Journal of the Earth and Space Physics, Vol. 46, No. 4, Winter 2021, P. 231-246 (Research)

DOI: 10.22059/jesphys.2020.299445.1007199

Data Fusion and Machine Learning Algorithms for Drought Forecasting Using Satellite Data

Mokhtari, R.[1] and Akhoondzadeh, M.[2*]

1. M.Sc. Graduated, Remote Sensing Division, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran

2. Assistant Professor, Remote Sensing Division, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran

(Received: 18 March 2020, Accepted: 29 Sep 2020)

Abstract

Drought is one of the natural disasters in the world, which is associated with various global factors, most of which can be observed using remote sensing techniques. One of the factors affecting agricultural drought is the vegetation associated with other drought-related factors. These parameters have a complicated relationship with each other, so machine learning algorithms can be used to predict better and model this phenomenon. Factors considered in this study include vegetation as the most critical factor, Land Surface Temperature (LST), Evapo Transpiration (ET), snow cover, rainfall, soil moisture these are derived from the active and passive sensors of satellite sensors as the products of LST, snow cover and vegetation using images of products of the MODIS sensor, rainfall using the images of the TRMM satellite, and soil moisture using the images of the SMOS satellite during a period from June 2010 to the end of 2018 for the central region of Iran. After that, primary processing was performed on these images. The vegetation index (NDVI) is modelled and predicted using an Artificial Neural Network algorithm (ANN), Support Vector Regression (SVR), Decision Tree (DT), Random Forest (RF) for monthly periods. By using these methods we have been able to present a model with desirable accuracy. The ANN approach has provided higher accuracy than the other three algorithms. Also, an average accuracy with RMSE=0.0385 and R^2=0.8740 was achieved.

Keywords: Drought, Machine learning, TRMM, MODIS, SMOS.

1. Introduction

Drought is a natural crisis that can occur intermittently in any area and any climate. This phenomenon, unlike other natural disasters, occurs gradually over a relatively long period, the effects of which can last for several years (Kogan, 2000). The drought phenomenon can have devastating and damaging effects on various factors, including human societies, the environment, and the climate of the region. Therefore, studying and monitoring this phenomenon in countries that experience frequent natural disasters is an obvious necessity (Zhang and Jia, 2013). Nowadays, the use of remote sensing techniques has closer oversight on this phenomenon. Images and remote sensing data from the meteorological data continuously obtain spatial information. Another advantage of this type of data is the spatial and temporal resolution of this type of data (Heumann, 2011). The drought phenomenon can be divided into four categories, including meteorological drought, agricultural drought, hydrological drought, and social drought (Wilhite and Buchanan-Smith, 2005). Agricultural drought refers to vegetation, and when the soil moisture content is lower than the amount of water needed for plant growth and health, and vegetation is weaker than previous periods in the area and drought occurs (Szalai and Szinell, 2000). There are different methods for studying and monitoring drought in time and space, one of which is the use of drought indices (Kogan, 1995). Numerous studies have been conducted on a variety of drought indices using satellite data, including vegetation and thermal data in various regions of the world. However, there are still significant challenges in increasing accuracy, better predicting this phenomenon. This phenomenon is mostly nonlinear, while most studies use linear models (Bai et al., 2018).

In a study, a time series of Nonlinear Aggregated Drought Index (NADI) was developed using precipitation data in meteorological stations in Australia, then by using two neural network methods, a direct

*Corresponding author: makhonz@ut.ac.ir

multistep neural network and recursive multistep neural network, have been forecasting this index for up to six months. The results have shown that both ways were the most accurate one-month prediction, and in the two-month and three-month periods of Direct Multistep Neural Network method has provided better accuracy (Barua et al., 2012). In another study, the Standardized Precipitation Index (SPI) obtained using precipitation data at meteorological stations, and then by using three machine learning methods including Artificial Neural Network algorithm (ANN) method, Support Vector Regression (SVR) and Wavelet-Transform Neural Network (WA-ANN) predicted this time series over three and six months (SPI) periods. The results showed that the wavelet neural network technique was better than the other two methods. Belayneh et al. (2014) produced SPI using precipitation data in Ethiopia and then implemented machine learning techniques including WA-SVR and WA-ANN. The results have shown that the model (WA-ANN) has shown better results (Belayneh and Adamowski, 2013). In another study, using MODIS data (land surface temperature, NDVI and evapotranspiration) and TRMM data (precipitation) from 2000 to 2012 in the United States, SPI was produced using random forest, boosted regression trees, and Cubist algorithms. The results show that the random forest method has been able to perform better modeling than the other two techniques. In this research, SPI is modeled as a meteorological drought index in meteorological stations. However, both Cubist algorithms and boosted regression trees have failed to perform the modeling well. Also, the type of study area has been influenced by the climate modeling process, so the type of study area is one of the essentials of this research (Park et al., 2017). In another study using meteorological precipitation data as well as satellite data such MODIS sensor and produce NDVI and Normalized Difference Water Index (NDWI) to predict the time series of SPI index by combined wavelet and neural network conversion method. Also, the combined method of wavelet transforms and support vector regression is performed in this study. The results have shown that both ways have been able to produce good results. Therefore, this study also forecasts precipitation in meteorological stations (Alizadeh and Nikoo, 2018). Another study used precipitation data using TRMM satellite data and MODIS snow cover to model and to predict the MODIS vegetation index. However, this study did not provide effective data such as soil moisture, evapotranspiration and land surface temperature. Also, in this study, time steps have been selected by trial and error, so this may not be for the appropriate time steps (Mokhtari and Akhoondzadeh, 2020).

Most of the studies using the meteorological data have been done to model and predict the precipitation in the meteorological stations. However, it is possible to increase the variety of data types to improve the accuracy of drought prediction using remote sensing data. An increased understanding of drought can be achieved by increasing the variety of satellite data types. The most crucial factor to consider in agricultural drought is vegetation indices, including NDVI. This indicator can show the status of vegetation, and when it is achieved over a long period, changes can be made in vegetation status in that area. The first type of drought is the meteorological drought. This is because reduced rainfall in a particular area causes other types of drought. Therefore, rainfall is one of the factors affecting the occurrence of agricultural drought. Soil moisture is an essential component of the water cycle that plays an important role in monitoring and predicting drought. Soil moisture is one of the causes of agricultural drought, and evapotranspiration is an essential component of the water and energy cycle. This factor can indicate the availability of water and the amount of moisture and its consumption by the plant. Therefore, evapotranspiration plays a very vital role in drought severity. Snow is a natural source of water, and snowmelt forms a significant part of the runoff. Shortage of snow in winter can lead to hydrological and agricultural drought; therefore, snow monitoring can be used in forecasting drought. Drought stress can be studied using surface brightness temperature. Land surface temperatures can provide valuable information on soil moisture content. Consequently, it can also affect the agricultural drought. Therefore, the factors of precipitation, snow, vegetation, land

surface temperature, evapotranspiration, and soil moisture are related and can affect each other. Satellite imagery and data can obtain said factors accurately worldwide.

In this study, TRMM satellite precipitation product, MODIS sensor snow cover product, MODIS sensor vegetation cover product, MODIS evapotranspiration product, MODIS land surface temperature product, and SMOS satellite soil moisture product for modelling, and prediction of the NDVI, which is one of the most important parameters of agricultural drought, is used. The above-mentioned data have been collected for approximately nine years from June 2010 to the end of 2018. As mentioned in previous studies, drought-related factors have deep and complex relationships. Therefore, machine learning methods can be used to model and predict this phenomenon. The machine learning methods used in this study is an ANN method, SVR, DT, and RF. To obtain a more accurate prediction of each time series before entering the machine learning algorithms using wavelet transform each mentioned signal time series is investigated.

2. Study Site and Materials

2-1. Study Site

The study area is the Islamic Republic of Iran. Geographically, Iran is located in the Middle East and borders the Persian Gulf, the Oman Sea, and the Caspian Sea. Iran is bordering with Iraq, Turkey, Armenia, Azerbaijan, Turkmenistan, Afghanistan, and Pakistan. The desert and semi-desert regions occupy more than half of the country. About one-third of Iran is mountainous, and a small part of Iran consists of the plain of the Caspian Sea and the plain of Khuzestan. Iran has high climate diversity. From the north to the south of the country, we are gradually facing different climate zones. In terms of rainfall, Iran is one of the arid and semiarid climates. In the west and northwest of Iran, winters are cold with heavy snowfall and rainfall, spring and fall are relatively mild, while summers are dry and hot. In the south and east of Iran, summers are scorching, and winters are mild. In northern Iran, the weather is different from other parts of Iran with heavy rainfall in almost all seasons of the year (Modarres, 2006). Figure 1 shows the geographical location of the study area.

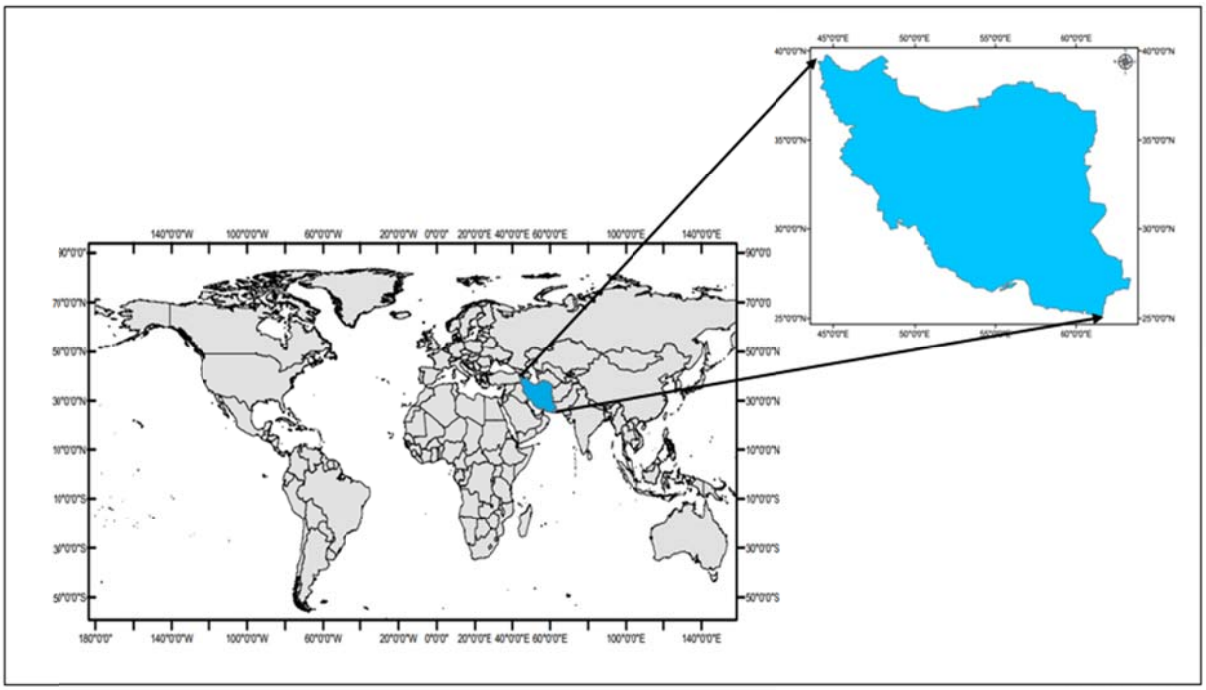

Figure 1. The geographical location of the study area.

2-2. Materials

In this study, the information and data of the MODIS sensor, SMOS satellite, and TRMM satellite were used. Therefore, a brief explanation of these three is given. Moderate Resolution Imaging Spectroradiometer (MODIS) is a sensor mounted on Terra satellite. This sensor captures the entire surface of the Earth's once every two days and obtains data in 36 spectral bands with different spatial resolutions (Sánchez et al., 2016). MODIS sees many of Earth's vital signs. Soil Moisture and Ocean Salinity (SMOS) is a satellite that forms part of ESA's Living Planet Programme launched in November 2009. The satellite's two main objectives are to monitor the surface soil moisture with a four percent accuracy at 35-50 km spatial resolution and monitor sea surface salinity with an accuracy of 0.1 PSU at 10- to 30-day average with a spatial resolution of 200 km (Kerr et al., 2010). The Tropical Rainfall Measuring Mission (TRMM) was a joint space mission between the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA) designed to monitor and study tropical rainfall. The TRMM was a joint mission between NASA and the Japan Aerospace Exploration Agency (JAXA) that designed to monitor and study tropical rainfall. The satellite was launched in Japan in November 1997 and is still in orbit. This satellite is a joint product of Japan and the USA. The satellite is 350 km above the Earth's surface, and its products are for a range from 50 degrees south to 50 degrees north. According to data provided by NASA, the product's spatial resolution is at least 0.25 by 0.25 degrees and maximum by five at five degrees (Duan and Bastiaanssen, 2013).

MODIS Vegetation Index Products (MOD13A3), produced on 1-month intervals and at multiple spatial resolutions, provides consistent spatial and temporal comparisons of vegetation canopy greenness, a composite property of leaf area, chlorophyll, and canopy structure. This product has a spatial resolution of 1 km, which is known as Level 3 product of MODIS. The MODIS Global Vegetation Indexes are designed to provide a spatial and temporal comparison of vegetation conditions. This product has a spatial resolution of 1 km, produced on 1-month intervals, which is known as Level 3 product of MODIS that provides regular spatial and temporal comparisons of the intensity of vegetation (Sánchez et al., 2016). The images were downloaded from https://earthexplorer.usgs.gov.

MODIS Land Surface Temperature (MOD11A2) is version 6 of the MODIS sensor, which is delivered every eight days with a resolution of one kilometer. The value of each pixel is obtained as a simple average of over eight-days (Sánchez et al., 2016). The images were downloaded from https://earthexplorer.usgs.gov.

MODIS Evapotranspiration Products (MOD16A2) is a universal product that evapotranspiration product can be used to calculate energy balance and regional water, soil water status, which is produced in eight days. This product has a spatial resolution of 500 m (Ramoelo et al., 2014). The images were downloaded from https://earthexplorer.usgs.gov.

MODIS Snow Cover (MOD10CM) is monthly average of snow covers in 0.05 degree (approx. 5 km) resolution Climate Modeling Grid (CMG) cells. The images were downloaded from https://search.earthdata.nasa.gov.

Monthly Precipitation Estimates TRMM (3B43) is TRMM satellite precipitation at latitude 50 degrees north and 50 degrees south. The spatial resolution of this data is 0.25 degrees by 0.25 degrees. The unit of data is in millimetres per hour. The images were downloaded from https://disc.gsfc.nasa.gov.

Monthly Soil Moisture L3 corresponds to the spatial average of the L2 soil moisture measurements in the Equal-Area Scalable Earth (EASE)-2 grid of 25 km and one-month temporal averaging periods. The images were downloaded from http://bec.icm.csic.es/land-datasets.

3. Methods

In this study, using satellite data and machine learning methods to model and predict agriculture drought in the study area has been done. Initially, satellite images and data from relevant sites were downloaded and pre-processed for constructing the images. The

correlation between the available data and the correlation between the different time steps obtained, and the best time steps are selected to predict the NDVI index. Each of these factors is known as a serial signal, which is used by signal decomposition of the wavelet transform to improve the performance of each signal. Then, the prediction and modeling of agricultural droughts are performed using an ANN, SVR, DT, and RF method. The flowchart in Figure 2 shows the research process.

3-1. Correlation Analysis

Correlation coefficient is one of the statistical property for correlation between two variables. There are several types of correlation multiplication and each have their own definitions. Pearson correlation coefficient is one of the methods for determining the correlation between two parametric variables. This coefficient varies between -1 to +1. When this coefficient is close to zero, it indicates that there is no relationship between the two variables (Benesty et al., 2009).

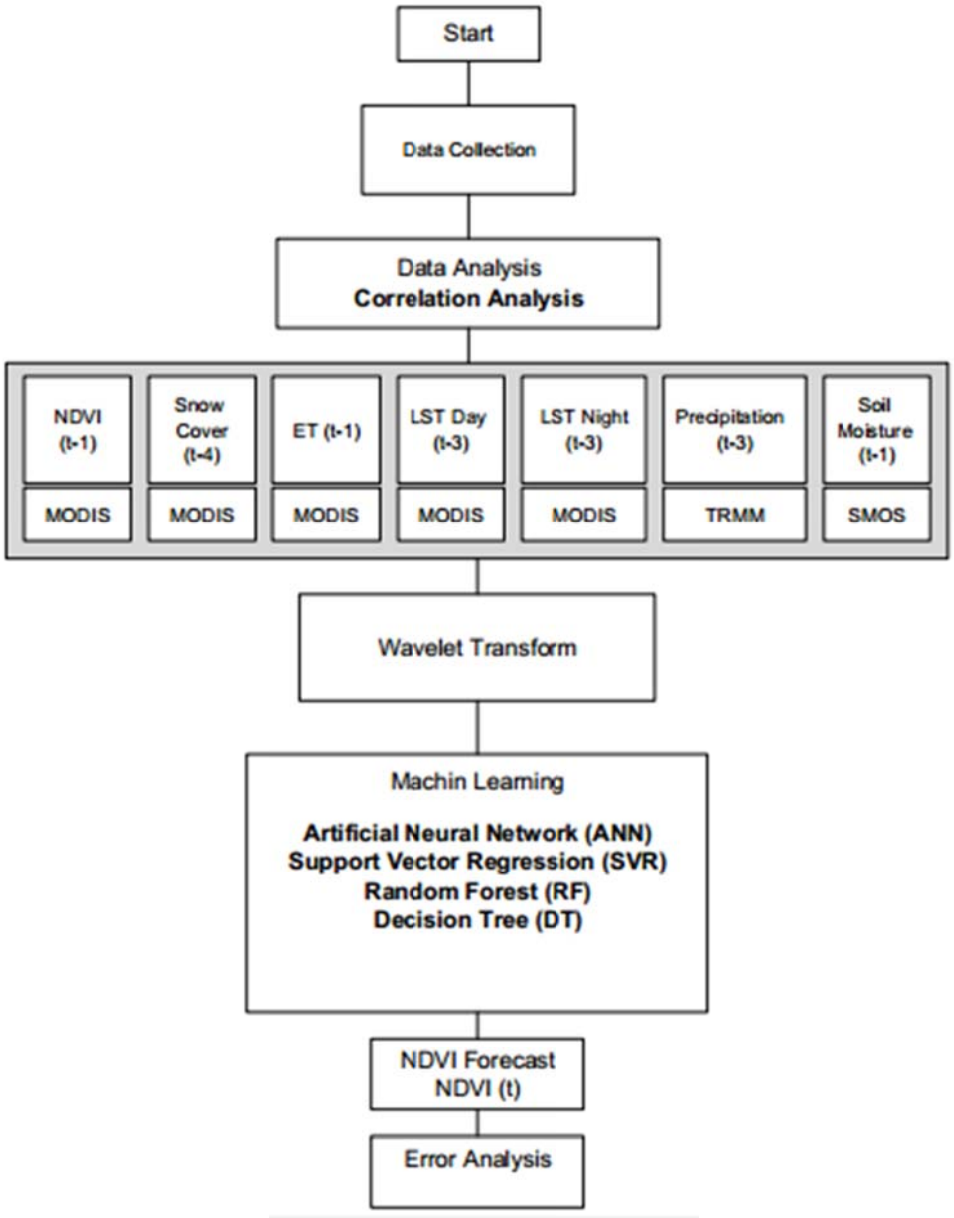

Figure 2. The flowchart of the research process.

3-2. Wavelet Analysis for Data Pre-Processing

One of the techniques used in signal processing is wavelet transform. Therefore, this conversion is used in one-dimensional, two-dimensional, and three-dimensional signal processing (Kim and Valdés, 2003). Wavelet transform is one of the achievements of mathematics that has important applications in engineering sciences today. This conversion is a more advanced type of Fourier transform. The wavelet function is a specified function with a mean of zero. The continuous wavelet transform (CWT) of a signal f(x) is defined as (Nason and Sachs, 1999):

$$\mathrm{CWT(a.b)}=\frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} f(x)\psi\left[\frac{x-b}{a}\right]dx \qquad (1)$$

where b is the scale parameter; a is the translation, ψ is the mother wavelet. Scale parameter means wavelet stretching or squeezing and denotes the amount or length of wavelength. Therefore, it is compacted on a small scale. When the scale has high values, it reduces the time resolution and increases the resolution of the frequency. To perform wavelet conversion on digital computers, it is necessary to discrete scale and translation parameters. This results in a discrete wavelet transform (DWT). the advantage DWT is that it is relatively simpler and requires less computational time. To create a DWT, it is sufficient to use discrete values of scale parameter and translation parameter (Nason and Sachs, 1999). Choosing an appropriate wavelet transform to solve a problem requires a sufficient understanding of the features of the candidate wavelet. In this research, the Daubechies wavelet transforms have been used. The Daubechies wavelets, as a family of orthogonal wavelets, they define a discrete wavelet transform. This type of wavelet transform is determined by the maximum number of vanishing moments for some given support (Daubechies, 1992).

3-3. Artificial Neural Network

The structure of neural networks is modeled from the biological network of the human brain. The neuron is the smallest unit of network processing. Each neuron consists of two parts, one inlet and the other in weight. Weight is a parameter whose value can be at most one and at least zero. The inputs are multiplied to the corresponding weight, then the weighted inputs are aggregated, and the result is transmitted as an input to all the neurons in the next layer. The ANN models used in this study have a feed-forward Multi-layer perceptron (MLP) architecture, which was trained with the back propagation algorithm (Belayneh et al., 2014). MLP consists of an input layer with multiple input elements, a hidden layer with multiple neurons, and an output layer called the target layer. The ANN used in this study can be represented by (Kim and Valdés, 2003):

$$Y_k'(t) = f_0\left[\sum_{j=1}^{m} w_{kJ}\, f_n\left(\sum_{i=1}^{n} w_{ji}\, x_i(t) + \left(w_{j0}\right)\right) + w_{k0}\right] \qquad (2)$$

where m is the number of neurons in the hidden layer, i is the input element, j is the hidden neuron, N is the number of samples, $x_i(t)$ is the input variable at time step t, w_{ji} is the weight that connects the i^{th} neuron in the input layer and the j^{th} neuron in the hidden layer; w_{j0} is bias for the jth hidden neuron; f_n is the activation function of the hidden layer; w_{kj} is the weight that connects the j^{th} neuron in the hidden layer and k^{th} neuron in the output layer; w_{k0} is bias for the k^{th} output neuron; f_0 is the activation function for the output neuron and $Y_k'(t)$ is the forecasted k^{th} output at time step t (Belayneh et al., 2014). The performance of the artificial neural network depends on the network architecture. The performance of the artificial neural network depends on the network architecture, different network architecture can provide different accuracy. For example, the number of neurons in the hidden layer needs to be optimized. This can be done by trial and error.

3-4. Support Vector Regression

Support vector machine (SVM) is a tool for classifying data in different classes. Therefore, the data can be separated linearly and non-linearly. Here we refer to another type of support vector machine algorithm that can be used in regression problems. Support Vector Regression (SVR) is a machine learning algorithm that can be used in regression problems. SVR models are

created based on the structural risk minimization principle. While neural network methods are based on the empirical risk minimization (Cortes and Vapnik, 1995). To solve a nonlinear regression problem, it can transmit the input data space to the feature space using the kernel, where linear regression can be applied:

$$f(x)=\sum_{i=1}^{n} w_i \phi_i(x)+b \quad (3)$$

where w_i is the weight factor, b is the bias term, ϕ_i denotes aset of non-linear transformation functions in the feature space. The purpose of the SVR algorithm is to estimate the regression function f(x). Detailed descriptions of SVR model development can be found in (Cimen, 2008). The SVR algorithm has two versions, including nu-SVR and epsilon-SVR. In the nu-SVR version, there is control over the number of data vectors that are considered as support vectors, whereas in the epsilon-SVR version, it is not possible. LIBSVM tool is one of the tools by which the problems can be predicted and modelled using the SVR algorithm. The LIBSVM tool supports both versions of SVR, including epsilon-SVR and nu-SVR. The nu-SVR model is used in time series and modelling problems (Chang and Lin, 2001). In this study, LibSVM tool and nu-SVR version were used for modeling and prediction using the SVR algorithm. The three basic parameters that are predicted in method nu-SVR are C, Gamma, and nu. For each leading month forecast, the parameters C, Gamma, and nu are optimized through a grid search method for getting the best RMSE for the validation data set (Zhang et al., 2014).

3-5. Decision Tree

Decision trees are commonly used in various research and operations, meaning that they can make a particular decision. In this study, the Classification and Regression Trees (CART) algorithm is used. CART is one of the algorithms in classification and regression, which acts as a tree hierarchy in the input space. This method use for discrete and continuous variables so it can be used for regression and classification applications. Each decision tree consists of a series of leaves, branches, and nodes. Depending on the data type of the experiment, a series of tests start along with the decision nodes from the root node and traverses the tree path to the leaf. In the leaf, the problem is predicted. The CART algorithm can easily be used for both nominal classification and regression problems (Ahmed et al., 2010).

3-6. Random Forest

Random forest is one of the ensemble learning methods for regression and classification problems (Breiman, 2001). The random forest is a set of decision trees that grow randomly beneath the feature space. Random forest is based on a set of decision trees CART. Each tree separately makes predictions for the regression and classification problems and finally on the classification problems based on the most votes and the regression problems based on the mean of the tree answers. In the random forest algorithm, the parameters affect the efficiency and accuracy of the algorithm. The most influential parameters can be the number of decision trees, the number of variables used in each node, and the maximum number of observations allowed in each node. Choosing inappropriate amounts of any of these parameters can lead to a reduction in the accuracy, and consequently, duplicate decisions are made. Because by increasing variables, the likelihood of using duplicate variables increases the likelihood of duplicate decision making (Breiman, 2017).

3-7. Performance Result

To assess the performance of the ANN, SVR, DT, and RF models, four statistical performance evaluation criteria are used: Coefficient of determination (R2), Root Mean Squared Error (RMSE), Mean Squared Error (MSE) and Mean Absolute Error (MAE). The R2 measures the degree of linear correlation between the observed variable and the predicted variable (Belayneh et al., 2014).

3-8. Pre-Processing

As mentioned above, to model and predict vegetation index (NDVI), which is one of the most important factors of drought agriculture, several parameters have been collected and used, including daytime and night-time land surface temperatures, evapotranspiration, soil moisture,

precipitation, snow cover from November 2010 to the last month of 2018. Initial pre-processing has been done on the above data and all data have been resampled with a spatial resolution of one kilometer.

4. Results and Discussion

After the pre-processing, the time series of each data such as vegetation index, land surface temperature during the day and night, evapotranspiration, soil moisture, precipitation and snow cover were obtained for the whole study area. Figure 3 shows the time series of the means of the total study area of each data as well as the linear regression of each of the graphs. According to these plots, the amount of vegetation index in the time period, land surface temperature, evapotranspiration and soil moisture have a positive growth, precipitation and snow cover have a negative growth.

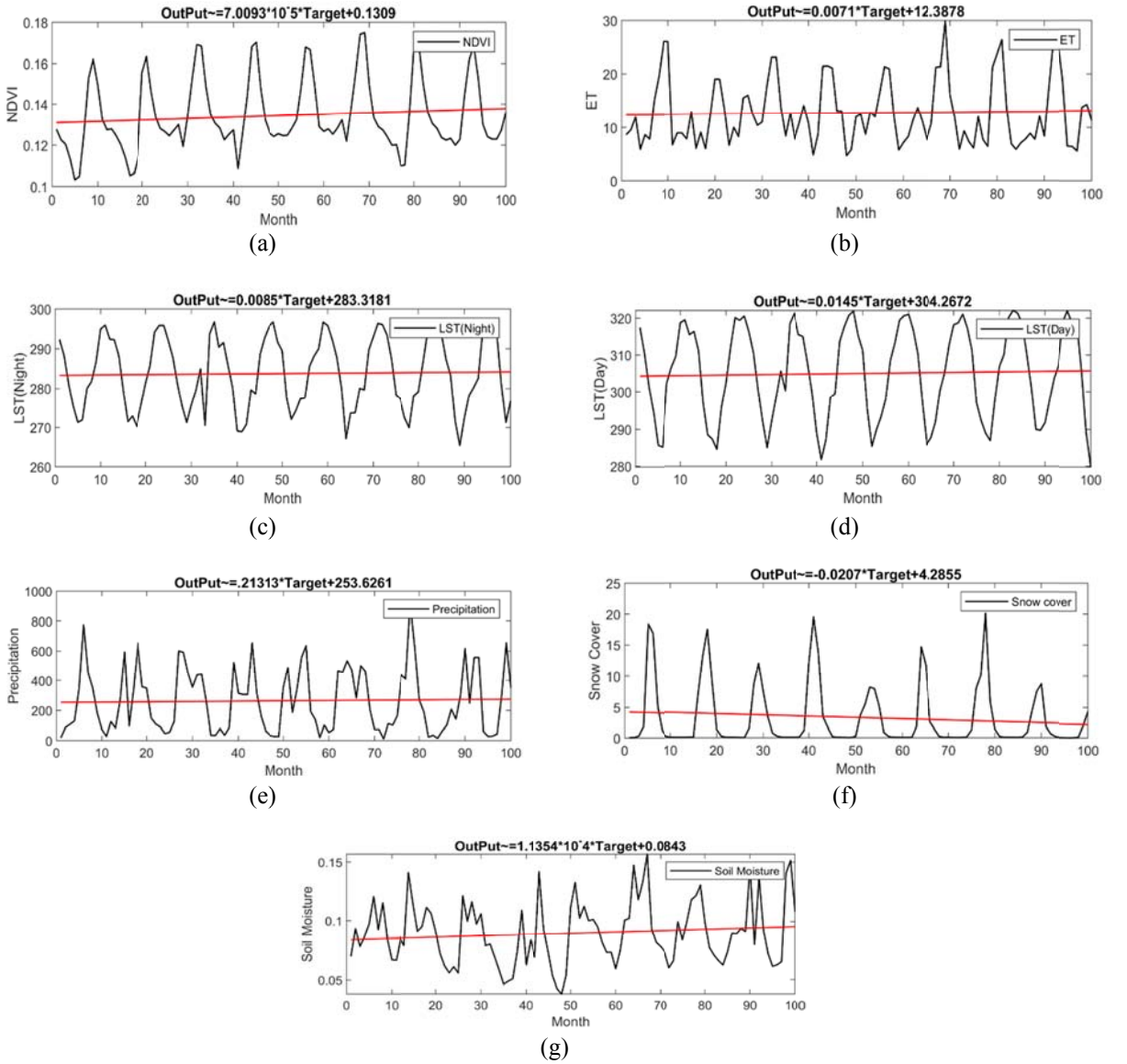

Figure 3. a) time series of NDVI index, b) the time series of ET, c) the time series of LST_Day, d) the time series of LST_Night, e) the time series of precipitation, f) the time series of snow cover, and g) the time series of soil moisture.

After that, the correlation between the data, and the time steps were performed. Then, using the wavelet transform to increase accuracy in modeling and prediction, signal decomposition is performed. Then different inputs are entered as inputs of machine learning algorithms such as ANN, SVR, DT, and RF, and the expected output is the (NDVI) index for the next month. The correlation coefficient between the time series steps of the parameters used is given in Table 1. According to Table 1, the correlation coefficients for the different steps are different, so that for the precipitation data, the highest correlation is related to the three-month step. However, the correlation coefficient for this factor is positive at all steps indicating that it has a direct effect on each factor with increasing vegetation index, and precipitation. In the snow cover data, the highest correlation coefficient in the four-month step is also negatively correlated in the one-month and two-month steps. In the presence of snow, vegetation index decreases sharply, especially in the cold months of the year. The correlation coefficient between the land surface temperature at day, night and vegetation are always negative, which is the highest in the three-month steps. It can also be explained that the correlation coefficient between these two parameters is negative, as vegetation decreases with increasing LST. The correlation between soil moisture and vegetation has a positive value and in the one-month step, has its highest value. The correlation between Evapotranspiration and vegetation has a positive value and in the one-month step has its highest value.

Using the wavelet transform, the primary signal can be hierarchically decomposed at N-level sub-series into subsets of approximation and detail. Therefore, in this study, each of the time-series data was analysed up to decomposed into 5-level sub-series using the wavelet transform, and then each sub-series was assigned to machine learning algorithms. Accuracy calculated for each step, and the results show that in the 1-level, the signal decomposition using wavelet transform is achieved with the highest accuracy. After one-level signal decomposition using wavelet transform, each of the inputs as shown in Figure 4 inputs to machine learning algorithms.

Table 1. The correlation coefficient between the time series steps of the parameters with NDVI.

Correlation coefficient with NDVI	One-month step	Two-month step	Three-month step	Four-month step	Five-month step	Six-month step
Precipitation	0.3473	0.5879	0.7189	0.4193	0.3526	0.1126
Snow Cover	-0.3847	-0.1573	0.7653	0.7823	0.3822	0.0200
Evapotranspiration	0.8324	0.3654	0.0365	0.1818	0.2120	0.2558
LST (day)	-0.3269	-0.4436	-.8326	-.7426	-.6024	-.2194
LST (night)	-0.3172	-0.3724	-0.5626	-0.5363	-0.4120	-0.1388
Soil Moisture	0.5224	0.3836	0.3324	0.2812	0.1211	0.0112

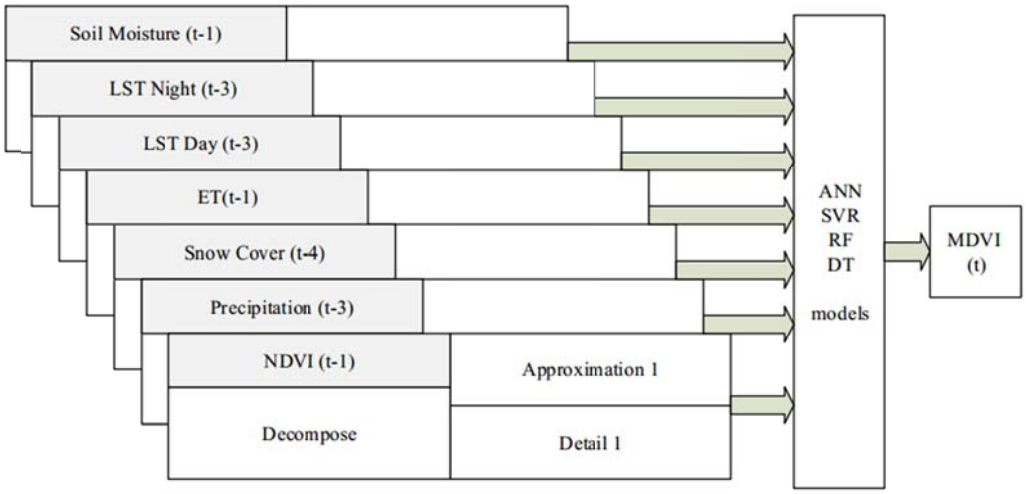

Figure 4. Inputs to machine learning algorithms.

Table 2 shows the accuracy of the artificial neural network algorithm in 2018. According to Table 2, the results show that the accuracy estimates for the forecast of the NDVI for the twelve months of 2018 are different. The results show that for May, it was the least accurate, and for October, it was the most accurate for the three parameters of RMSE, MSE, and MAE. However, the coefficient of R^2 was highest for October and the lowest for March.

Table 3 shows the accuracy of the support vector regression algorithm in 2018. According to Table 3, the results show that the accuracy estimates for the forecast of the NDVI for the twelve months of 2018 are different. The results also show that for May, it was the least accurate, and for October, it was the most accurate for the three parameters of RMSE, MSE, and MAE. However, the coefficient of R^2 was highest for October and the lowest for March.

Table 2. The accuracy of the algorithm the neural network method.

Month	RMSE	MSE	MAE	R^2	Month	RMSE	MSE	MAE	R^2
Jan	0.0364	0.0013	0.0157	0.8162	July	0.0445	0.0020	0.0210	0.8541
Feb	0.04	0.0016	0.0181	0.8177	Aug	0.0271	0.0007	0.0119	0.9374
Mar	0.0549	0.0030	0.0266	0.7795	Sept	0.0219	0.0004	0.0092	0.9565
Apr	0.0498	0.0025	0.0265	0.8742	Oct	0.0119	0.0003	0.0093	0.9655
May	0.0562	0.0032	0.0295	0.8722	Nov	0.033	0.0011	0.0152	0.8749
June	0.0495	0.0024	0.0231	0.8798	Dec	0.0373	0.0014	0.0184	0.8600

Table 3. The accuracy of the algorithm, the support vector regression method.

Month	RMSE	MSE	MAE	R^2	Month	RMSE	MSE	MAE	R^2
Jan	0.0393	0.0015	0.0174	0.7857	July	0.0479	0.0023	0.0225	0.8307
Feb	0.0432	0.0015	0.0195	0.7874	Aug	0.0294	0.0008	0.0129	0.9262
Mar	0.0594	0.0035	0.0288	0.7420	Sept	0.0236	0.0005	0.0099	0.9496
Apr	0.0535	0.0029	0.0284	0.8545	Oct	0.0206	0.0004	0.0101	0.9597
May	0.0598	0.0036	0.0315	0.8529	Nov	0.0353	0.0012	0.0170	0.8777
June	0.0533	0.0028	0.0248	0.8006	Dec	0.0402	0.0016	0.0198	0.8607

Table 4 shows the accuracy of the random forest algorithm in 2018. According to Table 4, the results show that the accuracy estimates for the forecast of the NDVI for the twelve months of 2018 are different. The results show that for May, it was the least accurate, and for October, it was the most accurate for the three indices of RMSE, MSE, and MAE. However, the coefficient of R^2 was highest for October and the lowest for March.

Table 5 shows the accuracy of the decision tree algorithm in 2018. According to Table 5, the results show that the accuracy estimates for the forecast of the NDVI for the twelve months of 2018 are different. The results show that for May, it was the least accurate and for October it was the most accurate for the three indices of RMSE, MSE, and MAE. However, the coefficient of R^2 was highest for October and the lowest for March.

Figure 5a is the NDVI index image of the vegetation product for the MODIS sensor for October 2018; Figure 5b is the NDVI index image obtained from the ANN algorithm model for October 2018; Figure 5c is the NDVI index image obtained from the SVR algorithm model for October 2018; Figure 5d is the NDVI index image obtained from the RF algorithm model for October 2018; and finally Figure 5e is the NDVI index image obtained from the DT algorithm model for October 2018. As mentioned earlier, for the four algorithms, October had the best accuracy over other months of the year.

Table 4. The accuracy of the algorithm, the random forest method

Month	RMSE	MSE	MAE	R^2	Month	RMSE	MSE	MAE	R^2
Jan	0.0419	0.0018	0.0186	0.7558	July	0.0516	0.0027	0.0243	0.8036
Feb	0.0462	0.0021	0.0209	0.7567	Aug	0.0316	0.0010	0.0139	0.9145
Mar	0.0640	0.0041	0.0311	0.7006	Sept	0.0252	0.0006	0.0106	0.9424
Apr	0.0577	0.0033	0.0307	0.8306	Oct	0.0221	0.0004	0.0108	0.9535
May	0.0664	0.0044	0.0345	0.8185	Nov	0.0382	0.0015	0.0184	0.8323
June	0.0571	0.0033	0.0267	0.8401	Dec	0.0433	0.0019	0.0213	0.8113

Table 5. The accuracy of the algorithm, the decision tree method.

Month	RMSE	MSE	MAE	R^2	Month	RMSE	MSE	MAE	R^2
Jan	0.0432	0.0019	0.0193	0.7407	July	0.0531	0.0028	0.0252	0.7925
Feb	0.0471	0.0022	0.0215	0.7472	Aug	0.0330	0.0011	0.0147	0.9069
Mar	0.0651	0.0042	0.0317	0.6902	Sept	0.0258	0.0006	0.0109	0.9395
Apr	0.0591	0.0035	0.0315	0.8227	Oct	0.0227	0.0005	0.0112	0.9508
May	0.0662	0.0044	0.0350	0.8198	Nov	0.0392	0.0015	0.0190	0.8239
June	0.0558	0.0034	0.0274	0.8319	Dec	0.0445	0.0020	0.0220	0.8008

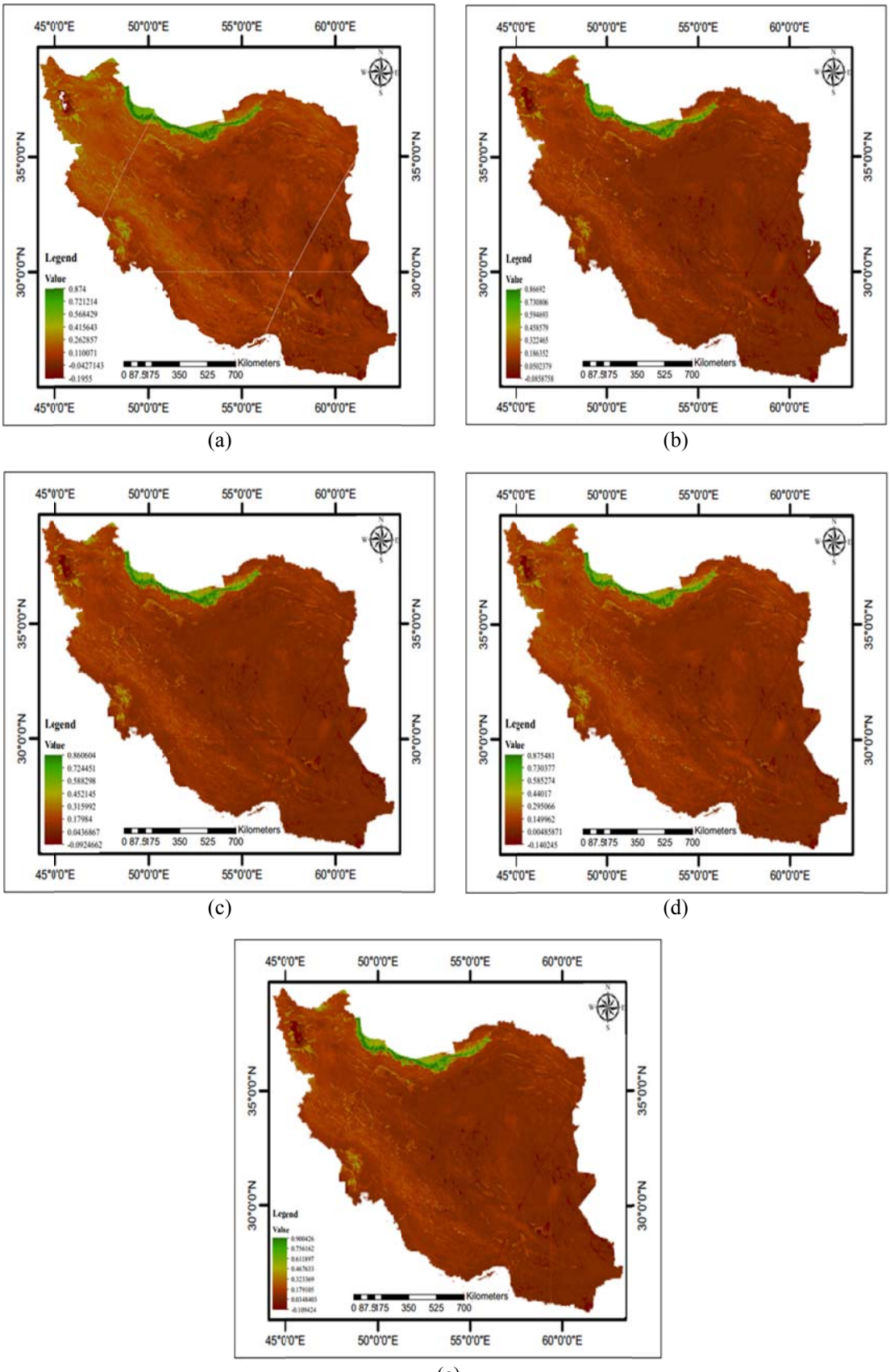

Figure 5. a) the NDVI index image of the vegetation product for the MODIS sensor, b) the NDVI index image obtained from the (ANN) algorithm, c) the NDVI index image obtained from the SVR algorithm, d) the NDVI index image obtained from the RF algorithm, e) the NDVI index image obtained from the DT algorithm for October 2018.

Figure 6a is a scatter plot of the observed and predicted NDVI values from the ANN model in October. Figure 6b is a scatter plot of the observed and predicted NDVI values from the SVR model. Figure 6c is a scatter plot of the observed and predicted NDVI values from the RF model. Figure 6b is a scatter plot of the observed and predicted NDVI values from the DT model. Each of the scatter plots shows that the true NDVI, and NDVI predicted values in each of the four algorithms are close to each other, also the coefficient of determination for the ANN method is better than the other three methods. However, three different methods have been able to provide high accuracy.

As the results show, the accuracy of all four algorithms is appropriate. However, the accuracy of the ANN algorithm is better than the other three algorithms. Therefore, the average accuracy for the twelve months from 2018 for ANN algorithm is RMSE = 0.0385 for SVR algorithm RMSE = 0.0421 for DR algorithm RMSE = 0.0454 and finally for RF algorithm is RMSE = 0.0462. The accuracy in October in the four algorithms is better than the other months of 2018 (Figure 7).

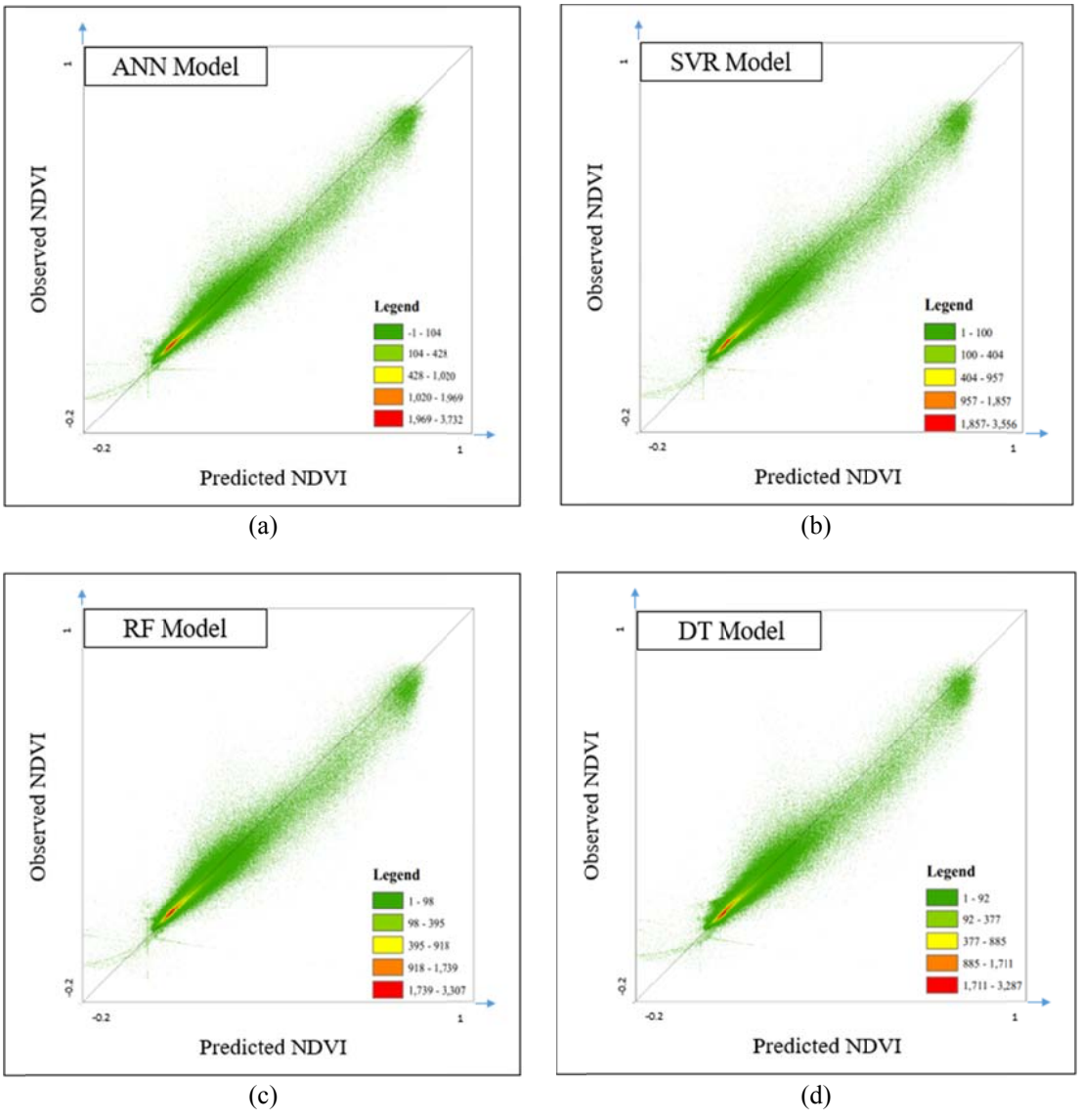

Figure 6. a) scatter plot of the observed and predicted NDVI values from the ANN model, b) from the SVR model, c) from the RF model, d) from the DT model.

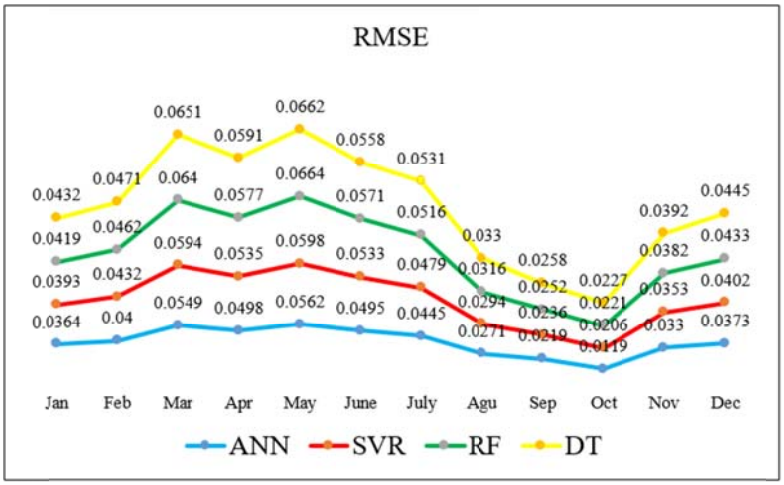

Figure 7. RMSE values in four algorithms.

5. Conclusion

Drought is a characteristic of the climate system, and this phenomenon occurs without regard to geographical boundaries as well as economic and political differences. Drought occurs in almost all climates. This phenomenon is a temporal anomaly and occurs in most parts of the world, especially in arid and semiarid regions. This can have devastating effects on crop yields, desertification, rangeland degradation, and vegetation degradation. Droughts can have adverse environmental impacts, including the loss of vegetation, deforestation, soil erosion, food shortages, and adverse social effects. Nowadays, many factors affect drought that can be investigated by remote sensing, and one can come to a better understanding of this phenomenon.

Since drought covers a wide area, monitoring it using remote sensing information provides a suitable tool for assessing drought crisis. Agricultural drought depends on a variety of factors such as land surface temperature, soil moisture, vegetation, precipitation, and evapotranspiration. These parameters have complex relationships and each has a different impact on the extent, and severity of the drought effect.

In this research, modeling of the NDVI index monthly was performed using TRMM rainfall data, SMOS satellite soil moisture, MODIS land surface temperature, snow cover, evapotranspiration with machine learning algorithms such as ANN algorithm, SVR algorithm, DT algorithm, and RF algorithm. One of the most critical problems we encountered in this study was the soil moisture images of the SMOS satellite available from mid-year 2010 to now. Therefore, all data used in this study are from mid-year 2010 to end-year of 2018. The correlation between the measured factors showed that the correlation coefficients were different at different time steps. Also, some factors, including surface temperature, are negatively correlated with vegetation; snow cover is positively correlated with vegetation over longer time intervals, and soil moisture, evapotranspiration, and precipitation are always positively correlated with vegetation cover. In this study, different time steps were used that were most correlated with vegetation. Wavelet transform has also been used to improve the accuracy of vegetation prediction. The ANN method, which is one of the methods of machine learning, has been able to provide acceptable results in this research. However, other methods of machine learning, such as support vector regression, random forest, have provided satisfactory results. The predicted results of vegetation cover for twelve months of the year 2018 showed that the use of different data such as land surface temperature, precipitation, soil moisture, and snow cover could play a significant role in vegetation prediction and modeling. However, the phenomenon of drought and vegetation depends on several factors, and identifying and measuring these factors is one of the most important challenges in drought studies. Suggestions can be made to influence the El Niño phenomenon and to incorporate its data

as inputs for modeling and predicting the drought in future studies.

Acknowledgements
The authors would like to thank European Space Agency (ESA), National Aeronautics and Space Administration (NASA), and USGS for providing TRMM data, SMOS data, and MODIS data.

References

Ahmed, N.K., Atiya, A.F. El Gayar, N. and El-Shishiny, H., 2010, An empirical comparison of machine learning models for time series forecasting, Econometric Reviews, 29 (5-6), 594-621.

Alizadeh, M.R. and Nikoo, M.R., 2018, A fusion-based methodology for meteorological drought estimation using remote sensing data. Remote Sensing of Environment, 211, 229-247.

Bai, J., Cui, Q., Chen, D., Yu, H., Mao, X., Meng, L. and Cai, Y., 2018, Assessment of the SMAP-Derived Soil Water Deficit Index (SWDI-SMAP) as an Agricultural Drought Index in China. Remote Sensing, 10(8), 1302.

Barua, S., Ng, A.W.M. and Perera, B.J.C., 2012, Artificial neural network–based drought forecasting using a nonlinear aggregated drought index. Journal of Hydrologic Engineering, 17(12), 1408-1413.

Belayneh, A., Adamowski, J., Khalil, B. and Ozga-Zielinski, B., 2014, Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. Journal of Hydrology, 508, 418-429.

Belayneh, A. and Adamowski, J., 2013, Drought forecasting using new machine learning methods/Prognozowanie suszy z wykorzystaniem automatycznych samouczących się metod. Journal of Water and Land Development, 18(9), 3-12.

Benesty, J., Chen, J., Huang, Y. and Cohen, I., 2009, Pearson correlation coefficient. In Noise Reduction in Speech Processing, 1-4, Springer.

Breiman, L., 2001, Random forests. Machine Learning, 45(1), 5-32.

Breiman, L., 2017, Classification and Regression Trees: Routledge.

Chang, C.-C. and Lin, C.J., 2001, LIBSVM: a library for support vector machines ACM Trans. Intell Syst Technol, 2(3).

Cimen, M., 2008, Estimation of daily suspended sediments using support vector machines. Hydrological Sciences Journal, 53(3), 656-666.

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

Daubechies, I., 1992, Ten Lectures on Wavelets. Vol. 61: Siam.

Duan, Z. and Bastiaanssen, W.G.M., 2013, First results from Version 7 TRMM 3B43 precipitation product in combination with a new downscaling–calibration procedure. Remote Sensing of Environment, 131, 1-13.

Heumann, B.W., 2011, Satellite remote sensing of mangrove forests: Recent advances and future opportunities. Progress in Physical Geography, 35(1), 87-108.

Kerr, Y.H., Waldteufel, P., Wigneron, J.-P., Delwart, S. Cabot, F. Boutin, J. Escorihuela, M.-J., Font, J., Reul, N. and Gruhier, C., 2010, The SMOS mission: New tool for monitoring key elements ofthe global water cycle. Proceedings of the IEEE, 98(5), 666-687.

Kim, T.-W. and Valdés, J.B., 2003, Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. Journal of Hydrologic Engineering, 8(6), 319-328.

Kogan, F.N., 1995, Droughts of the late 1980s in the United States as derived from NOAA polar-orbiting satellite data. Bulletin of the American Meteorological Society, 76(5), 655-668.

Kogan, F.N., 2000, Contribution of remote sensing to drought early warning. Early Warning Systems for Drought Preparedness and Drought Management, 75-87.

Modarres, R., 2006, Regional precipitation climates of Iran. Journal of Hydrology (New Zealand), 13-27.

Mokhtari Dehkordi, R. and Akhoondzadeh, M., 2020, Combining Neural Network and Wavelet Transform to Predict Drought in Iran Using MODIS and TRMM Satellite Data. Journal of

Geospatial Information Technology, 7(4), 175-191.

Nason, G.P. and von Sachs, R., 1999, Wavelets in time-series analysis. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 357(1760), 2511-2526.

Park, S., Im, J., Park, S. and Rhee, J., 2017, Drought monitoring using high resolution soil moisture through multi-sensor satellite data fusion over the Korean peninsula. Agricultural and Forest Meteorology, 237, 257-269.

Ramoelo, A., Majozi, N., Mathieu, R., Jovanovic, N., Nickless, A. and Dzikiti, S., 2014, Validation of global evapotranspiration product (MOD16) using flux tower data in the African savanna, South Africa. Remote Sensing, 6(8), 7406-7423.

Sánchez, N., González-Zamora, Á., Piles, M. and Martínez-Fernández, J., 2016, A new Soil Moisture Agricultural Drought Index (SMADI) integrating MODIS and SMOS products: a case of study over the Iberian Peninsula. Remote Sensing, 8(4), 287.

Szalai, S. and Szinell, C.S., 2000, Comparison of two drought indices for drought monitoring in Hungary—a case study. In Drought and Drought Mitigation in Europe, 161-166, Springer.

Wilhite, D.A. and Buchanan-Smith, M., 2005, Drought as hazard: understanding the natural and social context. Drought and Water Crises: Science, Technology, and Management Issues, 3-29.

Zhang, A. and Jia, G., 2013, Monitoring meteorological drought in semiarid regions using multi-sensor microwave remote sensing data. Remote Sensing of Environment, 134, 12-23.

Zhang, H., Chen, L., Qu, Y., Zhao, G. and Guo, Z., 2014, Support vector regression based on grid-search method for short-term wind power forecasting. Journal of Applied Mathematics.