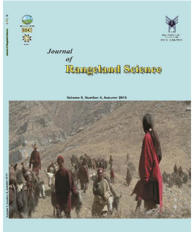
Contents available at ISC and SID

Journal homepage: www.rangeland.ir



Research and Full Length Article:

Application of Satellite Data and Data Mining Algorithms in Estimating Coverage Percent (Case study: Nadoushan Rangelands, Ardakan Plain, Yazd, Iran)

Zinab Mirshekari [A], Majid Sadeghinia [B]*, Saeideh Kalantari [C], Maryam Asadi [D]

[A] M. Sc. student Range Management, Department of Nature Engineering, Faculty of Agriculture & Natural Resources, Ardakan University, P.O. Box 184, Ardakan, Iran.

[B] Assistant Professor, Department of Nature Engineering, Faculty of Agriculture & Natural Resources, Ardakan University, P.O. Box 184, Ardakan, Iran *(Corresponding author), Email: msadeghinia@ardakan.ac.ir

[C] Assistant Professor, Department of Nature Engineering, Faculty of Agriculture & Natural Resources, Ardakan University, P.O. Box 184, Ardakan, Iran.

[D] Ph.D. student of Watershed Management, Department of Reclamation of Arid and Mountainous Regions Engineering, Faculty of Natural Resources, Tehran University, Tehran, Iran.

Received on: 11/11/2018
Accepted on: 02/23/2019

Abstract. Assessing and monitoring rangelands in arid regions are important and essential tasks in order to manage the desired regions. Nowadays, satellite images are used as an approximately economical and fast way to study the vegetation in a variety of scales. This research aims to estimate the coverage percent using the digital data given by ETM^+ Landsat satellite. In late May and early June 2018, the vegetation was measured in Ardakan plain, Yazd province, Iran. Information was obtained by 320 plots in 40 transects and also, the satellite images in terms of sampling time were downloaded and processed in USGS website. 16 indices involving NDVI, NIR, MSI, SS, IR1, MIRV1, NVI, TVI, RAI, SAVI, LWC, PD322, PD321, PD312, PD311 and IR2 were estimated. Through estimating the indices and extracting the values in order to conduct index-based predictions, six data mining models of Artificial Neural Network (ANN), the K Nearest Neighbor (KNN), Gaussian Process (GP), Linear Regression (LR), Support Vector Machine (SVM) and Decision Tree (DT M5) have been applied. Model assessment results indicated high vegetation estimate efficiency based on the indices but the model KNN with Root Mean Square Error (RMSE= 2.520) and Coefficient of determination (R^2= 0.94) and (RMSE= 2.872 and R^2= 0.96) had the highest accuracy in the training and data sets, respectively. As well, to determine the weight and importance of parameters, and to estimate the coverage percent, the weighing process were conducted based on support vector machine. Weighing results indicated that the KNN model and the Simple Subtraction (SS) index had higher weight and importance in terms of vegetation percent.

Key words: Coverage percent, Data Mining, Remote Sensing Indices, ETM^+ sensor

Introduction

Almost 86 million ha (55%) of Iran's regions are rangelands (Azarnivand *et al.*, 2012). Vegetation in each region is one of the most important nature phenomena and the best guide to judge the regional ecology (Kent, 2011). The coverage percent as one of important factors is the ratio of the surface covered with plants to the total surface expressed as percentage in order to study the plants on the ground surface quantitatively (Zhang *et al.*, 2003). Considering the importance of coverage percent determination and its applications, plenty of methods l ike field harvest technique and satellite images have been developed to estimate the variable. Coverage percent determination using conventional methods not only does not give insight into total region vegetation but also is costly and time consuming. In addition, it may be a large number of human errors in determining the coverage percent using conventional methods. Nowadays, satellite data provide a comprehensive perspective in terms of time and place (Darvishzadeh *et al.*, 2012). Thus, having no access to range data at some scales may explain the use of remote sensing techniques; it is able to provide the required information to assess the vegetation in all the regions. Vegetation studies are regarded as the first ones conducted by satellite data in natural resources management (Booth and Tueller, 2003).

One of the common ways to estimate the coverage percent is the use of satellite images and vegetation indices and several indices with different results based on the region type have been given (Matkan *et al.,* 2011). These indices are a mathematical function of few bands of digital satellite images; the significant difference of vegetation reflections is used in blue, red, green and infrared wavelengths. The indices are in the form of simple mathematical operations like subtraction and multiplication or other linear compounds and change the value of each pixel to a number in different bands (Matsushita *et al.,* 2007).

Literature review has shown that the application of indices can be an appropriate way to estimate the vegetation. One of the studies in this regard has been done by Lawrence and Ripple (1998) in order to compare the vegetation indices by the means of TM satellite images in Washington. Results have indicated that NDVI with R^2 (0.65) can be applied for the estimate of coverage percent.

Another study conducted by Baugh and Groeneveld (2009) using TM images in arid regions, Colorado, USA has shown that coverage percent could be estimated by NDVI and images time series with R^2 as 0.77. Darvishzadeh *et al.* (2012) estimate the coverage percent in Sheitoor Region, Bafgh, Yazd province, Iran and reported that plant indices which were computed by the soil line coefficients were of more precision (RMSE= 0.03 and R^2= 0.63) and were able to indicate the coverage percent in arid regions. Najafian *et al.* (2012) tried to identify the unknown spectral phenomena among integrated data of ALI+ ASTER satellite images and hyperspectral Hyperion ones based on correlation coefficient method in Sarcheshmeh, Iran and concluded that the NDVI index has had a critical role in assessing the vegetation of Iran's rangelands. Rahdari *et al.* (2012) studied the capability of satellite data in preparing vegetation crown percent maps in Mooteh habitat, Iran and concluded that the TSAV1, NDVI and RVA indices had high correlation coefficients. Alishah Aratboni *et al.* (2013) used satellite images to draw vegetation map of Sorkhabad rangelands in Mazandaran, Iran and reported that the used NDVI and RVI indices had the highest correlation coefficients as 0.70 and 0.74, respectively. Jabari *et al.* (2016) prepared the coverage percent map in Semirom, Isfahan, Iran using digital data of AWIFS sensor and suggested that the SAVI index was of the highest correlation coefficient given as 0.74. Lawrence and Ripple (1997) compared the indices by the means of TM satellite images in Washington. Results have indicated that NDVI and R^2 (0.65) were more likely to be used for the coverage percent estimate. Another study conducted by Baugh and Groeneveld (2009) using TM images in arid regions, Colorado, USA has shown that coverage percent could be estimated by NDVI and images time series with R^2 as 0.77.

One of the newest techniques to predict and estimate climate parameters, hydrology and natural phenomena is data mining technique (Chang *et al.*, 2015; Rasouli *et al.,* 2012). The technique is able to consider the relationship between various variables to classify or predict the desired parameters based on computational intelligence. Nowadays, using the above-mentioned technique in estimating the plant properties such as vegetation can be effective. For instance, using the artificial neural networks is one of the methods to compute the coverage percent. Carpenter *et al.* (1999) applied the artificial neural networks to prepare the vegetation map. Matkan *et al.* (2011) reviewed the usage efficiency of satellite images and artificial neural networks for coverage percent estimate in arid regions and concluded that the artificial neural networks may estimate the coverage percent with suitable precision and accuracy (RMSE= 0.02, R^2= 0.74).

Regarding the importance of coverage percent estimate and its variations trend review in different years, it is necessary to develop new and efficient methods without field operations in order to have an acceptable accuracy at minimal cost. Today, data mining techniques are considered as intelligence tools to communicate between different factors and prediction of vegetation parameters. The purpose of this study was to performance the evaluation of data mining techniques and satellite indicators in estimating percent vegetation cover. Therefore, in this study, two remote sensing and data mining techniques have been combined. At first, actual data in terms of coverage percent were gathered through field visits and operations and afterwards, satellite images based upon the given equations were extracted by satellite images in relation to vegetation indices sampling period. An appropriate data set involving six models such as artificial neural network, the nearest neighbor, Gaussian process, linear regression, support vector machine and decision tree was developed to model the coverage percent.

Materials and Methods

Study area

The study area is located at 53º 15´ to 54º 50´ E latitude and 31º 15´ to 32º 15´ N longitude in Ardakan- Yazd plain and Nadoushan, Yazd province, Iran. The plain is one of the widest plains in Yazd and surrounded by Shirkoh Mountains in west and southwest and Kharanaq Mountains in east. Mean precipitation is low and irregular estimated as 118 mm and evaporation is 2200-3200 mm. The plain area is almost 4117 km^2 and has the biggest ground water reservoir in Yazd province. Dominant species is *Artemisia sieberi*. The location of the study area is shown in Fig. 1.

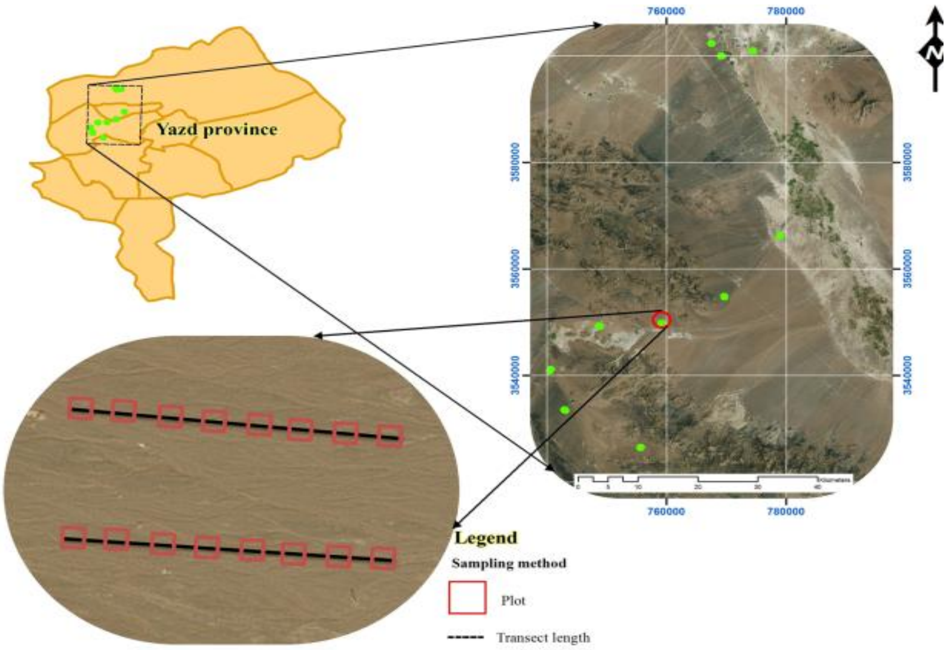


Fig. 1. Study area location

Sampling Method

Sampling has been done with respect to plant growth season in late April and early June 2018 when the coverage percent was at its maximum rate. First, the whole area was visited and then, 10 zones were sampled concerning the plant type and coverage percent. The zones have been selected with regard to the variety of Artemisia coverage percent. In each zone, four 150 m transects (2 transects in the slope direction and 2 transects perpendicular to slope) and eight $2m^2$ plots on each transect were established. Central coordinates of each plot were recorded by GPS and plot information including coverage percent was written concerning all the species existing in the plot in the related forms.

Satellite data

In this research, OLI[1] sensor images of Landsat8 taken on 27th May, 2018 were applied. Also, some atmospheric and radiometric corrections were implemented on the images to remove the errors.

Atmospheric correction

Atmospheric correction is necessary when the energy amounts emitted from the things or signal intensity sent from the things is less than atmospheric effects. It is the most important part before processing the remote sensing images, especially when comparing and analyzing the multiple time images; as well, it is considerably important for multi-spectrum recording systems like Landsat satellite sensing scanners which record the information in visible and infrared spectrum and mainly is affected by the atmospheric absorption on visible and infrared bands (Wei *et al.*, 2018). Atmospheric correction was performed by dimming algorithm, for this purpose, the Dark Subtraction function of ENVI[2] software was used.

Radiometric correction

Radiometric corrections involve those implemented on grey degrees and try to compensate the errors with the change of amounts separately (pixel to pixel). They must be implemented by varying the perspective, lighting, geometric view,

1_ Operational Land Imager

2_ The Environment for Visualizing Images weather conditions and sensing noise. They depend on sensors features and data collection conditions (Pons *et al.,* 2014). In order to perform radiometric correction, ENVI's radiometric calibration function was used.

Vegetation indices

Vegetation indices present a variety of multi-spectrum satellite data compounds to produce an image in relation to vegetation status (Eastman, 1995). Red and near infrared bands are mostly used to estimate the indices due to plant pigments absorption property like chlorophyll which leads to low reflection of plants in red band and intense reflection in the infrared band of electromagnet spectrum (Jafari *et al.,* 2011). Table 1 presents the equations related to plant indices in the study. Figs. 2 to 5 show the map of the calculated indices.

Table 1. The plant indices used in present research (Rock *et al.,* 1986; Leblon, 1993; Boyd *et al.,* 1996; Foody *et al.,* 2001; Pettorelli *et al.,* 2005; Arzani and King, 2008)

No	Index Name	Abbreviation	Formula
1	Near Infrared Ratio	NIR	TM4/TM3
2	Moisture Stress Index	MSI	TM5/TM4
3	Leaf Water Content (Mid- IR-Index)	LWC	TM5/TM7
4	Normalized Vegetation Index	NVI	(TM4-TM3)/ (TM4+TM3)
5	Transformed Vegetation Index	TVI	(TM5-TN3)/ (TM5+TM3)
6	Infrared Index	IR1	(TM4-TM5)/ (TM4+TM5)
7	Reflectance Absorption Index	RAI	TM4/ (TM3+TM5)
8	Normalized Difference Vegetation Index	NDVI	(TM5-TM4)/ (TM5+TM4)
9	Potential Different Index	PD311	TM3-TM1
10	Potential Different Index	PD312	(TM3-TM1)/ (TM3+TM1)
11	Potential Different Index	PD321	TM3-TM2
12	Potential Different Index	PD322	(TM3-TM2)/ (TM3+TM2)
13	Simple Subtraction Index	SS	TM4-TM3
14	Infrared Index	IR2	(TM4-TM7) /(TM4+TM7)
15	Modified Infrared Ratio Vegetation Index	MIRV1	(TM7-TM3)/ (TM7+TM3)
16	Soil Adjusted Vegetation Index	SAVI	[(TM5-TM4)/ (TM5+TM4+L)]*(1+L)

TM: Thematic Mapper

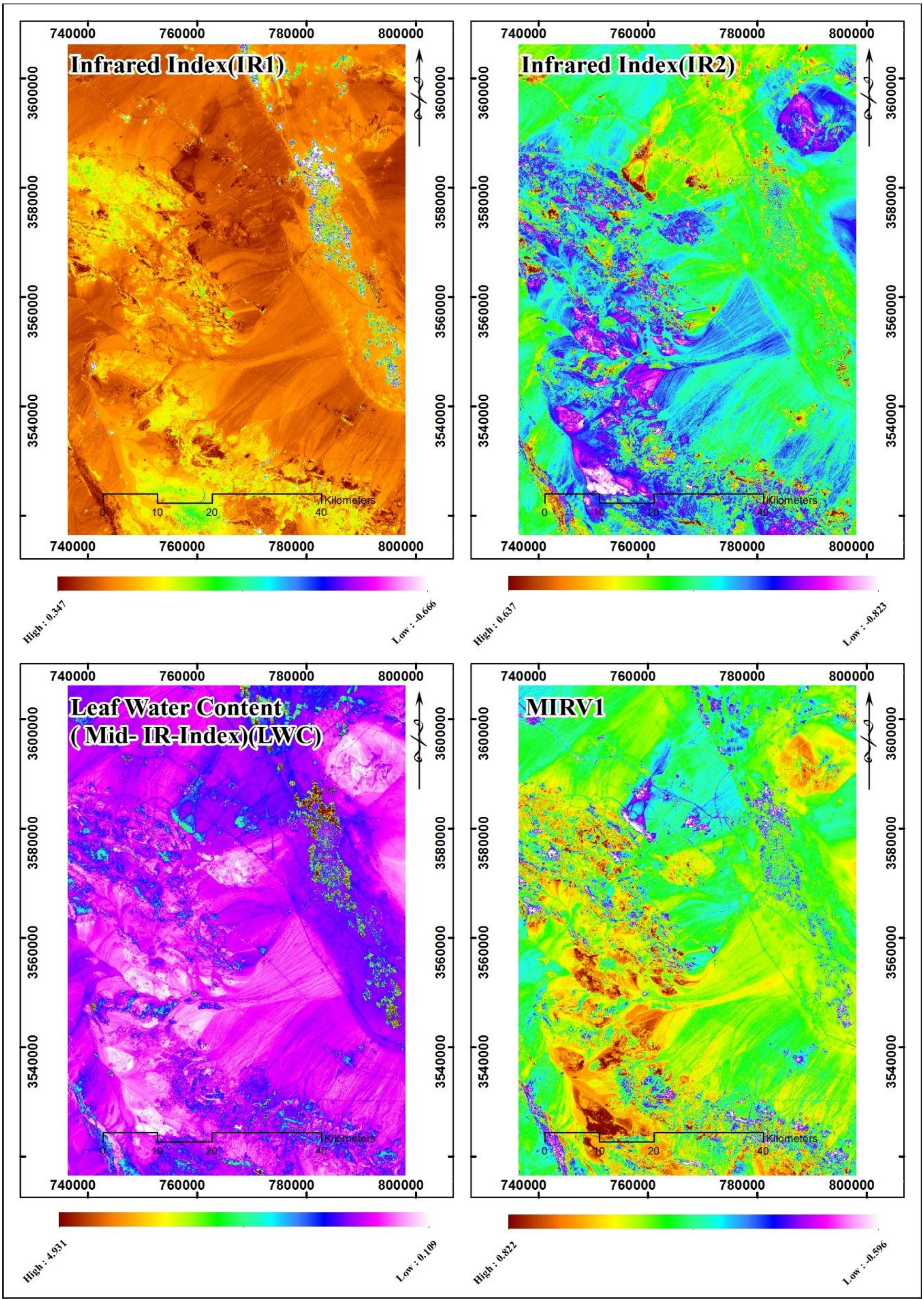


Fig. 2. Satellite Image Index Map (Infrared Index (IR1, IR2), LWC, MIRV1)

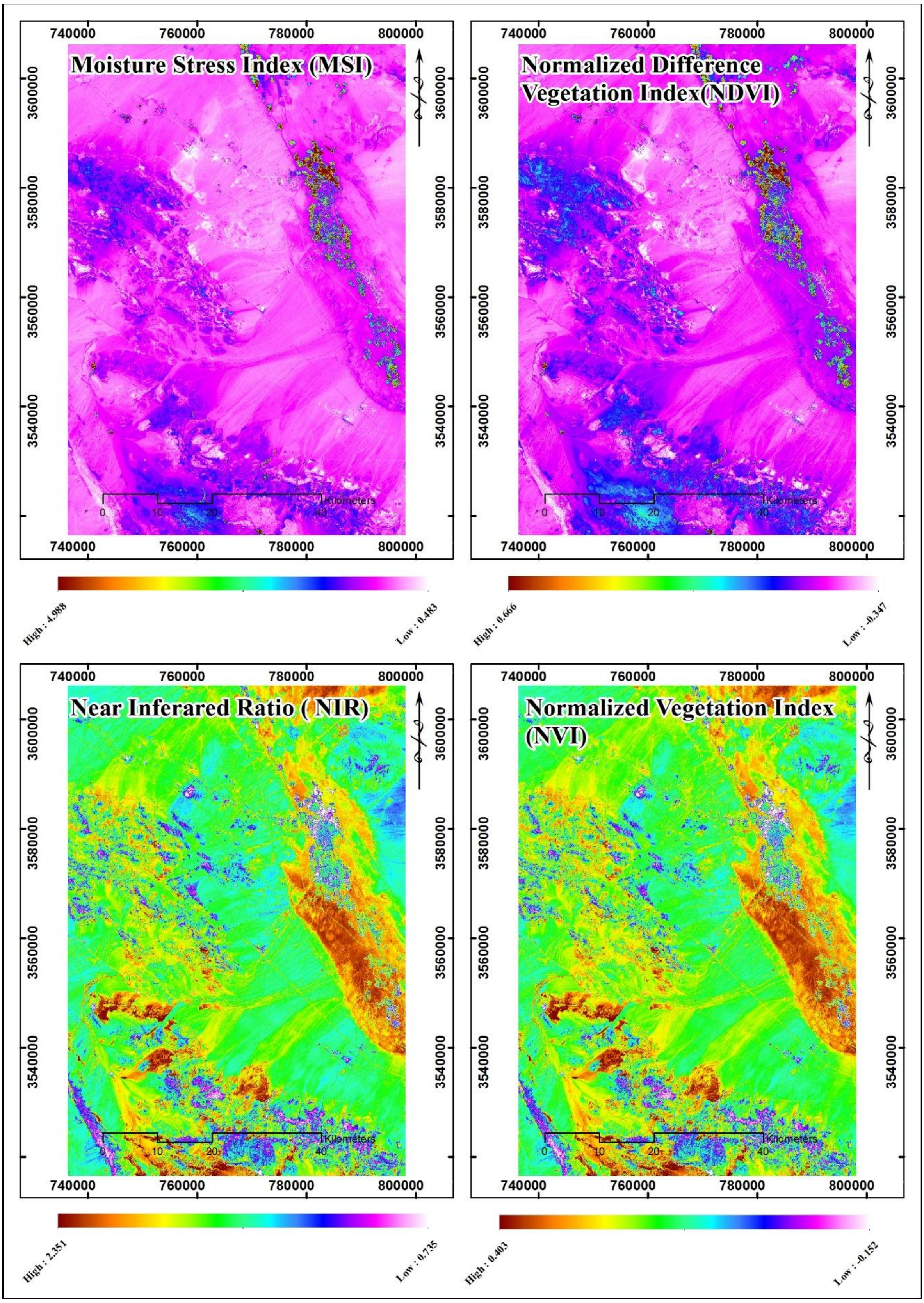


Fig. 3. Satellite Image Index Map (MSI, NDVI, NIR, NVI)

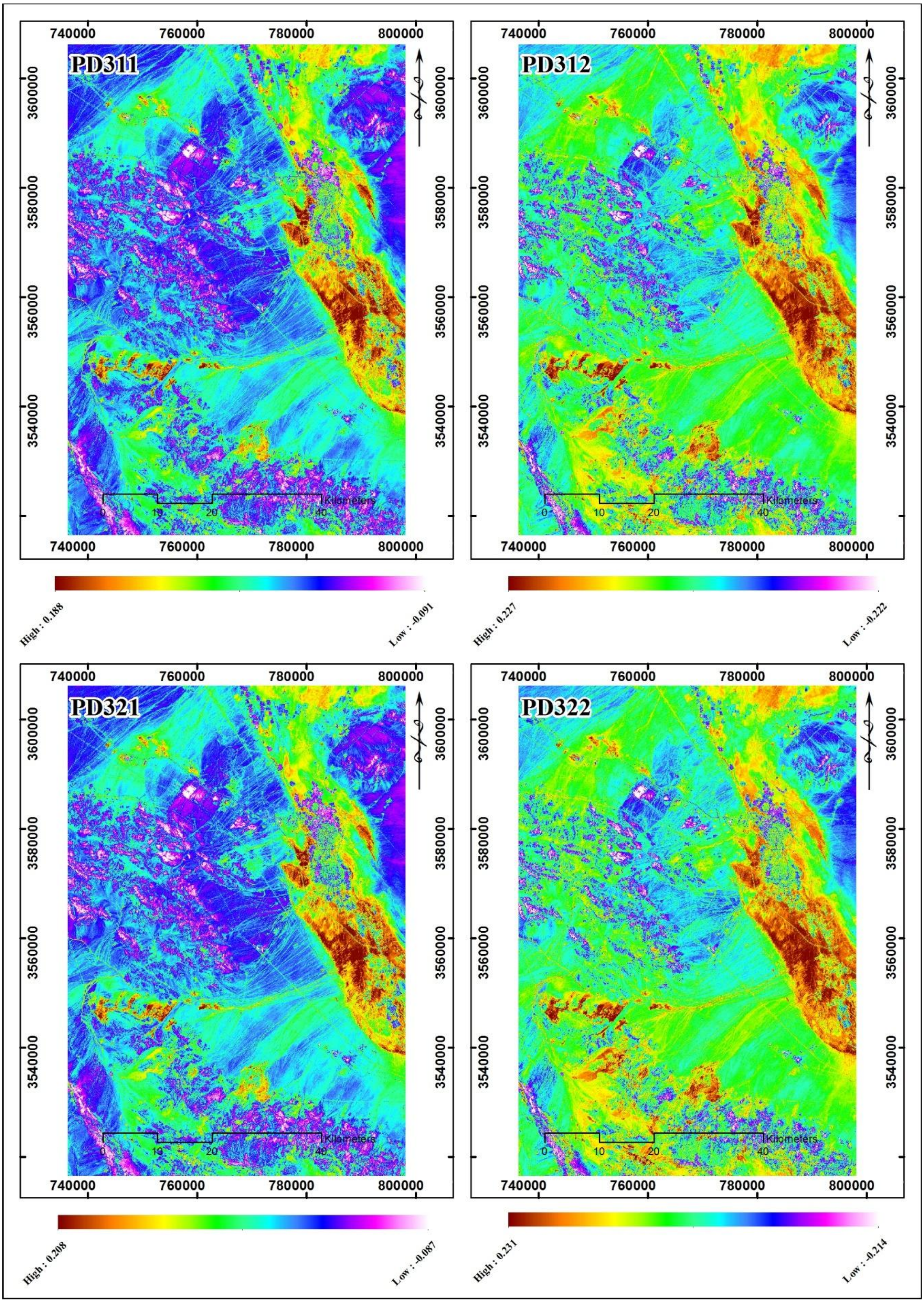


Fig. 4. Satellite Image Index Map (PD311, PD312, PD321, PD322)

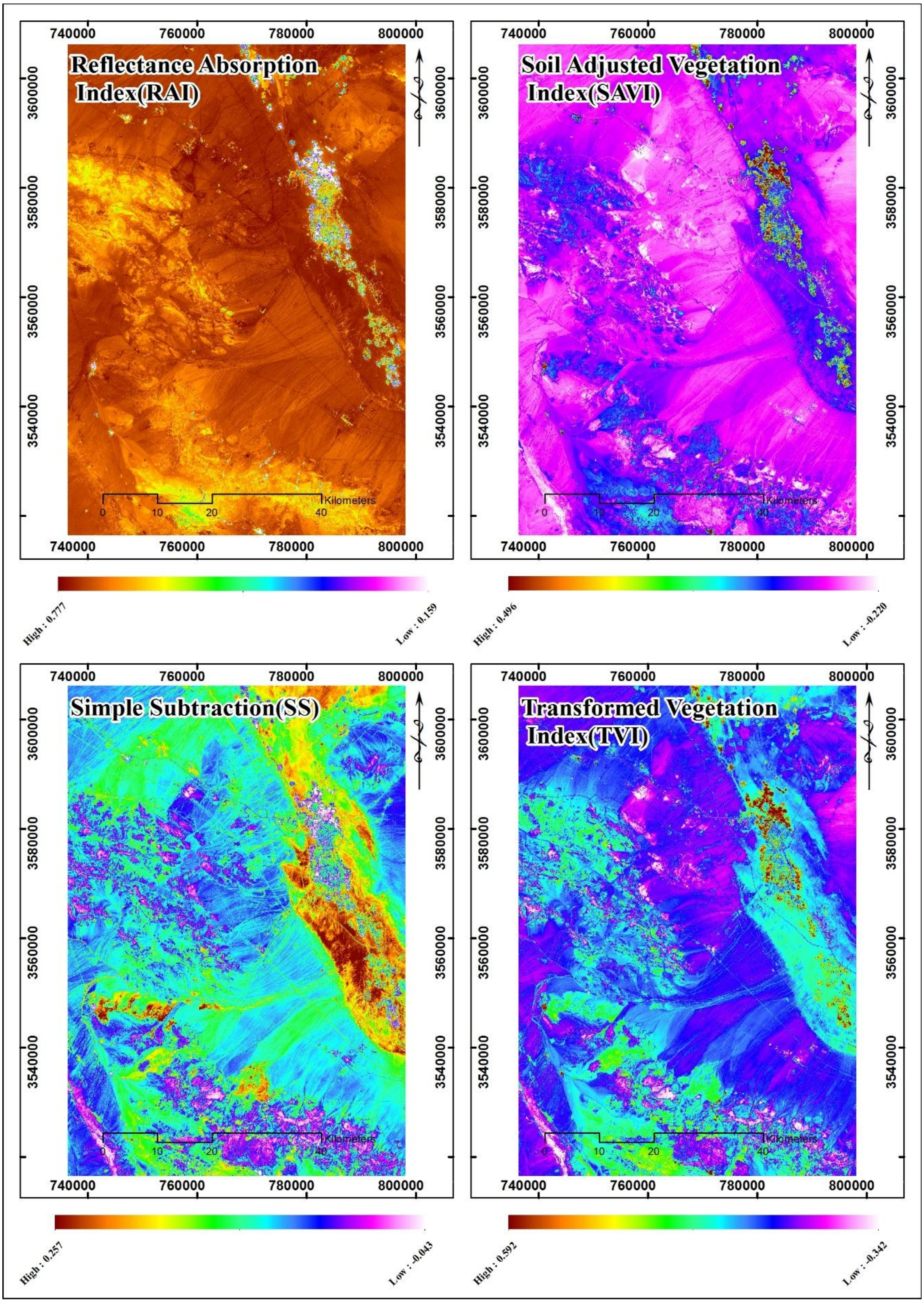


Fig. 5. Satellite Image Index Map (RAI, SAVI, SS, TVI)

Data mining technique

Data mining technique has been known as a stage of extracted knowledge process which is divided in to three phases including data preprocessing, data mining and extracted knowledge process. Often, before implementing data mining algorithms, it is essential to preprocess data such as data collection, data resources unifying, data clearing, continuous to discrete quantities conversion, suitable features selection and data volume reduction (Han *et al.*, 2011). In this study, data have been divided into two training and test sets and models were examined in two modes. In order to implement the models in the training data set, indices of satellite images as model input and data of coverage percent as model output were first introduced to the model. Then, in each model, a suitable structure was addressed to estimate the plant coverage percent. Each model has a different structure and regulating parameters. Afterwards, determining the best structure concerning each model contributes to introduce the test data set into the model, which only involves the inputs. Based on the evaluation criteria, the results have been investigated. Each model structure and the evaluation criteria have been summarized. Applied data mining algorithms are as follows:

a) Nearest neighbor

The nearest neighbor is one of the variable classification methods based on their similarities, which classifies data patterns without any need to predetermined patterns. It seeks to specify the data point features with respect to the nearest neighbor characteristics. One of the best classifications is the nearest neighbor k classification which attributes the test sample to a class with the most votes among the nearest neighbors K. KNN classification is a simple and applicable method due to high understanding and no need to data assumptions (Bashiri *et al.*, 2018). In the conducted study, the model was the best model due to the number of k given as 3 based on Mixed Euclidean distance to do predictions.

b) Artificial neural networks

Neural networks are dynamic systems which transfer knowledge or hidden law beyond data to the network structure through processing experimental data. The systems learn the general laws by new and primary data computations and are regarded as intelligent systems. Information processing in neural networks is done by a method similar to human brain. The network is consisted of lots of processing elements called neurons working in a parallel form (Zarkogianni *et al.*, 2011). Intelligent neural networks can discover the aspects hidden in science and relations hidden in data by analyzing data and relationships between actual inputs and outputs. The knowledge is inferred by attributing weights to neural network (Su, 1994). In the study model, 6 hidden layers and 700 training cycles with memento estimated as 0.1 had the best performance.

c) Gaussian process

The Gaussian process is a strong non-parametric machine learning method to develop comprehensive probabilistic models by real world issues. It is a random process consisted of random amounts in each point in one time or location domain so that each random variable has a normal distribution. Furthermore, each finite set involving random variables has multivariate normal distribution. Gaussian processes expand multivariate Gaussian distributions to infinite dimensions. Formally, a Gaussian process produces data located in this spectrum so that each finite domain subset follows a multivariate Gaussian distribution (Rasmussen and Williams, 2006; Asadi *et al.*, 2017). In the Gaussian process model, it has been shown that the model with 200 max-basis vectors and Laplace Kernel function was of the best structure.

d) Linear regression

Linear regression model is applied in the most time series predictions. Prediction variables involve a linear function of one or more independent variables (predictor) and an error. By applying a linear relation between independent variables (predictor), prediction process will be possible with the lowest error (Rajaee *et al.,* 2011).

e) Decision tree

One of the popular classifications is decision tree which is simple and possible to interpret the results. It is able to produce understandable descriptions out of relations existing in a data set and enables the predictions in the form of laws with suitable fitness concerning the statistical parameters. It is applied for discrete functions and error data and contributes to discover knowledge (Balouchi *et al.,* 2015). A decision tree is an appropriate classification model using data sets (David *et al.,* 2013).

f) Support vector machine

This algorithm is a non-parametric statistical learning method which was first developed by Vapnik (1995). In fact, it is a binary classifier which tries to create hyperplane with the most distance from two classes using an optimum algorithm in learning stage. One of the most important aspects in summarizing the algorithm is that all the training points are not applied to determine the decision making border and only a few points called support vector with the lowest distance to ultrasound are used to define the border (Mountrakis *et al.,* 2011). The model with radial kernel and kernel Gama 1 functions as well as 200 kernel cycles has the best structure.

Model performance assessment

Investigating the model performance and comparing the prediction power are necessary to assess the performance of networks. In this research, in addition to drawing the graphs of observation values against prediction ones by models, quantitative indices like Correlation Coefficient (R), Coefficient of Determination (R^2) and Root Mean-Square Error (RMSE) are used.

Parameters and indices weighing

All the input parameters do not have uniform impact and importance. Some parameters have more correlation and relation to model output and considerably affect the predictions. Parameters weighing is one of data mining techniques. In this study, the support vector machine is used for weighing and the algorithm determines the normal vector coefficients of one linear support vector machine as the weight of property (Sani Abade *et al.,* 2017).

Results

According to the conducted samplings, the most important range species are as follows.

Artemisia sieberi, *Seidlitzia rosmainus*, *Alhagi Camelorum*, *Noaea mucronate*, *Fortuynia bungei*, *Pteropyrum aucheri*, *Convolvulus sp.*, *Astragalus sp.*, *Scariola orientalis*, *Eurotia ceratoides*

Reviewing the statistical summary of training data and test data has indicated that the IR^2 index had the most coefficients of variations and the RAI index was of the least coefficient. Results have been summarized in Table 2.

Table 2. Descriptive statistical summary of training and test data

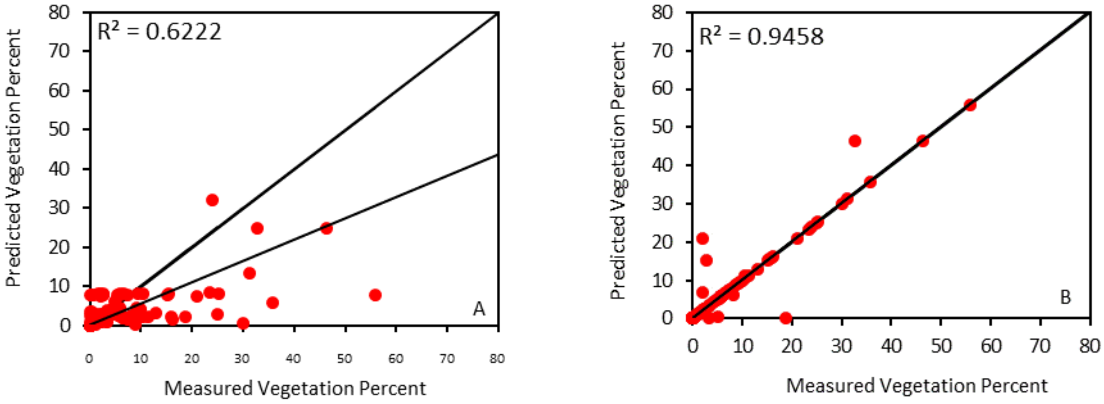
Indices	Training set			Test set		
	Mean	SD	CV	Mean	SD	CV
Soil Adjusted Vegetation Index (SAVI)	0.070	0.0197	0.281	0.069	0.0211	0.307
Reflectance Absorption Index (RAI)	0.497	0.0104	0.021	0.498	0.0116	0.023
Potential Different Index (PD322)	0.059	0.0178	0.301	0.059	0.0170	0.290
Potential Different Index (PD321)	0.024	0.0099	0.408	0.024	0.0094	0.392
Potential Different Index (PD312)	0.059	0.0282	0.483	0.058	0.0272	0.472
Potential Different Index (PD311)	0.024	0.0140	0.578	0.024	0.0135	0.566
Infrared Index (IR1)	-0.088	0.0202	-0.231	-0.086	0.0220	-0.256
Infrared Index (IR2)	-0.051	0.0650	-1.268	-0.048	0.0592	-1.239
Leaf Water Content (Mid- IR-Index) (LWC)	1.087	0.1652	0.152	1.091	0.1513	0.139
Modified Infrared Ratio Vegetation Index (MIRV1)	0.150	0.0617	0.412	0.146	0.0567	0.388
Normalized Difference Vegetation Index (NDVI)	0.088	0.0202	0.231	0.086	0.0220	0.256
Moisture Stress Index (MSI)	1.193	0.0492	0.041	1.189	0.0538	0.045
Near Infrared Ratio (NIR)	1.222	0.0310	0.025	1.221	0.0300	0.024
Normalized Vegetation Index (NVI)	0.100	0.0124	0.125	0.100	0.0120	0.121
Transformed Vegetation Index (TVI)	0.186	0.0279	0.150	0.184	0.0287	0.157
Simple Subtraction Index (SS)	0.048	0.0127	0.267	0.047	0.0121	0.255

At first, models were implemented on training data set and through regulating the parameters; the best structure was specified for models suggesting more appropriate results. Results achieved by implementing the models on training data have been presented in Table 3. Results have indicated that the KNN model with the RMSE= 2.520 and R^2= 0.94 had higher precision and the SVM model with RMSE= 9.389 and R^2= 0.39 had the lowest precision in estimating the coverage percent.

Table 3. Training results of different models

MODEL	Abbreviation	RMSE	R^2	R
Artificial Neural Network	ANN	7.215	0.622	0.788
K Nearest Neighbor	KNN	2.520	0.945	0.972
Gaussian process	GP	6.816	0.679	0.824
Linear Regression	LR	6.769	0.485	0.696
Support Vector Machine	SVM	9.389	0.399	0.631
Decision Tree (M5)	DT M5	4.430	0.826	0.909

Distribution of estimated and actual values has been given in Fig. 6.



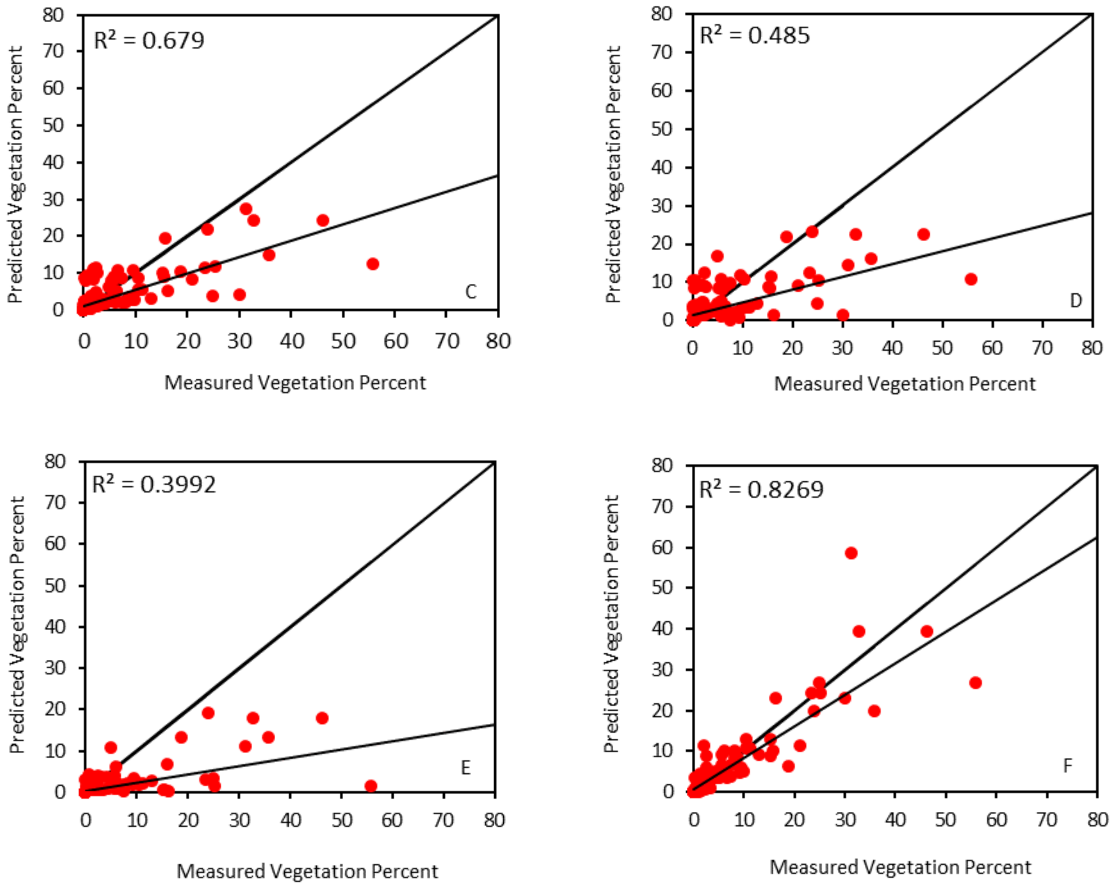


Fig. 6. Distribution of estimated and actual values in terms of coverage percentage for training results: A) Artificial Neural Network, B) the Nearest Neighbor, C) Gaussian process, D) Linear Regression, E) Support Vector Machine and F) Decision Tree M5

Results of models implementation on test data have been presented in Table 4 indicating that the KNN model with RMSE= 2.872 and R^2= 0.96 was of the highest precision and the SVM model with RMSE= 11.423 and R^2= 0.36 had the lowest precision in estimating the coverage percent.

Table 4. Test results for different models

MODEL	Abbreviation	RMSE	R2	R
Artificial Neural Network	ANN	7.869	0.796	0.892
K Nearest Neighbor	KNN	2.872	0.960	0.979
Gaussian process	GP	8.242	0.691	0.831
Linear Regression	LR	8.316	0.706	0.840
Support Vector Machine	SVM	11.423	0.368	0.606
Decision Tree (M5)	DT M5	5.838	0.771	0.878

Distribution of estimated and actual values has been given in Fig. 7.

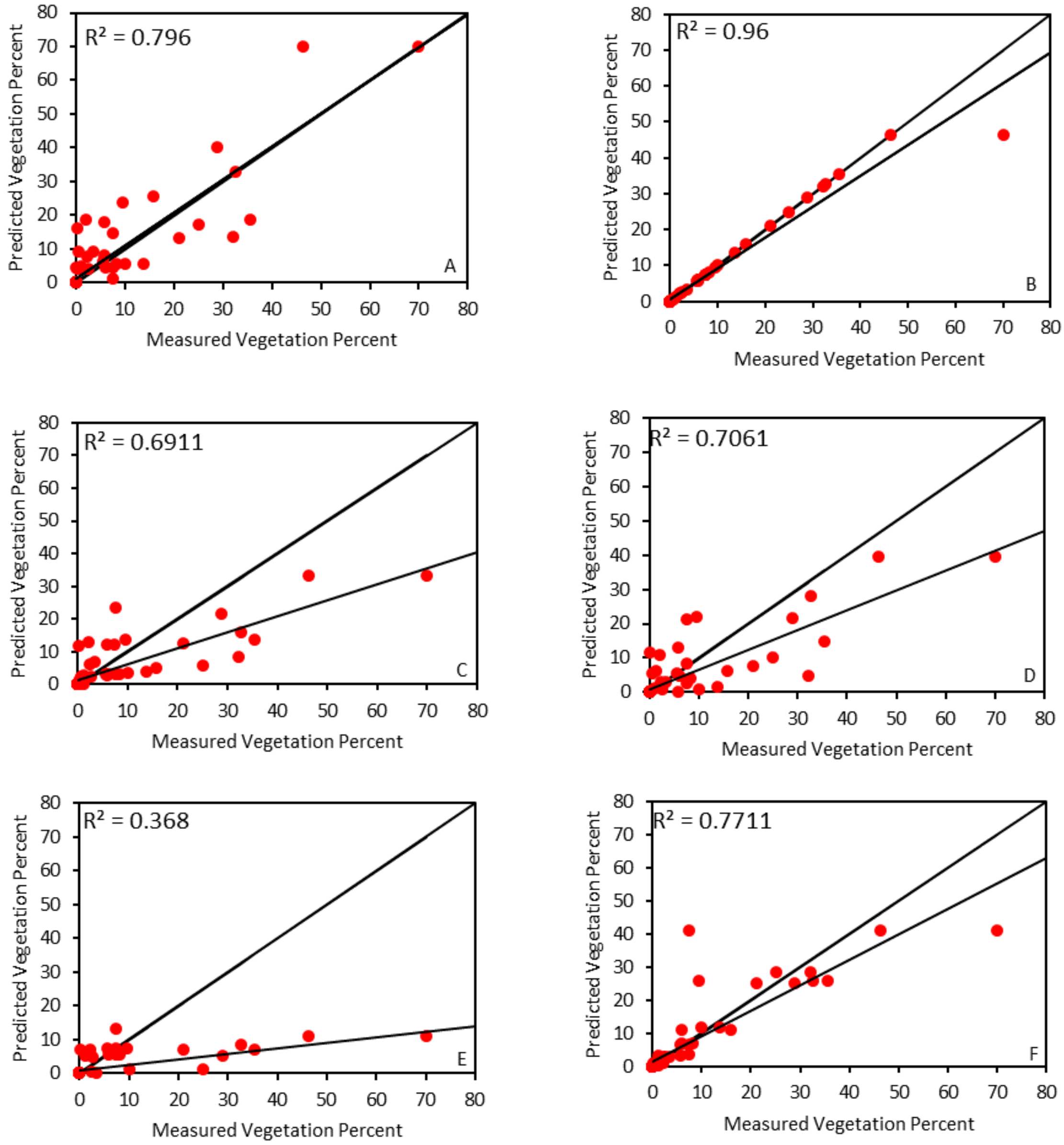


Fig. 7. Distribution of estimated and actual values in terms of coverage percentage for test result: A) Artificial Neural Network, B) the Nearest Neighbor, C) Gaussian process, D) Linear Regression, E) Support Vector Machine and F) Decision Tree M5

Weighing results

Applying various indices on satellite images in the region has suggested that among the studied indices, the SS and VNIR1 indices had the highest and lowest weight and impact on measuring the coverage percent, respectively.

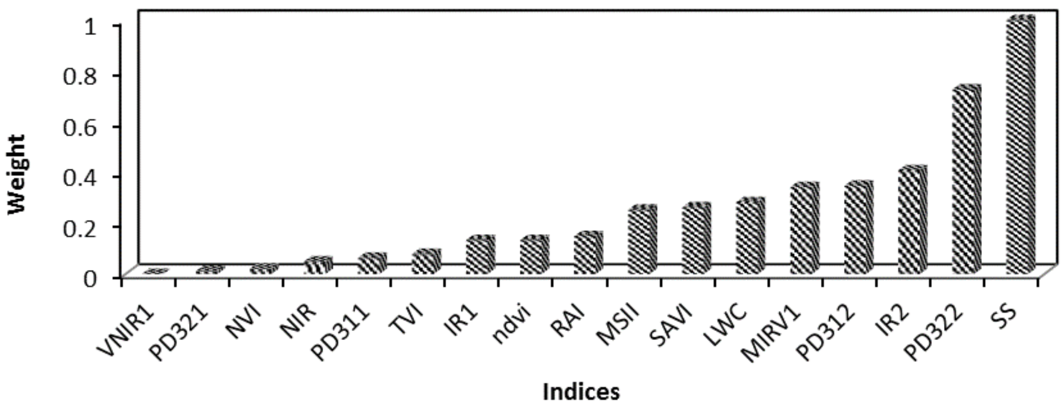


Fig. 8. Weighing graph using support vector machine (See Table 1 for the full name of indices)

Discussion

Vegetation is one of the most important ecosystem components and knowing the coverage percent is necessary to conduct the environmental and natural resources studies. In this paper, field harvest was first done and satellite images were taken from USGS website coincidentally. This coincidence considerably affected the accuracy of maps drawn by satellite images and the images were corrected atmospherically and radiometrically while obtaining vegetation indices. Studying the data weighing has shown that the SS index was of the most precision in coverage percent estimate. Using vegetation indices in various regions has given various results. Najafian *et al.* (2012) suggested using the NDVI index to assess the vegetation. Jabari *et al.* (2016) have shown that the SAVI index had the highest correlation coefficient. Lawrence and Ripple (1998) applied the satellite TM images and reported that the NDVI index with (R^2=0.65) could estimate the coverage percent. Therefore, it is necessary to assess the different indices in different fields to estimate the plant coverage.

Another method to compute coverage percent is data mining technique. Results indicated that k nearest neighbor was of the best performance and most correlation; Carpenter *et al.* (1999) utilized the ANN to give a vegetation map and also, Matkan *et al.* (2011) found that the ANN was able to estimate coverage percent with high precision. Since timely access to field data is difficult and limited for studying and monitoring vegetation at global and regional scales and coverage percent estimate with the conventional method involving total vegetation estimate is time consuming and gives no accurate information, the remote sensing technology is greatly useful and superior to other methods due to providing a comprehensive and uniform perspective, repeatability, information collection ease, high information precision and time saving. On the other hand, in lots of regions with respect to inaccessible former vegetation information, it is not plausible to investigate the vegetation variations using conventional methods but remote sensing data pave the way to study the desired variations and indicate the importance of the data usage in the range researches. Data mining technique has not been considerably utilized in vegetation estimate but this paper demonstrates that the technique is able to estimate the coverage percent with high precision; therefore, it is recommended to apply the mentioned technique to estimate the coverage percent.

References

Alishah Aratboni, F., Arzani, H., Hosseini, S. Z., Babaie Kafaki, S. and Mirakhorlou, K., 2013. Rangelands vegetation cover mapping using IRS-LissIII data image processing (Case study: Sorkh Abad Basin, Mazandaran). Iranian Journal of Range and Desert Research, 20(3): 454- 462 (In Persian).

Arzani, H., and King, G. W., 2008. Application of remote sensing (landsat TM data) for vegetation parameters measurement in western division of NSW. International Grassland Congress. Hohhot, China. ID NO. 1083.

Asadi, M., Fatzhadeh, A. and Taghizadeh Mehrjerdi, R., 2017. Optimization suspended load estimation models by using geo-morphometric parameters and attribute reduction technique. Iranian Journal of Soil and Water Research, 48(3): 669- 678 (In Persian).

Azarnivand, H., Alizadeh, E., Sour, A. and Hajibeglo, A., 2012. The Effects of Range Management Plans of Soil Properties and Rangeland Vegetation (Case Study: Eshtehard Rangelands). Journal of Rangeland Science, 2(4): 625- 634.

Baugh, W. M. and Groeneveld, D. P., 2009. Broadband vegetation index performance evaluated for a low-cover environment. International Journal of Remote sensing, 27(21): 4715- 4730.

Bashiri, M., Ariapour, M. and Golkarian, A., 2018. Using Data Mining Algorithms in Separation of Sediment Sources in Nodeh Watershed, Gonabad. Desert Ecosystem Engineering Journal, 7(19): 81- 94 (In Persian).

Balouchi, B., Nikoo, M. R. and Adamowski, J. 2015. Development of expert systems for the prediction of scour depth under live-bed conditions at river confluences: application of different types of ANNs and the M5P model tree. Applied Soft Computing, 34: 51- 59.

Boyd, D. S., Foody, G. M., Curran, P. J., Lucas, R. M. and Honzak, M., 1996. An assessment of radiance in Landsat TM middle and thermal infrared wavebands for the detection of tropical forest regeneration. International Journal of Remote Sensing. 17(2): 249-261.

Booth, D. T. and Tueller, P. T., 2003. Rangeland monitoring using remote sensing. Arid Land Research and Management, 17: 455- 467.

Chang, F. J., Tsai, Y. H., Chen, P. A., Coynel, A. and Vachaud, G., 2015. Modeling Water Quality in an Urban River Using Hydrological Factors– Data Driven Approaches. Journal of Environmental Management, 151: 87- 96.

Carpenter, G. A., Gopal, S., Macomber, S., Martens, S., Woodcock, C. E., and Franklin, J., 1999. A Neural Network method for efficient vegetation mapping. Remote Sensing of Environment 70(3): 326–338.

Darvishzadeh, R., Matkan, A. A., Hosseiniasl, A. and Ebrahimi Khusefi, M., 2012. Estimation of vegetation fraction in the Central arid region of Iran using satellite images (Case study: Sheitoor basin, Bafgh). Arid Biome Scientific and Research Journal, 2(1): 25-38 (In Persian).

David, S. K., Saeb. A. T. M., and Al Rubeaan, K., 2013. Comparative Analysis of Data Mining Tools and Classification Techniques using WEKA in Medical Bioinformatics. Computer Engineering and Intelligent Systems, 4(13): 28- 38.

Eastman, J. R., 1995. IDRISI for Windows: Student Manual, Version 1.0. Clark University Publication. USA.

Foody, G. M., Cutler, M., McMorrow, J., Pelz, D., Tangki, H., Boyd, D. S., and Douglas, I., 2001. Mapping the biomass of Bornean tropical rain forest from remotely sensed data. Global Ecology and Biogeography. 10(4): 379–387.

Han, J., Pei, J. and Kamber, M., 2011, Data Mining: Concepts and Techniques (3nd Edition). Published by Elsevier, 744p.

Jabari, S., Khajedin, S. J., Jafari, R. and Soltani. S., 2016. Application of AWIFS digital data to determine vegetation cover (Case Study: Semirom- Isfahan). Journal of Rangeland. 9(4): 333- 343 (In Persian).

Jafari, H. R., Hamzeh, M., Nasiri, H. and Rafii, Y., 2011. Developing Decision Tree and Data Mining Based Conceptual Model for Detecting Land Cover Changes using TM images and Ancillary Data (Study area: Centeral section of Bouyerahmad County). Environmental Science. 8(3): 1-20 (In Persian).

Kent, M., 2011. Vegetation Description and Data Analysis: A Practical Approach (2nd Edition). Wiley-Blackwell. 428p

Lawrence, R. L. and Ripple, W. J., 1998. Comparisons among Vegetation Indices and Bandwise Regression in a Highly Disturbed, Heterogeneous Landscape: Mount St. Helens, Washington. Remote Sensing of Environment, 64(1): 91- 102.

Leblon, B., 1993. Soil and vegetation optical properties. In: Applications in Remote Sensing, Volume 4, The International Center for Remote Sensing Education. Wiley Press, New York, USA.

Matkan, A. A., Darvishzadeh, R., Hosseiniasl, A. and Ebrahimi Khusefi, M., 2011. Capability using satellite images and neural network in estimating vegetation percentage in arid region. Journal of Environmental Erosion, 1: 7- 27 (In Persian).

Matsushita. B., Yang. W., Chen. J., Onda. Y. and Qiu. G., 2007. Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to topographic effects: A case study in high-density Cypress forest. Sensors. 7(11): 2636-2651.

Mountrakis, G., Im, J. & Ogole, C., 2011, Support Vector Machines in Remote Sensing: A Review, ISPRS Journal of Photogrammetry and Remote Sensing, 66: 247 259.

Najafian, T., Fathianpour, N., Ranjbar, H. and Bakhshpour, R., 2012. Detecting Unknown Spectral Features from ALI, ASTER and Hyperion Images using Correlation Coefficient Method. Case Study: Sarcheshmeh copper mine, IRAN. Journal of Advanced Applied Geology 2(5): 59-68 (In Persian).

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J. M., Tucker, C. J. and Stenseth, N. C., 2005. Using the satellite – derived NDVI to assess ecological responses to environmental change. Journal of Trends in Ecology and Evolution. 9(20): 503-510.

Pons, X., Pesquer, L., Cristobal, J. and Gonzalez-Guerrero, O., 2014. Automatic and improved radiometric correction of Landsat imagery using reference values from MODIS surface reflectance images. International Journal of Applied Earth Observation and Geoinformation. 33: 243- 254.

Rahdari, V., Maleki Najafabadi, S., Afsari, K.H., Abtin, E., Piri, H. and Fakhireh, A., 2012. Change detection of Hamoun wild life refuge using RS & GIS. Remote Sens. GIS Journal Iran. 3(2): 59-70 (In Persian).

Rajaee, T., Nourani, V., Zounemat-Kermani, M. and Kisi, O. 2011. River Suspended Sediment Load Prediction: Application of ANN and Wavelet Conjunction Model. Journal of Hydrologic Engineering, 16(8): 613- 627.

Rasmussen, C. E. and Williams, C. K. I., 2006. Gaussian Processes for Machine Learning. The MIT Press, ISBN 026218253X. c. 2006. Massachusetts Institute of Technology. USA.

Rasouli, K., Hsieh, W. W. and Cannon, A. J., 2012. Daily Streamflow Forecasting by Machine Learning Methods with Weather and Climate Inputs. Journal of Hydrology, 414: 284-293.

Rock, B.N., Vogelmann, J. E., Williams, D. L., Vogelmann, A. F., and Hoshizaki, T., 1986. Remote Detection of Forest Damage, 36(7): 439-445.

Sani Abade, M., Mahmoudi, S. and Taherparvar, D., 2017. Data Mining Applications (3nd Edition), Niaz-e-Danesh Publication. Tehran (In Persian).

Su, M. C., 1994. Use of neural networks as medical diagnosis expert systems. Computers in Biology and Medicine, 24(6): 419- 429.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer Verlag, New York, USA.

Wei, J., Lee, Z., Garcia, R., Zoffoli, L., Armstrong, R. A., Shang, Z., Sheldon, P. and Chen, R. F., 2018. An assessment of Landsat-8 atmospheric correction schemes and remote sensing reflectance products in coral reefs and coastal turbid waters. Remote Sensing of Environment, 215: 18- 32.

Zarkogianni, K., Vazeou, A., Mougiakakou, S. G., Prountzou, A. and Nikita, K. S., 2011. An insulin infusion advisory system based on autotuning nonlinear model-predictive control. IEEE Transaction Biomedical Engineering, 58(9): 2467- 2477.

Zhang, Y. X., Li, X. B. and Chen, Y. H., 2003. Overview of field and multi-scale remote sensing measurement approaches to grassland vegetation coverage. Advanced Earth science, 18: 85-93 (in Chinese with English abstract).

کاربرد داده‌های رقومی و الگوریتم‌های داده‌کاوی در تخمین درصد پوشش (مراتع ندوشن و دشت یزد اردکان)

زینب میرشکاری[الف]، مجید صادقی‌نیا[ب*]، سعیده کلانتری[ج]، مریم اسدی[د]
[الف] دانشجوی کارشناسی ارشد مرتعداری، گروه مهندسی طبیعت، دانشکده کشاورزی و منابع طبیعی، دانشگاه اردکان، اردکان، ایران.
[ب] استادیار، گروه مهندسی طبیعت، دانشکده کشاورزی و منابع طبیعی، دانشگاه اردکان، اردکان، ایران. *(نگارنده مسئول)، پست الکترونیک: msadeghinia@ardakan.ac.ir
[ج] استادیار، گروه مهندسی طبیعت، دانشکده کشاورزی و منابع طبیعی، دانشگاه اردکان، اردکان، ایران.
[د] دانشجوی دکتری آبخیزداری، گروه مهندسی احیا مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، تهران، ایران.

تاریخ دریافت: ۱۳۹۷/۰۸/۲۰
تاریخ پذیرش: ۱۳۹۷/۱۲/۰۴

چکیده. ارزیابی و پایش مراتع مناطق خشک یکی از موارد مهم و ضروری جهت مدیریت این مناطق به حساب می‌آید. امروزه از تصاویر ماهواره‌ای به عنوان روشی نسبتا کم هزینه و سریع جهت مطالعه پوشش گیاهی در مقیاس‌های متفاوت استفاده می‌شود. هدف از این تحقیق، تخمین میزان پوشش با استفاده از داده‌های رقومی سنجنده ETM^+ ماهواره لندست ۸ می‌باشد. درصد پوشش گیاهی در اواخر اردیبهشت ماه و اوایل خرداد ماه سال ۱۳۹۷ در منطقه دشت یزد- اردکان مورد اندازه‌گیری قرار گرفت. برداشت اطلاعات در قالب ۳۲۰ پلات استقرار یافته در طول ۴۰ ترانسکت انجام گرفته و همچنین تصاویر ماهواره‌ای مربوط به زمان نمونه‌برداری از سایت USGS دانلود و پردازش‌های لازم بر روی تصاویر انجام شد. ۱۶ شاخص رایج شامل NDVI، NIR، MSI، SS، IR1، MIRV1، NVI، TVI، RAI، SAVI، LWC، PD322، PD321، PD312، PD311 و IR2 محاسبه گردید. با محاسبه شاخص‌ها و استخراج مقادیر آنها، جهت پیش‌بینی بر مبنای شاخص‌ها از شش مدل داده‌کاوی شبکه عصبی مصنوعی (ANN)، نزدیک‌ترین همسایه (KNN)، فرآیند گوسی (GP)، رگرسیون خطی (LR)، ماشین بردار پشتیبان (SVM) و درخت تصمیم‌گیری (DT) استفاده شد. نتایج ارزیابی مدل‌ها نشان دهنده کارایی بالای برآورد پوشش بر مبنای شاخص‌ها بوده اما مدل KNN در مجموعه داده‌های آموزش با میانگین مربعات خطا RMSE برابر با ۲/۵۲۰ و ضریب تعیین برابر با ۰/۹۴ و در مجموعه داده‌های تست با RMSE برابر با ۲/۸۷۲ و ضریب تعیین برابر با ۰/۹۶ دارای دقت بالاتری نسبت به سایر مدل‌ها بودند. همچنین جهت تعیین وزن و اهمیت هریک از پارامترها جهت برآورد درصد پوشش فرآیند وزن‌دهی بر مبنای مدل ماشین‌بردار پشتیبان صورت گرفت. نتایج وزن‌دهی نشان داد که مدل KNN و شاخص SS دارای وزن و اهمیت بیشتری در برآورد درصد پوشش منطقه داشته است.

کلمات کلیدی: درصد پوشش، داده‌کاوی، شاخص‌های سنجش از دور، سنجنده ETM^+