," 31(6), 1371–1387. https://doi.org/10.1007/s00477-016-1223-9

Zhu, A. X., Gu, N. N., Liu, J., Qin, C. Z., & Zhou, C. H. (2018). Spatial prediction based on Third Law of Geography. "Annals of GIS," 24, 225–240. https://doi.org/10.1080/19475683.2018.1534890

Zhu, A. X., Zhang, G. M., Wang, W., Xiao, W., Huang, Z. P., Dunzhu, G. S., ... Yang, S. T. (2015). A citizen data-based approach to predictive mapping of spatial variation of natural phenomena. "International Journal of Geographical Information Science," 29(10), 1864–1886. https://doi.org/10.1080/13689929.2015.1058387

Zhu, Z. D., & Liu, S. (1984). The concept of desertification and the differentiation of its development. "Journal of Desert Research," 4(3), 2–8. (In Chinese with English abstract)

Zomer, R. J., Trabucco, A., Bossio, D. A., van Straaten, O., & Verchot, L. V. (2008). Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. "Agriculture, Ecosystems & Environment," 126, 67–80. https://doi.org/10.1016/j.agee.2008.01.014

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

---

**How to cite this article:** Yue Y, Min L, Lin W, A-Xing Z. A data-mining-based approach for aeolian desertification susceptibility assessment: A case-study from Northern China. "Land Degrad Dev." 2019;1–16. https://doi.org/10.1002/ldr.3393