Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

The Feasibility of Machine-Learning Methods to Extract the Surface Evaporation Quantity Using Satellite Imagery

E. Norouzi, S. Behzadi*

Surveying Engineering Department, Faculty of Civil Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran.

Article Info

Article History:
Received 27 October 2020
Reviewed 25 December 2020
Revised 19 January 2021
Accepted 21 March 2021

Keywords:
Climatic phenomena
Remote sensing
Machine learning
Decision tree
GIS

*Corresponding Author's Email Address:
behzadi.saeed@gmail.com

Abstract

Background and Objectives: Climate phenomena such as quantity of surface evaporation are affected by many environmental factors and parameters, which makes modeling and data mining difficult. On the other hand, the estimation of surface evaporation for a target station can be difficult as a result of partial or complete lack of local meteorological data under many conditions. In this regard, satellite imagery can play a special role in modeling and data mining of climatic phenomena, because of their significant advantages, including availability and their potential analysis. Therefore, addressing the improvement and expansion of machine learning methods and modeling algorithms along with remote sensing data is inevitable.
Methods: In this research, we intend to study the ability of 11 machine-learning modeling algorithms to model data and surface evaporation phenomena using satellite imagery. We used two methods to prepare the database: PCA and its opposite method using standard deviation and correlation.
Results: The calculation of the Root Mean Squared Error (RMSE) indicated that, in general, the use of the PCA method has a better result in preparing and reducing the dimensions of large databases for all methods of machine learning. The SEGPR model was ranked first with the least error (93.49%) in the Principal Component Analysis (PCA) method, and the Artificial Neural Network (ANN) model performed well in both data preparation methods (93.42, 93.38), and the Classification-Tree-Coarse model had the highest error in both methods (92.66, 92.67).
Conclusion: Consequently, it can be said that by changing the methods of database preparation in order to train models, the modeling results can be changed effectively.

Introduction

The loss of water from lakes, rivers, oceans, vegetation, and the earth, as well as man-made structures such as reservoirs and irrigation conduits, is a major concern of hydrologists and irrigation specialists. This loss, compounded by the lack of usable water in some areas, indicates a need for field and laboratory research that will contribute to the understanding of the processes and parameters that comprise and contribute to evaporation [1]. Since the role of environmental variables in the amount and distribution of surface evaporation is undeniable, and the difference in the surface evaporation of neighboring points in small basins is justified by the involvement of local factors, consideration of these factors are important in modeling and analysis. In this regard, remote sensing technology plays an unparalleled role in obtaining surface evaporation data, because satellite images have advantages and privileges that the availability and analysis of them are the most important points [2]. Considering the importance of changing the future climate of the planet and the wide-reaching effects on various aspects of meteorological and hydrological issues on the planet, extensive efforts have been made to extract climate data more accurately as well as less costly, and fewer human and physical errors in measuring climate data in the future. On the other hand, evaporation is a non-linear process which occurs in nature due to temperature differences [3]; the changes in climate variables in regional scales are also not explicitly identified. Therefore, addressing the improvement and extension of machine learning methods and modeling algorithms along with remote sensing data is inevitable.

In interpolation methods, a good number of scattered terrestrial stations is required to be used in the target area, so in areas where these conditions are not available, one cannot expect an appropriate approximation for the result of interpolation and modeling [4]. In contrast, recently satellite imagery is widely available at very low cost and even free of charge [5]. Therefore, using remote sensing data to model climate data is a convenient and economical way. In recent times, the artificial intelligence approaches such as coactive neuro-fuzzy inference system (CANFIS), adaptive neuro-fuzzy inference system (ANFIS) which is hybrid of artificial neural networks (ANN) and fuzzy inference system (FIS), fuzzy-logic (FL) [6], radial basis neural network (RBNN) which is a type of ANN, support vector machines (SVM), generalized regression neural network (GRNN) which is a type of ANN, genetic algorithm (GA) [7], wavelet transformation (WT) and multi-layer perceptron neural network (MLPNN) have been significantly utilized in diverse fields such as modeling daily evaporation [3, 8-11]. In order to model the target data, the researchers used terrestrial station data as inputs, which have many limitations. So many researchers tend to use satellite imagery [2, 12-14]. In studying model trees and the neural network for modeling the rainfall, [15] concluded that tree models could be a suitable substitute for the neural network for precipitation modeling. Meike Kühnlein, in a study [16], on improving the accuracy of rainfall rates from optical satellite sensors with machine learning, showed that using machine learning methods can accurately improve the rates of precipitation even up to hourly rates. Doña et al [17] used remote sensing to estimate the temporal variation of the flooded area and their associated hydrological patterns related to the seasonality of precipitation and evapotranspiration. He applied several inverse modeling methods, such as two-band and multispectral indices, single-band threshold, classification methods, artificial neural network, support vector machine algorithm and genetic programming and the genetic programming approach yielded the best results, with a kappa value of 0.98 and a total error of omission-commission of 2%. Xu et al [13] up-scaled evapotranspiration from eddy covariance flux tower sites to the regional scale with machine learning algorithms. Five machine learning algorithms were employed for evapotranspiration upscaling including artificial neural network, Cubist, deep belief network, random forest, and support vector machine. They demonstrated that the artificial neural network, Cubist, random forest, and support vector machine algorithms have almost identical performance in estimating evapotranspiration and have slightly lower root mean square error than deep belief network at the site scale. On the other hand, in spite of the considerable importance of preparing the initial data to enter the model, in many cases it is not taken into account as much as required.

In this research, we intend to investigate the relationship between Landsat 8 satellite imagery and terrestrial station data for surface evaporation quantity using different machine learning methods, including Artificial Neural Network, Neuro-Fuzzy (ANFIS), Classification-Tree-Coarse, Classification-KNN-Coarse, and Regression-Tree-Medium, Regression-interactions linear, Fine Gaussian SVM, Medium Gaussian SVM, Gaussian Process Regression- SEGPR, Ensemble-Boosted Trees, and Ensemble-Bagged Trees. Then we compare the different methods to find the optimal method. This study was conducted with the following objectives: (i) selecting appropriate input variables combination for the models using two ways, one of them is Principal Component Analysis (PCA) and the other one is applying standard deviation and correlation as an indicator; (ii) calibrating and validate the heuristic models with selected input variables; and (iii) comparing the results from the listed models with those of the interpolation based models, IDW.

Technical Work Preparation

Study Area and Data Acquisition

The study area includes three provinces of Tehran, Alborz and Qazvin in Iran, which are located between 48 ° 43 ′ 38.83 ″E to 53 ° 09 ′ 11.70 ″E longitude and 34 ° 50 ′ 14.29 ″N to 36 ° 47 ′ 1.33 ″N latitude at an altitude of 1495.9m above MSL (Mean Sea Level) Fig. 1. In Fig. 1, the study area and the position of stations, which have the ability to measure surface evaporation, are shown on the region map. Daily surface evaporation data are collected from Meteorological Organization and Water Resources Management Organization of Iran.

The primary data source was a series of Landsat-8 satellite images. The spatial resolution of the band 8 for the satellite is 15 m; other bands (1,2,3,4,5,6,7,9,10,11) have a spatial resolution of 30 m. The selected Landsat-8 scenes are path 164 / row 35 and path 165/ row 35, which covers the extent of the area of interest. Scenes that were mostly cloud-free from 2013 to 2016 are obtained from the United States Geological Survey Earth Explorer website [18]. All images were Level 1T products, which have been precision and terrain corrected in the GeoTIFF format and are in the UTM Zone 39S projection and WGS datum [19]. The resulting dataset comprised 65 full scenes.

Preparing the Initial Datasets for Machine Learning

Machine Learning (ML) depends heavily on data. It is the most crucial aspect that makes algorithm training possible and explains why machine learning became so popular in recent years. Regardless of actual terabytes of information and data science expertise, if data records are not prepared and organized, a machine will be nearly useless or perhaps even harmful. All datasets are always needed correction. That's why data preparation is such an important step in the machine learning process. In a nutshell, data preparation is a set of procedures that helps make the dataset more suitable for machine learning. In broader terms, the data preparation also includes establishing the right data collection mechanism. These procedures consume most of the time spent on machine learning. Sometimes it takes months before the first algorithm is built [20, 21].

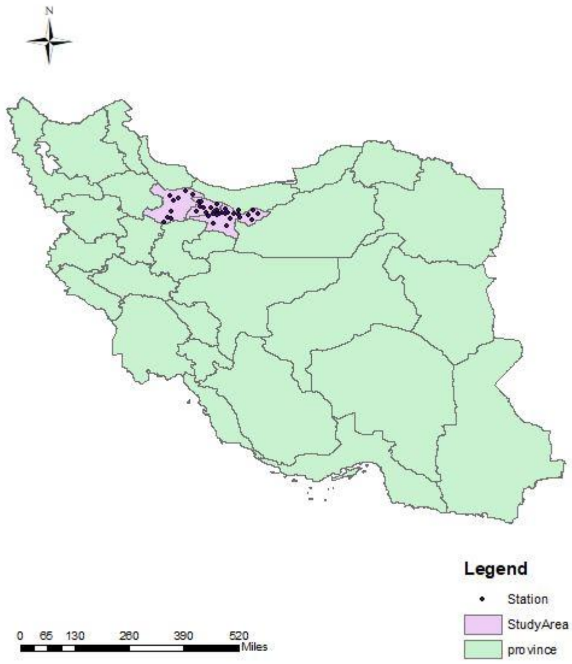

Fig. 1: Location map of the study area and Distribution of terrestrial stations.

Knowing what must be modeled or estimated will help to decide which data may be more valuable to collect. When formulating the problem, data exploration must be conducted, and it must be tried to think in the categories of classification, clustering, regression, and ranking. For instance, when an algorithm needs to answer binary yes-or-no questions, classification is the best method, or when it comes to finding the rules of classification and the number of classes, clustering is a suitable choice etc. Generally since the surface evaporation dataset is formed by numerical values, regression algorithm is more beneficial to this case [20, 21]; however, (i) it must be considered that surface evaporation depends on numerous factors which makes it too complicated to be formulated accurately; (ii) On the other hand, It is tempting to include as much data as possible. Since the target attribute (what value you want to model) is known, common sense will guide the further. It can be assumed which values are critical and which are going to add more dimensions and complexity to the dataset without any useful contribution. This approach is called attribute sampling. (iii) Since missing values can tangibly reduce prediction accuracy, this issue must be addressed as a priority. In terms of machine learning, assumed or approximated values are "more right" for an algorithm than just missing ones. Hence in this study, all the well-known methods of machine learning are almost applied with two statistical data preparation techniques, one of them is PCA and the other one is the way which uses standard deviation and correlation, in order to find the best algorithm for the purpose of the study.

A. Determining Optimal Bands Using PCA Technique

Major databases are increasingly expanding and publicizing, while making them more difficult to interpret. Principal Component Analysis (PCA) is a technique for reducing the size of such databases, increasing the capability of interpreting, and simultaneously minimizing data problem. The PCA technique does this by creating a new variable that maximizes the variance successively [22]. Table 1 illustrates the result of PCA technique. All bands of Landsat-8 images except for bands 8 and 11 were used to extract 4 optimal bands.

B. Determine Optimal Bands Using Standard Deviation and Correlation

Using standard deviation and correlation statistics, parameters that have greater correlation and amplitude than each other are determined, and then a number of optimal parameters is used instead of using all of the parameters. In this study it is done according to (1). This formula is set for three parameters but it is applied for five parameters (including four bands of satellite imagery and surface evaporation values):

$$x = \frac{(std(B1) + std(B2) + std(B3))}{|corr(B1, B2)| + |corr(B1, B3)| + |corr(B2, B3)|} \quad (1)$$

where x is the benchmark for optimization, B_i is parameter (bands and surface evaporation value), std and corr are standard deviation and correlation respectively. The x index in (1) was calculated for all possible states in selecting 4 among 9 bands of satellite images. The state which had maximum value of the optimization index is considered as the best quadr-combination of 9 bands of landsat-8 images. The selected state was the combination of bands 1, 2, 9, and 10. Table 2 shows the result of calculations.

Table 1: The result of the PCA technique. Extraction of four optimal parameters using nine satellite image bands and with respect to surface evaporation values (where B1, B2, … are bands of satellite images, SE is surface evaporation values and P1,P2,.. are PCA method outputs).

B1	B2	B3	B4	B5	B6	B7	B9	B10	SE(mm)
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
89.901	90.358	89.790	84.962	65.623	18.771	5.453	0.171	10.915	6.6
81.208	83.540	87.441	86.594	68547	19.050	5.144	0.144	9.718	3.1
85.637	86.940	83.352	76.810	68.990	15.933	4.429	0.165	10.383	10.9
75.756	75.191	70.831	74.459	64.291	15.270	3.951	0.114	8.206	3.5
100.999	104.870	111.148	111.205	80.569	23.728	7.140	0.190	12.126	11
108.146	110.718	107.014	97.934	66.238	17.571	5.264	0.199	12.254	15
108.220	106.200	97.357	83.510	80.723	14.908	3.963	1.700	8.051	8.7
79.493	82.053	85.722	82.855	65.064	17.944	4.952	0.208	8.961	2.6
74.560	72.920	71.325	73.404	58.903	14.413	3.984	0.116	11.620	10.6
61.609	58.271	53.319	52.317	40.409	10.218	2.820	0.082	9.775	5.4
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…

P1	P2	P3	P4	SE(mm)
…	…	…	…	…
…	…	…	…	…
-0.739	-0.756	-0.067	-0.061	6.6
6.164	-3.217	1.999	0.643	3.1
-0.287	1.138	-0.058	0.234	10.9
-0.413	0.443	-0.104	-0.069	3.5
-0.295	1.289	0.069	-0.171	11
-0.905	0.713	-0.082	0.743	15
-1.460	-0.177	-0.056	0.508	8.7
-1.549	-0.456	0.103	-0.158	2.6
-1.482	0.178	0.012	0.617	10.6
-1.680	0.234	0.261	0.051	5.4
…	…	…	…	…
…	…	…	…	…

Table 2: The result of calculations using equation (1). Extraction of four optimal bands using nine satellite image bands and with respect to surface evaporation values (where B1, B2, … are bands of satellite images, SE is surface evaporation values and the item specified in the middle table is a suitable combination of satellite imagery bands).

B1	B2	B3	B4	B5	B6	B7	B9	B10	SE(mm)
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
300.92	310.34	275.27	242.53	156.84	26.02	7.14	5.85	4.11	1.7
198.18	198.89	169.72	146.19	93.33	13.43	3.89	3.99	4.26	3.2
265.91	272.65	228.21	200.82	130.33	16.66	4.32	3.70	4.19	1.3
267.38	240.27	235.58	206.36	133.30	17.62	4.62	5.08	3.88	1.3
236.44	273.45	205.74	179.50	115.60	14.53	4.26	5.37	3.94	2.4
268.55	285.21	232.88	203.74	130.84	15.71	4.29	6.72	3.85	1.8
279.17	267.14	247.85	216.60	139.04	16.83	4.76	8.01	3.38	3
262.84	291.40	227.43	200.02	128.73	8.10	3.09	12.61	2.76	3.3
322.41	327.09	281.54	244.48	156.40	7.56	2.78	6.82	3.52	3.8
228.78	232.34	196.86	172.96	111.20	6.55	2.35	5.26	3.44	0.8
…	…	…	…	…	…	…	…	…	…

State	X
…	…
B1,B2,B4,B10,SE	197.45
B1,B2,B5,B6,SE	52.66
B1,B2,B5,B7,SE	48.90
B1,B2,B5,B9,SE	46.20
B1,B2,B5,B10,SE	167.80
B1,B2,B6,B7,SE	47.97
B1,B2,B6,B9,SE	55.83
B1,B2,B6,B10,SE	177.32
B1,B2,B9,B10,SE	282.454
B1,B3,B4,B5,SE	48.133
…	…

B1	B2	B9	B10	SE(mm)
…	…	…	…	…
…	…	…	…	…
300.92	310.34	5.85	4.11	1.7
198.18	198.89	3.99	4.26	3.2
265.91	272.65	3.72	4.19	1.3
267.38	272.16	3.70	3.88	1.3
236.44	240.27	5.08	3.94	2.4
268.55	273.45	5.37	3.85	1.8
279.17	285.21	6.72	3.38	3
262.84	267.14	8.01	2.76	3.3
322.41	327.09	6.82	3.52	3.8
228.78	232.34	5.26	3.44	0.8
…	…	…	…	…

Machine Learning Algorithms

In general, 11 machine learning methods were used to model the surface evapotranspiration using satellite imagery. In this section, each of which is briefly described:

A. Artificial Neural Network (ANN)

Neural network is one of the techniques of machine learning that its application in modeling and predicting many phenomena including climate phenomena has been proved in numerous studies. Sulaiman and Wahab [10] describe the modeling and prediction of heavy rainfall. Although it is difficult to model and predict climatic phenomena, machine learning methods, especially artificial neural networks, are reliable and it can be used for climate phenomena such as precipitation and surface evaporation [4].

B. Neuro-Fuzzy (ANFIS)

An artificial neural network based on the Takagi-Sugeno fuzzy system [23]. Since this system combines neural networks and fuzzy logic concepts, both of them can be used in the same frame [24].

C. Classification-Tree-Coarse

This technique is a tool to support decisions that use trees to model. The Decision Tree is commonly used in various research and operations. Specifically, in the decision analysis, it is used to identify the strategy that is most likely to reach the goal. Another use of Decision Trees is the description of conditional probability calculations [25].

D. Classification-KNN-Coarse

The KNN algorithm is one of the simplest data mining and classification algorithms. This algorithm performs simple classification operations and returns reliable results as predictions. In a literal sense, this method chooses the tracks in which the selected neighborhood has the highest number of records attributed to them. Therefore, traces that are more closely related to each other in the K nearest neighbor are considered as the new record category [26].

E. Regression-Tree-Medium

This method is one of the machine learning regression techniques that uses a decision tree to predict and model, and acts as a combination. Classification and regression tree algorithm is one of the widely used algorithms in water resources management related fields since it is easy to understand and interpretable prediction model [27, 28].

F. Regression-Interactions Linear

Linear regression is a modeling method and the relationship between an associated variable and one to several other variables that can be used to solve many real-world problems. Regression interactions allow for the detection of effect heterogeneity but require that heterogeneity is a linear function of an observed and reliably measured predictor of heterogeneity [29].

G. Fine Gaussian SVM

One of the most common methods in the data classification domain is the SVM algorithm or support vector machine. In simple terms, support vectors are a collection of points in the n-dimensional data that defines the boundaries of the categories, and the classification of the data is based on them, and by moving one of them, the output of the classification may be Change [30].

H. Medium Gaussian SVM

SVM is basically a binary separator. A multi-class pattern recognition can be achieved by combining two-class vector machines [30].

I. Gaussion Process Regression- SEGPR

Gaussian process consists of a set of random variables as one of the new methods of data mining, with its normal characteristics and using kernel functions, has a high ability to solve nonlinear problems. The Gaussian regression models are based on the assumption that the regulatory observation should carry information about each other. Gaussian processes are a way to specify the priority directly on the function space [31].

J. Ensemble-Boosted Trees

Instead of using only one decision tree, Ensemble methods use the combination of multiple decision trees to predict better performance. The main idea behind these models is that weaker learners combine to form a stronger learner [32]. Boosting is an ensemble technique to create a collection of predictors. In this technique, learners are learned sequentially with early learners fitting simple models to the data and then analyzing data for errors. In other words, consecutive trees (random sample) are fitted at every step, and the goal is to solve for net error from the prior tree.

K. Ensemble-Bagged Trees

Bagging is another Ensemble technique that is used when the goal is to reduce the variance of a decision tree [32]. Here idea is to create several subsets of data from training sample chosen randomly with replacement. Now, each collection of subset data is used to train their decision trees. As a result, we end up with an ensemble of different models. Average of all the predictions from different trees are used which is more robust than a single decision tree.

Results and Discussion

In this study, after preparing (or so-called GIS-Ready) the surface evaporation data for 40 stations, along with corrected reflectance values of their Landsat-8 satellite imagery, four optimal bands of the nine bands (1,2,3,4,5,6,7,9,10) are selected in Landsat-8 satellite imagery to continue the work. In this study two ways are applied: first one through the PCA method and the other one with using (1).

In each of these two methods the data is divided into two categories: Train and Test.

In the next step, in order to model the relationship between satellite images and surface evaporation values and generate a decision function, train data is introduced into each of the modeling methods described briefly in the previous section, then the simulator or decision function is obtained for each of them.

Since 11 machine learning methods have been applied, 11 functions are obtained for the PCA method, and then 11 functions for the second method (Optimal Index).

Models' accuracies were evaluated according to (i) the Root Mean Square Error (RMSE) statistic and (ii) compared with the generated map of the interpolation method.

The RMSE can be expressed as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(SE_{i,observed} - SE_{i,predicted})^2} \quad (1)$$

where N is the number of data points used in the study and SE represents surface evaporations.

A. Modeling Daily Surface Evaporations Using PCA Technique

In this part, mentioned machine learning methods were applied for estimating surface evaporations of 40 stations on the study area using 4 optimal bands obtained by the PCA method. In PCA method, all satellite image bands are used to select the optimal bands for modeling. Each band has a coefficient, and the stronger the bands are related to the data, the larger the coefficient. Table 3 gives the test results of the applied models in estimating surface evaporation data. RMSE values range from 7.9417 to 8.9458 mm for the 11 models. The minimum RMSE value were found for SEGPR algorithm (test set) while the Classification-Tree-Coarse model provided the worst accuracy. Therefore, according to RMSE values in this study, SEGPR is known as the best method among other applied algorithms for surface evaporation modeling. Figure 2 illustrates the error variation of the test results spatially for the SEGPR models. From the figure, it is clear that all the models generally provided similar accuracy.

Table 3: Test results of the applied models in estimating surface evaporation by using PCA technique

Machine Learning Algorithms	RMSE (mm)	100 - Normalized RMSE (%)
Artificial Neural Network	8.017072	93.42863
Neuro-Fuzzy (ANFIS)	8.158462	93.31274
Classification-Tree-Coarse	8.945852	92.66733
Classification-KNN-Coarse	8.77615	92.80643
Regression Interactions Linear	8.3724	93.13738
Regression-Tree-Medium	7.9729	93.46484
Fine Gaussian SVM	8.1792	93.29574
Medium Gaussian SVM	7.9944	93.44721
Squared Exponential GPR	7.9417	93.49041
Ensemble Boosted Trees	7.9928	93.44852
Ensemble Bagged Trees	8.0257	93.42156

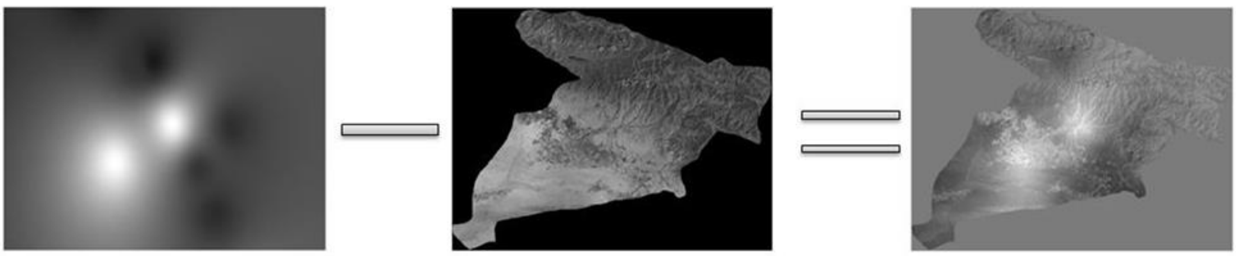

Fig. 2: The error variation of the test results spatially for the SEGPR model. From the left, the map obtained from the IDW interpolation method and the SEGPR machine learning algorithm and spatial distribution of errors.

B. Modeling Daily Surface Evaporations Using the X Index in (1)

At this stage, all the steps taken for output of the PCA technique are performed in the previous section for the optimal bands obtained from the index X (introduced in Section 2) method. The 11 machine learning methods are applied for estimating surface evaporations of 40 stations in the studied area using 4 optimal bands obtained by the index X. Table 4 shows test results of the applied models in estimating surface evaporation data.

The minimum and maximum RMSE values are 8.0649 and 8.9417 for Artificial Neural Network and Classification-Tree-Coarse methods, respectively. Therefore, according to RMSE values in this study, the neural network method is known as the best method among other methods for modeling surface evaporation.

Figure 3 illustrates the error variation of validation results spatially for the Artificial Neural Network model. It is also clear from the results of this section that all models generally offer the same precision.

Table 4: Test results of the applied models in estimating surface evaporation by using the x index

Machine Learning Algorithms	RMSE (mm)	100 - Normalized RMSE (%)
Artificial Neural Network	8.064963	93.38937
Neuro-Fuzzy (ANFIS)	8.120741	93.34365
Classification-Tree-Coarse	8.941777	92.67067
Classification-KNN-Coarse	8.482914	93.04679
Regression Interactions Linear	8.64715	92.91217
Regression-Tree-Medium	8.0694	93.38574
Fine Gaussian SVM	8.208877	93.27141
Medium Gaussian SVM	8.189189	93.28755
Squared Exponential GPR	8.1486	93.32082
Ensemble Boosted Trees	8.270296	93.22107
Ensemble Bagged Trees	8.17696	93.29757

Generally, calculating Root Mean Squared Error (RMSE) indicated that the use of PCA technique in preparing and reducing the dimensions of large databases has better results for all methods. On the other hand, in the method of Optim index factor (OIF), normalization is performed neither for remote sensing data nor the surface evaporations, but in the PCA, the normalization has been taken into account for only remote sensing data, and surface evaporation values are intact into modeling algorithms. Therefore, significant changes are seen in the mean square error of the PCA method compared to the OIF. Therefore, it is easy to see the effect of normalization and the use of the PCA method in preparing and reducing the size of large databases in Table 3 and Table 4. Most models have very close RMSE values, which proves that machine learning decision making models are valid in modeling climatic phenomena such as surface evaporation using remote sensing data, and applying these decision models for modeling and data mining is inevitable in the future.

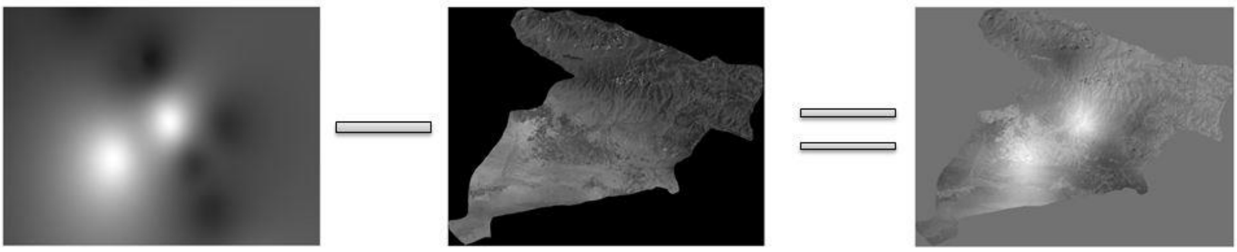

Fig. 3: The error variation of the test results spatially for the ANN model. From the left, the map obtained from the IDW interpolation method and the ANN machine learning algorithm and spatial distribution of errors.

In the method of OIF, the least error belongs to Artificial Neural Network method and the maximum error is related to Classification-Tree method. In the opposite manner, using the PCA technique, SEGPR with the least error and the Classification-Tree model still with the highest RMSE have the highest and lowest ratings, respectively. Artificial Neural Network (ANN) model has a good performance in both methods, and the Classification-Tree model has the highest error in both methods. Therefore, it is clear that by changing the methods of database preparation in order to train the models, the modeling results can be changed effectively.

Conclusion

With regard to the importance of changing the future climate of the planet and the wide effects on the various aspects of meteorological and hydrological issues, extensive efforts have been made in order to extract climatic data, more accurate and at the same time less costly and without human and physical errors in the future. On the other hand, the changes in climate variables in regional scales are not explicitly identified, they depend on a large number of local factors. Hence, addressing the improvement and expansion of machine learning methods and modeling algorithms using remote sensing data is inevitable. One of the outputs of this research is simulation models for data mining through satellite imagery, which is shown in Fig. 4. This figure shows an example of these products.

In this research, in order to study the appropriate methods for modeling and data mining for the surface evaporation, we employed the important methods of the machine learning and the time series of remote sensing and meteorological data and their integration, as well as the impact of the use of the methods like PCA and OIF which were used to prepare data due to training the models.

According to the results obtained in the previous section, the Artificial Neural Network model had acceptable result in both methods and it was quite evident that the impact of the methods of database preparation could be impressively significant. Since the discussion of data preparation in order to training modeling algorithms has not yet been sufficiently considered, and given the significant effect of this on the results of the obtained models, it is suggested to pay more attention to this issue in future studies.

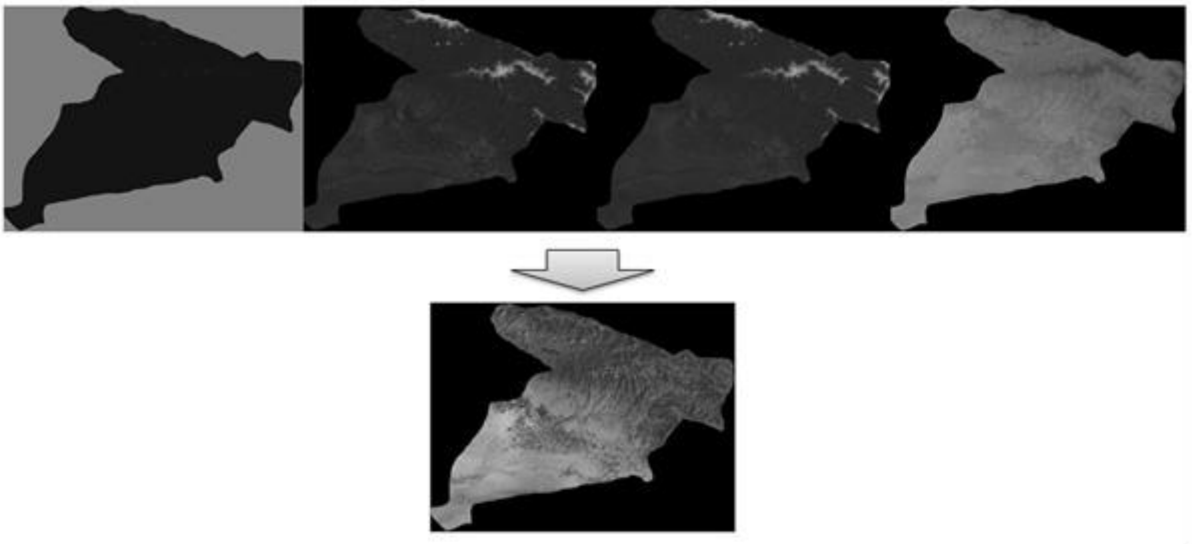

Fig. 4: An example of this study's products. Top, four optimized bands selected by using PCA technique, Down, Surface evaporation map prepared from these optimal bands through the ANN algorithm.

Author Contributions

E. Norouzi and S. Behzadi, designed the methodology. E. Norouzi collected the data. S. Behzadi carried out the data analysis. S. Behzadi and E. Norouzi interpreted the results and wrote the manuscript.

Acknowledgment

The authors gratefully thank the anonymous reviewers and the editor of JECEI for their useful comments and suggestions.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Abbreviations

PCA	Principal component analysis
ANFIS	Adaptive neuro fuzzy inference system
ANN	Artificial Neural Network
KNN	K-Nearest Neighbors
SVM	Support Vector Machines

GPR Gaussian Process Regression

RMSE Root Mean Square Error

References

[1] F.E. Jones, Evaporation of water with emphasis on applications and measurements: CRC Press, 2018.

[2] G.A. Martin, A.C. Kidd, S. Tsim, P. Halford, A. Bibby, N. A. Maskell, et al., "Inter-observer variation in image interpretation and the prognostic importance of non-expansile lung in malignant pleural effusion," Respirology, 25(3): 298-304, 2020.

[3] A. Malik, A. Kumar, O. Kisi, "Monthly pan-evaporation estimation in Indian central Himalayas using different heuristic approaches and climate based models," Comput. Electron. Agric., 143: 302-313, 2017.

[4] S. Behzadi, A. Jalilzadeh, "Introducing a novel digital train model using artificial neural network algorithm," Civ. Eng. Dimen., 22(2): 47-51, 2020.

[5] A. Jalilzadeh, S. Behzadi, "Machine learning method for predicting the depth of shallow lakes using multi-band remote sensing images," Soft Comput. Civ. Eng., 3(2): 59-68, 2019.

[6] S. Behzadi, Z. Mousavi, E. Norouzi, "Mapping historical water-supply qanat based on fuzzy method. an application to the Isfahan Qanat (Isfahan, Iran)," Int. J. Numer. Methods in Civ. Eng., 3(4): 24-32, 2019.

[7] S. Behzadi, M. Kolbadinejad, "Introducing a novel method to solve shortest path problems based on structure of network using genetic algorithm," Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., 2019.

[8] A. Malik, A. Kumar, S. Kim, M. H. Kashani, V. Karimi, A. Sharafati, et al., "Modeling monthly pan evaporation process over the Indian central Himalayas: application of multiple learning artificial intelligence model," Eng. Appl. Comput. Fluid Mech., 14(1): 323-338, 2020.

[9] L. Wang, B. Hu, O. Kisi, M. Zounemat-Kermani, and W. Gong, "Prediction of diffuse photosynthetically active radiation using different soft computing techniques," Q. J. R. Meteorolog. Soc., 143(706): 2235-2244, 2017.

[10] J. Sulaiman, S.H. Wahab, "Heavy rainfall forecasting model using artificial neural network for flood prone area," in IT Convergence and Security 2017, ed: Springer: 68-76, 2018.

[11] X. Lu, Y. Ju, L. Wu, J. Fan, F. Zhang, Z. Li, "Daily pan evaporation modeling from local and cross-station data using three tree-based machine learning models," J. Hydrol., 566: 668-684, 2018.

[12] S. Ghimire, R.C. Deo, N.J. Downs, N. Raj, "Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and Reanalysis atmospheric products in solar-rich cities," Remote Sens. Environ., 212: 176-198, 2018.

[13] T. Xu, Z. Guo, S. Liu, X. He, Y. Meng, Z. Xu, et al., "Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale," J. Geophys. Res.: Atmos., 123: 8674-8690, 2018.

[14] Q. Zhou, A. Flores, N.F. Glenn, R. Walters, B. Han, "A machine learning approach to estimation of downward solar radiation from satellite-derived data products: An application over a semi-arid ecosystem in the US," PLoS One, 12: e0180239, 2017.

[15] D.P. Solomatine, K.N. Dulal, "Model trees as an alternative to neural networks in rainfall—runoff modelling," Hydrol. Sci. J., 48: 399-411, 2003.

[16] M. Kühnlein, T. Appelhans, B. Thies, T. Nauss, "Improving the accuracy of rainfall rates from optical satellite sensors with machine learning—A random forests-based approach applied to MSG SEVIRI," Remote Sens. Environ., 141: 129-143, 2014.

[17] C. Doña, N.-B. Chang, V. Caselles, J. M. Sánchez, L. Pérez-Planells, M.D. M. Bisquert, et al., "Monitoring hydrological patterns of temporary lakes using remote sensing and machine learning models: Case study of la Mancha Húmeda Biosphere Reserve in central Spain," Remote Sens., 8(8): 618, 2016.

[18] E. Norouzi, S. Behzadi, "Evaluating machine learning methods and satellite images to estimate combined climatic indices," Int. J. Numer. Methods Civ. Eng., 4(1): 30-38, 2019.

[19] F. Ling, G. M. Foody, H. Du, X. Ban, X. Li, Y. Zhang, et al., "Monitoring thermal pollution in rivers downstream of dams with Landsat ETM+ thermal infrared images," Remote Sens., 9(11): 1175, 2017.

[20] M. Kaidan, T. Maksymyuk, V. Andrushchak, M. Klymash, "Intelligent data flow aggregation in edge nodes of optical label switching networks," in Proc. 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT): 145-148, 2019.

[21] R. Mitchell, J. Michalski, T. Carbonell, An artificial intelligence approach: Springer, 2013.

[22] I.T. Jolliffe, J. Cadima, "Principal component analysis: a review and recent developments," Philos. Trans. Roy. S. A: Math. Phy. Eng. Sci., 374: 20150202, 2016.

[23] D. Karaboga, E. Kaya, "Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey," Artif. Intell. Rev. 52(4): 2263-2293, 2019.

[24] J. Wu, Y. Su, Y. Cheng, X. Shao, C. Deng, C. Liu, "Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system," Appl. Soft Comput., 68: 13-23, 2018.

[25] S.J. Narayanan, I. Paramasivam, R.B. Bhatt, "On the estimation of optimal number of clusters for the induction of fuzzy decision trees," Int. J. Data Sci., 2(3): 221-245, 2017.

[26] K. Khamar, "Short text classification using kNN based on distance function," Int. J. Adv. Res. Compu. Commun. Eng., 2(4): 1916-1919, 2013.

[27] O. Kisi, O. Genc, S. Dinc, M. Zounemat-Kermani, "Daily pan evaporation modeling using chi-squared automatic interaction detector, neural networks, classification and regression tree," Comput. Electron. Agric., 122: 112-117, 2016.

[28] O. Genç, B. Gonen, M. Ardıçlıoğlu, "A comparative evaluation of shear stress modeling based on machine learning methods in small streams," J. Hydroinf., 17(5): 805-816, 2015.

[29] M.L. Van Horn, T. Jaki, K. Masyn, G. Howe, D.J. Feaster, A.E. Lamont, et al., "Evaluating differential effects using regression interactions and regression mixture models," Educ. Physiol. Meas., 75(4): 677-714, 2015.

[30] H. Jafarian, S. Behzadi, "Evaluation of PM2. 5 emissions in Tehran by means of remote sensing and regression models," Pollution, 6(3): 521-529, 2020.

[31] M. Pal, S. Deswal, "Modelling pile capacity using Gaussian process regression," Comput. Geotech., 37(7-8): 942-947, 2010.

[32] R.U. Maheswari, S. Rajalingam, T. Senthilkumar, "Condition monitoring of coal mine using ensemble boosted tree regression model," in Intelligent Communication Technologies and Virtual Mobile Networks: 19-29, 2019.

Biographies

Esmail Norouzi has a B.Sc. in Geomatics engineering, and currently studying at master degree in Geographic Information Systems at Shahid Rajaee teacher training University. His professional and research interests include Geospatial Data mining, Geoinformatics (GIS).

Saeed Behzadi is an Assistant Professor in Surveying Engineering Department of Shahid Rajaee Teacher Training University, Iran. He received his M.Sc. degree in Geospatial Information Science in 2008 and a Ph.D. degree in 2013, and has worked within the industry and academia in Iran since 2013. He teaches in the Department of Surveying Engineering and Civil Engineering in the area of GIS, Remote sensing, and Computer Science.

How to cite this paper:

E. Norouzi, S. Behzadi, "The feasibility of machine-learning methods to extract the surface evaporation quantity using satellite imagery," J. Electr. Comput. Eng. Innovations, 9(2): 229-238, 2021.

DOI: 10.22061/JECEI.2021.7563.406

URL: https://jecei.sru.ac.ir/article_1541.html