Land subsidence spatial modeling and assessment of the contribution of geo-environmental factors to land subsidence: comparison of different novel ensemble modeling approaches

Alireza Arabameri¹, Peyman Yariyan², M. Santosh³,⁴,⁵

¹Department of Geomorphology, Tarbiat Modares University, Jalal Ale Ahmad Highway, Tehran 9821, Iran
²Department of Surveying Engineering, Saghez Branch, Islamic Azad University, Saghez 66819-73477, Iran;
P.Yariyan@iausaghez.ac.ir
³School of Earth Sciences and Resources, China University of Geosciences Beijing, Beijing, China
⁴Department of Earth Sciences, University of Adelaide, Adelaide, South Australia, Australia
⁵State Key Laboratory of Continental Dynamics, Department of Geology, Northwest University, Xi'an, China
Abstract: Land subsidence is a worldwide threat. In arid and semiarid land, groundwater depletion is the main factor that induce the subsidence and results in environmental damages, with high economic losses. To foresee and prevent the impact of land subsidence is necessary to develop accurated maps of the magnitude and evolution of the subsidences. Land subsidence susceptibility maps (LSSMs) provide one of the effective tools to manage vulnerable areas, and to reduce or prevent land subsidence. In this study, we used a new approach to improve Decision Stump Classification (DSC) performance and combine it with machine learning algorithms (MLAs) of Naive Bayes Tree (NBTree), J48 decision tree, alternating decision tree (ADTree), logistic model tree (LMT) and support vector machine (SVM) in land subsidence susceptibility mapping (LSSSM). We employ data from 94 subsidence locations, among which 70% were used to train learning hybrid models, and the other 30% were used for validation. In addition, the models’ performance was assessed by ROC-AUC, accuracy, sensitivity, specificity, odd ratio, root-mean-square error (RMSE), Kappa, frequency ratio and F-score techniques. A comparison of the results obtained from the different models, reveal that the new DSC-ADTree hybrid algorithm has the highest accuracy (AUC = 0.983) in preparing LSSSMs as compared to other learning models such as DSC-J48 (AUC = 0.976), DSC-NBTree (AUC = 0.959), DSC-LMT (AUC = 0.948), DSC-SVM (AUC = 0.939) and DSC (AUC = 0.911). The LSSSMs generated through the novel scientific approach presented in our study provide reliable tools for managing and reducing the risk of land subsidence.

Keywords: Land subsidence; Machine learning; Hybrid modeling; GIS; Iran.
1. Introduction

Land subsidence is widespread around the world. It affects many regions of the world (Hu et al., 2004; Bhattarai et al., 2017) and is well-known by the damage that cause in historical sites (Burbey, 2002; Teatini et al., 2005; Brambati et al., 2003). Land subsidence is a phenomenon that can occur gradually in some locations (Wang et al., 2009) or abruptly in some others (Wang et al., 2012; Caramanna et al., 2008) under the influence of diverse natural or human factors, including groundwater extraction and mining (Gutiérrez et al., 2014; Deng et al., 2017). Land subsidence and human activities are connected along the history and the present times (Rahmati et al., 2019).

Under arid and semiarid ecosystems, the main reason for the subsidence is the excessive use of groundwater (Pacheco et al., 2006). Also, the presence of faults and seams create permeable zones eventually leading to land subsidence susceptibility (LSS) (Perrin et al., 2015). Other factors, such as soil compaction, rock dissolution, and tectonic movements in large areas, will also influence LSS (Zhu et al., 2015). The extent of subsidence for every 10 m surface fall of groundwater is between 1 and 50 cm. Thus, the thickness and density of the layers can affect the intensity of these changes (Lofgren, 1969). Therefore, with the increase in the population rate, subsidence will gradually damage the infrastructure, gas and water transmission lines and of structures, railways and roads, leading to severe environmental economic problems (Abidin et al., 2013; Yin et al., 2016; Lee and Park, 2013). Most of the observations of the LSS phenomena around the world have been related to dry and low rainfall areas. This phenomenon has been observed in many parts of the world, such as Arizona and California in USA, the cities of Osaka and Tokyo in Japan, Venice in Italy, Bangkok in Thailand, and Mexico City (Larson et al., 2001).

About the methodology applied, it has been shown that a database of geographical information by combining different topographic, human and hydrological factors can aid in modelling the prediction of LSS (Park et al., 2012). This information would contribute to design a better management of the land subsidence with environmental programs to prevent destruction (Ghorbanzadeh et al., 2018). So far, no
consensus has been reached among researchers on choosing the best model for mapping the sensitivity of land subsidence. However, the use of spatial models in the GIS environment using different factors and parameters has provided relatively favorable results for modeling this phenomenon (Feizizadeh and Blaschke, 2013). There are various models for preparing LSSSMs (LSSSMs) (Navas et al., 2012). In recent years, various studies have been conducted to map the potential areas of LSS in terms of modeling, as well as to map the sensitivity of LSS based on the accuracy of the forecast (Ghorbanzadeh et al., 2018) assessed the LSS forecast by ANFIS, EBF (Pradhan et al., 2014), ANN (Li et al., 2009), SVM (Zhi-xiang et al., 2009), GM (Tang et al., 2008), etc., by GIS. Multi-criteria decision-making (MCDM) techniques have also been used to assess the risk of LSS (Bhattarai and Kondoh, 2017). But in these methods for decision making, a pairwise comparison matrix should be created based on the opinions of experts. This can be a problem for MCDM methods because uncertainty about determination of the importance of each factor will lead to a decrease in accuracy (Oh and Lee, 2011) produced an LSS risk map, using the LR and ANN methods. ANN need a lot of data to learn, which can be a problem for them to analyze (Biswa jee and Saro, 2007). Perrin et al., (2015) applied weight of evidence (WOE) for mapping the risk of land subsidence. In their study, after determining the relationship between subsidence factors and data, dependent and independent factors were combined. The WofE method, with the main assumptions available, is limited in terms of application. At best, the WofE index patterns predict the conditional independence (CI) of each factor according to the independent variables (Zhang and Agterberg, 2018). (Lyu et al., 2019) assessed the risk of urban infrastructure related to land subsidence and used a fuzzy analytical hierarchical process by designing questionnaires for both AHP and FAHP models. Their research results show that FAHP model is more accurate than AHP model. However, the interference of personal opinions by questionnaire in assessing natural hazards cannot provide reliable results for decision makers. Several authors (Calderhead et al., 2011; Deng et al., 2017; Zhang et al., 2019) analyzed the LSS forecast by InSAR technique. Although the results of the InSAR method carry good accuracy, the need for at least two pairs of SAR images in this technique is one of the notable limitations (Pradhan et al., 2014). In addition to the aforementioned studies, data-based models and MLA have recently been
used to address nonlinear problems in the LSS mapping (Oh et al., 2019; Tien Bui et al., 2018). Despite the use of some methods to prepare LSSM, it seems that by comparing the performance of different techniques, more accurate results can be achieved. Therefore, the group of combined learning models has been shown to possess the potential to increase the accuracy of the classification of individual models (Pham et al., 2017b).

Semnan sedimentary plain of Iran is highly prone to land subsidence. Nearly 25% of this basin has susceptibility to land subsidence. Inhabitants of this basin are suffering from the effects of land subsidence that are causing damages to housing, infrastructures, agricultural areas, and gardens. However, identifying the hazardous areas in this most extremely land subsidence affected area is most important for reducing the damages. In this research, we introduced five ensemble hybrid artificial intelligence approaches of Naïve Bayesian tree (NBTree), J48, ADTree, LMT and SVM. as a Meta classifiers based on decision stump classification model (DSC) as a base classifier called DSC-NBTree, DSC – J48, DSC – ADTree, DSC- LMT and DSC - SVM for land subsidence susceptibility mapping in the Semnan sedimentary plain, Semnan province, Iran. DSC benchmark model was used for comparison of the designed models. The difference between the present study and other related research is combining decision stump classification model (DSC) with five new hybrid MLAs such as Naïve Bayesian tree NBTree, J48, AD Tree, LMT and SVM. Also, another difference between this study and other studies is the use of effective factors in land subsidence, including the content of clay and the content of sand, which is one of the other innovations in this research.

2. Matrial and methods

2.1 Study area

In recent years, the Semnan plain in Iran has been facing severe drought and subsidence due to the excessive use of groundwater for agricultural purposes, and if the current trend continues, the reduction of groundwater levels in this region will emerge as a serious crisis. Since land improvement and restoration after subsidence is costly (Park et al., 2014), it is essential to adopt precautionary measures.
The Semnan city in Iran, with an area of about 22,120 Km², is characterized by dry and temperate climate with an average altitude of 1130 meters above sea level (Figure1). The city is located in the South of Alborz mountain range and north of Kavir plain with a longitude of 53° 11’ and a latitude of 35° 34’. The study area is the Semnan plain, which is located in the southern part to of the east of this city. Due to unfavorable natural conditions, lack of suitable water and soil, this area is uninhabited. Also, the scattering of precipitation has been reduced by less than 100 mm due to the impact of air currents in this area the plain of Semnan is relatively smooth, which is limited by several faults. Geologically, the northern part of Semnan plain exposes volcanic rock formations including rhyodacite and andesite, whereas the southern part is composed of marl, shale, sandstone and conglomerate (termed as Upper red Fm). According to hydrographic information, the groundwater level of Semnan plain shows a descending trend during the last 24-year period, indicating a decrease in groundwater resources. The northeastern part has the largest share of water level drop due to the high density of agricultural wells. The water drop decreases from east to west and to the south, thus increasing grain pressure and the density of the soil layers and eventually leading to permanent subsidence of the land.

2.2 **Methodology**

The methodology followed in the study includes six main stages (Figure 2). These involved preliminary data collection (previous land subsidence) through a sequence of field surveys with the support of high-resolution satellite images. After the LS inventory is entered, the next step was the preparation of the LSSCFs. The third step is to find out the correlation between the predictor variables. Following this, land subsidence susceptibility modeling is done applying the Decision Stump Classification (DSC) model and its ensemble with SVM, LMT, NBTree, J48, and ADTree approaches. Next, the validation is performed by applying ROC-AUC, accuracy, sensitivity, specificity, odd ratio, RMSE, Kappa, frequency ratio and F-score values. The sixth and concluding step is finalizing those models and their applicability in the study area. We used data from 94 land subsidence locations, among which 28 (30 %) were used for validation and the remaining 66 (70%) for training.
2.2.1 Data Acquisition

2.2.1.1 LSSIM

Preparation of LSSIM is necessary for hazard analysis (Guzzetti et al., 2000). The LSSIM (LSIM) of Semnan plain was prepared by the following sources: Interpretation of satellite images, LSSmap (Semnan Natural Resources Department), geological map with a scale of 1: 50,000 (Iranian Geological organization), and field impressions. Figure 1 shows the location of subsidence in the study area. Accordingly, the number (94) of available subsidence locations were randomly assigned to 70% (66 land subsidence) for training and 30% (28 land subsidence) for accreditation. The LSIM has been prepared by ArcGIS software with pixel size (12.5). Some of the land subsidences in the study area are shown in figure 3.

2.2.1.2 land subsidence conditioning factors (LSSCFs)

There are many types of LSSCFs that can affect its occurrence. Based on previous studies, and field studies in the study area, 12 LSSventilation agents were selected for this study, which include the following factors: content of clay, content of sand, groundwater withdraw, topographic wetness index (TWI), plan curvature, elevation, slope, distance to road (DR), distance to stream (DtS), drainage density (DD), lithology, LU/LC (LU/LC).

The above mentioned factors were prepared through digital elevation model (DEM) with 12.5 special resolution, geological maps, land use and satellite imagery. Maps of each layer have been prepared using ArcGIS 10.4.2 and SAGAGIS 3.2 software with a spatial resolution of 12.5 * 12.5 m for the entire study area (Table 1). Also, in order to analyze the occurrence of land subsidence, the layers are classified according to the Natural Break method (Figure 4).

Many land subsidences occur due to water leaking from the thin soils. One of the effective soil factors in LSSrelate to the engineering characteristics of clay and sediments (Ma et al., 2006). Clay due to its high contraction and impermeability has a high ability to retain a moisture for a long time (Gong et al.,
2009). Therefore, the thickness of the clay layer is directly related to the rate of subsidence of the land
(Changxing et al., 2007). Also the content of sand is one of the most effective factors in direct relation
with land subsidence. Thus, uncontrolled extraction of sand and gravel from an area will increase the
LSSintensity (Ma et al., 2019). The map of soil factors used in this study was prepared through data from
the Geological Survey of Iran with a scale of 1: 100000 (Figure 4 h-i). One of the hydrological factors
used in this study is groundwater withdraw. Excessive groundwater withdraw is one of the main causes of
subsidence (Wang et al., 2019). LSSoccurs mainly in areas where groundwater is extracted more than
they are fed during the year (Mahmoudpour et al., 2016). The groundwater withdraw map was prepared
using data from 500 wells and their harvest during the years 2000 to 2019 related to the Regional Water
Organization (RWO) of Semnan Province in five classes (Figure 4 j).

In addition, maps of topographic factors effective in subsidence such as topographic wetness index
(TWI), plan curvature, elevation and slope using DEM (PALSAR Satellite) with a resolution of 12.5 m
and geomorphology factors of DR, DtS and DD with euclidean distance interval and Kernel density in
ArcGIS 10.4.2 were also prepared (Figure 4 a-g). The type of lithology is also considered as one of the
effective geological factors in the occurrence of subsidence with digitization in ArGIS 10.4.2 with a scale
of 1: 100000 in 10 classes (Table 2). LU/LC map is classified into 4 groups through Landsat 8 OLI
information: Agriculture, Bareland, Rangeland, Residential area (Figure 4 k).

2.2.2 Land subsidence modeling (LSM)

2.2.2.1 Decision stump classifier (DSC)

Decision Stump Classification (DSC) is a type of MLA and is connected to other nodes (its leaves) as
a one-dimensional decision tree with an internal node (root) (Iba and Langley, 1992). The DSC algorithm
as a subsidiary classifier method only uses a specific attribute for segmentation (Chen et al., 2019). The
DSC is commonly used in other learning algorithms for classification and this type of classification can
be a subset for building a stronger classification (Jun et al., 2014). This algorithm is adjustable by the
following function:
\[ y(x, i, t) = 2I(x_i = t) - 1 \] 

where I is a function. If the value of z is true, I(Z) = 1, but if I(Z) = 0, x is the molecular fingerprint vector, also i is the value of the fingerprint and t is the target (Oliver and Hand, 1994).

2.2.2.2 Alternating decision tree (AD Tree)

The decision trees provide high speed and accuracy in forecasting and interpreting data set (Sikandar et al., 2018). ADTree is one of the algorithms that can be used with strong classification in data mining (Demirpolat and Das, 2019). The ADTree algorithm was introduced in 1999 by (Freund and Mason, 1999), and is one of the base classifiers that, in addition to providing classification results, can provide an assessment of the accuracy of the results (Hong et al., 2015). This algorithm consists of a series of decision in order to divide the predictive nodes in the educational data set (Chen et al., 2018). Assuming that the prediction nodes are divided, the following equation will occur:

\[ Z(c) = 2 \left( \sqrt{W_+(c)W_-(c)} + \sqrt{W_+(c)W_-(c)} \right) + W^\prime \]  

where \( W_+(c) \) and \( W_-(c) \) are a set of positive and negative weights that meet the demand for c, \( W^\prime \) is tuples’ weighted sum. By finding the lowest value of Z, can be achieved the split testing. It was also first introduced by Pfahringer (Pfahringer et al., 2001), \( Z_{\text{pure}} \) pruning technology in the optimal ADTree algorithm (Equation 3).

\[ Z_{\text{pure}} = 2 \left( \sqrt{W_+} + \sqrt{W_-} \right) + W^\prime \]  

where \( Z_{\text{pure}} \) is the lowest value of Z and is used to reduce the evaluation of predictive nodes.

2.2.2.3 Decision tree J48 (DTJ)

The DT J48 (DTJ) is a C4.5 decision tree open source program (Quinlan, 1993a) that can be provided using Waikato Environment for Knowledge Analysis (Hall et al., 2009). In the Decision tree J48
algorithm, prediction is done with several different conditions, and is a type of control phrase is in the programming language (Maulana and Defriani, 2020). Indeed DTJ is a type of forecasting method for modeling that can be used in statistical relations, data mining and MLA (Nguyen et al., 2020). In this technique, a feature in internal nodes will be used for segmentation (Bhargava et al., 2013). In the training phase, the J48 recognizes a new trait by a decision tree and creates a new sample with high accuracy for classification (Panigrahi and Borah, 2018).

2.2.2.4 Naïve Bayesian tree (NBTree)

The NBTree method was first proposed (Kohavi, 1996) as a hybrid algorithm, and is a type of classification algorithm used in data mining (Rahmati et al., 2019). The NBT model is very popular due to its simplicity in construction, short time to implement it, and use of low-ranking training data (Pham et al., 2017a) (Saha et al., 2020). Therefore, the first step in modeling in the NBT algorithm is tree growth based on entropy (degree of disorder) (Nhu et al., 2020b), such that, if S is set of training, and \(|S|\) is the total number of factors, they can be classified in \(n\) classes \(S_i (i = 1, 2, ..., n)\), \(|S_i|\) is a factor belong to classes \(S_i\). As a result, the expected classification can be calculated as follows:

\[
\text{Entropy} (S) = - \sum_{i=1}^{n} \left( \frac{|S_i|}{|S|} \right) \log_2 \left( \frac{|S_i|}{|S|} \right) \tag{4}
\]

in addition, if attribute \(A\) is considered in set \(S\), entropy is as follows:

\[
\text{Entropy}_A (S) = - \sum_{i=1}^{n} \left( \frac{|S_i|}{|S|} \right) \text{Info}(S_i) \tag{5}
\]

also information gain ratio (IGR) is used to show the difference between entropy (S) and \(\text{Entropy}_A (S)\):

\[
\text{Gain Ratio} (A) = \frac{\text{InfoGain} (A)}{\text{Split Info} (A)} = \frac{\text{Entropy} (S) - \text{Entropy}_A (S)}{- \sum_{i=1}^{n} \left( \frac{|S_i|}{|S|} \right) \log_2 \left( \frac{|S_i|}{|S|} \right)} \tag{6}
\]
finally, after assuming the independence of the attributes, the performance of the NBT algorithm is calculated as follows:

$$t_{NB} = \arg\max_{Z_i} PP(t_i) \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi} \epsilon} e^{-\frac{(t_i - \sigma)^2}{2\epsilon^2}}$$  (7)

where $PP(t_i)$ is the probability of output of the variables $t_i = (1,0)$, $\sigma$ and $\epsilon$ showed the mean and standard deviation of $r_1$, respectively (Murphy KP, 2006).

2.2.2.5 Logistic model tree (LMT)

LMT is a type of supervised classification model that uses the C4.5 algorithm to split (Quinlan, 1993b). The LMT algorithm combines LR and decision tree learning (Landwehr et al., 2005). In the logistics type, the LogitBoost algorithm is used to generate LR in each group, and then the segmentation begins with the C4.5 criterion (Sumner et al., 2005). Also the C4.5 algorithm uses the entropy technique to achieve optimal classification accuracy (Lim et al., 2000). In the LMT classification method for dividing the tree into nodes and leaves, the IGR technique will be formulated as follows:

$$\text{Gain Ratio} (A) = \frac{\text{gain} (A)}{\text{Split Info} (A)}$$  (8)

where gain (A) a type of experiment for classifying instructional examples, and Split Info (A) is the information generated when classifying instructional examples (Quinlan, 1993c).

2.2.2.6 SVM

SVM is a MLA based on the optimal separating hyper-plane theory (Cortes and Vapnik, 1995). The SVM training algorithm assigns new examples to other groups, and is able to create a kind of binary linear classification (Ben-Hur et al., 2001); (Suykens and Vandewalle, 1999). Also, SVMs have the ability to structural risk minimization (SRM), which reduces the problem of overuse (Seifi and Riahi, 2020). SVM creates an optimal cloud page by separating two classes (Figure 5) using Equation 9.
\[ Min_{\omega, b, \xi} : \frac{1}{2} \omega^T \omega + c \sum_{i=1}^{\xi_i} \]  

(9)

where \( \omega \) is a coefficient vector, \( b \) is the offset of the hyperplane, \( \xi_i \) is the variable positive and \( c \) determines the error parameters.

2.2.3 Model validation methods

2.2.3.1 Statistical validation methods

Statistical validation indicators have good accuracy for evaluating the results of learning models (Nhu et al., 2020a). Therefore, in this study, five statistical validation methods such as RSME, Accuracy (ACC), Sensitivity (SST), Specificity (SPF) and Kappa (K) were used to evaluate the accuracy of LSS maps. The relationships of these methods are summarized in Table 3.

| Outcome Status |
|----------------|
|                |

2.2.3.2 Odds ratio (OR)

The OR is a statistical indicator that determines the strength of the relationship between events A and B (Szumilas, 2010). The OR will be defined as the ratio of the odds of A in the presence and absence of B or the ratio of chance B in the presence and absence of A (Morris and Gardner, 1988). Therefore, if the OR is greater than 1, A and B are related, but if the OR is less than 1, A and B are negatively correlated, and the presence of one event reduces the chances of other event (Viera, 2008). Table 4 shows the values of odd ratios (OR):

| Outcome Status |
|----------------|
|                |
\[ OR = \frac{a/c}{b/d} = \frac{ad}{bc} \quad (10) \]

**2.2.3.3 Receiver operating characteristic (ROC)**

The ROC curve is the technique of displaying the true positive rate versus the false positive rate and a diagnostic test (Yariyan et al., 2020). The ROC curve evaluation for models is shown by the area under the curve (AUC) (Yariyan et al., 2019). The AUC has values between 0.5 and 1.0 (Chang et al., 2020).

**2.2.3.4 Seed cell area index (SCAI)**

The SCAI (Süzen and Doyuran, 2004) is actually the percentage of the area of each class of susceptibility to the percentage of points of occurrence in each class (Wang and Li, 2017). In general, low SCAI values for high and very high-susceptibility classes and high values in low and very low-susceptibility classes indicate the accuracy of the maps produced (Pourghasemi et al., 2014). The performance evaluation of the models in the SCAI index is calculated as follows:

\[ SCAI = \frac{\text{Aerial extent of susceptibility classes} \, (\%)}{\text{Inventory of training and testing set in each class} \, (\%)} \quad (11) \]

**2.2.3.5 FR**
The FR determines the small relationship between the occurrence of LSS and the various factors influencing it (Pradhan et al., 2014). FR was used to determine the quantitative relationship between subsidence positions and susceptibility classes. The FR values will be calculated through the following equation:

\[ FR_i = \left( \frac{A_i}{B_i} \right) / \left( \frac{H_i}{L} \right) = \frac{P_i}{K} \]  

where \( FR_i \) is frequency ratio value of each susceptibility class of the models, \( A_i \) is area of a model class, \( B_i \) is total area of the map of each model, \( H_i \) is number of pixels in each of the susceptibility classes, \( L \) total number of pixels in each map, \( P_i \) is percentage of area according to the sensitivity class for each model and \( K \) is the percentage for the entire domain.

### 2.2.3.6 F-Score

The F-score statistical index in statistical analysis is a measure of the accuracy of a test (Sasaki, 2007). In this index, the F value of the accuracy measurement score is related to a test. In this index, the F value of the accuracy measurement score is related to a test and commonly used in information retrieval and classification performance (Derczynski, 2016). This score is based on the following will be calculated relationship:

\[ F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision}^{-1} + \text{recall}^{-1}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \]  

where the value of the score of \( F_1 \) is the harmonic mean of precision and recall (Chicco, 2020).

### 3. Results

#### 3.1 frequency ratio analysis

To determine the relationship between the location of the subsidence and the conditioning factors, the frequency ratio (FR) method has been used. The calculated FR values for each class of factors are
presented in Figure 6. In the case of the lithology layer with 10 classes, the highest FR values were calculated in Qcf group (2.50) and followed by the Mur (0.48) and Qft2 (0.26) groups. Also, the value of FR in other groups is zero, which indicates that they do not affect the subsidence compared to the previous three groups. The Agriculture class had the highest FR (3.90) at LU / LC factor, followed by Rangeland (0.52) and Bare land (0.13), have the highest value, respectively. In the case of the Groundwater withdraw layer, class 0 <28.72 has the highest FR value (2.18). Accordingly, the lower the groundwater exploitation, the lower the FR. So that the lowest FR value is related to class 8.95 (0.35). Content of sand soil layer with 5 classes from 0.61% to 0.49%, has the highest FR calculated in class 0.49% (1.92). In addition, as in the previous layer, the FR value increases with increasing clay percentage; thus, the highest FR value is in class 20% (>4.75) and the lowest in class 14-16% (0.24). In the case of DR, high FR values are associated with the greatest distance. For example, class 1057-2320m has a value of FR 1.27. For DD, the highest FR values were calculated in class 1.31-1.98 (2.17) and the lowest values were calculated in class 0.59 (>0.08). Areas with a distance of 100-500m from the stream have higher FR values. So that, 100-200 m class with FR (1.47) is the most effective class. For topographic factors, high FR values were calculated in TWI layers (class 7.14-1.75 = 1.21), plan curvature layer (class Flat = 1.09), slope layer (class 2 = >1.19) and elevation (class 1031-1089m = 3.01) were calculated. Hence, there is a special relationship between classes of slope and elevation for FR values, so that as elevation and slope increase, FR decreases.

3.2 Multi-collinearity analysis

In order to ensure the independent effect of each of the conditioning factors, multicollinearity testing is required (Chen et al., 2020). If there is collinearity between the factors, the accuracy of the model prediction decreases (Arabameri et al., 2018). In this study, tolerance (TOL) and variance inflation factor (VIF) indices have been used for to perform multicollinearity testing. If VIF <5 or 10 and TOL >0.1 or 0.2, then there is no-collinearity between the conditioning factors (Roy et al., 2019). The highest value of
VIF is 4.395 and the lowest value of TOL is 0.228 (Table 4). However, the results show that there is no-
colinearity between the factors and they can all be used to modeling the susceptibility of land subsidence.

3.3 Importance of LSCF by Ad boost model

Adaptive Boosting (Ad boost) is a type of MLA (Schapire, 2003), and has been used to determine the
relative importance of the factors influencing LSS (Figure 7). Based on the results, it can be seen that the
groundwater withdrawal is the most important factor for LSS because it has the highest score (0.095)
compared to other factors. Following that, the following factors are of the highest importance,
respectively:

LU/LC (Score = 0.093), lithology (Score = 0.09), elevation (Score = 0.088), DtS (Score = 0.0.086),
DD (Score = 0.0.084), content of sand (Score = 0.076), content of clay (Score = 0.073), DR (Score =
0.071), slope (Score = 0.069), TWI (Score = 0.067) and plan curvature (Score = 0.063).

4. Discussion

4.1 Land subsidence susceptibility modeling (LSSM)

The LSSSMs were been prepared using Decision stump classifier model (DSC) and SVM, LMT,
NBTree, J48 and ADTree hybrid MLA in R software. In order to prepare LSSSMs (LSSM), subsidence
ventilation factors were used as a classification source for combined learning models. The classification
of DSC individual model and DSC-SVM, DSC-LMT, DSC-NBTree, DSC-J48 and DSC-ADTree hybrid
models group varies from 0 to 1. Thus, the pixels of the images associated with each map are shown with
two values of 0 and 1, which 0 indicates stable conditions and 1 indicates the high probability of land
subsidence. After that, six LSSSMs (LSSMs) were classified in the ArcGIS10.4.2 software environment
into five categories of subsidence sensitivity: very low, low, moderate, high and very high (Figure 8 a-f).
The spatial relationship of the subsidence location with the sensitivity maps is shown in Table 5.
Therefore, the correlation between the sensitivity classes of the combined and individual models group
with the position of the subsidence in the very high class for DSC-NBTree models (85.11%), DSC-LMT
(81.91%), DSC-AD Tree (80.55%), DSC (% 78.72), DSC-J48 (77.66) and DSC-SVM (77.66) indicate an
appeal table spatial relationship.

4.2 Validation and comparison of the models

The performance of combined and individual learning models was evaluated using nine popular
techniques: AUC, ACC, SST, SPF, odd ratio, RMSE, K, FR and F-score. First, AUC values were
obtained using the ROC curve for the training and validation data set (Figure 9). Accordingly, in the
training group, the highest AUC values in DSC-ADTree model (AUC = 0.950), DSC-J48 (AUC = 0.948),
DSC-NBTree (AUC = 0.928), DSC-LMT (AUC = 0.922), DSC-SVM (AUC = 0.9) and DSC (AUC =
0.902) are obtained, respectively (Figure 10). Also in the validation group, the combined model DSC-
ADTree with AUC = 0.983 has the highest accuracy and then other hybrid and individual models DSC-
J48 (AUC = 0.976), DSC-NBTree (AUC = 0.959), DSC-LMT (AUC = 0.948) DSC-SVM (AUC = 0.939)
and DCS (AUC = 0.911) have good results. In addition, the values obtained for the kappa and accuracy
indicators also indicate the high accuracy of the DSC-ADTree hybrid model (ACC = 0.93, K = 0.86)
compared to other models (Figure 10). Also, the results of the evaluation of four statistical techniques of
SST, SPF, F-score and RMSE for DSC-ADTree models (SST = 0.92, SPF = 0.94, F-score = 0.93, RMSE
= 0.39), DSC-J48 (SST = 0.91, SPF = 0.88, F-score = 0.90, RMSE = 0.4), DSC-LMT (SST = 0.86, SPF =
0.85, F-score = 0.86, RMSE = 0.44), DSC-NBTree (SST = 0.86, SPF = 0.88, F-score = 0.87, RMSE =
0.43), DSC-SVM (SST = 0.82, SPF = 0.83, F-score = 0.82, RMSE = 0.46) and DSC (SST = 0.82, SPF =
0.80, F-score = 0.81, RMSE = 0.49). These results indicate the high accuracy of the DSC-ADTree model
compared to the other models (Figure 10). However, other models also have acceptable results for
mapping susceptibility to land subsidance.

The odd ratio (OR) technique was used to compare the relative chance of subsidence in the models
used in the study. Therefore, the ratio between the correct and incorrect classified values for both training
and validation groups was calculated (Figure 11). Based on the results of odd ratios for the training group,
DSC-ADTree model (OR = 109.10), DSC-J48 (OR = 72.50), DSC-NBTree (OR = 45.92), DSC-LMT
(OR = 35.47), DSC - SVM (OR = 22.50) and DSC (OR = 18.35) and validation group DSC-ADTree (OR = 129), DSC-J48 (OR = 78), DSC-NBTree (OR = 27.60), DSC-LMT (OR = 16.87), DSC-SVM (OR = 7.50) and DSC (OR = 5.28) respectively have higher values of OR (higher chance of subsidence) and higher accuracy.

To evaluate the accuracy of classification and comparison of sensitivity classes to LSS models, the FR and SCAI methods were used. Based on the logical relationship between subsidence levels and sensitivity classes, for very low to high susceptibility classes, the FR values increase and SCAI decreases. The results of model evaluation using SCAI and FR indicators are shown in Figure 12. In the whole group of models, with increasing susceptibility from very low to very high, the FR value of the upward trend and the SCAI index decreases. This can indicate a significant correlation between the susceptibility classes and the location of the subsidence. The high values of FR index in very high susceptibility classes for DSC-ADTree (FR = 7.80), DSC-J48 (FR = 7.20), DSC-NBTree (FR = 5.78), DSC-LMT (FR = 7.35), DSC-SVM (FR = 6.35) and DSC (FR = 6.50) show a strong correlation of DSC-ADTree model with subsidence zones (Figure 12).

4.2 Implications

Creating a reliable map of LSS susceptibility has remained as a challenge for landuse planning. Many researchers have proposed various models to address this challenge, but there is no consensus among them on a specific model. In this study, we proposed and evaluated five new combined learning model groups based on the Decision Stump Classification (DSC) model for mapping LSS susceptibility. First, based on the spatial relationship between the location of the subsidence and the conditioning factors, the modeling process was determined and the group of combined learning models DSC-SVM, DSC-LMT, DSC-NBTree, DSC-J48 and DSC-ADTree based on DSC model were developed. Using analyzes, we showed that factors such as Groundwater withdraw, LU / LC, lithology, elevation, distance to flow, DD, content of sand and content of clay had the greatest impact on LSS in the study area, respectively. In contrast, DR, slope, TWI, and plan curvature, have less effect as compared to the other factors. Also,
groundwater withdrawal as introduced in previous studies (Chaussard et al., 2014; Arabameri et al., 2020b) was confirmed as an important factor in our study. The hybrid techniques used in this study (DSC-SVM, DSC-LMT, DSC-NBTree, DSC-J48 and DSC-ADTree) improve the performance of the DSC model. Among this group of hybrid models, DSC-ADTree accuracy better than the other hybrid groups in both training phase (AUC = 0.950) and validation (AUC = 0.983). In addition, the DSC-ADTree group learning technique works better than other techniques in reducing RMSE (0.39). The superiority of the ADTree learning technique in combination with DSC and the improvement of the modeling process are among the characteristics of this model (Wu et al., 2020); (Arabameri et al., 2020a).

Although for the first time in our study Decision Stump Classification (DSC) was used in combination with group learning techniques to model land subsidence, the approach of using DSC as a basic model to combine with learning algorithms, as previously illustrated (Pham et al., 2019). In their study, the Rotation Forest based Decision Stump model provided better results in modeling. The similarity of their study with that of the present one is the optimal performance of the DSC model as a basis for learning techniques. Comparing the results of this study with the results from other studies is difficult, because the type of factors and study, educational data, regional conditions, etc. are among the factors that can directly affect the results. However, the results of our study show a high accuracy of the group of combined models for land subsidence. Based on the different validation methods, all combined models and individual DSC model are shown to achieve good results for LSSmapping. The results of this study show that most of the subsidence is recorded in areas with altitudes of 1031 to 1089 and slope of less than 8 degrees. Also, the LSSrisk map based on the optimal model (DSC-ADTree) shows that 15% of the study area is in a high and very high risk range, which indicates a critical situation. In general, despite agricultural lands and excessive use of groundwater resources, drought and numerous wells for agricultural use has caused the discharge of groundwater inventory and its failure to compensate.

Our research is of great help to achieve the Sustainable Development Goals of the United Nations and prevent the degradation of the natural resources. Following the State-of-the-Art by Keesstra et al., (2016)
and Visser et al., (2019) the contribution of a circular economy and the rational use of the resources in the Planet will avoid the desertification of the land. It is relevant that in semiarid climatic conditions new managements will contribute to achieve the sustainability and will contribute to achieve the Land Degradation Neutrality challenge (Keesstra et al., 2018).

5 Conclusions

Land subsidence is a land degradation process when induced by human misuse, and an economical threat. Hence, identification, modeling, evaluation and analysis are of great importance for effective land management. MLA based on data mining have recently been able to provide good results for modeling and mapping environmental hazards. In the present study, the subsidence of Semnan plain in Iran was evaluated using a new modeling approach (DSC-SVM, DSC-LMT, DSC-NBTree, DSC-J48 and DSC-ADTree). The results show the high accuracy of hybrid learning algorithms for preparing LSSSMs (LSSMs). Our study shows that the new DSC-ADTree hybrid model, as the optimal model in this study, can help managers and planners in assessing LSS risk, and to adopt necessary measures.

Finding: The authors did not receive support from any organization for the submitted work.

Conflict of interest: All the authors declare that there is no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Code availability: The models and code that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contribution: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Alireza Arabameri]. The first draft of the manuscript was
written by [Peyman Yariyan], [M. Santosh] and [Alireza Arabameri] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Consent to participate: All authors are consent.

Consent for publication: All authors are consent.

References

Abidin, H., Andreas, H., Gumilar, I., Sidiq, T.P., Fukuda, Y., 2013. Land subsidence in coastal city of Semarang (Indonesia): characteristics, impacts and causes. Geomatics, Natural Hazards and Risk, 4(3): 226-240.

Arabameri, A., Pradhan, B., Rezaei, K., Yamani,M., Pourghasemi, H.R., Lombardo, L., 2018. Spatial modelling of gully erosion using evidential belief function, logistic regression, and a new ensemble of evidential belief function–logistic regression algorithm. Land Degradation & Development, 29(11): 4035-4049.

Arabameri, A., Chen, W.; Blaschke, T., Tiefenbacher, J.P., Pradhan, B., Tien Bui, D., 2020a. Gully Head-Cut Distribution Modeling Using Machine Learning Methods—A Case Study of NW Iran. Water, 12(1): 16.

Arabameri., Saha, S., Roy, J., Tiefenbacher, J., Cerda, A., Biggs, T., Pradhan, B., ThaoThi Ngo, P., Collin, A., 2020b. A novel ensemble computational intelligence approach for the spatial prediction of land subsidence susceptibility. Science of The Total Environment: 138595.

Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V., 2001. Support vector clustering. Journal of machine learning research, 2(Dec): 125-137.

Bhargava, N., Sharma, G., Bhargava, R., Mathuria, M., 2013. Decision tree analysis on j48 algorithm for data mining. Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering, 3(6).
Bhattarai, R., Alifu, H., Maitiniyazi, A., Kondoh, A., 2017. Detection of land subsidence in Kathmandu Valley, Nepal, using DInSAR technique. Land, 6(2): 39.

Bhattarai, R., Kondoh, A., 2017. Risk Assessment of Land Subsidence in Kathmandu Valley, Nepal, Using Remote Sensing and GIS. Advances in Remote Sensing, 6(02): 132.

Biswaieet, P., Saro, L., 2007. Utilization of optical remote sensing data and GIS tools for regional landslide hazard analysis using an artificial neural network model. Earth Science Frontiers, 14(6): 143-151.

Brambati, A., Carbognin, L., Quaia, T., Teatini, P., Tosi, L., 2003. The Lagoon of Venice: geological setting, evolution and land subsidence. Episodes, 26(3): 264-268.

Burbey, T.J., 2002. The influence of faults in basin-fill deposits on land subsidence, Las Vegas Valley, Nevada, USA. Hydrogeology Journal, 10(5): 525-538.

Calderhead, A., Therrien, R., Rivera, A., Martel, R., Garfias, J., 2011. Simulating pumping-induced regional land subsidence with the use of InSAR and field data in the Toluca Valley, Mexico. Advances in Water Resources, 34(1): 83-97.

Caramanna, G., Ciotoli, G., Nisio, S., 2008. A review of natural sinkhole phenomena in Italian plain areas. Natural hazards, 45(2): 145-172.

Chang, Z., Du, Z., Zhang, F.; Huang, F., Chen, J., Li, W., Guo, Z., 2020. Landslide susceptibility prediction based on remote sensing images and gis: Comparisons of supervised and unsupervised machine learning models. Remote Sensing, 12(3): 502.

Changxing, S., Dian, Z., Lianyuan, Y., Bingyuan, L., Zulu, Z., Ouyang, Z., 2007. Land subsidence as a result of sediment consolidation in the Yellow River Delta. Journal of Coastal Research, 2007(231): 173-181.

Chaussard, E., Wdowinski, S., Cabral-Cano, E., Amelung, F., 2014. Land subsidence in central Mexico detected by ALOS InSAR time-series. Remote sensing of environment, 140: 94-106.
Chen, S., Shen, B., Wang, X., Yoo, S.-J., 2019. A Strong Machine Learning Classifier and Decision Stumps Based Hybrid AdaBoost Classification Algorithm for Cognitive Radios. Sensors, 19(23): 5077.

Chen, W., Fan, L., Li, C., Pham, B.T., 2020. Spatial prediction of landslides using hybrid integration of artificial intelligence algorithms with frequency ratio and index of entropy in nanzheng county, china. Applied Sciences, 10(1): 29.

Chen, W., Zhang, S., Li, R., Shahabi, H., 2018. Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. Science of the total environment, 644: 1006-1018.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine learning, 20(3): 273-297.

Demirpolat, A.B., Das, M., 2019. Prediction of viscosity values of nanofluids at different pH values by alternating decision tree and multilayer perceptron methods. Applied Sciences, 9(7): 1288.

Deng, Z., Ke, Y., Gong, H., Li, X., Li, Z., 2017. Land subsidence prediction in Beijing based on PS-InSAR technique and improved Grey-Markov model. GIScience & Remote Sensing, 54(6): 797-818.

Derczynski, L., 2016. Complementarity, F-score, and NLP Evaluation, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16), pp. 261-266.

Feizizadeh, B., Blaschke, T., 2013. Land suitability analysis for Tabriz County, Iran: a multi-criteria evaluation approach using GIS. Journal of Environmental Planning and Management, 56(1): 1-23.

Freund, Y., Mason, L., 1999. The alternating decision tree learning algorithm, icml, pp. 124-133.

Ghorbanzadeh, O., Blaschke, T., Aryal, J., Gholaminia, K., 2018. A new GIS-based technique using an adaptive neuro-fuzzy inference system for land subsidence susceptibility mapping. Journal of Spatial Science: 1-17.

Gong, S.L., Li, C., Yang, S.L., 2009. The microscopic characteristics of Shanghai soft clay and its effect on soil body deformation and land subsidence. Environmental Geology, 56(6): 1051-1056.
Gutiérrez, F., Parise, M., De Waele, J., Jourde, H., 2014. A review on natural and human-induced geohazards and impacts in karst. Earth-Science Reviews, 138: 61-88.

Hall, M., Frank, E., Holmes, G., Pfahringer Peter Reutemann, B., Witten, I., 2009. The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1): 10-18.

Hong, H., Pradhan, B., Xu, C., Bui, D.T., 2015. Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. Catena, 133: 266-281.

Hu, R., Yue, Z., Wang, L.u., Wang, S., 2004. Review on current status and challenging issues of land subsidence in China. Engineering Geology, 76(1-2): 65-77.

Iba, W., Langley, P., 1992. Induction of one-level decision trees, Machine Learning Proceedings 1992. Elsevier, pp. 233-240.

Jun, X., Lu, Y., Lei, Z., Hui, X., 2014. Boosting decision stumps to do pairwise classification. Electronics letters, 50(12): 866-868.

Keesstra, S. D., Bouma, J., Wallinga, J., Tittonell, P., Smith, P., & Bardgett, R. D. 2016. The significance of soils and soil science towards realization of the United Nations Sustainable Development Goals. Soil, 2, 111-128

Keesstra, S., Mol, G., de Leeuw, J., Okx, J., de Cleen, M., Visser, S. 2018. Soil-related sustainable development goals: Four concepts to make land degradation neutrality and restoration work. Land, 7(4), 133.

Kohavi, R., 1996. Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid, Kdd, pp. 202-207.

Landwehr, N., Hall, M., Frank, E., 2005. Logistic model trees. Machine learning, 59(1-2): 161-205.

Larson, K., Başağaoğlu, H., Marino, M., 2001. Prediction of optimal safe ground water yield and land subsidence in the Los Banos-Kettleman City area, California, using a calibrated numerical simulation model. Journal of hydrology, 242(1-2): 79-102.
Lee, S., Park, I., 2013. Application of decision tree model for the ground subsidence hazard mapping near abandoned underground coal mines. Journal of environmental management, 127: 166-176.

Li, H., Zhao, X., Chi, H., ZHANG, J.-j., 2009. Prediction and analysis of land subsidence based on improved BP neural network model. Journal of Tianjin University, 1(42): 60-64.

Lim, T.-S., Loh, W.-Y., Shih, Y.-S., 2000. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Machine learning, 40(3): 203-228.

Lofgren, B.E., 1969. Field Measurement of Aquifer-system Compaction, San Joaquin Valley, California, USA. US Geological Survey.

Lyu, H.-M., Shen, S.-L., Zhou, A., Yang, J., 2019. Risk assessment of mega-city infrastructures related to land subsidence using improved trapezoidal FAHP. Science of The Total Environment: 135310.

Ma, K., Zhang, Y., Ruan, M., Guo, J., Chai, T., 2019. Land Subsidence in a Coal Mining Area Reduced Soil Fertility and Led to Soil Degradation in Arid and Semi-Arid Regions. International Journal of Environmental Research and Public Health, 16(20): 3929.

Ma, R., Wang, Y., Ma, T., Sun, Z., Yan, S., 2006. The effect of stratigraphic heterogeneity on areal distribution of land subsidence at Taiyuan, northern China. Environmental geology, 50(4): 551-568.

Mahmoudpour, M., Khamehchiyan, M., Nikudel, M.R., Ghassemi, M.R., 2016. Numerical simulation and prediction of regional land subsidence caused by groundwater exploitation in the southwest plain of Tehran, Iran. Engineering Geology, 201: 6-28.

Maulana, M.F., Defriani, M., 2020. Logistic Model Tree and Decision Tree J48 Algorithms for Predicting the Length of Study Period. PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic, 8(1): 39-48.

Morris, J.A., Gardner, M.J., 1988. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. British Medical Journal (Clinical Research Edition), 296(6632): 1313-1316.
Navas, J.M., Telfer, T.C., Ross, L.G., 2012. Separability indexes and accuracy of neuro-fuzzy classification in geographic information systems for assessment of coastal environmental vulnerability. Ecological informatics, 12: 43-49.

Nguyen, V.-N., Yariyan, P., Amiri, M., Dang Tran, A., Pham, T.D., Do, M.P., Thi Ngo, P.T., Nhu, V.-H., Quoc Long, N., Tien Bui, D. A., 2020. A New Modeling Approach for Spatial Prediction of Flash Flood with Biogeography Optimized CHAID Tree Ensemble and Remote Sensing Data. Remote Sensing, 12(9): 1373.

Nguyen, V.-N., Yariyan, P., Amiri, M., Dang Tran, A., Pham, T.D., Do, M.P., Thi Ngo, P.T., Nhu, V.-H., Quoc Long, N., Tien Bui, D. A., 2020. A New Modeling Approach for Spatial Prediction of Flash Flood with Biogeography Optimized CHAID Tree Ensemble and Remote Sensing Data. Remote Sensing, 12(9): 1373.

Nhu, V.-H., Janizadeh, S., Avand, M., Chen, W., Farzin, M., Omidvar, E., Shirzadi, A., Shahabi, H. J., Clague, J., Jaafari, A., Mansoorypoor, F., Thai Pham, B., Ahmad, B.B., Lee, S., 2020a. GIS-based gully erosion susceptibility mapping: A comparison of computational ensemble data mining models. Applied Sciences, 10(6): 2039.

Nhu, V.-H., Shirzadi, A., Shahabi, H., Singh, S.K., Al-Ansari, N., Clague, J.J., Jaafari, A., Chen, W., Miraki, S., Dou, J., Luu, C., Górski, K., Thai Pham, B., Nguyen, H.D., Ahmad, B.B., 2020b. Shallow Landslide Susceptibility Mapping: A Comparison between Logistic Model Tree, Logistic Regression, Naïve Bayes Tree, Artificial Neural Network, and Support Vector Machine Algorithms. International Journal of Environmental Research and Public Health, 17(8): 2749.

Oh, H.-J., Lee, S., 2011. Integration of ground subsidence hazard maps of abandoned coal mines in Samcheok, Korea. International Journal of Coal Geology, 86(1): 58-72.

Oh, H.-J., Syifa, M., Lee, C.-W., Lee, S., 2019. Land subsidence susceptibility mapping using bayesian, functional, and meta-ensemble machine learning models. Applied Sciences, 9(6): 1248.

Oliver, J., Hand, D., 1994. David, “Averaging over decision stumps, in machine learning” ECML-94, European Conference on Machine Learning, Catania, Italy, pp. 231-241.

Pacheco, J. Arzate, J., Rojas, E., Arroyo, M., Yutsis, V., Ochoa, G., 2006. Delimitation of ground failure zones due to land subsidence using gravity data and finite element modeling in the Querétaro valley, México. Engineering Geology, 84(3-4): 143-160.
Panigrahi, R., Borah, S., 2018. Rank allocation to J48 group of decision tree classifiers using binary and multiclass intrusion detection datasets. Procedia computer science, 132: 323-332.

Park, I., Lee, J., Saro, L., 2014. Ensemble of ground subsidence hazard maps using fuzzy logic. Central European Journal of Geosciences, 6(2): 207-218.

Perrin, J., Cartannaz, C., Noury, G., Vanoudheusden, E., 2015. A multicriteria approach to karst subsidence hazard mapping supported by weights-of-evidence analysis. Engineering Geology, 197: 296-305.

Pfarhringer, B., Holmes, G., Kirkby, R., 2001. Optimizing the induction of alternating decision trees, Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, pp. 477-487.

Pham, B.T., Jaafari, A., Prakash, I., Singh, S., Quoc, N.K., Tien Bui, D., 2019. Hybrid computational intelligence models for groundwater potential mapping. Catena, 182: 104101.

Pham, B.T., Bui, D.T., Pourghasemi, H.R., Indra, P., Dholakia, M., 2017a. Landslide susceptibility assessment in the Uttarakhand area (India) using GIS: a comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. Theoretical and Applied Climatology, 128(1-2): 255-273.

Pham, B.T., Bui, D.T., Prakash, I., Dholakia, M., 2017b. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. Catena, 149: 52-63.

Pourghasemi, H., Moradi, H., Aghda, S.F., Gokceoglu, C., Pradhan, B., 2014. GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran). Arabian Journal of Geosciences, 7(5): 1857-1878.

Pradhan, B., Abokharama, M.H., Jebur, M.N., Tehrany, M.S., 2014. Land subsidence susceptibility mapping at Kinta Valley (Malaysia) using the evidential belief function model in GIS. Natural hazards, 73(2): 1019-1042.

Quinlan, J., 1993a. The Morgan Kaufmann series in machine learning. San Mateo.

Quinlan, J.R., 1993b. C4.5: Programming for machine learning. Morgan Kauffmann, 38: 48.
Quinlan, J.R., 1993c. The Morgan Kaufmann series in machine learning.

Rahmati, O., Ghorbanzadeh, O., Teimurian, T., Mohammadi, F., Tiefenbacher, J.P., Falah, F., Pirasteh, S., Ngo, P.-T.T., Bui, D.T., 2019. Spatial Modeling of Snow Avalanche Using Machine Learning Models and Geo-Environmental Factors: Comparison of Effectiveness in Two Mountain Regions. Remote Sensing, 11(24): 2995.

Rahmati, O., Golkarian, A., Biggs, T., Keesstra, S., Mohammadi, F., & Daliakopoulos, I. N. (2019). Land subsidence hazard modeling: Machine learning to identify predictors and the role of human activities. Journal of environmental management, 236, 466-480.

Saha, S., Roy, J., Arabameri, A., Blaschke, T., Tien Bui, D., 2020. Machine Learning-Based Gully Erosion Susceptibility Mapping: A Case Study of Eastern India. Sensors, 20(5): 1313.

Sasaki, Y., 2007. The Truth of the F-Measure. 2007.

Schapire, R.E., 2003. The boosting approach to machine learning: An overview, Nonlinear estimation and classification. Springer, pp. 149-171.

Seifi, A., Riahi, H., 2020. Estimating daily reference evapotranspiration using hybrid gamma test-least square support vector machine, gamma test-ANN, and gamma test-ANFIS models in an arid area of Iran. Journal of Water and Climate Change, 11(1): 217-240.

Sikandar, A., Anwar, W., Bajwa, U.I., Wang, X., Sikandar, M., Yao, L., 2018. Decision tree based approaches for detecting protein complex in protein protein interaction network (PPI) via link and sequence analysis. IEEE Access, 6: 22108-22120.

Sumner, M., Frank, E., Hall, M., 2005. Speeding up logistic model tree induction, European conference on principles of data mining and knowledge discovery. Springer, pp. 675-683.

Suykens, J.A., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural processing letters, 9(3): 293-300.

Süzen, M.L., Doyuran, V., 2004. A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. Environmental geology, 45(5): 665-679.
Szumilas, M., 2010. Explaining odds ratios. Journal of the Canadian academy of child and adolescent psychiatry, 19(3): 227.

Tang, Y.-Q., Cui, Z.-D., Wang, J.-X., Yan, L.-P., Yan, X.-X., 2008. Application of grey theory-based model to prediction of land subsidence due to engineering environment in Shanghai. Environmental Geology, 55(3): 583-593.

Teatini, P., Ferronato, M., Gambolati, G., Bertoni, W., Gonella, M., 2005. A century of land subsidence in Ravenna, Italy. Environmental Geology, 47(6): 831-846.

Tien Bui, D., Shahabi, H., Shirzadi, A., Chapi, K., Pradhan, B., Chen, W., Khosravi, K., Panahi, M., Bin Ahmad, B., Saro, L., 2018. Land subsidence susceptibility mapping in south korea using machine learning algorithms. Sensors, 18(8): 2464.

Viera, A.J., 2008. Odds ratios and risk ratios: what's the difference and why does it matter? Southern medical journal, 101(7): 730-734.

Visser, S., Keesstra, S., Maas, G., De Cleen, M. 2019. Soil as a Basis to Create Enabling Conditions for Transitions Towards Sustainable Land Management as a Key to Achieve the SDGs by 2030. Sustainability, 11(23), 6792.

Wang, G., You, G., Shi, B., Yu, J., Tuck, M., 2009. Long-term land subsidence and strata compression in Changzhou, China. Engineering Geology, 104(1-2): 109-118.

Wang, J., Gao, W., Xu, S., Yu, L., 2012. Evaluation of the combined risk of sea level rise, land subsidence, and storm surges on the coastal areas of Shanghai, China. Climatic change, 115(3-4): 537-558.

Wang, Q., Li, W., 2017. A GIS-based comparative evaluation of analytical hierarchy process and frequency ratio models for landslide susceptibility mapping. Physical Geography, 38(4): 318-337.

Wang, Y.-Q., Wang, Z.-F., Cheng, W.-C., 2019. A review on land subsidence caused by groundwater withdrawal in Xi’an, China. Bulletin of Engineering Geology and the Environment, 78(4): 2851-2863.
Wu, Y. Ke, Y., Chen, Z., Liang, S., Zhao, H., Hong, H., 2020. Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. Catena, 187: 104396.

Yariyan, P., Avand, M., Soltani, F., Ghorbanzadeh, O., Blaschke, T., 2020. Earthquake vulnerability mapping using different hybrid models. Symmetry, 12(3): 405.

Yariyan, P., Karami, M., Ali Abbaspour, R., 2019. Exploitation of MCDA to Learn the radial base neural network (RBFNN) aim physical and social vulnerability analysis versus the earthquake (Case study: Sanandaj city, IRAN). International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.

Yin, J., Yu, D., Wilby, R., 2016. Modelling the impact of land subsidence on urban pluvial flooding: A case study of downtown Shanghai, China. Science of the Total Environment, 544: 744-753.

Zhang, D., Agterberg, F., 2018. Modified Weights-of-Evidence Modeling with Example of Missing Geochemical Data. Complexity, 2018.

Zhang, Y., Liu, Y., Jin, M., Jing, Y., Liu, Y., Liu, Y., Sun, W., Wei, J., Chen, Y., 2019. Monitoring land subsidence in Wuhan city (China) using the SBAS-InSAR method with radarsat-2 imagery data. Sensors, 19(3): 743.

Zhi-xiang, T., Pei-xian, L., Li-li, Y., Ka-zhong, D., 2009. Study of the method to calculate subsidence coefficient based on SVM. Procedia Earth and Planetary Science, 1(1): 970-976.

Zhu, J., Yang, Y., Yu, J., Gong, X., 2015. Land subsidence of coastal areas of Jiangsu Province, China: historical review and present situation. Proceedings of the International Association of Hydrological Sciences, 372: 503.