International Journal of Environmental Science and Technology (2024) 21:9527–9540
https://doi.org/10.1007/s13762-024-05574-z

ORIGINAL PAPER

Estimation of soil erodability parameters based on different machine algorithms integrated with remote sensing techniques

F. Saygın[1] · H. Aksoy[2] · P. Alaboz[3] · M. Birol[4] · O. Dengiz[5]

Received: 28 April 2023 / Revised: 23 November 2023 / Accepted: 6 March 2024 / Published online: 6 April 2024

Abstract
Erosion causes significant damage to life and nature every year; therefore, controlling erosion is of great importance. Therefore, maintaining the balance between soil, plants, and water plays a vital role in controlling erosion. Aim of this study was to estimate some erodability parameters (structural stability index—SSI, aggregate stability—AS, and erosion ratio—ER) with indices and reflectance obtained via TripleSat satellite imagery using machine learning algorithms (support vector regression—SVR, artificial neural network—ANN, and K-nearest neighbors—KNN) in Samsun Province, Vezirköprü, Türkiye. Various interpolation methods (inverse distance weighting—IDW, radial basis function—RBF, and kriging) were also used to create spatial distribution maps of the study area for observed and predicted values. Estimates were made using NDVI, SAVI, and ASVI indices obtained from satellite images and NIR reflectance. Accordingly, the ANN algorithm yielded the lowest MAE (2.86%), MAPE (9.46%), and highest R^2 (0.82) for SSI estimation. For AS and ER estimation, SVR had the highest predictive accuracy. Given the RMSE values in spatial distribution maps for observed and estimated values (SSI 7.861–7.248%, AS 14.485–14.536%, and ER 4.919–3.742%), the highest predictive accuracy was obtained with kriging. Thus, it was concluded that erosion parameters can be successfully estimated with reflectance and index values obtained from satellite images using SVR and ANN algorithms, and low-error distribution maps can be created using the kriging method.

Keywords Soil properties · Erodibility factors · Machine-learning algorithm · Remote sensing · Kriging

Introduction

Soil is viewed by most people as a production material, whereas scientists define it as a dynamic being that lives and sustains (Saygin et al. 2019a). Erosion is defined as the abrasion, decomposition, transport, and accumulation of soils, all of which occur as soils move from one environment to another. Erosion is the most significant type of land degradation that limits the sustainable use of soil and water resources (Lal 1988; Ozdemir et al. 2018). Erosion physically moves the topsoil, which is rich in nutrients, organic matter, and soil microorganisms, decreasing its productivity (Dabral et al. 2016). Given its climate and topographic conditions, Turkey has a high tendency for erosion (Kanar and Dengiz 2015). The slope factor has a particularly high impact on erosion. The Black Sea region of Türkiye with sloping lands experiences heavy rainfall every year, and this destruction of vegetation breaks the soil's resistance to abrasion, exposing it to erosion (Saygin et al. 2019b).

Particularly resulting from human activities, erosion causes an irreversible loss of soil function, which takes the

Editorial responsibility: Gobinath Ravindran.

P. Alaboz
pelinalaboz@isparta.edu.tr

1 Plant Production and Technology Department, Faculty of Agriculture Sciences and Technology, Sivas University of Science and Technology, Sivas, Turkey

2 Department of Forestry and Forest Products, Ayancık Vocational School, Sinop University, Sinop, Turkey

3 Department of Soil Science and Plant Nutrition, Faculty of Agriculture, Isparta University of Applied Sciences, Isparta, Turkey

4 Black Sea Agricultural Research Institute, Soil and Water Resources Department, Ministry of Agriculture and Forestry, Samsun, Turkey

5 Plant Nutrition and Soil Science Department, Agricultural Faculty, Ondokuz Mayıs University, Samsun, Turkey

soil centuries to achieve (Saygin et al. 2019b). The severity of erosion is based on climate, topography, soil, and land use (El-Swaify 1997). Hence, soil degradation due to precipitation and wind erosion is a natural, geomorphological phenomenon that poses a serious problem for the ecosystem. In addition to natural causes, this process is intensified and accelerated by human-induced degradation, negatively affecting the quality of the environmental ecosystem. Research indicates that soil erosion—which includes chemical, physical, and biological degradation, excess salts, wind, and water erosion—is a crucial part of land degradation, which results in environmental issues (Luleva et al. 2012). Thus, playing a leading role in agricultural production, soil loses its productive capacity and cannot fulfill its other functions in the ecosystem. Thus, soil erosion negatively impacts the environment by reducing soil fertility and damaging agricultural lands and land topography (Rodrigo-Comino et al. 2018). The same amount of precipitation falling on different soil properties may cause different rates of fragmentation, abrasion, and sedimentation (Askin et al. 2016). Erosion estimates are made based on various models (Saini et al. 2015); research also investigates some parameters to evaluate erosion susceptibility, including aggregate stability, dispersion rate, erosion rate, clay rate, and structural stability of soils (Alaboz et al. 2021).

In order to detect, investigate, assess, and estimate soil erosion, remote sensing (RS) with a geographic information system (GIS) provides crucial information about the dynamics and intensity of erosion over time and space. (Wang and Shao 2013). Technological advances in computers, such as image processing, Global Positioning System (GPS), and GIS, are effective tools for processing spatial data, soil, precipitation, temperature, etc., and socioeconomic data at different scales, for making integrated analysis of a region's resources, and for finding optimum solutions for a range of problems (Venkataratnam and Sankar 1996). Recent studies have shown that satellite spectral data (band ratios, band transforms, and spectral band) and vegetation indices with area measurements correlate well with soil properties (Lu 2006; Viana et al. 2012; Bulut et al. 2022). Yaghouti et al. (2019) calculated the NDVI, SAVI, LAI, DVI, and RVI indices using RGB and NIR band values obtained from Landsat 7ETM + and estimated high-yield rice fields. Another study calculated the NDVI, EVI, ARVI, SAVI, VARI, and NDWI indices using MODIS satellite image band values and estimated the volumetric soil water content in the root zone of annual grasslands and oak savannas (Liu et al. 2008). These approaches indicate correlations between index and band values and soil properties.

A number of recent studies have used pedotransfer functions, to try and determine the properties of soil using various other physical and chemical characteristics. These algorithms include artificial neural networks (Yamaç et al. 2020), support vector regression (Saygın et al. 2023), random forests (Gunarathna et al. 2019) and K-nearest neighbors (Alaboz et al. 2021). Machine-learning algorithms are widely used to transform the complex and multivariate structure of soil and earth science problems into solutions (Shahin 2016; Abedini et al. 2019). There are numerous studies on estimating soil erosion using machine-learning algorithms (Kashani et al. 2020; Sahour et al. 2020). Using the CART algorithm, Alaboz et al. (2021) were able to successfully estimate the USLE-K factor and aggregate stability from decision trees. Saygin et al. (2021) estimated erosion susceptibility with a MAPE value of 5.76% using the random forest algorithm. Other authors have successfully determined soil properties with low error rates using spectral reflectance, indices, and reflectance data (Mishra et al. 2022; Garosi et al. 2019). The different error rates for different regions and soils in these studies reveal the dynamic characteristics of soil (Alaboz and Isildar 2019). Hence, these studies should be conducted and investigated on a regional basis.

The present research aimed to (i) investigate some erosion susceptibility parameters of a 6018.2 ha study area in Samsun, Vezirkopru, (ii) estimate AS, SSI, and ER properties with NDVI, SAVI, and ASVI indices and band reflectance values calculated from TripleSat satellite image band (Red, Green, Blue, and NIR) values using various machine algorithms (SVR, ANN, K-NN) and evaluate the accuracy of these estimations, and (iii) evaluate the spatial distributions of actual and estimated data. Evaluation of the predictability of erosion sensitivity parameters with different algorithms. Creating spatial distribution maps with estimated values reveals the originality of this study. Forecast models are applied in many studies, but in this study, an integrated evaluation is made with different geostatistical approaches. This constitutes the novelty of the study.

Materials and methods

Description of the study area

The study area is located within Samsun province, Vezirkopru, in the Central Black Sea Region of Türkiye, between coordinates 710,400–712,400 E and 4,549,000–4,552,000 N (WGS-84, Zone 36N, UTM-m) (Fig. 1).

The study area covers 6018.2 ha with an average altitude ranging from 486 to 720 m above sea level. The area mostly consists of steep lands, with some slight sloping lands in the eastern, northeastern, and southeastern parts (Fig. 2). This area is mostly used for dry farming and pasture lands.

In terms of climatic conditions, Vezirköprü is classified between the humid, temperate climate of the Black Sea coastal zone and the continental climate of the inland regions. The area experiences colder winters and warmer

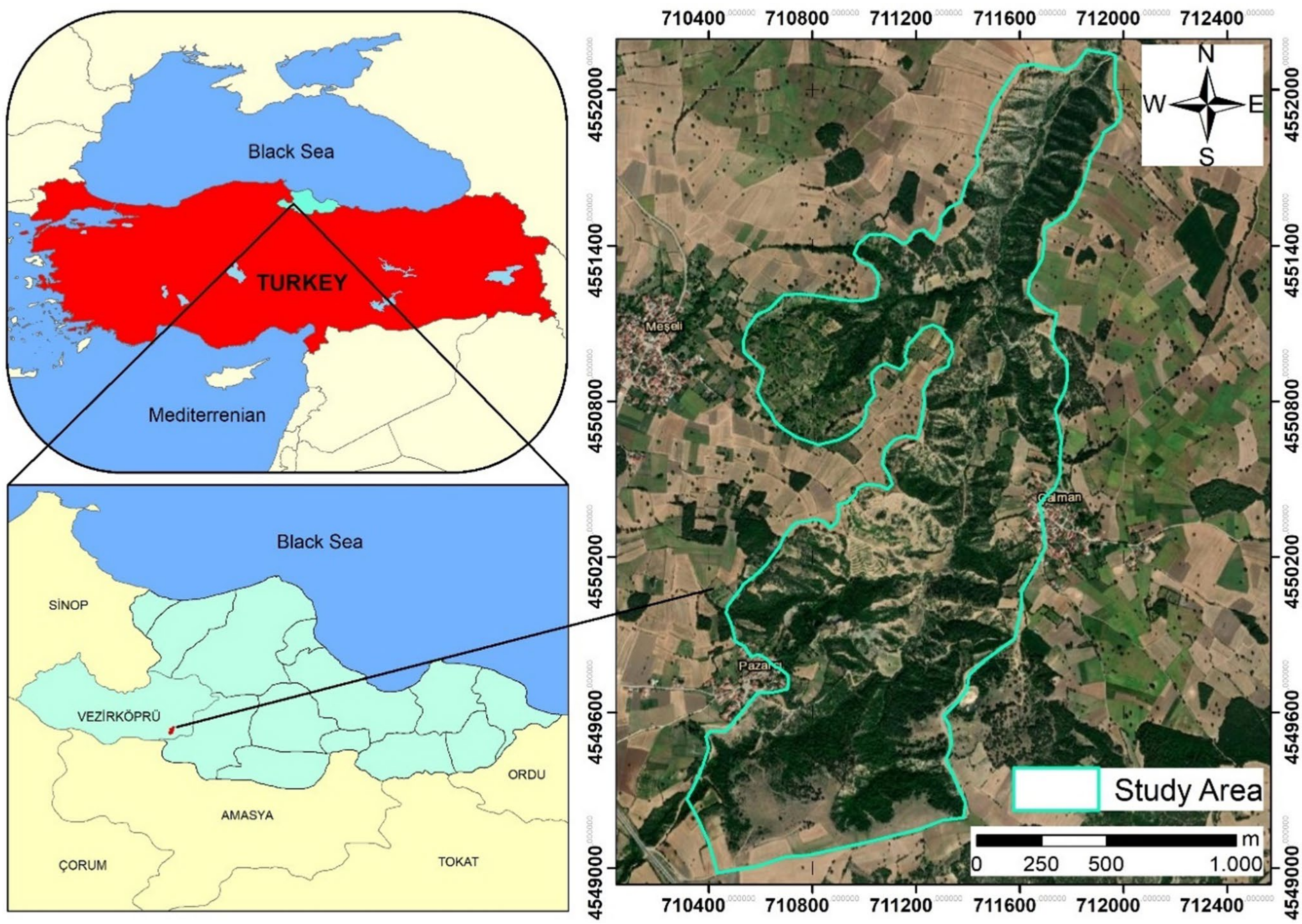


Fig. 1 Location of the study area

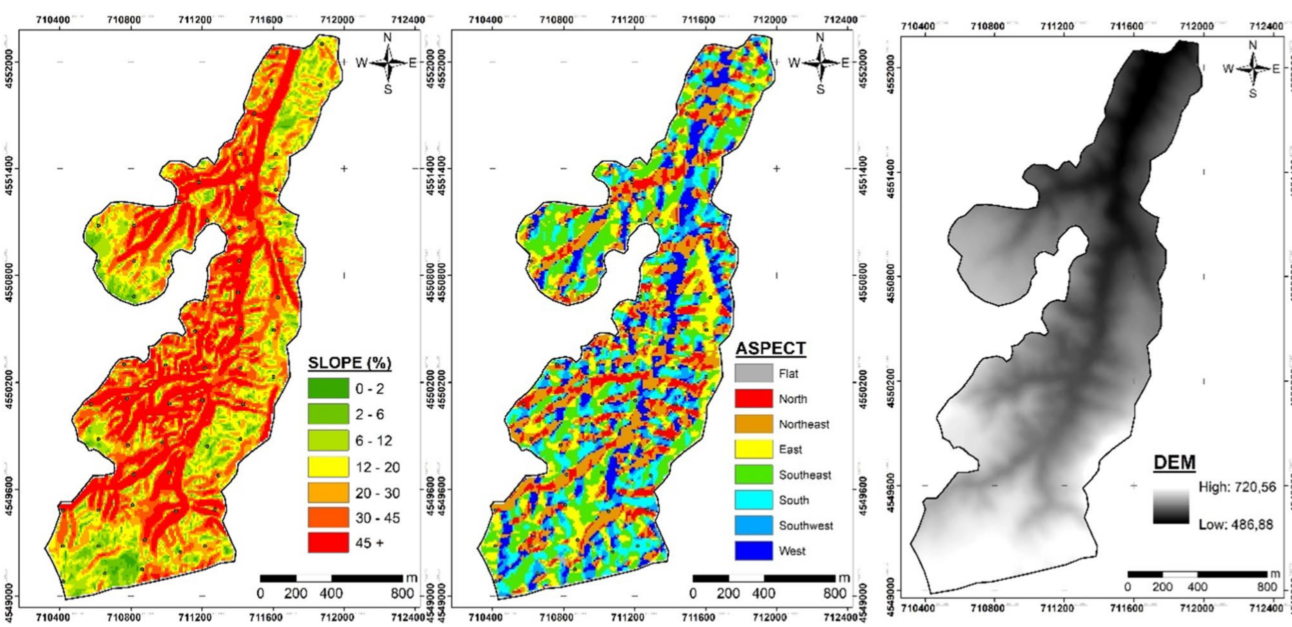


Fig. 2 Slope, aspect and DEM maps of the study area

summers than the coast. The region has 12.5 °C annual average temperature and 724.5 mm of yearly average precipitation. The Newhall simulation model indicates that the soils in Vezirkopru exhibit a xeric moisture regime (in subgroup typical Xeric) in semi-arid zones and a mesic soil temperature regime (Van Wambeke 2000; Turan et al. 2018).

Material

During this study, soil samples were disturbed at a depth of 0–30 cm from 59 points in the study area (200 m × 200 m) (Fig. 3). First, the samples were separated from their coarse particles in the laboratory and passed through a 2-mm sieve. The procedures for analyzing soil samples are displayed in Table 1.

Satellite images and their properties

Band separations were made using an image from the TripleSat satellite. The TripleSat satellite image's technical specifications are displayed in Table 2.

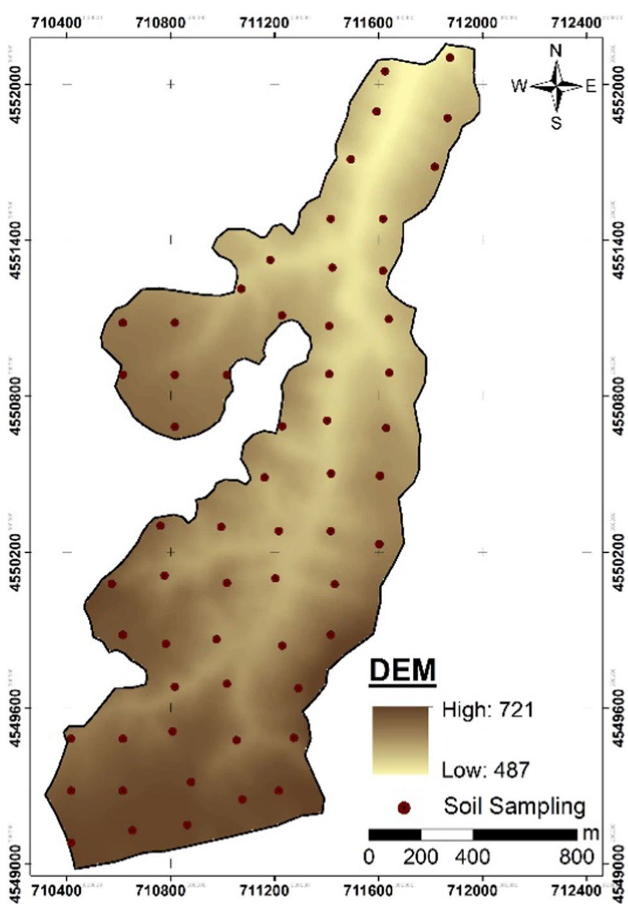


Fig. 3 Soil sampling pattern in the study area

Table 1 Soil analysis and used literature

Analysis	References
Soil texture (clay, silt, sand—(%)	Bouyoucos (1951)
Organic matter—(%)	Jackson (1958)
Structural stability index—(%)	Atalay (2006)
Erosion ratio—(%)	Ngatunga et al. (1984)
Aggregate stability—(%)	Demiralay (1993)
Electrical conductivity—(μmhos cm^{-1})	Richards (1954)

The date of the TripleSat satellite image is June 4, 2018. It has five bands with spatial resolutions of 0.8 and 3.2 m, respectively, including panchromatic and multispectral bands. The satellite photographs have a temporal resolution of 97.7 min and a stripe width of 23.4 km. It was calculated that the index and reflectance values of the terrestrial points were in two stages. In the first stage, we performed atmospheric corrections and calibrated the TripleSat satellite image using QGIS Desktop 3.8.1 software. In this way, atmospheric errors were removed from the satellite image, and reflectance images were obtained. In the second stage, each band was converted into a "vector" layer (ArcGIS Desktop 10.7.1). The boundaries of the sample points were determined with these vector transformations. The layer of average reflectance values was converted to "raster" data (zonal statistic). It was created NDVI, SAVI, and ASVI index maps were created using the band (Red, Green, Blue, and NIR) reflectance values of each sample area and the "zonal statistics" command for each sample point to obtain the index values.

Estimation models

In the current study, it was estimated some soil erodability parameters (aggregate stability, structure stability index, and erosion rate) were estimated with NDVI, SAVI, and ASVI indices and NIR band reflectance values obtained from a TripleSat satellite image using various machine-learning algorithms (support vector machine, artificial neural network, and K-nearest neighbors). In the models, NDVI, ASVI,

Table 2 Technical properties of TripleSat satellite image

Bands	Electromagnetic area	Wave length (nm)	Spatial resolution (m)
B1	Blue (B)	440 – 510	3.2
B2	Green (G)	510 – 590	3.2
B3	Red (R)	600 – 670	3.2
B4	Near infrared (NIR)	760 – 910	3.2
B5	Panchromatic	-	0.8

SAVI, and NIR features were used as input data, while AS, SSI, and EO features were used as output.

Support vector regression

Support vector machines (SVMs) are algorithms for supervised machine learning that are employed in regression analysis and classification. Based on the kind of data, the algorithm employs kernel functions to estimate both linear and nonlinear classification operations. SVM modeling aims to identify the hyperplane that provides the maximum margin by partitioning vector sets with categorical states on one side of the plane and variables on the other (Cortes and Vapnik 1995). As a result, the SVM is an N-dimensional dataset to identify the optimal classification hyperplane. The reference points, or support vectors, are used to draw this hyperplane. Smola and Schölkopf (2004) note that SVM does not always have a linear plane, despite the fact that finding hyperplanes in linear planes is much simpler and this line helps us to make classifications. In the case of nonlinear planes, the data are given an extra dimension; thus, the hyperplane that will separate the classes can be determined. The kernel function can help transform a nonlinear problem into a linear problem. In the current study, it was used support vector regression (SVR) and a radial basis function as the kernel function (Ballabio 2009). Because a radial basis classification is nonlinear, it moves the sample into a higher-dimensional space. For estimation, the R software's "e1071" and "Caret" packages were utilized. In addition, we applied a fivefold cross-validation to the dataset.

Artificial neural network

Multi-layer sensors, an input layer, one or more intermediate layers, and an output layer make up artificial neural networks (Cakir 2019). The model output layer is the third layer, the hidden intermediate layer is the second, and the model input layer is the first. For estimations with ANNs, it was used the feed-forward back propagation technique and the 4:10:1 architecture of the Levenberg–Marquardt algorithm (LM), which is a different learning algorithm. For ANN estimations, it was used the “nntool” package in the MATLAB software. In ANNs, 70% of the dataset was used to create the model, 15% for validation, and 15% for testing.

K-nearest neighbors

The K-nearest neighbors algorithm determines the class where the data point is located and the K value of the nearest neighbor. The K value indicates the number of nearest observations required to estimate the new observation. Euclidean distance is most often used as the neighbor distance (İlarslan 2016). K-NN is used for classification and regression in supervised learning. MATLAB software was used to estimate the K-NN algorithm. Additionally, fivefold cross-validation was applied to the dataset and the K value was set to 6.

A Taylor diagram was also used to evaluate the actual and forecast data using the “chron, lattice, ggplot2, plotrix, graphics” packages in the RStudio software. The diagram is shown in Fig. 4. In addition, mean absolute error (MAE), the root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (*R*2) features were analyzed in order to assess the prediction accuracy of the models.

Interpolation analysis and creation of spatial distribution maps

Using soil samples taken from coordinates in the grid system at 200-m intervals, spatial distribution maps were produced for SSI, AS, and ER analysis results using interpolation models. While creating these spatial distribution maps for the values derived from the models of estimation, various interpolation models were used in the ArcGIS 10.5v software. Before creating the maps, data that did not show normal distribution were transformed according to their characteristics. Inverse distance weighting (IDW), radial basis function (RBF), and kriging methods were used to create the maps, and the approach with the lowest RMSE was identified.

The estimated value from each grid data point is selected in the IDW approach to decrease with increasing distance from the current point. The most accurate result is obtained when the distance power is set to 1, as this power is weak for datasets with low coefficients of variation (CV); for those datasets, it can be set to 2 or 3. (Gotway et al. 1996). RBF is another method for interpolating multidimensional data, often used to estimate difficult-to-predict points or those with limited data (Taşan and Demir 2017). The biggest advantage of RBF is that it involves few restrictions and can therefore be used comfortably in any size (Wright 2003).

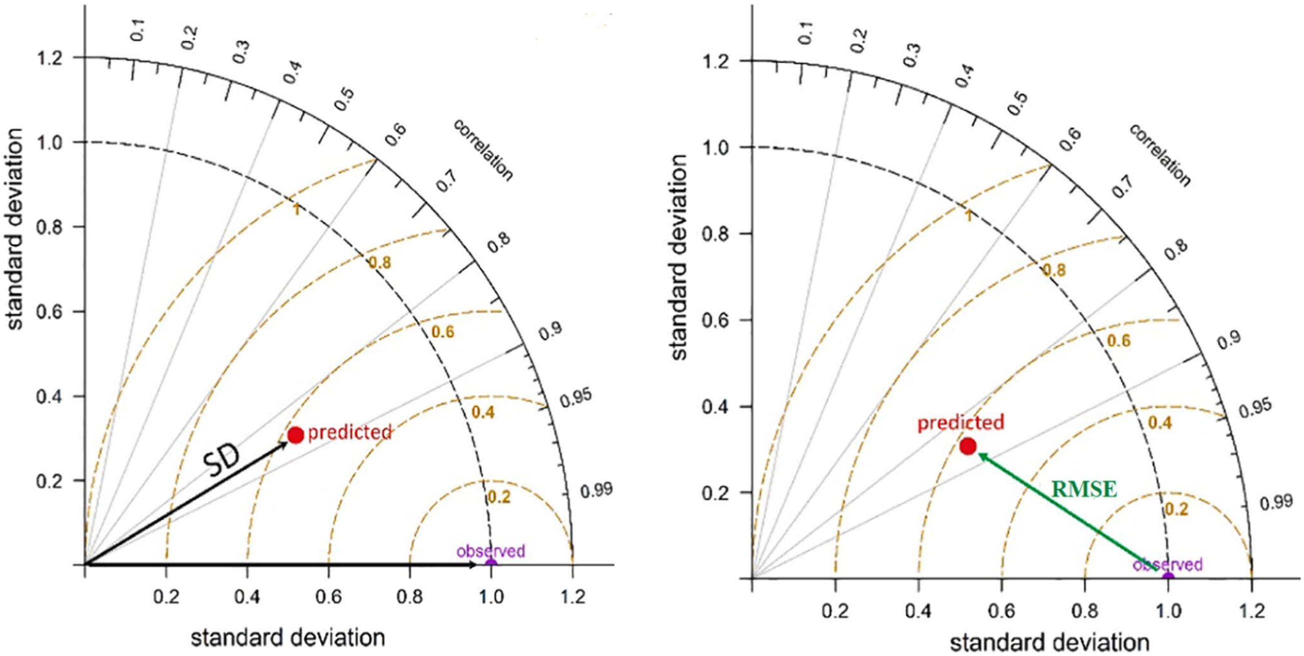
Semivariograms are used to calculate the distance in the kriging method. To reduce the estimate's mean squared error, the weights are selected independently (Gotway et al. 1996). Although kriging semivariograms are challenging to forecast and model, scientific research frequently finds this geostatistical method to be highly beneficial (Gotway et al. 1996).

The workflow diagram of this study is shown in Fig. 5.

Results and discussion

Soil properties and descriptive statistics

Table 3 shows the basic descriptive statistics of the soil physicochemical properties. Accordingly, the sand content



SD: standard deviation, RMSE: root mean squarer error

Fig. 4 Taylor diagram

of the soils varied from 17.56 to 63.83%, the silt content from 15.24 to 41.15%, and the clay content from 15.81 to 51.91%. The clay loam (CL) group was dominant in the study area. 46.43% of the sampling points were clay loam (CL), 16.07% were sandy clay loam (SCL), 14.29% were clay (C), 12.50% were sandy loam (SL), 8.93% were loam (L), and 1.79% were silty clay loam (SiCL). The OM contents of the soils ranged from 0.18 to 5.91%.

Aggregate stability values ranged from 19.95 to 78.02%. Aggregate size distribution and stability measurements are some of the significant elements that indicate soil quality (Ozdemir et al. 2015; Eraslan et al. 2016). In addition, aggregate stability is a crucial parameter regarding the internal resistance of soil aggregates to environmental factors that cause degradation (Hillel 1982; Saygin et al. 2019b). Aggregate stability affects the physical and chemical processes of soils, including water movement in the soil, plant growth, and biological activity (Zhang and Miller 1996; Six et al. 2000; An et al. 2008). Aggregate stability is also inversely proportional to soil erosion tendencies.

In this study, the ER content of the soils ranged from 1.65 to 24.65%, and SSI content from 13.2 to 52.25%. Kanar and Dengiz (2015) observed the DR, ER, AS, and erodibility properties of soils in the Madendere Basin and evaluated their erosion tendencies. In this present study, it was found that the ER values ranged from 37.81 to 86.30% and AS values ranged from 17.45 to 73.21%. Dabral et al. (2016) analyzed soil samples from areas under different land-use types in the northern Indian town of Nirjuli and found a dispersion rate of 15% and an erosion rate of 10%, reporting that the soils were highly erodible. Bryan (1968) defined the limit of the erosion rate as 10% for soils that are sensitive to erosion. In this study, it was evaluated that the erosion tendency was high in regions where the erosion rate was above 10%. Given the distribution of soil properties, the erosion ratio was the furthest from the normal distribution, whereas the other properties were close to the normal distribution.

Predictive accuracy of models

Table 4 lists the algorithms used to estimate the SSI, AS, and ER properties of soils and the estimation accuracy of the entire dataset. NDVI, SAVI and ASVI indices and NIR band reflectance values were taken into account in the estimation of the models. Mohammadifar et al. (2021) investigated various machine-learning algorithms to estimate terrain susceptibility to water erosion hazards. The authors used a Landsat 8 image to extract three indices for soil, water, and vegetation, including MNDWI, NDVI, and SAVI. Using a boosted generalized additive model, they found the relative importance of NDVI and SAVI values to be higher than land use and soil texture. While estimating SSI, the lowest MAE (2.86%) and MAPE (9.46%) values and the highest R^2 (0.82) were obtained using the ANN algorithm. However, while

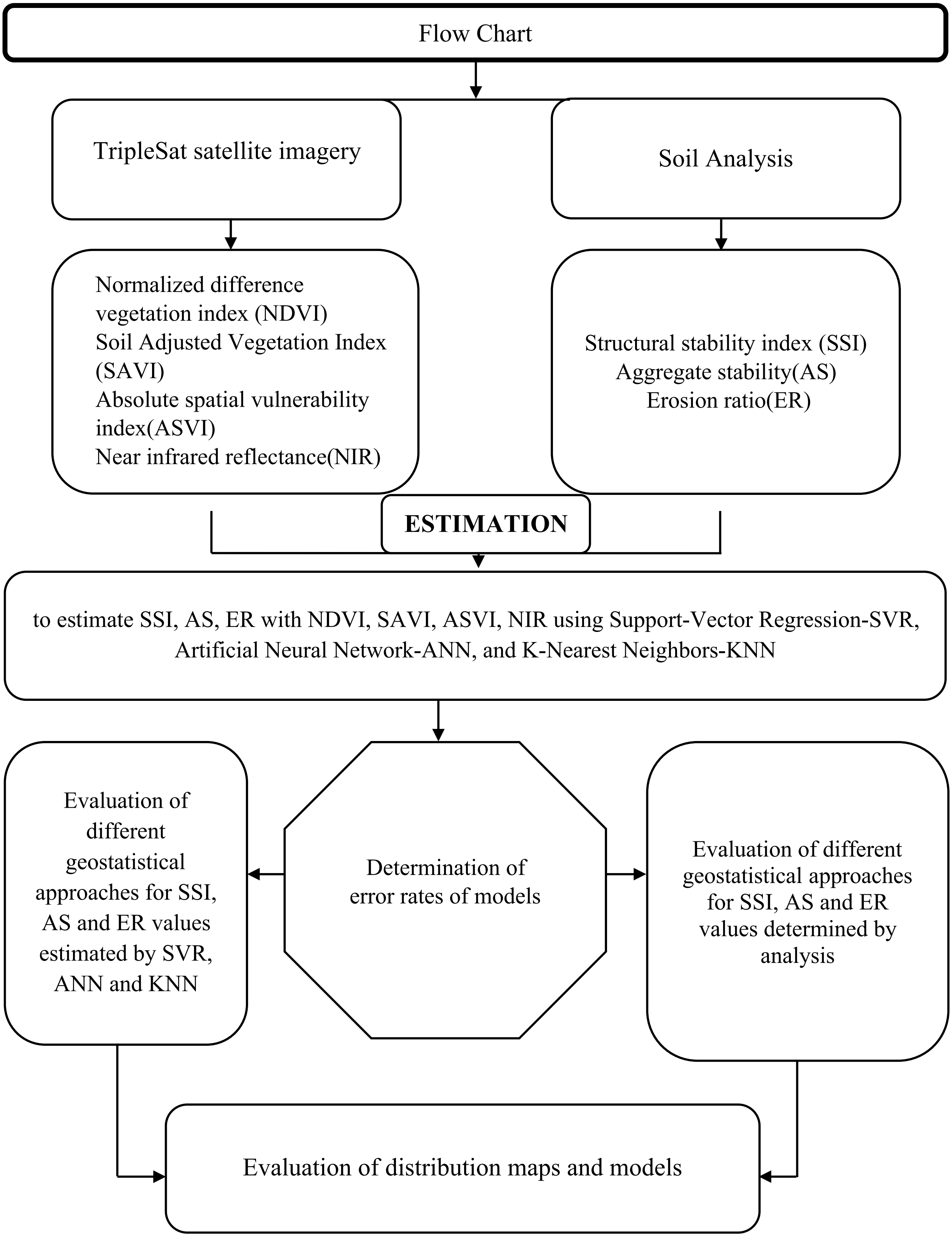


Fig. 5 Flow chart

Table 3 Descriptive statistics values of parameters

Variable	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
Sand %	39.4	12.4	17.56	63.83	0.3	−0.96
Clay %	32.07	8.48	15.81	51.91	−0.11	−0.64
Silt %	28.531	5.559	15.24	41.15	−0.26	−0.44
OM %	2.766	1.523	0.18	5.91	0.33	−0.74
AS %	52.42	14.72	19.95	78.02	−0.38	−0.8
ER %	9.282	4.767	1.65	24.65	1.23	1.59
SSI %	31.62	8.65	13.2	52.25	0.2	0.01
NDVI	0.1725	0.2193	−0.2114	0.4956	−0.21	−1.24
SAVI	0.2583	0.3285	−0.3167	0.7423	−0.21	−1.24
ASVI	191.3	43.33	101.54	255.59	−0.11	−0.65
NIR	190.73	43.35	101	255	−0.11	−0.65

Table 4 The predictive accuracy of different algorithms

		MAE	MAPE	R^2
SSI	ANN	2.86	9.46	0.82
	SVR	5.98	24.45	0.68
	KNN	7.89	35.45	0.52
AS	ANN	5.88	25.62	0.68
	SVR	3.78	7.50	0.89
	KNN	8.41	44.33	0.33
EO	ANN	6.01	36.45	0.54
	SVR	2.04	27.73	0.75
	KNN	8.12	49.52	0.41

estimating AS and ER, SVR displayed the highest predictive accuracy.

Given the values obtained from the estimation of SSI with ANN, the RMSE and SD values are given in the Taylor diagram below (Fig. 6). The RMSE obtained for observed and estimated data was 3.82%, whereas the standard deviation was 7.86% for actual data and 8.25% for estimated data.

During AS estimating, the most successful algorithm was SVM. Figure 7 shows the Taylor diagram generated from these estimated values. The RMSE value obtained from AS estimation was 4.89%, and the standard deviation was 14.72% for observed data and 12.63% for estimated data.

ER estimation was the most successful with the SVM algorithm. Figure 8 shows the Taylor diagram for the RMSE and SD values from these estimations. The RMSE value between the observed and estimated data was 2.40%, and the standard deviation was 3.68% for the observed data and 4.76% for the estimated data.

Using the dataset obtained from the satellite image and band values, the SVR algorithm gave estimates with low error rates for AS and ER. SVR has also been shown to yield effective results and good predictive accuracy in previous research. Mishra et al. (2022) estimated the cation exchange capacity of soils for different land uses using

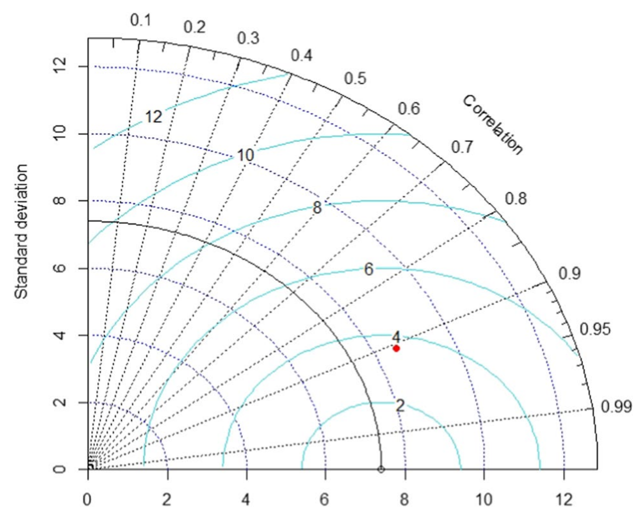


Fig. 6 Estimation accuracy of SSI with ANN algorithm

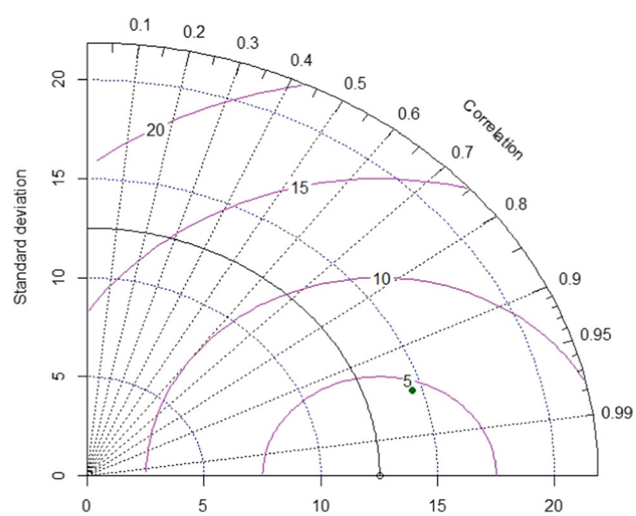


Fig. 7 Estimation accuracy of AS using SVR algorithm

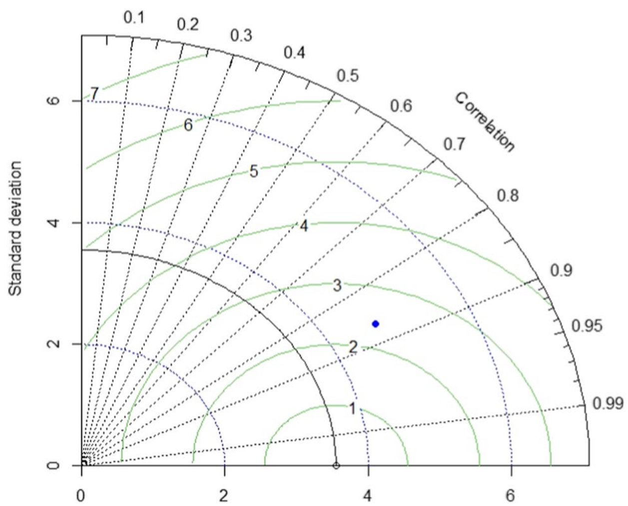


Fig. 8 Estimation accuracy of ER with SVR algorithm

various machine-learning algorithms. The authors reported successful results on small datasets using SVR. Vu Dinh et al. (2021) estimated the erodibility using fuzzy k-nearest neighbor (FKNN), ANN, SVM, least-squares support vector machine (LSSVM), and relevance vector machine (RVM) algorithms and found that the RVM model gave the most successful results in both training and testing phases. Garosi et al. (2019) used different machine-learning algorithms to create a susceptibility map for gully erosion in some parts of the Ekbatan Basin in western Iran. The authors concluded that the RF model displayed the highest efficiency, kappa coefficient, and AUC (92.4%) while yielding the lowest MAE and RMSE values.

Spatial distribution

The best techniques for interpolation are displayed in Table 5. Their RMSE values for the observed soil properties and the estimated data obtained from various algorithms using different interpolation methods. These interpolation methods consisted of 15 models, including the IDW, RBF, and kriging submodels, and the spatial distribution maps were produced considering the smallest RMSE values.

The Kriging method showed the lowest error rate considering the spatial distributions of the estimated and observed data. For the spatial variability of sand, silt, and clay contents, regression kriging yielded lower RMSE values than multilinear regression, according to Mitran et al. (2019). According to Seyedmohammadi et al. (2019), fuzzy logic and geostatistics were used to provide spatial variability in soil texture, and the kriging method produced distribution maps with an acceptable degree of accuracy.

Figure 9 shows the NDVI, SAVI, and ASVI maps produced by analyzing the TripleSat satellite image.

The soil-adjusted vegetation index (SAVI) is used very frequently in vegetation estimates because it gives results regarding soil moisture content compared with the NDVI (Ren et al. 2018). SAVI has also been used to evaluate vegetation growth in semi-arid regions because it is associated with NDVI (Bannari et al. 1995; Hu et al. 2021). The atmospheric soil vegetation index (ASVI) is an index that can adjust soil brightness and atmospheric scattering; therefore, it gives more successful results than NDVI in estimating vegetation in areas with poor vegetation (Lopes et al. 2020). Therefore, ASVI is used more frequently than NDVI in regions with poor vegetation.

Among the machine algorithms evaluated in this study, the SVR model resulted in the closest values to reality. The map produced with SVR showed relatively reduced AS values toward the northern part of the study area, with a high rate of abrasion-resistant aggregates in the interior parts (Fig. 10).

Given the structural stability values (Fig. 11), ANN gave the highest values among the estimated models. The distribution map produced with ANN showed lower reflectance values in the southern part of the area and higher reflectance values in the northwestern part.

Considering the erosion rate (ER) analysis, the SVR model gave the closest result to reality. According to the distribution map produced with SVR, the northern and northeastern parts of the area had relatively higher ER values, indicating higher susceptibility to erosion (Fig. 12).

Regarding the parameters investigated here, we observed a similar distribution pattern between the actual maps and those created by forecast data. A similar distribution pattern was obtained because of the low error rates in the model.

Table 5 Preferred models and RMSE values for interpolation analysis

Properties	Determined mode	RMSE	Properties	Determined mode	RMSE
AS	Kriging-ordinary-Gaussian	14.49	AS_SVM	Kriging-simple-spherical	14.54
SSI	Kriging-simple-spherical	7.86	SSI_ANN	Kriging-simple-spherical	7.25
ER	Kriging-universal-Gaussian	4.92	ER_SVM	Kriging-ordinary-Gaussian	3.74
NDVI	Kriging-simple-spherical	0.21	SAVI	Kriging-simple-spherical	0.32
ASVI	Kriging-ordinary-Gaussian	42.87			

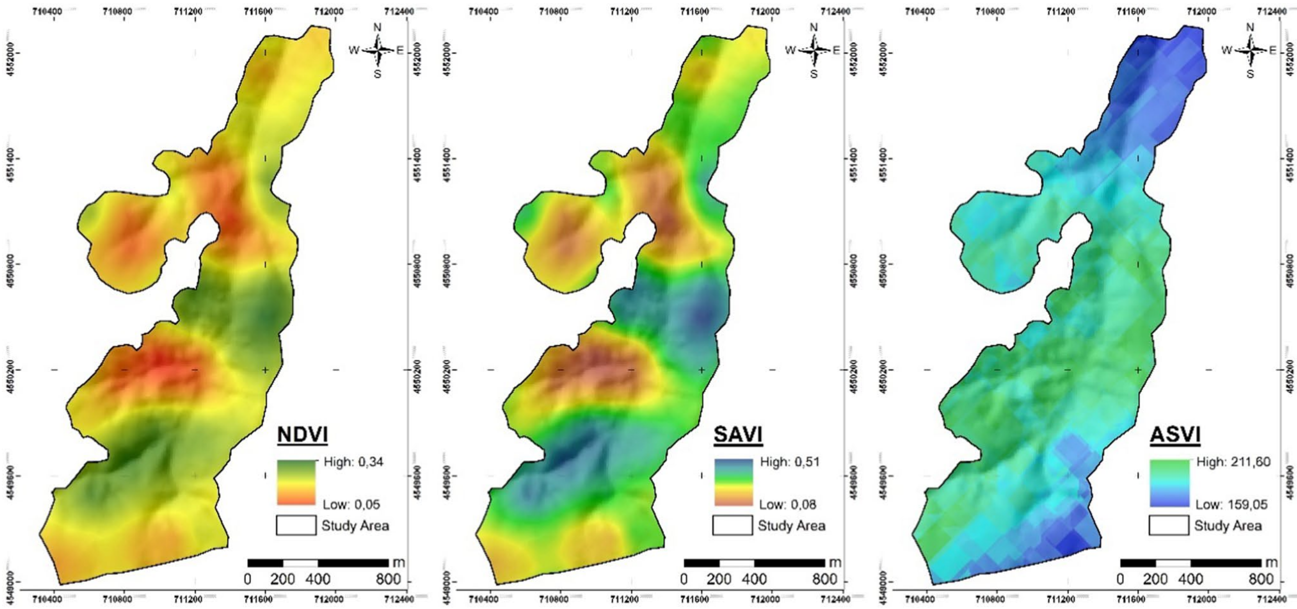


Fig. 9 NDVI, SAVI, ASVI distribution maps

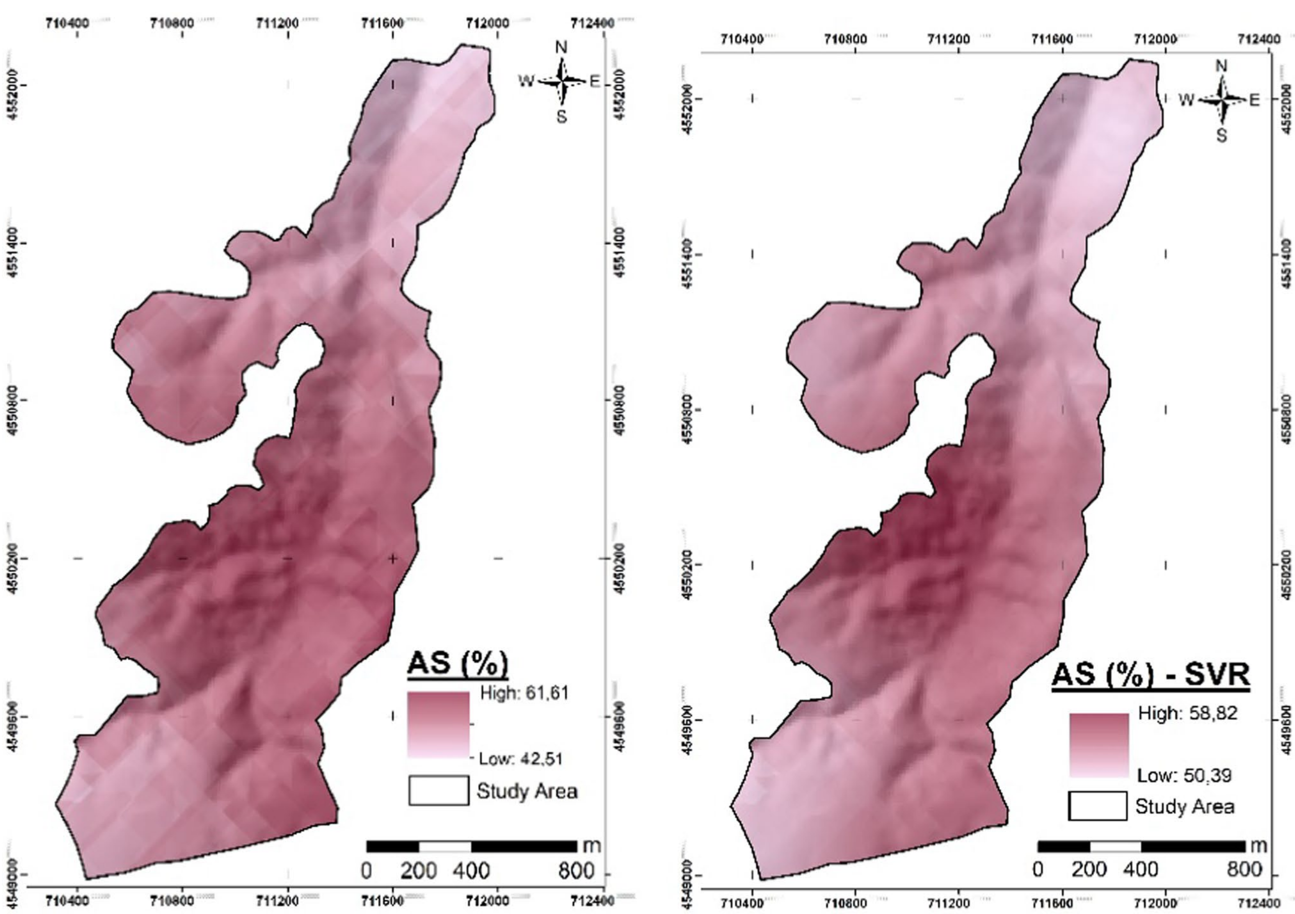


Fig. 10 Actual (left) and estimated (right) data of aggregate stability

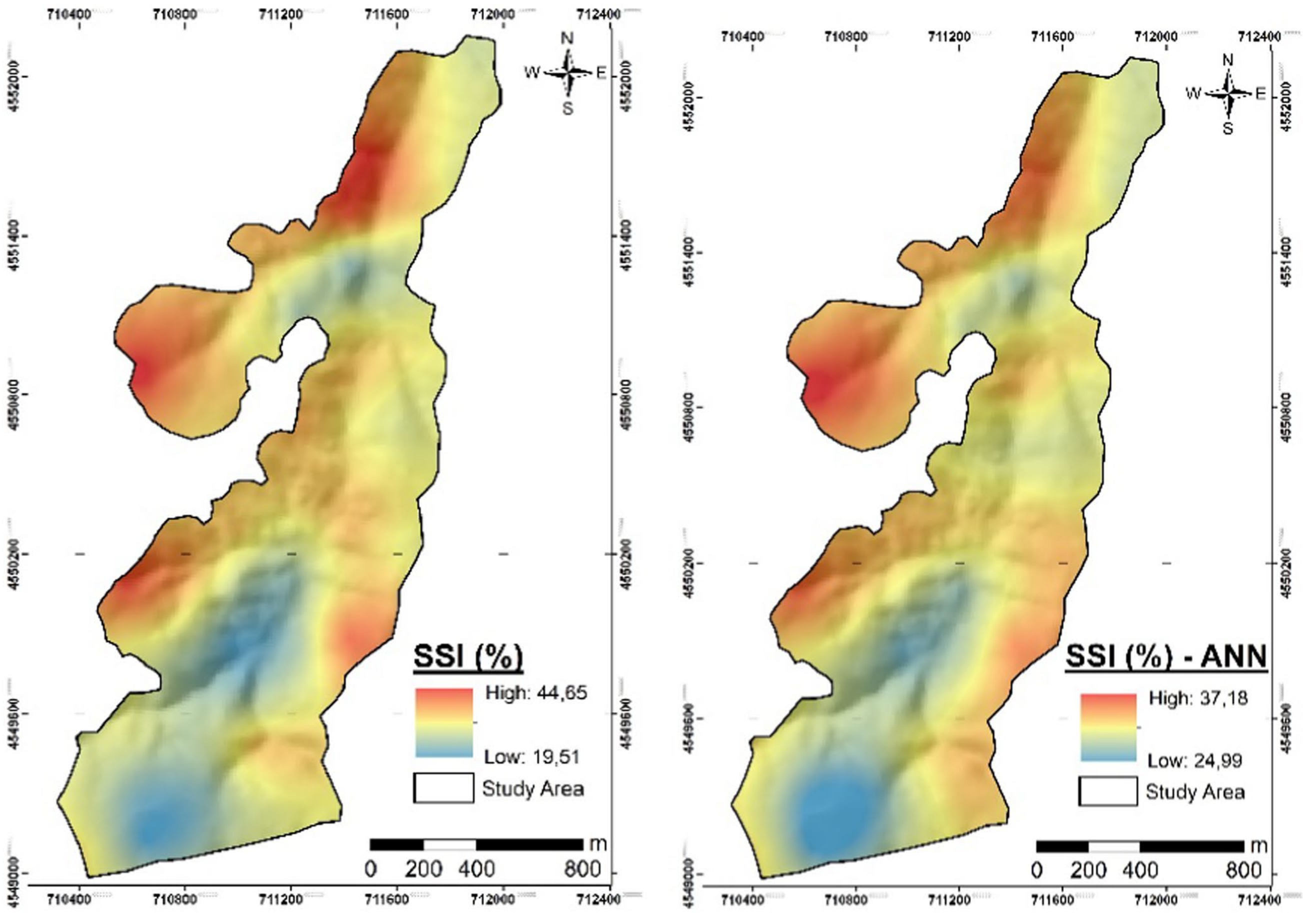
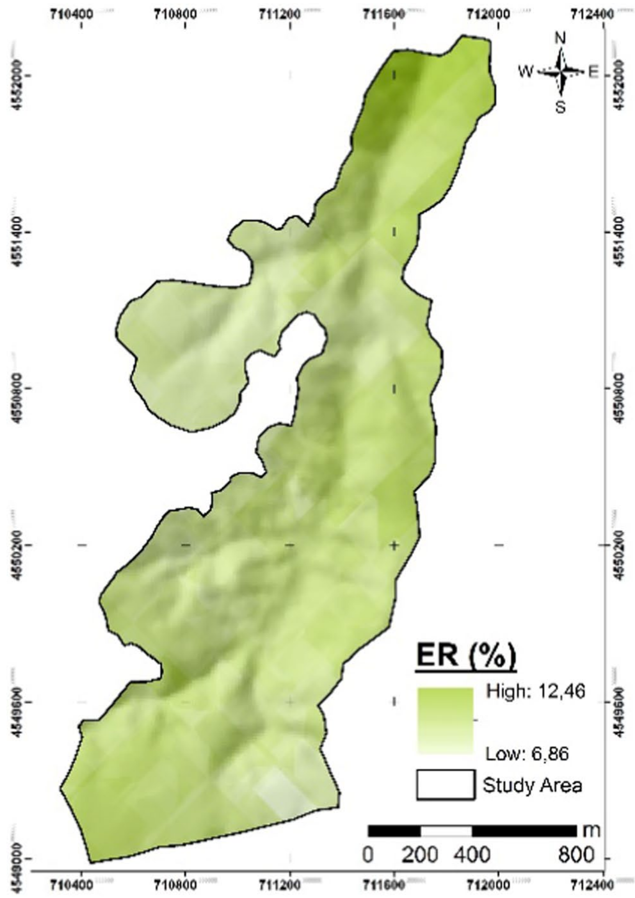


Fig. 11 Actual (left) and estimated (right) data of structural stability index

Conclusion

In the present study, a high-resolution satellite image was used to interpret erosion susceptibility analyses with terrestrial sampling, and estimations were made using various machine-learning algorithms, thus achieving a high sensitivity for identifying areas susceptible to erosion. Therefore, the current study yielded positive outcomes for estimating erosion susceptibility parameters using machine algorithms. Among the algorithms analyzed here, SVR and ANN had relatively lower error rates. For distribution maps, lower error rates were obtained by using the kriging method. The results of the current study reveal that ANN and SVR algorithms can make predictions with high accuracy by using index and reflectance features as input data. Among different interpolation methods, it has been shown that the distribution maps of the estimated values with the kriging method exhibit a pattern close to reality. It is determined that using it in other studies can be done safely. These kinds of maps can yield significant benefits at the national and international levels.

Determining the parameters that impact soil erosion and verifying them by terrestrial measurements is of critical importance for controlling and managing erosion, which is a major problem worldwide. With this research, we conclude that it is possible to make determinations and estimations for risks against sustainable land management using satellite images and algorithms, thereby preventing the dangers that may arise in advance. This study will be considered as a basis and will shed light on more comprehensive studies. In the next studies, the predictability of parameters that are difficult to determine, such as USLE-K, will be examined with different algorithms, and it is planned to evaluate the autocorrelation of spatial maps with field controls.



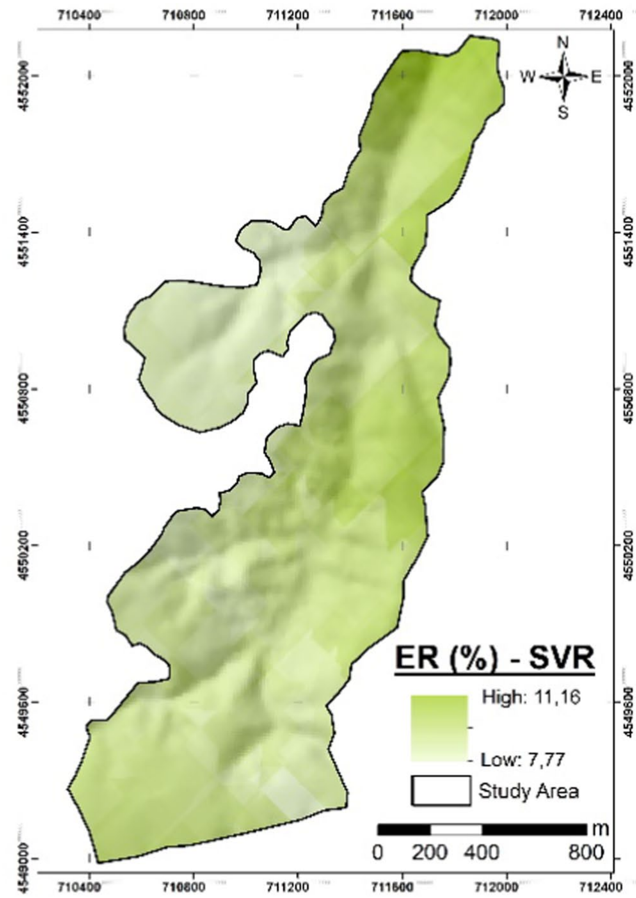


Fig. 12 Actual (left) and estimated (right) data of erosion ratio

Acknowledgements I would like to thank dear Assist. Prof. Muhammad Azhar NADEEM from Sivas University Science and Technology for editing the manuscript.

Authors contributions All authors contributed to the study conception and design. Methodology, conceptualization, and resources were performed by OD, FS and MB. Supervision was performed by OD, FS and PA. Material preparation, investigation, and analysis were performed by MB and FS. All maps were created by FS and HA.

Funding The authors declare that they have received no funding, grants or other support.

Data availability statement The datasets are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest All authors declare that they have no conflict interests.

Ethical approval We guarantee that the article is original work that has not been considered for publication, in whole or in part, in any other government publication. I declare that I give due credit to the studies cited in this article.

References

Abedini M, Ghasemian B, Shirzadi A, Shahabi H, Chapi K, Pham BT, Ahmad BB, Bui DT (2019) A novel hybrid approach of bayesian logistic regression and its ensembles for landslide susceptibility assessment. Geocarto Int 34(13):1427–1457

Alaboz P, Işıldar AA (2019) Evaluation of pedotransfer functions (PTFs) for some soil physical properties. Turk J Sci Eng 1(1):28–34 ((**In Turkish**))

Alaboz P, Dengiz O, Demir S, Şenol H (2021) Digital mapping of soil erodibility factors based on decision tree using geostatistical approaches in terrestrial ecosystem. CATENA 207:105634

An SS, Huang YM, Zheng FL, Yang JG (2008) Aggregate characteristics during natural revegetation on the loess plateau. Pedosphere 18:809–816. https://doi.org/10.1016/S1002-0160(08)60077-6

Aşkın T, Türkmen F, Tarakçıoğlu C (2016) Assessment of erosion risk in soils of Ordu central district by geostatistical techniques. J Soil Sci Plant Nutr 4(2):69–77

Ballabio C (2009) Spatial prediction of soil properties in temperate mountain regions using support vector regression. Geoderma 151(3–4):338–350

Bannari A, Morin D, Bonn F, Huete AR (1995) A review of vegetation indices. Remote Sens Rev 13:95–120

Bouyoucous GJ (1951) A recalibration of the hydrometer for making mechanical analysis of soil. Agron J 43:434–438

Bryan RB (1968) The development, use and efficiency of indices of soil erodibility. Geoderma 2:5–26

Bulut S, Sivrikaya F, Günlü A (2022) Evaluating statistical and combine method to predict stand above-ground biomass using remotely sensed data. Arab J Geosci 15(9):1–14

Çakir, F.S. (2019) Artificial neural networks. Nobel Publications 2nd Edition. Ankara (in Turkish)

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

Dabral PP, Choudhury A, Barman M, Pandey PK (2016) Determination of erodibility under different landuses in the vicinity of Nirjuli, Arunachal Pradesh. J Soil Water Conserv 15(4):292–295. https://doi.org/10.5958/2455-7145.2016.00018.7

Demiralay, İ., 1993. Physical analyzes in soil. Atatürk University Faculty of Agriculture, Publication No:143,:90–95, Erzurum.

El-Swaify SA (1997) Factors affecting soil erosion hazards and conservation needs for tropical steeplands. Soil Technol 11(1):3–16

Eraslan S, İmamoğlu A, Coşkun A, Saygın F, Dengiz O (2016) Determination of soil erodobility based on aggregate and structure stabilities in İnebolu catchment and their relationship with land cover. In: International geography symposium, 13–14 October, TÜCAUM, 779–794. Ankara.

Garosi Y, Sheklabadi M, Conoscenti C, Pourghasemi HR, Van Oost K (2019) Assessing the performance of GIS-based machine learning models with different accuracy measures for determining susceptibility to gully erosion. Sci Total Environ 664:1117–1132

Gotway CA, Ferguson RB, Hergert GW, Peterson TA (1996) Comparison of kriging and inverse-distance methods for mapping soil parameters. Soil Sci Soc Am J 60(4):1237–1247

Gunarathna MHJP, Sakai K, Nakandakari T, Momii K, Kumari MKN (2019) Machine learning approaches to develop pedotransfer functions for tropical Sri Lankan soils. Water 11:1940

Hillel D (1982) Introduction to soil physics, 2nd edn. Academic Press, San Diego, CA

Hu P, Sharifi A, Tahir MN, Tariq A, Zhang L, Mumtaz F, Shah SHIA (2021) Evaluation of vegetation indices and phenological metrics using time-series MODIS data for monitoring vegetation change in Punjab. Pakistan Water 13(18):2550

İlarslan K (2016) The prediction of stock prices wıth K-NN (K- nearest neigbour) algorithm: a sample application from bist. Asos Journal 4(30):375–392

Kanar E, Dengiz O (2015) Determination of relatinship between land use/land cover and some erodibility indexes in madendere watershed soils. Turk J Agric Res 2(1):15–27

Kashani MH, Ghorbani MA, Shahabi M, Naganna SR, Diop L (2020) Multiple AI model integration strategy—application to saturated hydraulic conductivity prediction from easily available soil properties. Soil Tillage Res 196:104449. https://doi.org/10.1016/j.still.2019.104449

Lal R (1988) Soil degradation and the future of agriculture in sub-Saharan Africa. J Soil Water Conserv 43(6):444–451

Liu S, Chadwick O, Roberts D (2008). Relations between root-zone soil moisture and MODIS-derived vegetation indices in oak savanna and open grassland in California. In: AGU fall meeting abstracts, vol. 2008, pp B43C-0444

Lopes CL, Mendes R, Caçador I, Dias JM (2020) Assessing salt marsh extent and condition changes with 35 years of Landsat imagery: Tagus Estuary case study. Remote Sens Environ 247:111939

Lu D (2006) The potential and challenge of remote sensing-based biomass estimation. Int J Remote Sens 27:1297–1328

Luleva MI, Van Der Werff H, Van Der Meer F, Jetten V (2012) Gaps and opportunities in the use of remote sensing for soil erosion assessment. Chemistry 21(5):748–764

Mishra G, Sulieman MM, Kaya F, Francaviglia R, Keshavarzi A, Bakhshandeh E, Rawat D (2022) Machine learning for cation exchange capacity prediction in different land uses. In: CATENA, 216, 106404

Mitran T, Solanky V, Janakirama Suresh G, Sujatha G, Sreenivas K, Ravisankar T (2019) Predictive mapping of surface soil texture in a semiarid region of India through geostatistical modeling. Model Earth Syst Environ 5(2):645–657

Mohammadifar A, Gholami H, Comino JR, Collins AL (2021) Assessment of the interpretability of data mining for the spatial modelling of water erosion using game theory. CATENA 200:105178

Newhall F (1972) Calculation of soil moisture rejime from climatic records. Inpublished. Soil Conservation Service, USDA. Rev. 4. Washington, D.C

Ngatunga ENR et al (1984) Effect of surface management on runoff and soil erosion from some plot milangano. Geoderma 33:1–12

Özdemir N, Öztürk E, Durmuş ÖTK (2018) The effects of organic conditioner applications on some physical properties of soils and their relations with soil loss in artificial rain conditions. Turk J Agric Res 5(3):191–200

Özdemir N, Öztürk E, Kop Durmuş ÖT, Ekberli İ (2015) Effects of organic and inorganic amendments on soil erodibility. Eurasian J Soil Sci 4(4):266–271

Ren H, Zhou G, Zhang F (2018) Using negative soil adjustment factor in soil-adjusted vegetation index (SAVI) for aboveground living biomass estimation in arid grasslands. Remote Sens Environ 209:439–445

Rodrigo-Comino J, Keesstra S, Cerdà A (2018) Soil erosion as an environmental concern in vineyards: the case study of Celler del Roure, Eastern Spain, by means of rainfall simulation experiments. Beverages 4(2):31

Sahour H, Gholami V, Vazifedan M (2020) A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer. J Hydrol 591:125321

Saini SS, Jangra R, Kaushik SP (2015) Vulnerability assessment of soil erosion using geospatial techniques-A pilot study of upper catchment of Markanda river. Int J Adv Remote Sens Gis Geogr 2(1):9–21

Saygın F, Aksoy H, Alaboz P, Dengiz O (2023) Different approaches to estimating soil properties for digital soil map integrated with machine learning and remote sensing techniques in a sub-humid ecosystem. Environ Monit Assess 195(9):1061

Saygın F, Alaboz P, İç S, Çağlar A, Dengiz O (2021) Estimation of erosion erodobility index with random forest algorithm. In: 3rd international cukurova agriculture and veterinary congress, 603–619, 9–10 October, Adana/Turkey

Saygın F, Dengiz O, İmamoğlu A (2019a) Assessment of the relationship between some physico-chemical properties of soil and some erodibility parameters in micro basin scale. Artvin Coruh Univ J For Fac 20(1):82–91

Saygin F, Dengiz O, İç S (2019b) Relationships between erodibility and some soil properties of soils in micro chatchment scale. Soil Water J 15–23

Seyedmohammadi J, Navidi MN, Esmaeelnejad L (2019) Geospatial modeling of surface soil texture of agricultural land using fuzzy logic, geostatistics and GIS techniques. Commun Soil Sci Plant Anal 50(12):1452–1464

Shahin MA (2016) (2016) State-of-the-art review of some artificial intelligence applications in pile foundations. Geosci Front 7(1):33–44

Six J, Elliott ET, Paustian K (2000) Soil macro-aggregate turnover and micro-aggregate formation: a mechanism for c sequestration under no-tillage agriculture. Soil Biol Biochem 32:2099–2103. https://doi.org/10.1016/S0038-0717(00)00179-6

Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14(3):199–222

Taşan M, Demir Y (2017) Determination of spatial distribution of iron and manganese contents with different interpolation methods at rice cultivated areas. Anadolu J Agric Sci 32(1):64–73

Turan M, Dengiz O, Turan İD (2018) Samsun ilinin Newhall modeline göre Toprak sıcaklık ve nem rejimlerinin belirlenmesi. Türkiye Tarımsal Araştırmalar Dergisi 5(2):131–142

Van Wambeke AR (2000) The Newhall simulation model for estimating soil moisture & temperature regimes. Department of Crop and soil Science. Cornell University, Ithaca, NY

Venkataratnam, L., Sankar, T. (1996). Remote sensing and GIS for assessment, monitoring, and management of degraded lands. Surveillance des Sols dans l'Environment par Télédétection et Systèmes d'Information Géographiques (Monitoring Soils in the Environment with Remote Sensing and GIS), 503–516.

Viana H, Aranha J, Lopes D, Cohen W (2012) Estimation of crown biomass ofPinus pinaster stands and shrubland above-ground biomass using inventory data, remotely sensed imagery and spatial prediction models. Ecol Model 226:22–35

VuDinh T, Hoang ND, Tran XL (2021) Evaluation of different machine learning models for predicting soil erosion in tropical sloping lands of Northeast Vietnam. Appl Environ Soil Sci 2021

Wang YQ, Shao MA (2013) Spatial variability of soil physical properties in a region of the Loess Plateau of PR China subject to wind and water erosion. Land Degrad Dev 24(3):296–304

Wright GB (2003) Radial basis function interpolation: numerical and analytical developments. University of Colorado at Boulder

Yaghouti H, Pazira E, Amiri E, Masihabadi MH (2019) The feasibility of using vegetation indices and soil texture to predict rice yield. Polish J Environ Stud 28(4):2473–2481

Yamaç SS, Şeker C, Negiş H (2020) Evaluation of machine learning methods to predict soil moisture constants with different combinations of soil input data for calcareous soils in a semi arid area. Agric Water Manag 234:106121

Zhang XC, Miller WP (1996) Polyacrylamide effect on information and erosion in furrows. Soil Sci Soc Am J 60:866–872. https://doi.org/10.2136/sssaj1996.03615995006000030027x