Received: 11 July 2022 | Revised: 1 December 2022 | Accepted: 13 February 2023

DOI: 10.1002/ldr.4655

RESEARCH ARTICLE

WILEY

Performance evaluation of machine learning algorithms to assess soil erosion in Mediterranean farmland: A case-study in Syria

Safwan Mohammed[1,2] | Ali Jouhra[3] | Glory O. Enaruvbe[4] | Bashar Bashir[5] | Mona Barakat[3] | Firas Alsilibe[6] | Luc Cimusa Kulimushi[7,8] | Abdullah Alsalman[5] | Szilárd Szabó[9]

[1]Institute of Land Use, Technical and Precision Technology, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, Debrecen, Hungary

[2]Institutes for Agricultural Research and Educational Farm, University of Debrecen, Debrecen, Hungary

[3]Department of Soil and Water Science, Faculty of Agriculture, Tishreen University, Lattakia, Syria

[4]African Regional Institute for Geospatial Information Science and Technology, Obafemi Awolowo University, Ile-Ife, Nigeria

[5]Department of Civil Engineering, College of Engineering, King Saud University, Riyadh, Saudi Arabia

[6]Department of transport infrastructure and water resources, Széchenyi István University, Gyor, Hungary

[7]Faculty of Agronomy, Université Catholique de Bukavu, Bukavu, Democratic Republic of Congo

[8]Centre Régional d'Etudes Interdisciplinaires Appliquées au Développement Durable (CEREIAD), Université Catholique de Bukavu, Bukavu, Democratic Republic of Congo

[9]Department of Physical Geography and Geoinformatics, Faculty of Science and Technology, University of Debrecen, Debrecen, Hungary

Correspondence
Safwan Mohammed, Institute of Land Use, Technical and Precision Technology, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, Debrecen 4032, Hungary.
Email: safwan@agr.unideb.hu

Funding information
Project no. TKP2021-NKTA-32 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NKTA funding scheme. Also, this research was supported by the Researchers Supporting Project number (RSP2023R296), King Saud University, Riyadh, Saudi Arabia. Szabo S. was supported by the KKP 144068 project.

Abstract

The development of new techniques, such as machine learning (ML), can provide better insight into the processes and drivers of soil erosion and runoff. However, the performance of these techniques to assess soil erosion in agricultural landscapes is poorly understood. The aim of this study was to evaluate the performance of four machine learning algorithms, generalized linear model (GLM), Random Forest (RF), elastic net regression (EN) and multiple adaptive regression splines (MARS), in predicting soil erosion and runoff in Syria. Soil erosion and runoff were measured on three experimental plots (2.25 m × 1.50 m × 0.50 m, 0.10 m depth in the soil), combined with three different slopes and land use types: RS (8%, olive), SS (12%, citrus), KS (20%, pomegranate). Both erosion and runoff were determined after rainfall events of >10 mm between October 2019 and April 2020. Based on 24 effective rainfall events, the average soil erosion was 0.18 ± 0.14 kg m^{-2} per event in KS, 0.14 ± 0.11 kg m^{-2} per event in SS, and 0.12 ± 0.10 kg/m^2 per event in RS. Regression analysis indicated strong relationship between the rainfalls and the runoff, the highest connection was recorded in the KS plot ($r^2 = 0.85$; $p < 0.05$ $n = 24$). The analysis of covariance indicated that only the runoff had a significant impact on soil erosion

($p = 0.02$) with a medium effect ($\varepsilon^2 p = 0.26$). However, the impacts of rainfall events and slope categories on soil erosion were limited ($\varepsilon^2 p < 0.01$) and not significant ($p > 0.05$). ML techniques were usually efficient in the prediction, the RF and MARS models were the most accurate: RF had the strongest correlation with the measured values ($r = 0.85$) with a low estimation error (0.06 kg m^{-2}), but MARS's standard deviation (SD) was closer to the recorded values' SD. GLM and EN were the weakest predictor models. Modeled values of the slightest slope (8%) had the worst accuracies, and the predictions of the 12% slope were the best in all models. This study provides important insights into the usefulness of machine learning techniques and algorithms in predicting the rate of soil erosion and runoff in agricultural dominated landscapes. We highlighted that the RF and MARS algorithms were better predictors of soil erosion and runoff in the coastal region of Syria.

KEYWORDS
Entisols, GLM, ML-algorithms, soil properties, Syria

1 | INTRODUCTION

Soil erosion is by far the most important type of land degradation, exerting negative effects on the environment and societies worldwide (Nasir & Selvakumar, 2018; van Leeuwen et al., 2019). Soil erosion is recognized as a key environmental issue as water-induced soil erosion remains a dominant agent of degradation globally (Lal et al., 2018). It is a major threat to the sustainable agro-ecosystem, especially in places where high erosion rates alter natural cycles of crops production (Muluneh et al., 2017). This makes soil erosion a critical issue that humanity must address if sustainable food production is desired (García-Ruiz et al., 2015). Soil erosion is characterized by three actions: soil detachment, transport, and deposition. The process of soil relocation is influenced by three main components: (i) rainfall (amount, intensity, duration and erosivity), (ii) soil properties (texture, moisture, organic matter content, aggregate stability), and (iii) local conditions (i.e., land use, land cover, slope) (Dunne et al., 2010; Kulimushi, Bashagaluke, et al., 2021; Mohammed, Al-Ebraheem, et al., 2020; Mohammed, Alsafadi, et al., 2020).

Since the introduction of the first equation to predict soil erosion (Zingg, 1940), there have been significant advances in soil erosion modeling. These advances have led to the development of numerous models to estimate soil erosion under different ecological and landscape conditions (see Dutta (2016) for a list of common soil erosion models). However, while empirical approaches can be useful for specific cases or for model calibration, they are limited by large uncertainties and high variability between predicted and observed soil erosion values (Alewell et al., 2019).

Machine learning (ML) techniques provide a useful alternative for analyzing the complexities inherent in soil science and other branches of geoscience (Abedini et al., 2019; Mosaffaei & Jahani, 2021; Shahin, 2016). ML is based on the idea that systems can learn from data, find similarities, and make decisions with limited human intervention (Rahmati et al., 2017; Sahour et al., 2020). Several studies have applied ML-algorithms to predict soil erosion (Arabameri et al., 2018; Rahmati et al., 2017), soil classification (Brungard et al., 2015), soil moisture (Bertalan et al., 2022), and soil salinity (Bakacsi et al., 2019; Szatmári et al., 2020; Wang et al., 2020). The use of ML algorithms has recently gained wide acceptance for predicting soil erosion (Angileri et al., 2016; de Farias & Santos, 2014; Licznar & Nearing, 2003) (Table 1).

The Mediterranean region is considered one of the most erosion prone regions in the world. García-Ruiz et al. (2013) noted that the Mediterranean basin is a soil erosion hotspot. This is supported by numerous studies that have reported rapid soil erosion in various countries in the region (Khallouf et al., 2021). Many reasons have been put forward to explain the accelerated soil loss in the Mediterranean basin. These include: (i) poor soil structure that accelerates the mineralization process of organic matter; (ii) shallow soil; and (iii) poor management of vegetation cover, as well as traditional agricultural practices (García-Ruiz, 2010; Keesstra et al., 2019). In addition, climate change (mainly variability in rainfall duration and intensity) and changes in the agricultural system have aggravated soil degradation (especially soil erosion) (Benchettouh et al., 2017; Cerda et al., 2018; de Hipt et al., 2018) in the region.

Soil erosion is a threat to the economic development of Syria, and a large amount of productive soil is lost annually due to natural and anthropogenic factors (Masri et al., 2015; Mohammed et al., 2021). In general, wind erosion is dominant in the eastern and central regions of Syria (Masri et al., 2015), while the western and northern parts, as well as the mountains, are more prone to water-induced soil erosion (Mohammed et al., 2016; Mohammed, Al-Ebraheem, et al., 2020; Mohammed, Alsafadi, et al., 2020). Despite the challenges posed by soil erosion in the Mediterranean region, only a few studies have examined the question. In addition, the few studies focusing on soil erosion mostly adopted the revised universal soil loss equation (RUSLE) in a geographic information system (GIS) environment or remote sensing techniques (Abdo, 2021; Abdo & Salloum, 2017;

TABLE 1 Examples of machine learning algorithms applied for soil erosion and runoff prediction.

Acronyms	Algorithms	References	Study area
ANN	Artificial Neural Network	Licznar & Nearing (2003), Arif & Danoedoro (2017); Rahmati et al. (2017)	United States, Indonesia, Iran
KNNs	Kohonen Neural Network	de Farias & Santos (2014)	Brazil
SVM	Support Vector Machine	Dinh et al. (2021), Meshram et al. (2020); Al-Mukhtar (2019)	Vietnam, Iran, India, Iraq
BRT	Boosted Regression Trees	Arabameri et al. (2019); Saha et al. (2020)	Iran, India
RF	Random Forest	Phinzi, Ngetar, and Ebhuoma (2021), Mohammed, Al-Ebraheem, et al. (2020); Cheng et al. (2018); Zeraatpisheh et al. (2022)	South Africa, Syria, China, Iran
NB	Naïve Bayes	Hossein Alizadeh et al. (2019)	Iran
SGT	Stochastic Gradient Tree	Angileri et al. (2016)	Italy

Almohamad, 2020; Alsafadi et al., 2022). However, it is essential to use field-based observations to validate the model outputs. The development of new techniques, such as machine learning algorithms, can be used for model validation and provide better insight into the processes and drivers of soil erosion and runoff. However, the performance of these techniques for assessing soil erosion and runoff in agriculture dominated landscapes is poorly understood. The aim of this study is to evaluate the performance of four machine learning algorithms, generalized linear model (GLM), Random Forest (RF), elastic net regression (EN), and multiple adaptive regression splines (MARS) in predicting soil erosion and runoff in Syria. This evaluation is important because soil erosion and runoff pose a major threat to agricultural development in many countries including Syria. The objectives of the study were to: (i) measure soil erosion and runoff in three different slopes in the coastal region of Syria, (ii) predict soil erosion and runoff by using four ML algorithm, and (iii) evaluate the performance of four ML algorithms for soil erosion and runoff estimation. Results can provide information for environment-oriented non-governmental organizations (NGO), land planners, and policy makers to better understand the causes and processes of soil erosion and runoff in the eastern Mediterranean (Pal et al., 2022; Phinzi, Abriha, & Szabó, 2021) and to choose better alternatives for soil erosion control and management.

2 | MATERIALS AND METHODS

2.1 | Study area

The study was conducted in the Al-Abrash River basin (36°27′, 35°95′ E, 34°69′ and 34°91′ N), which is one of the most important basins in Tartous Governorate, Syria (Figure 1). It is part of the Syrian coastal basin that extends to the Mediterranean Sea. The region is characterized by the Mediterranean climate with cold, rainy winter and dry summer. Average temperatures range from 8°C in January to 28°C in August, which is the hottest month of the year with the highest average temperatures. Rainfall in the study area ranges from 800 mm to 1400 mm yr^{-1}, occurring mainly in winter (November–March). September is generally the onset of precipitation, and December/January is when it peaks (Figure 1). According to data collected from Syrian Ministry of Agriculture for the Safita agro-climate station (central of Al-Abrash River Basin), the average rainfall was 1106 mm during 1960–2020. Records show: 21.95% is received in autumn/fall, 54.79% in winter, 22.45% in spring, and 0.98% in summer (Figure 1).

The topography of the region is dominated by mountain ranges in the upper part of the basin, while the plateau predominates in the central part. Because of this landscape, land use and land cover are distributed differently within the basin. Agricultural land accounts for about 65% of the basin's land, followed by forest land with 21%, urban areas with 5%, water bodies with 5%, and others with 4%. Due to the large variation in temperature with altitude above sea level, there is a discrepancy between the type of crops existing in this study area and the type of natural vegetation cover prevails. In this typical agriculture basin, protected crops such as tomato and cucumber (4%) dominate the downstream region and are mainly irrigated. The middle part of basin has predominantly rain-fed agriculture, mainly olive trees (57%) and crops (25%). In the upstream mountains, fruit plantation such as citrus (7%), apple (7%) is common (Mohammed, Alsafadi, et al., 2020). The common soil orders are Entisols and Inceptisols according to the American taxonomy of soils, where the parent material is calcareous rocks (Mohammed, Al-Ebraheem, et al., 2020).

2.2 | Experimental sites and data collection

An initial soil survey was conducted throughout the study area to determine soil types. Three major soil types were identified in the study area. Based on soil type and slope, three experimental locations were marked for this study. The first location was Kafar-Snief (KS) (*x*: 36.17, *y*: 34.812) with a slope of 20%, the second was Snisinia (SS) (*x*: 36.133, *y*: 34.776) with a slope of 12%, while the last was Rawas (RS) (*x*: 36.025, *y*: 34.748) with a slope of 8% (Table 2). Surface soil samples (0–10 cm) were randomly collected from each study and thoroughly mixed to obtain a homogenous sample before being

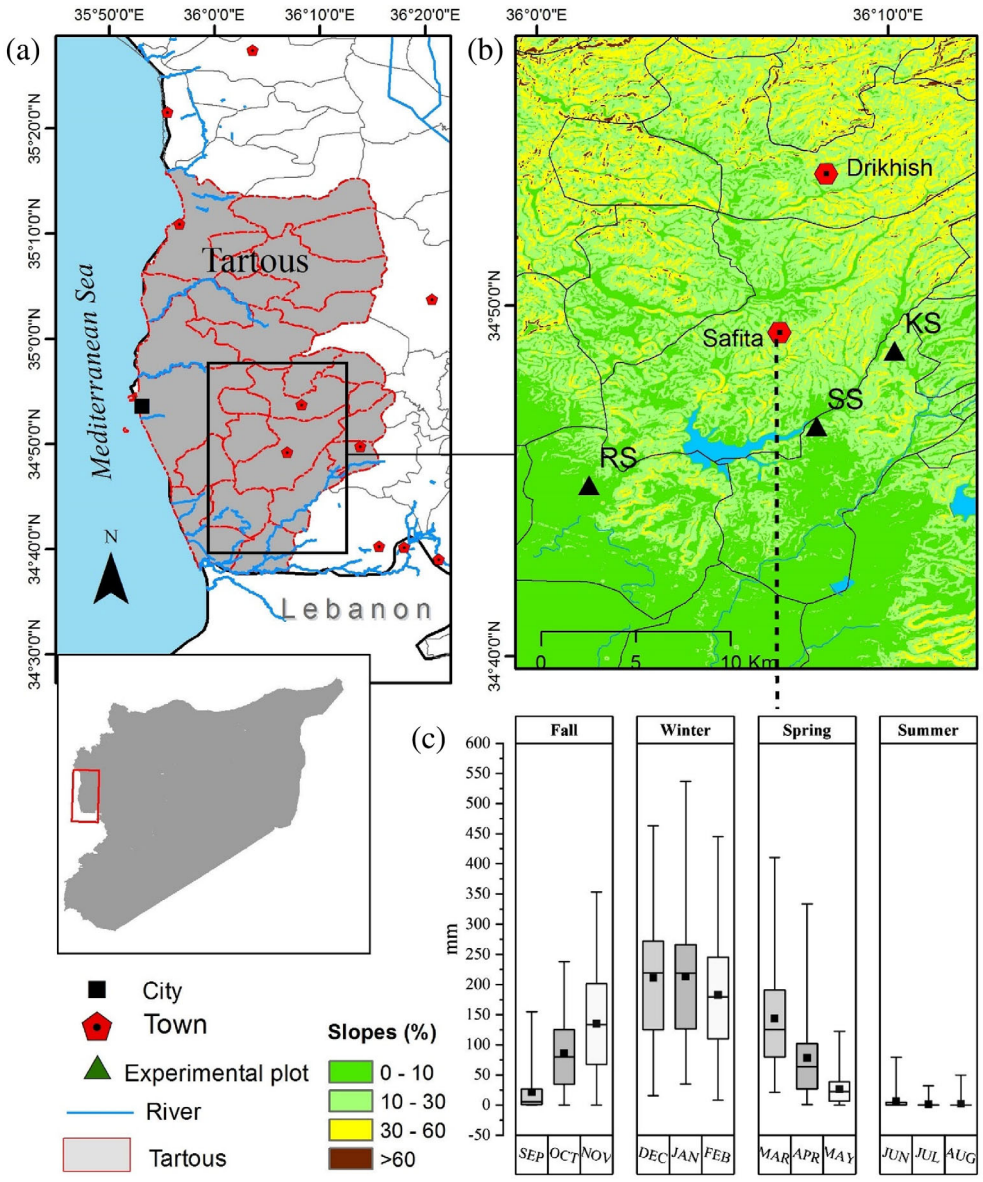

FIGURE 1 Map of the study area: (a) location of the study area, (b) experimental plots (RS, SS, and KS) within Al-Abrash River basin, (c) box plot of the monthly rainfall in the study area (196–2020). [Colour figure can be viewed at wileyonlinelibrary.com]

transported to the laboratory for routine soil analysis. Analysis included soil texture, electrical conductivity (EC) (ds/m), and organic matter (OM%) (Table 2).

Experimental plots (2.25 m × 1.50 m × 0.50 m; 0.10 m depth in the soil) were established at each site to measure the volume of runoff and eroded soil. Each plot was equipped with a 250-L container (tank) (Figure 2) to trap soil sediment and eroded debris. A rain gauge was placed adjacent to experimental plots to measure the depth of rainfall after the rains. This approach was previously implemented and verified in Syria by Meshram et al. (2020) and Mohammed et al. (2021).

Based on the field observation of Kbibo et al. (2017) and the national recommendation of the Ministry of Environment, soil erosion samples should be collected after each rainfall event exceeding 10 mm (i.e., effective rain). Thus, in this research, we followed this recommendation, where a 1-L sample was collected from each container. A total of 72 samples (24 events × 3 locations) were collected during the monitoring period. Before collecting the samples, the volume of water (i.e., runoff) in each container was accurately measured (mm). Then, the contents (water + eroded soil) of each container were carefully mixed to ensure sample homogeneity (1 L). Containers were emptied, thoroughly cleaned, and covered in preparation for the next rain event. This process was repeated each time it rained during the study period (between 10/24/2019 and 4/5/2020). The soil from each sample was separated from the water by the sedimentation process, and then dried in a soil drying oven for 24 hr at 105°C. The dried soil was weighed and assigned to the accumulated runoff volume in the tank (kg m^{-2}).

2.3 | Machine learning models

In this study, the predictive power of four ML models was tested for rainfall, runoff, and the slope factors.

TABLE 2 Land use, slope conditions, and basic soil properties of the studied locations (means, $n = 3$).

Location	X	Y	Slope %	Land use	Silt %	Clay %	Sand %	EC (ds m^{-1})	OM%
RS	36.025	34.748	8	Olive	23.42	34.31	42.27	1.1	1.97
SS	36.133	34.776	12	Citrus	35.96	38.05	26.02	0.95	2.11
KS	36.17	34.812	20	Pomegranate	24.01	38.12	37.87	1	1.31

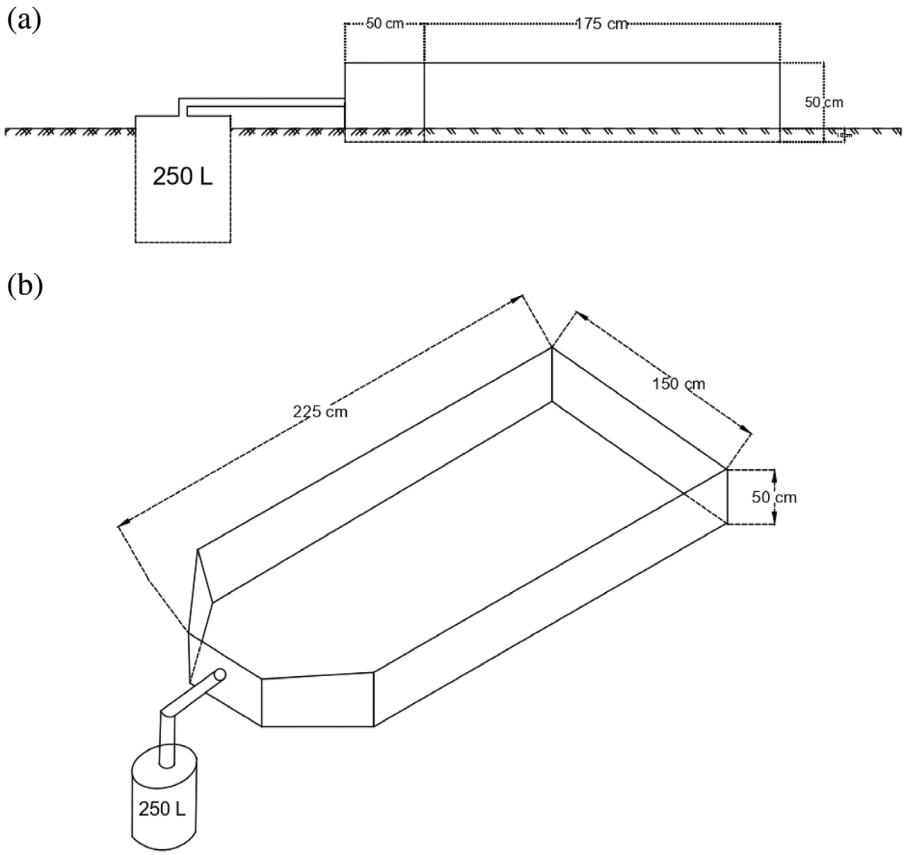

FIGURE 2 Design of the experimental plot in each location: (a) section, (b) 3D view.

2.3.1 | Generalized linear model (GLM)

The GLM was selected as a linear model with prerequisites on the normal distribution of residuals and the homoscedasticity. The model was developed by McCullagh and Nelder (1989), and the algorithm is used in cases where a target variable is explained by ***n*** independent variables and hence creates a linear relationship between the variables (Pourghasemi et al., 2020). In this study, the GLM was the analysis of covariance (ANCOVA model) with an ordinal predictor (slope) and two covariates (rainfall and runoff). We investigated the role of the variables and reported the result of the hypothesis tests, and beside the significance, the effect sizes ($\epsilon^2 p$) were also calculated.

GLM was also used as a predictor function and as a statistical function known as LOGIT that is expressed as Equation (1):

$$\boldsymbol{Y} = \mathbf{Pr}(\boldsymbol{Y} = \mathbf{1}) = \frac{e^{C_0 + C_1 X_1 \ldots + C_n X_n}}{\mathbf{1} + e^{C_0 + C_1 X_1 \ldots + C_n X_n}} \quad (1)$$

$\boldsymbol{Y}$ is the dependent variable and x_1, x_2,… x_n are independent observations.

2.3.2 | Random Forest

The RF is a robust algorithm that relies on bootstrapping (i.e., random sampling) and decision trees (Breiman, 2001). Many decision trees are involved in classification and prediction. The output of RF could be represented as Equation (2):

$$H(x) = \arg_z^{\max} \sum_{i=1}^{k} I(h_i x = Z) \quad (2)$$

Where: $H(x)$: RF model, h_i: the model of single decision tree, Z:output.

In this study, 500 decision trees were defined, and each decision tree was built based on bootstrapped samples (i.e., repetitions), with random selection of variables with replacement. Due to its predictive performance, ability to reduce variance in the statistical learning method, and feature importance ranking, RF has become a widely applied technique in many application areas, including gully erosion, landslides, and runoff susceptibility (Pásztor, 2021; Phinzi et al., 2020; Sun et al., 2020). For variable importance ranking, the RF model uses the mean decrease accuracy (%IncMSE; based on repetitions

providing the average increase in squared residuals) and the mean decrease Gini (IncNodePurity; the quality of the decision trees splits). Both %IncMSE and IncNodePurity reflect the variable importance, with higher values indicating higher importance (Kuhn et al., 2008).

2.3.3 | Elastic net regression

The elastic net regression (EN) is a type of regularized regression, which overcomes the limitations of LASSO (least absolute shrinkage and selection operator) and ridge regression (Zou & Hastie, 2005). Ordinary least square (OLS) regression has problems when the independent variables are correlated, and this leads to a decline in the accuracy of the resulting model. The ridge and LASSO methods apply a penalty term (L1-norm and L2-norm, respectively) to the OLS equation to obtain a better model with lower bias and variance. However, both LASSO and ridge have limitations, EN applies a weighted combination of the two regularizations, that is, penalizing the correlating variables, and those that contribute less to the explained variance. Rainfall and runoff had a correlation of $r = 0.79$ ($p < 0.001$), and thus could bias the results, EN was an effective method to address this problem. Following Zou and Hastie (2005), the EN algorithm is expressed in the Equation 3:

$$(\beta) = \frac{1}{2}\frac{\sum_{i=1}^{n}\left(b_i - x^i\beta\right)^2}{n} + \lambda\left(\frac{1-\alpha}{2}\sum_{j=1}^{m}\beta^2 + \alpha\sum_{j=1}^{m}|\beta_j|\right) \quad (3)$$

α and λ are the turning parameters while α combines the other two regulations alternatives.

2.3.4 | Multiple adaptive regression splines (MARS)

The multiple adaptive regression splines (MARS) was first introduced by Friedman and Roosen (1995). It is as a non-parametric version of multiple linear regression without prerequisites on the variables (Mosavi, Golshanc, et al., 2020; Rotigliano et al., 2018). MARS models are developed on segmented sections of the multivariate space defined by the dependent and independent variables. Linear regression models are then run on all segments. Each segment has a linear function with its own function parameters and the connections of lines are called the knots. All knots have two basic functions (BFs). Next, BFs will be the independent variables of the final model. MARS model is expressed in Equation 4:

$$f(x) = sign\left(C_0 + \sum_{m=1}^{k} C_m b_m(x)\right) \quad (4)$$

$f(x)$ is the output regression model; C_0 is the constant coefficient; k is the number of basic functions in the model; c_m is the coefficient of the m basis function; $b_m(x)$ is the m the basis function.

All models were built with the caret package (Kuhn et al., 2020) of R 4.04 (Pinheiro et al., 2021) combined with hyper-parameter tuning (fine tuning of model parameters with the built-in procedure of the caret package) aimed at achieving the smallest RMSE.

2.4 | Principal component analysis

We applied standardized principal component analysis (PCA) using the correlation matrix to study the effect of slope classes on the multidimensional space defined by erosion, rainfall and runoff. Model fit was assessed using the root mean square residual (RMSR). RMSR<0.05 indicates very good, <0.1 good model fit while higher values were excluded because they do not support the application of PCA (Basto & Pereira, 2012). The principal components (PCs) scores were visualized in a scatterplot diagram and slope classes were delineated.

2.5 | Model validation

Model validation was performed using an independent dataset obtained by splitting the dataset into a training and a testing part using a ratio of 70:30 stratified random sampling on slope classes. The training data were used to train the models and the model prediction was based on the test data. This accuracy assessment approach was the independent accuracy testing (IAT). Differences between the predicted and observed data were evaluated using Taylor diagrams. Taylor diagrams show the correlation between observed and predicted values, the squared root of the difference between observed and predicted values (root mean square error, RMSE), and report the standard deviation of observed values and predicted values. We visualized the results by the regression algorithms and slope steepness. Furthermore, Nash-Sutcliffe coefficients (NSE) were also determined by models (Nash & Sutcliffe, 1970).

The models were developed using the repeated k-fold cross-validation (RKCV) method. Hence, fivefolds and 10 repetitions were applied: data had been randomly split fivefold and four were used to train the model while onefold was kept testing the predicted values with independent data. This procedure was repeated until all folds had become test data. Then, the entire process was repeated 10 times, and accuracy data from 50 models was available. R^2, and RMSE were reported in the evaluation, and we calculated the basic statistics (LQ: lower quartile, UQ: upper quartile, IQR: inter-quartile range, median, minimum, and maximum) of the 50 models, we also determined the relative RMSE (rRMSE) as the ratio of RMSE and mean erosion.

Although both IAT and RKCV are reliable measures for assessing model accuracy, there are some differences. The IAT evaluates predictions based on different factors (in this case by slope categories), but relies on a single random split, while RKCV uses 50 models to provide prediction accuracy, but only as an overall evaluation. The RKCV can be considered an independent measure of accuracy. However, we examined the model efficacy on a completely independent dataset (testing data) as an extrapolation. In addition, the combined application of the approaches allowed us to exploit the advantages of these approaches.

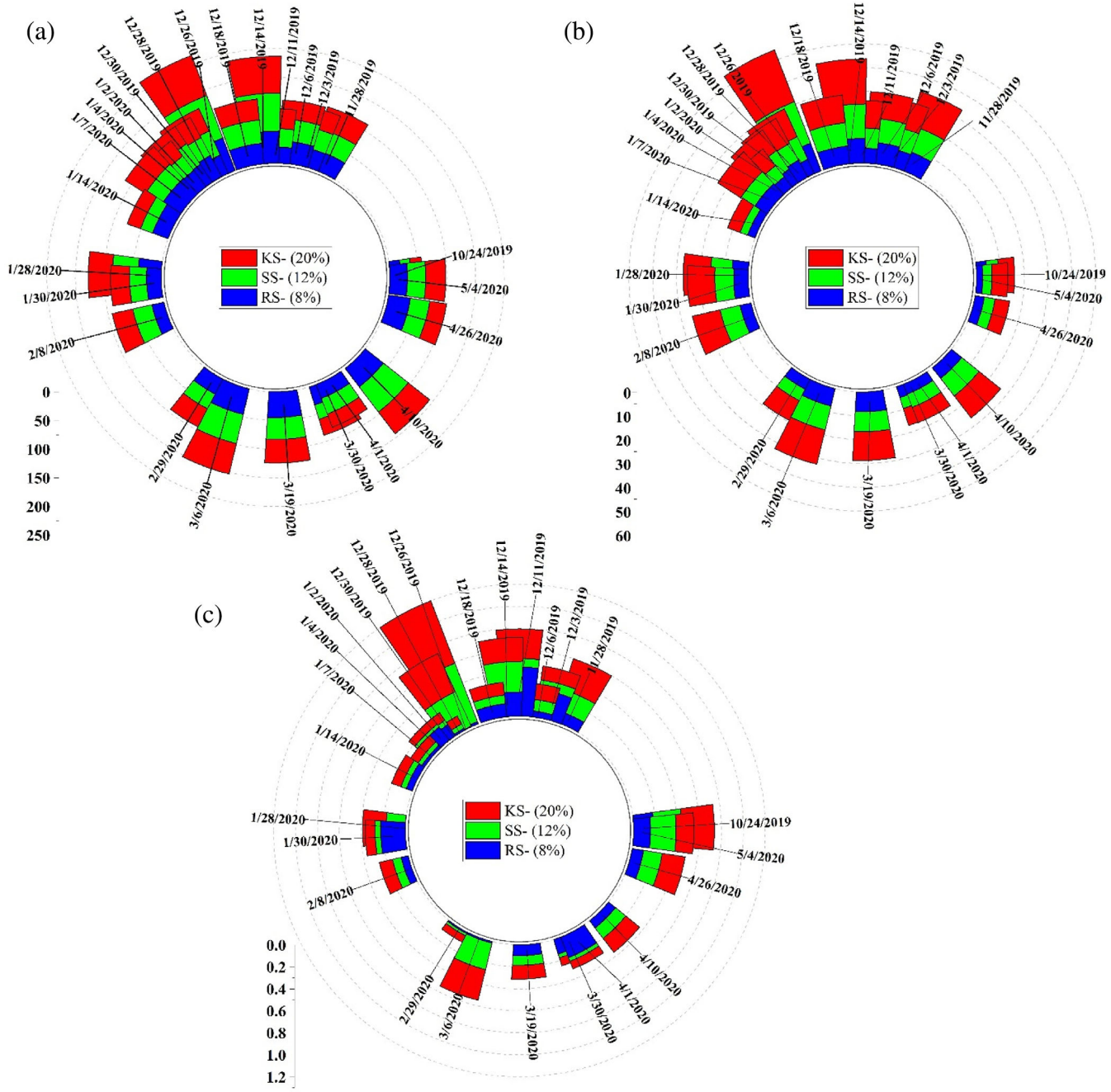

FIGURE 3 Measured rainfall (mm) (a), runoff (mm) (b), and soil erosion (kg m^{-2}) (c) after each rainfall event in the RS, SS, and KS experimental plots. [Colour figure can be viewed at wileyonlinelibrary.com]

3 | RESULTS

3.1 | Soil erosion and runoff in the coastal region of Syria

Cumulative rainfall during the study period was 946 ± 30.86 mm measured during 24 rainstorm events (rainfall >10 mm) on each plot (Figure 3). The highest runoff per rainfall event that was recorded at RS (8%) was 12.33 mm in December 2019 (26/12/2019) while the highest erosion, 0.44 kg m^{-2}, was observed at the beginning of the rainy season (24/10/2019). The highest values of runoff (18.60 mm and 23.31 mm) and soil erosion (0.56 kg m^{-2}, and 0.60 kg/m^2) at SS (12%) and KS (20%), respectively, also occurred at the end of December (26/12/2019) (Figure 3).

Although the average amount of rainfall in the study area was similar at all study sites, the average amount of rainfall was slightly higher at KS compared to the other locations (Figure 4 and Table 3). A similar pattern was observed for mean runoff and mean soil erosion. The regression models showed a positive and significant ($p < 0.05$) relationship between rainfall and runoff. The relationship was strongest at the KS plot ($R^2 = 0.85$; $p < 0.05$ $n = 24$). In terms of eroded soil, the impact of rainfall was not significant at the RS plot ($R^2 = 0.1$;

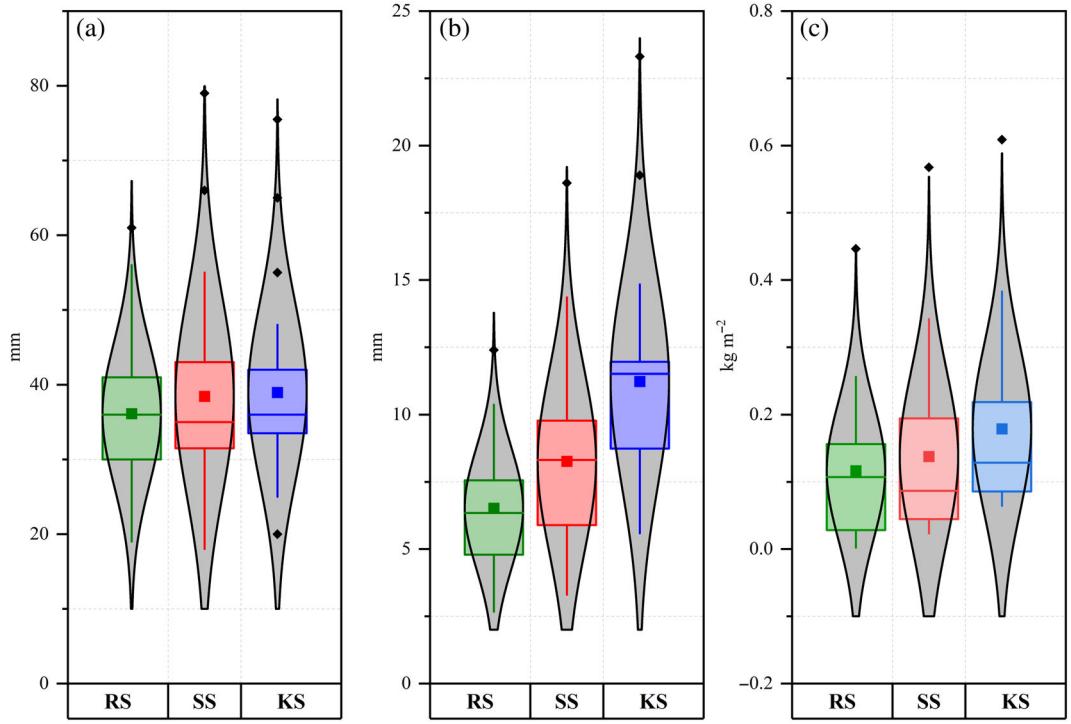

FIGURE 4 Boxplot ($n = 24$) of observed rainfall (mm) and measured soil erosion (kg m^{-2}) and runoff (mm) in the studied locations (RS, SS, KS): (a) rainfall, (b) runoff, and (c) soil erosion [median (-----); mean (•)]. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Basic descriptive statistics for the studied variables ($n = 24$).

Statistic	Minimum	Maximum	Range	Median	Mean	Standard deviation (n)	Skewness (Pearson)	Kurtosis (Pearson)	Standard error of the mean
RS Rainfall (mm)	19.00	61.00	42.00	36.00	36.12	9.37	0.69	0.70	1.91
SS Rainfall (mm)	18.00	79.00	61.00	35.00	38.44	13.13	1.36	1.98	2.68
KS Rainfall (mm)	20.00	75.50	55.50	36.00	38.96	11.81	1.38	2.21	2.41
RS Runoff (mm)	2.67	12.40	9.73	6.35	6.52	2.19	0.60	0.34	0.45
SS Runoff (mm)	3.30	18.61	15.31	8.31	8.26	3.29	1.08	1.94	0.67
KS Runoff (mm)	5.58	23.31	17.73	11.51	11.23	3.87	1.15	1.99	0.79
RS Erosion (kg m^{-2})	0.00	0.45	0.44	0.11	0.12	0.10	1.38	2.40	0.02
SS Erosion (kg m^{-2})	0.02	0.57	0.54	0.09	0.14	0.13	1.73	3.13	0.03
KS Erosion (kg m^{-2})	0.06	0.61	0.54	0.13	0.18	0.12	1.79	3.47	0.03

$p > 0.05$ $n = 24$), however, it was significant at the other locations (Table 4 and Figure 5).

The analysis of covariance (ANCOVA) showed that runoff had a significant effect on soil erosion ($p = 0.02$) with a medium effect ($\varepsilon^2 p = 0.26$). However, rainfall events and slope categories were not significant ($p > 0.05$) and had a small impact ($\varepsilon^2 p < 0.01$) (Table 5).

3.2 | Multivariate evaluation of erosion, runoff, and rainfall

PCA was confirmed by the RMSR of 0.06% (good fit) with two PCs explaining 93% of the total variance. PC1 explained 58% of the variance and was correlated with rainfall and runoff. PC2 explained 35% of the variance and correlated with erosion. The performance of the slope classes was similar in the multivariate space (Figure 6), thus, based on slope classes, PCA was not efficient in discriminating slopes. Although erosion and runoff at RS (8%) were different from the results at SS (12%) and KS (20%), the differences were not significant for PC1 ($F = 2.952$, df = 2.72, $p = 0.058$) and PC2 ($F = 1.323$, df = 2.72, $p = 0.273$).

3.3 | Performance of machine learning models

According to the RKCV, the MARS algorithm had the best performance, $R^2 = 0.406$, with respect to the median of the 50 models. The

Location	Regression equation	r^2	p
RS	Runoff (mm) = 0.047 + 0.17 Rainfall depth (mm)	0.58	0.00
	Erosion (kg m^{-2}) = 0.24–0.0034 × Rainfall depth (mm)	0.1	0.15
SS	Runoff (mm) = − 0.40 + 0.22 × Rainfall depth (mm)	0.8	0.00
	Erosion (kg m^{-2}) = − 0.11 + 0.006 × Rainfall depth (mm)	0.47	0.00
KS	Runoff (mm) = − 0.6 + 0.30 × Rainfall depth (mm)	0.85	0.00
	Erosion (kg m^{-2}) = − 0.069 + 0.0063 × Rainfall depth (mm)	0.37	0.00

TABLE 4 Regression equation, correlation, and significant of correlation between observed rainfall and both measured runoff and soil erosion in the three studied locations (RS, SS, KS) (n = 24).

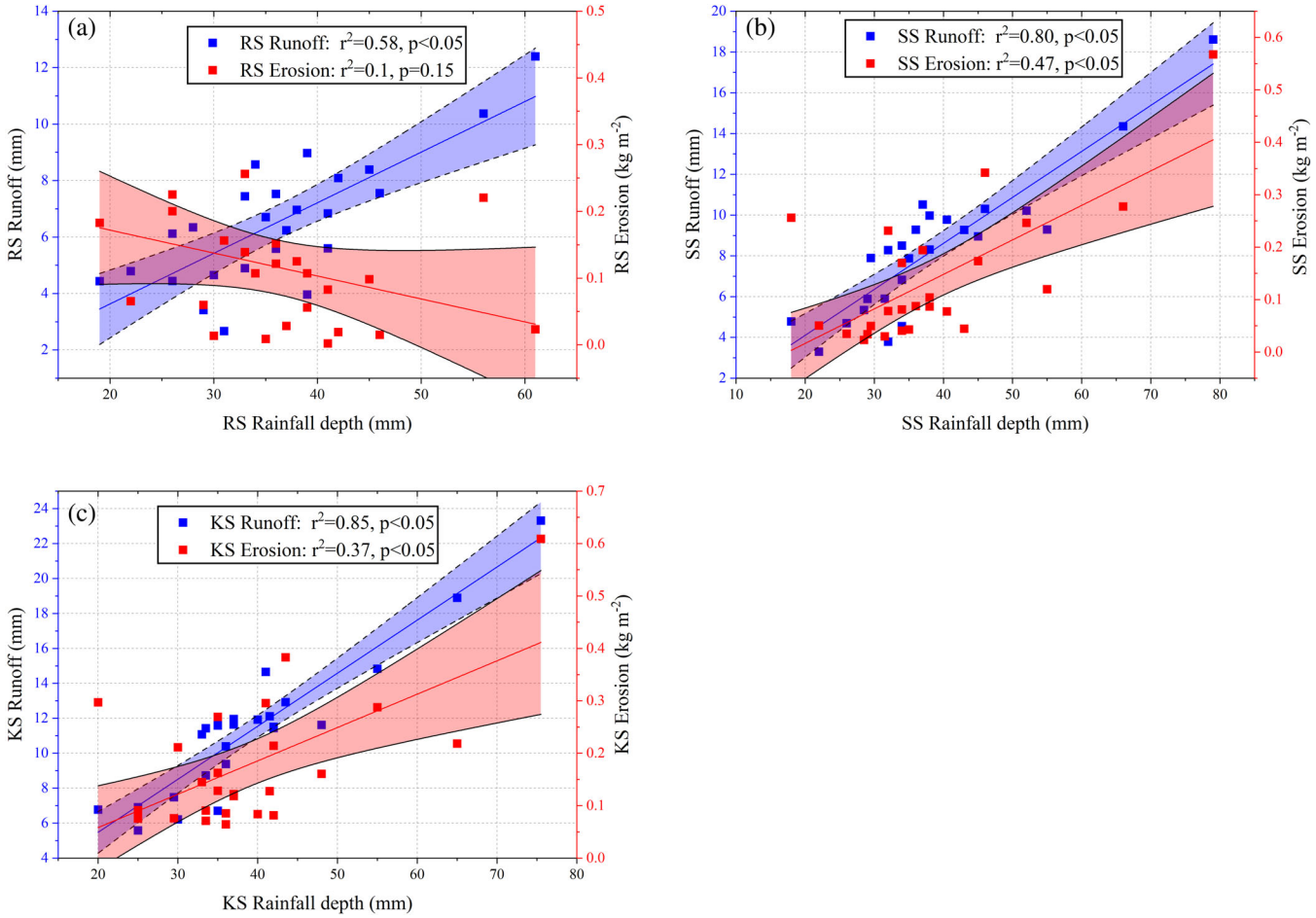

FIGURE 5 Scatter plot (n = 24) between observed rainfall and measured soil erosion and runoff in an event scale in (a) RS, (2) SS, and (3) KS. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Summary of ANCOVA analysis for soil erosion data and other factorial variables (i.e., observed rainfall and both measured runoff and slope (RS, SS, KS).) during the monitoring period (SS: Sum of Squares, df: degree of freedom, F: F-statistic, p: significance, $\varepsilon^2 p$: effect size).

	SS	Df	F	p	$\varepsilon^2 p$
Model	0.06437	4	6.68663	< 0.001	0.235
Runoff	0.06083	1	5.49486	0.022	0.263
Rainfall	8.04 e−5	1	0.00726	0.932	−0.013
Slope inclination	0.00346	2	0.15604	0.856	−0.024
Residuals	0.77499	70			
Total	2.40242	75			

RMSE was also the smallest for MARS, the mean rRMSE was 52%. Compared to the other models, upper quartile of RMSE was lower than the medians for GLM, EN, and RF. Based on the 50 models, MARS and EN had the largest IQR of the R^2 (0.563), but the MARS's LQ was the highest (0.097) and EN's LQ was only 0.001 better (i.e., 0.057) than that of GLM. The EN regression was the weakest model, with the median R^2 of 0.167; nevertheless, the largest rRMSE belonged to GLM (74.3%). The RF provided the smallest IQR (0.302) but the UQ was only 0.36 indicating a poor fit with high rRMSE values (82%) (Figure 7).

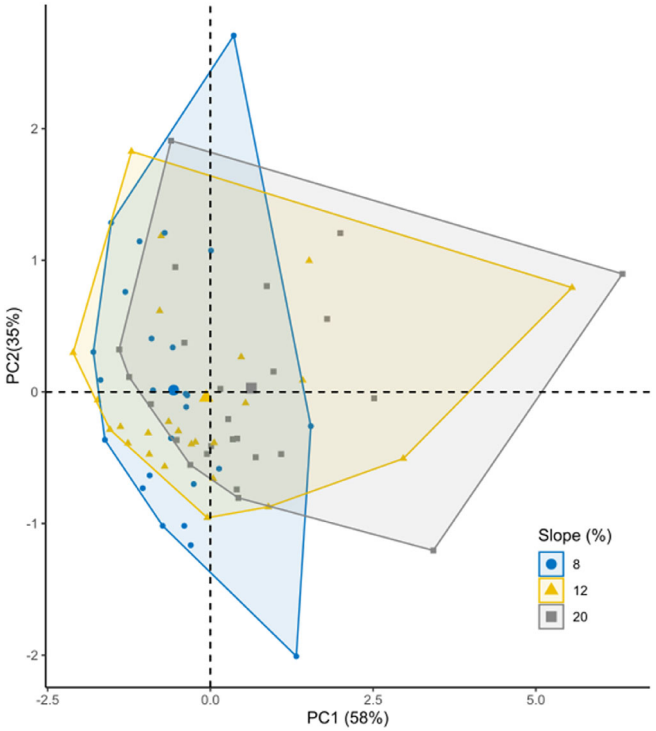

FIGURE 6 Principal component analysis for studied variables (soil erosion, runoff, rainfall) based on slope classes. [Colour figure can be viewed at wileyonlinelibrary.com]

The IAT provided similar results to the RKCV in terms of model performance, RF and MARS had the strongest correlation with the observed data and had the smallest RMSE. RF had better correlation and lower RMSE than all other models, but MARS's SD was the closest to the real values (Figure 8). NSEs also confirmed that RF and MARS were the most efficient models (GLM: 0.15;EN: 0.19; MARS: 0.31; RF: 0.42).

We also analyzed the effect of slope inclination on the predicted values, which was only possible with the IAT approach. We found that the slightest slope (8%) had the worst predicted values, in some cases this slope category did not provide reasonable outcome and could not be plotted on the Taylor diagrams, the correlation tended towards zero and RMSEs were high (Figure 9). The best predictions were observed with the 12% slope category, and even the steepest slope (20%) category made the prediction efficacy worse.

4 | DISCUSSION

This research used an experimental approach to validate the results of the implemented machine learning algorithms. Samples from three experimental locations were used to extract soil properties as well as slope inclination and rainfall data, which were then used to calculate soil erosion. It is crucial to define the most significant soil parameters that are most related to soil erosion. Parameters that provide information about soil structure and formation are most relevant, such as

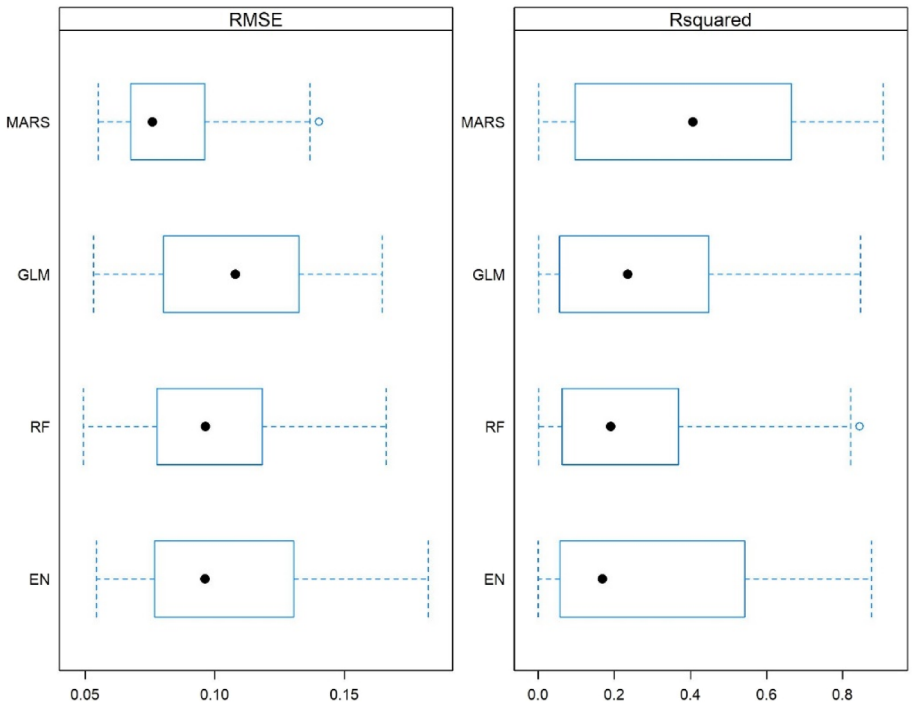

FIGURE 7 Accuracy metrics (RMSE, R^2) of performance of machine learning models (MARS: multiple adaptive regression splines, GLM: generalized liner model, RF: Random Forest, EN: elastic net regression). [Colour figure can be viewed at wileyonlinelibrary.com]

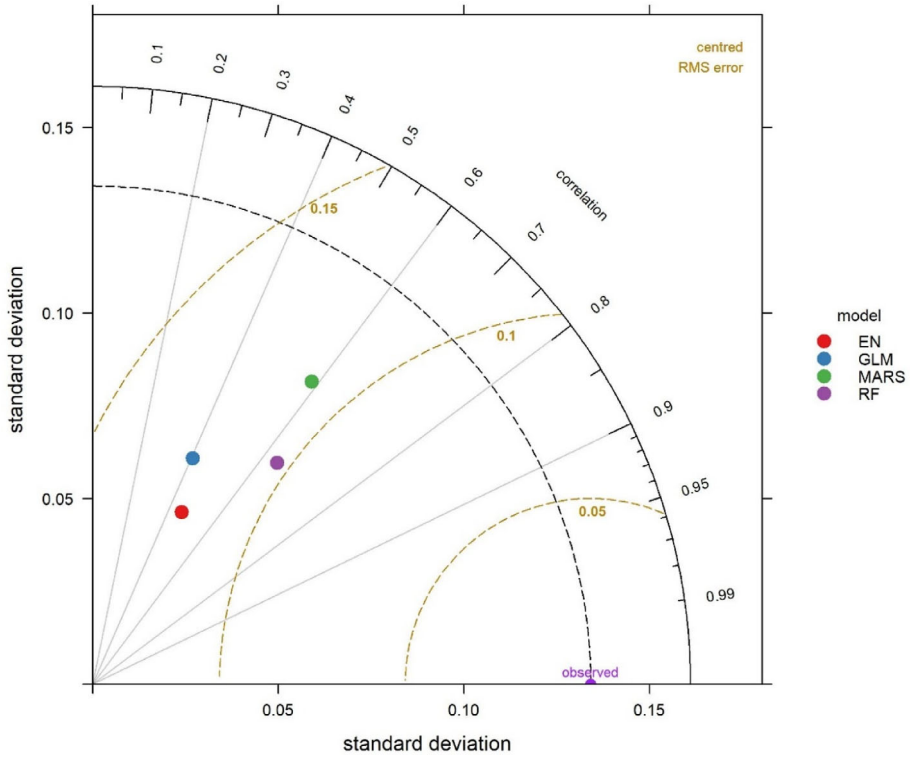

FIGURE 8 Independent accuracy testing (IAT) for machine learning models (MARS: multiple adaptive regression splines, GLM: generalized liner model, RF: Random Forest, EN: elastic net regression). [Colour figure can be viewed at wileyonlinelibrary.com]

texture, moisture, organic matter content, and aggregate stability. These parameters describe how soil particles aggregate to come together, as the size of aggregated particles affects the ease of soil dispersion, such as dispersion in runoff water (Diaz-Gonzalez et al., 2022). In this sense, soil characteristics, including moisture, organic matter content, aggregate stability, and soil particle size distribution, play a critical role in influencing the rate and pattern of soil erosion, as they control the rate of water infiltration, runoff generation, and erosion processes. For instance, soil structure, soil water holding capacity, and the nature of the landform (slope) in the landscape strongly affect the rate of water infiltration, runoff generation, soil mechanical behavior, ease of soil detachment, and soil erosion. These processes have been shown to exhibit distinct spatio-temporal characteristics that may determine the nature of erosion at the plot, hillslopes, and catchment scales (Auzet et al., 2004; Mosavi, Sajedi-Hosseini, et al., 2020).

4.1 | Soil erosion and runoff in the eastern Mediterranean

The susceptibility to erosion and runoff in the costal part of Syria results from the interaction among the main erosive factors, including soil characteristics, heavy rainfall, and severe inclination (Masri et al., 2015; Mohammed et al., 2021; Mohammed, Al-Ebraheem, et al., 2020; Mohammed, Alsafadi, et al., 2020). Although we also confirmed these observations, we found that the triggering (slopes and rainfalls) and modifying (soil characteristics and vegetation cover) factors had different weights. Erosion and rainfall had a positive and significant relationship (R^2 = 0.10–0.47) from the bivariate regression models at the SS and KS sites, but this was not the case for the RS (Table 4, Figure 5). However, the multivariate model (ANCOVA) indicated that neither rainfall nor slope inclination had a significant relationship with soil erosion (Table 5); that is, involving all sites and variables in a single model, it was found that the importance of the main trigger was not significant in this case. The frequency of intense rainfalls events (i.e., erosivity) is significant in the coastal part of Syria (Abdo, 2020), which could influence the sediment transport capacity of runoff. Thus, the fact that rainfall events were equal in quantity, but the intensities were different, resulted in variations in runoff and soil erosion (Table 3 and Figure 4).

With respect to modifying factors, the results agreed with Lal (1994), who reported that regardless of rainfall amount, vegetation coverage dissipates the kinetic energy of raindrops. The identical findings of Kulimushi, Choudhari, et al. (2021) and Kulimushi, Maniragaba, et al. (2021) indicated that under constant conditions of rainfall intensity, poor land use practices and land topography are the most important factors that determine the intensity of runoff and soil erosion. Rainfalls and runoff were highly correlated (r = 0.79 p < 0.001), but the correlation between runoff and soil erosion was moderate (r = 0.52 p < 0.001), accordingly, land cover and soil properties got higher relevance, which justified the findings of Nasir and Selvakumar (2018). The eastern part along the Mediterranean is characterized by clay rich soils, including clay and clay loam. In general, clay soils have a naturally well-developed soil structure, which holds the particles together and increases their cohesion. Favourable soil texture combined with organic matter increases aggregate stability, soil permeability, and water infiltration rate, which considerably tolerates the aggressiveness effect of rainfall on soil particles (Kulimushi, Bashagaluke, et al., 2021; Zeraatpisheh et al., 2021). We can hence

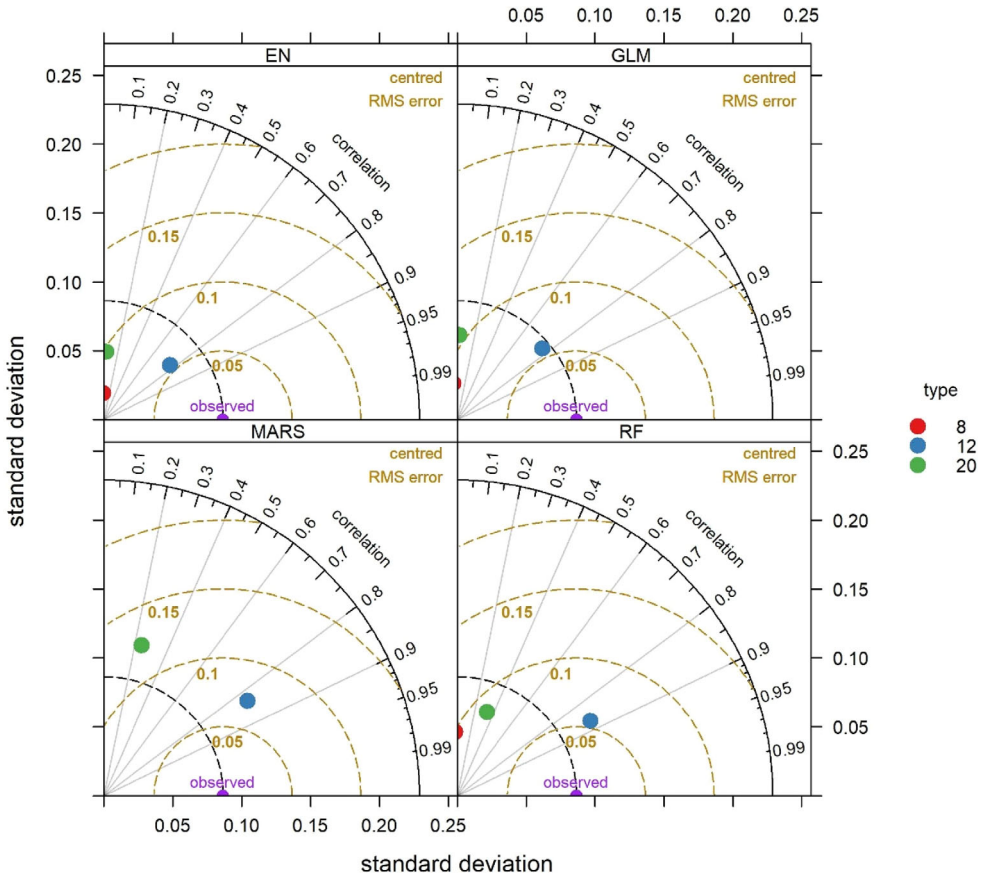

FIGURE 9 Effect of slope inclination on the performance of the machine learning algorithms, which depicted by using IAT approach (MARS: multiple adaptive regression splines, GLM: generalized liner model, RF: Random Forest , EN: elastic net regression). [Colour figure can be viewed at wileyonlinelibrary.com]

confirm that the observed variations in runoff and soil erosion are not due to the difference in texture classes. For instance, RS (8%) recorded low runoff and soil erosion while the site is characterized by clay loam classified as medium erodible soil texture, while KS (20%) a site characterized by heavy soil (clay) showed higher values of runoff and soil erosion compared to RS (8%). In addition, SS (12%) site has good OM content (2.11%) but runoff and soil erosion were not lower than RS (8%). Thus, the observed runoff and soil erosion could then be explained by other factors that require further investigation.

4.2 | Applicability of machine learning algorithms for soil erosion estimation

Although machine leaning is increasing in popularity in many fields, it has only recently been widely applied for soil erosion estimation and mapping (Dinh et al., 2021; Javidan et al., 2019; Nguyen et al., 2021; Sahour et al., 2021). However, the performance of many algorithms for water-induced soil erosion is mostly poorly understood and requires detailed analysis. We revealed that RF and MARS models are more reliable in predicting soil erosion than GLM and EN. Similar findings were reported by Previously, Garosi et al. (2019) concluded that RF models performed better in predicting soil erosion where soil texture ranged from sandy clay loam to sandy loam and land use types were of orchard rangeland and farmland. However, in contrast to our results which showed that MARS apparently performed better than RF ($R^2_{RKCV} = 40.6\%$), but the $RMSE_{IAT}$ was more favourable with RF. This result is consistent with the findings of Nguyen et al. (2021): although RF performed better overall, MARS also provided good modeled values. These results draw attention to the importance of using an independent test dataset, as RKCV was efficient in model building, but as a measure of accuracy, IAT provided a more reliable outcome.

According to ANCOVA, the relevance of slope steepness did not directly influence erosion (i.e., slope was not significant and had a small effect size); however, we identified that the predictive ability of the machine learning models was dependent on slopes: an 8% slope had no correlation with observed values, and MARS was not able to predict its values correctly; furthermore, a 20% slope almost doubled the RMSE compared to the well-predicted values of a 12% slope (Figure 9). This means that, in general, the models predicted erosion with a reasonable error, and especially the RF had the performance (Figure 8), but, with respect to the slope categories, the 12% provided the best modeled values. For the 8% and 20% slopes, the role of soil properties and land use biased the overall trend, which also caused the small effect sizes for slopes and rainfalls (Table 5). Another cause of the ambiguous role of slopes was the multivariate similarity in rainfall, runoff, and erosion characteristics of the three types of steepness (Figure 6), which justifies the relevance of soil and land use.

4.3 | Research limitation and future outlook

Although the models evaluated in this study produced reasonable results, resource limitations constrained us to limit data collection to three experimental plots. This did not allow for adequate variation in landscape properties and other processes capable of influencing model performance. Therefore, future studies should seek to include more experimental plots that are more representative of landscape features such as land use and topographical features. The plurality of machine learning algorithms implies that future evaluation of machine learning algorithms should include more diverse algorithms applied under different environmental, social, and economic conditions to gain more insight into their performance.

5 | CONCLUSIONS

We aimed to reveal the effectiveness of four ML algorithms in predicting soil erosion using experimental data from three plots in the Eastern Mediterranean. We made the following observations:

- Slope inclination and rainfall were not the most important factors of soil erosion. Although runoff was the most important, its relationship with erosion was only moderate, and the R^2 was only 0.27. Therefore, runoff triggered erosion but rainfall intensity, soil properties (OM and clay content), and vegetation cover may have reduced its erosivity.
- The robust RF and MARS algorithms were superior to the linear GLM and EN models in predicting soil erosion. However, even the MARS, with the best performance, had a median R^2 (of 50 models) of 40.6, with a larger RMSE (52%), which is a consequence of similar erosion rate at the SS and KS sites.
- RKCV and IAT highlighted important features of the models, and we found that using the independent dataset (IAT), the RF model was the best with a high correlation between the modeled and observed data (0.85) and low RMSE (0.06). However, the slope inclination of 8% was not a successful part of the model having unusable results.
- The results draw attention to the fact that, although slopes are important factors in soil erosion, other environmental factors (e.g., land use, soil properties) can overwrite this general rule, and site-specific characteristics require special attention. Soil conservation practices need to be re-evaluated in terms of runoff reduction.

The output of this research calls the attention to the need of careful consideration of the erosion factors, which requires a site-specific evaluation including the soil properties and land cover beside rainfall and slope inclination.

ACKNOWLEDGMENTS

Authors would like to thank Debrecen University for supporting open access publishing. Additionally, the authors would like to thank the anonymous reviewers and editors for their valuable feedback.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Safwan Mohammed https://orcid.org/0000-0003-2311-6789
Glory O. Enaruvbe https://orcid.org/0000-0002-8235-0834
Firas Alsilibe https://orcid.org/0000-0003-3327-262X
Luc Cimusa Kulimushi https://orcid.org/0000-0003-2228-0042
Szilárd Szabó https://orcid.org/0000-0002-2670-7384

REFERENCES

Abdo, H., & Salloum, J. (2017). Spatial assessment of soil erosion in Alqerdaha basin (Syria). *Modeling Earth Systems and Environment*, *3*(1), 26. https://doi.org/10.1007/s40808-017-0294-z

Abdo, H. G. (2020). Evolving a total-evaluation map of flash flood hazard for hydro-prioritization based on geohydromorphometric parameters and GIS–RS manner in Al-Hussain River basin, Tartous, Syria. *Natural Hazards*, *104*(1), 681–703. https://doi.org/10.1007/s11069-020-04186-3

Abdo, H. G. (2021). Estimating water erosion using RUSLE, GIS and remote sensing in Wadi-Qandeel River basin, Lattakia, Syria. *Proceedings of the Indian National Science Academy*, *87*, 1–10. https://doi.org/10.1007/s43538-021-00047-0

Abedini, M., Ghasemian, B., Shirzadi, A., Shahabi, H., Chapi, K., Pham, B. T., Ahmad, B., & Tien Bui, D. (2019). A novel hybrid approach of bayesian logistic regression and its ensembles for landslide susceptibility assessment. *Geocarto International*, *34*(13), 1427–1457. https://doi.org/10.1080/10106049.2018.1499820

Alewell, C., Borrelli, P., Meusburger, K., & Panagos, P. (2019). Using the USLE: Chances, challenges, and limitations of soil erosion modelling. *International Soil and Water Conservation Research*, *7*(3), 203–225. https://doi.org/10.1016/j.iswcr.2019.05.004

Almohamad, H. (2020). Impact of land cover change due to armed conflicts on soil erosion in the basin of the northern Al-Kabeer River in Syria using the RUSLE model. *Water*, *12*(12), 3323. https://doi.org/10.3390/w12123323

Al-Mukhtar, M. (2019). Random forest, support vector machine, and neural networks to modelling suspended sediment in Tigris River-Baghdad. *Environmental Monitoring and Assessment*, *191*(11), 1–12. https://doi.org/10.1007/s10661-019-7821-5

Alsafadi, K., Bi, S., Abdo, H. G., Al Sayah, M. J., Ratonyi, T., Harsanyi, E., & Mohammed, S. (2022). Spatial–temporal dynamic impact of changes in rainfall erosivity and vegetation coverage on soil erosion in the eastern Mediterranean. *Environmental Science and Pollution Research*, 1–19. https://doi.org/10.1007/s11356-022-24012-6

Angileri, S. E., Conoscenti, C., Hochschild, V., Märker, M., Rotigliano, E., & Agnesi, V. (2016). Water erosion susceptibility mapping by applying stochastic gradient treeboost to the Imera Meridionale River basin (Sicily, Italy). *Geomorphology*, *262*, 61–76. https://doi.org/10.1016/j.geomorph.2016.03.018

Arabameri, A., Pradhan, B., & Lombardo, L. (2019). Comparative assessment using boosted regression trees, binary logistic regression, frequency ratio and numerical risk factor for gully erosion susceptibility modelling. *Catena*, *183*, 104223. https://doi.org/10.1016/j.catena.2019.104223

Arabameri, A., Pradhan, B., Pourghasemi, H. R., Rezaei, K., & Kerle, N. (2018). Spatial modelling of gully erosion using GIS and R programing:

A comparison among three data mining algorithms. *Applied Sciences*, *8*(8), 1369. https://doi.org/10.3390/app8081369

Arif, N., & Danoedoro, P. (2017). Analysis of artificial neural network in erosion modeling: A case study of Serang watershed. In *IOP conference series: Earth and environmental science* (Vol. 98, 012027). London, UK: IOP Publishing. https://doi.org/10.1088/1755-1315/98/1/012027

Auzet, A.-V., Poesen, J., & Valentin, C. (2004). Editorial: Soil surface characteristics: Dynamics and impacts on soil erosion. *Earth Surface Processes and Landforms*, *29*(9), 1063–1064. https://doi.org/10.1002/esp.1101

Bakacsi, Z., Tóth, T., Makó, A., Barna, G., Laborczi, A., Szabó, J., Szatmári, G., & Pásztor, L. (2019). National level assessment of soil salinization and structural degradation risks under irrigation. *Hungarian Geographical Bulletin*, *68*(2), 141–156. https://doi.org/10.15201/hungeobull.68.2.3

Basto, M., & Pereira, J. M. (2012). An SPSS R-menu for ordinal factor analysis. *Journal of Statistical Software*, *46*(4), 1–29. https://doi.org/10.18637/jss.v046.i04

Benchettouh, A., Kouri, L., & Jebari, S. (2017). Spatial estimation of soil erosion risk using RUSLE/GIS techniques and practices conservation suggested for reducing soil erosion in Wadi Mina watershed (northwest, Algeria). *Arabian Journal of Geosciences*, *10*(4), 79. https://doi.org/10.1007/s12517-017-2875-6

Bertalan, L., Holb, I., Pataki, A., Szabó, G., Szalóki, A. K., & Szabó, S. (2022). UAV-based multispectral and thermal cameras to predict soil water content–a machine learning approach. *Computers and Electronics in Agriculture*, *200*, 107262. https://doi.org/10.1016/j.compag.2022.107262

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf

Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., & Edwards, T. C., Jr. (2015). Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, *239*, 68–83. https://doi.org/10.1016/j.geoderma.2014.09.019

Cerda, A., Rodrigo-Comino, J., Novara, A., Brevik, E. C., Vaezi, A. R., Pulido, M., Giménez-Morera, A., & Keesstra, S. D. (2018). Long-term impact of rainfed agricultural land abandonment on soil erosion in the Western Mediterranean basin. *Progress in Physical Geography: Earth and Environment*, *42*(2), 202–219. https://doi.org/10.1177/0309133318758521

Cheng, Z., Lu, D., Li, G., Huang, J., Sinha, N., Zhi, J., & Li, S. (2018). A random forest-based approach to map soil erosion risk distribution in Hickory plantations in western Zhejiang Province, China. *Remote Sensing*, *10*(12), 1899. https://doi.org/10.3390/rs10121899

de Farias, C. A. S., & Santos, C. A. G. (2014). The use of Kohonen neural networks for runoff–erosion modeling. *Journal of Soils and Sediments*, *14*(7), 1242–1250. https://doi.org/10.1007/s11368-013-0841-9

de Hipt, F. O., Diekkrueger, B., Steup, G., Yira, Y., Hoffmann, T., & Rode, M. (2018). Modeling the impact of climate change on water resources and soil erosion in a tropical catchment in Burkina Faso, West Africa. *Catena*, *163*, 63–77. https://doi.org/10.1016/j.catena.2017.11.023

Diaz-Gonzalez, F. A., Vuelvas, J., Correa, C. A., Vallejo, V. E., & Patino, D. (2022). Machine learning and remote sensing techniques applied to estimate soil indicators–review. *Ecological Indicators*, *135*, 108517. https://doi.org/10.1016/j.ecolind.2021.108517

Dinh, T. V., Nguyen, H., Tran, X. L., & Hoang, N. D. (2021). Predicting rainfall-induced soil erosion based on a hybridization of adaptive differential evolution and support vector machine classification. *Mathematical Problems in Engineering*, *2021*, 1–20. https://doi.org/10.1155/2021/6647829

Dunne, T., Malmon, D. V., & Mudd, S. M. (2010). A rain splash transport equation assimilating field and laboratory measurements. *Journal of Geophysical Research: Earth Surface*, *115*(F1), 1–6. https://doi.org/10.1029/2009JF001302

Dutta, S. (2016). Soil erosion, sediment yield and sedimentation of reservoir: A review. *Modeling Earth Systems and Environment*, *2*(3), 1–18. https://doi.org/10.1007/s40808-016-0182-y

Friedman, J. H., & Roosen, C. B. (1995). An introduction to multivariate adaptive regression splines. *Statistical Methods in Medical Research*, *4*, 197–217. https://doi.org/10.1177/096228029500400303

García-Ruiz, J. M. (2010). The effect of land uses on soil erosion in Spain: A review. *Catena*, *81*(1), 1–11. https://doi.org/10.1016/j.catena.2010.01.001

García-Ruiz, J. M., Beguería, S., Nadal-Romero, E., González-Hidalgo, J. C., Lana-Renault, N., & Sanjuán, Y. (2015). A meta-analysis of soil erosion rates across the world. *Geomorphology*, *239*, 160–173. https://doi.org/10.1016/j.geomorph.2015.03.008

García-Ruiz, J. M., Nadal-Romero, E., Lana-Renault, N., & Beguería, S. (2013). Erosion in mediterranean landscapes: Changes and future challenges. *Geomorphology*, *198*, 20–36. https://doi.org/10.1016/j.geomorph.2013.05.023

Garosi, Y., Sheklabadi, M., Conoscenti, C., Pourghasemi, H. R., & Van Oost, K. (2019). Assessing the performance of GIS-based machine learning models with different accuracy measures for determining susceptibility to gully erosion. *Science of the Total Environment*, *664*, 1117–1132. https://doi.org/10.1016/j.scitotenv.2019.02.093

Hossein Alizadeh, M., Kariminejad, N., Chen, W., Pourghasemi, H. R., Alinejad, M., Behbahani, A. M., & Tiefenbacher, J. P. (2019). Gully headcut susceptibility modeling using functional trees, naïve Bayes tree, and random forest models. *Geoderma*, *342*, 1–11. https://doi.org/10.1016/j.geoderma.2019.01.050

Javidan, N., Kavian, A., Pourghasemi, H. R., Conoscenti, C., & Jafarian, Z. (2019). Gully erosion susceptibility mapping using multivariate adaptive regression splines—Replications and sample size scenarios. *Water*, *11*(11), 2319. https://doi.org/10.3390/w11112319

Kbibo, I., Ibrahim, J., & Bou-Issa, A. (2017). Studying the effect of soil erosion for eight different systems with different slopes in the coastal area under forests, burned forest and planted soil system. *ishreen University Journal for Research and Scientific Studies Biological Sciences Series*, *39*, 25–38 (in Arabic).

Keesstra, S. D., Rodrigo-Comino, J., Novara, A., Giménez-Morera, A., Pulido, M., Di Prima, S., & Cerda, A. (2019). Straw mulch as a sustainable solution to decrease runoff and erosion in glyphosate-treated clementine plantations in eastern Spain. An assessment using rainfall simulation experiments. *Catena*, *174*, 95–103. https://doi.org/10.1016/j.catena.2018.11.007

Khallouf, A., Talukdar, S., Harsányi, E., Abdo, H. G., & Mohammed, S. (2021). Risk assessment of soil erosion by using CORINE model in the western part of Syrian Arab Republic. *Agriculture & Food Security*, *10*(1), 1–15. https://doi.org/10.1186/s40066-021-00295-9

Kuhn, J., Neumann, S., & Egert, B. (2008). Building blocks for automated elucidation of metabolites: Machine learning methods for Nmr prediction. *BMC Bioinformatics*, *9*, 400. https://doi.org/10.1186/1471-2105-9-400

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., & Kenkel, B. (2020). *Caret: Classification and Regression Training. R package version 6.0-86*. Astrophysics Source Code Library: Cambridge, MA, USA.

Kulimushi, L. C., Bashagaluke, J. B., Choudhari, P., Masroor, M., & Sajjad, H. (2021). Novel combination of analytical hierarchy process and weighted sum analysis for watersheds prioritization. A study of Ulindi catchment, Congo River basin. *Geocarto International*, *37*, 1–39. https://doi.org/10.1080/10106049.2021.2002426

Kulimushi, L. C., Choudhari, P., Mubalama, L. K., & Banswe, G. T. (2021). GIS and remote sensing-based assessment of soil erosion risk using RUSLE model in south-Kivu province, eastern, Democratic Republic of Congo. *Geomatics, Natural Hazards and Risk*, *12*(1), 961–987. https://doi.org/10.1080/19475705.2021.1906759

Kulimushi, L. C., Maniragaba, A., Choudhari, P., Elbeltagi, A., Uwemeye, J., Rushema, E., & Singh, S. K. (2021). Evaluation of soil erosion and sediment yield spatio-temporal pattern during 1990–2019. *Geomatics, Natural Hazards and Risk*, *12*(1), 2676–2707. https://doi.org/10.1080/19475705.2021.1973118

Lal, R. (1994). Water management in various crop production systems related to soil tillage. *Soil and Tillage Research*, *30*(2–4), 169–185. https://doi.org/10.1016/0167-1987(94)90004-3

Lal, R., Mokma, D., & Lowery, B. (2018). Relation between soil quality and erosion. In R. Ratta, & R. Lal (Eds.), *Soil quality and soil erosion* (pp. 237–258). CRC Press.

Licznar, P., & Nearing, M. A. (2003). Artificial neural networks of soil erosion and runoff prediction at the plot scale. *Catena*, *51*(2), 89–114. https://doi.org/10.1016/S0341-8162(02)00147-9

Masri, Z., Van Donk, S., Bruggeman, A., & Turkelboom, F. (2015). Post-harvest summer tillage to control wind erosion in the Khanasser Valley, Syria. *Aeolian Research*, *17*, 219–229. https://doi.org/10.1016/j.aeolia.2015.03.004

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed., p. 532). Chapman and Hall/CRC.

Meshram, S. G., Singh, V. P., Kisi, O., Karimi, V., & Meshram, C. (2020). Application of artificial neural networks, support vector machine and multiple model-ANN to sediment yield prediction. *Water Resources Management*, *34*(15), 4561–4575. https://doi.org/10.1007/s11269-020-02672-8

Mohammed, S., Alaa, K., Omran, A., Quoc, B. P., Nguyen, T. T. L., Van, N. T., Duong, T. A., & Endre, H. (2021). Predicting soil erosion hazard in Lattakia Governorate (W Syria). *International Journal of Sediment Research*, *36*(2), 207–220. https://doi.org/10.1016/j.ijsrc.2020.06.005

Mohammed, S., Al-Ebraheem, A., Holb, I. J., Alsafadi, K., Dikkeh, M., Pham, Q. B., Linh, N. T. T., & Szabo, S. (2020). Soil management effects on soil water erosion and runoff in Central Syria—A comparative evaluation of general linear model and Random Forest regression. *Water*, *12*(9), 2529. https://doi.org/10.3390/w12092529

Mohammed, S., Alsafadi, K., Talukdar, S., Kiwan, S., Hennawi, S., Alshihabi, O., Sharaf, M., & Harsanyie, E. (2020). Estimation of soil erosion risk in southern part of Syria by using RUSLE integrating geo informatics approach. *Remote Sensing Applications: Society and Environment*, *20*, 100375. https://doi.org/10.1016/j.rsase.2020.100375

Mohammed, S., Kbibo, I., Alshihabi, O., & Mahfoud, E. (2016). Studying rainfall changes and water erosion of soil by using the WEPP model in Latakia, Syria. *Journal of Agricultural Sciences, Belgrade*, *61*(4), 375–386. https://doi.org/10.2298/JAS1604375M

Mosaffaei, Z., & Jahani, A. (2021). Modeling of ash (Fraxinus excelsior) bark thickness in urban forests using artificial neural network (ANN) and regression models. *Modeling Earth Systems and Environment*, *7*(3), 1443–1452. https://doi.org/10.1007/s40808-020-00869-9

Mosavi, A., Golshanc, M., Janizadehd, S., Choubine, B., Melesse, A. M., & Dineva, A. A. (2020). Ensemble models of GLM, FDA, MARS, and RF for flood and erosion susceptibility mapping: A priority assessment of sub-basins. *Geocarto International*, *37*, 2541–2560. https://doi.org/10.1080/10106049.2020.1829101

Mosavi, A., Sajedi-Hosseini, F., Choubin, B., Taromideh, F., Rahi, G., & Dineva, A. (2020). Susceptibility mapping of soil water erosion using machine learning models. *Water*, *12*(7), 1995. https://doi.org/10.3390/w12071995

Muluneh, A., Bewket, W., Keesstra, S., & Stroosnijder, L. (2017). Searching for evidence of changes in extreme rainfall indices in the central Rift Valley of Ethiopia. *Theoretical and Applied Climatology*, *128*(3), 795–809. https://doi.org/10.1007/s00704-016-1739-4

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nasir, N., & Selvakumar, R. (2018). Influence of land use changes on spatial erosion pattern, a time series analysis using RUSLE and GIS: The cases of Ambuliyar sub-basin, India. *Acta Geophysica*, *66*(5), 1121–1130. https://doi.org/10.1007/s11600-018-0186-2

Nguyen, K. A., Chen, W., Lin, B.-S., & Seeboonruang, U. (2021). Comparison of ensemble machine learning methods for soil erosion pin measurements. *ISPRS International Journal of Geo-Information*, *10*(1), 42. https://doi.org/10.3390/ijgi10010042

Pásztor, L. (2021). Advanced GIS and RS applications for soil and land degradation assessment and mapping. *ISPRS International Journal of Geo-Information*, *10*(3), 128. https://doi.org/10.3390/ijgi10030128

Phinzi, K., Abriha, D., Bertalan, L., Holb, I., & Szabó, S. (2020). Machine learning for gully feature extraction based on a pan-sharpened multispectral image: Multiclass vs. binary approach. *ISPRS International Journal of Geo-Information*, *9*(4), 252. https://doi.org/10.3390/ijgi9040252

Phinzi, K., Abriha, D., & Szabó, S. (2021). Classification efficacy using k-fold cross-validation and bootstrapping resampling techniques on the example of mapping complex gully systems. *Remote Sensing*, *13*(15), 2980. https://doi.org/10.3390/rs13152980

Phinzi, K., Ngetar, N. S., & Ebhuoma, O. (2021). Soil erosion risk assessment in the Umzintlava catchment (T32E), Eastern Cape, South Africa, using RUSLE and random forest algorithm. *South African Geographical Journal*, *103*(2), 139–162. https://doi.org/10.1080/03736245.2020.1716838

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2021). R Core Team (2021). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-152. https://cran.r-project.org/web/packages/nlme/

Pourghasemi, H. R., Sadhasivam, N., Kariminejad, N., & Collins, A. L. (2020). Gully erosion spatial modelling: Role of machine learning algorithms in selection of the best controlling factors and modelling process. *Geosciences Frontiers*, *11*, 2207–2219. https://doi.org/10.1016/j.gsf.2020.03.005

Rahmati, O., Tahmasebipour, N., Haghizadeh, A., Pourghasemi, H. R., & Feizizadeh, B. (2017). Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. *Geomorphology*, *298*, 118–137. https://doi.org/10.1016/j.geomorph.2017.09.006

Rotigliano, E., Martinello, C., Agnesi, V., & Conoscenti, C. (2018). Evaluation of debris flow susceptibility in El Salvador (CA): A comparison between multivariate adaptive regression splines (MARS) and binary logistic regression (BLR). *Geography Bulletin*, *67*, 361–373. https://doi.org/10.15201/hungeobull.67.4.5

Saha, S., Roy, J., Arabameri, A., Blaschke, T., & Tien Bui, D. (2020). Machine learning-based gully erosion susceptibility mapping: A case study of eastern India. *Sensors*, *20*(5), 1313. https://doi.org/10.3390/s20051313

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. *Journal of Hydrology*, *591*, 125321. https://doi.org/10.1016/j.jhydrol.2020.125321

Sahour, H., Gholami, V., Vazifedan, M., & Saeedi, S. (2021). Machine learning applications for water-induced soil erosion modeling and mapping. *Soil and Tillage Research*, *211*, 105032. https://doi.org/10.1016/j.still.2021.105032

Shahin, M. A. (2016). State-of-the-art review of some artificial intelligence applications in pile foundations. *Geoscience Frontiers*, *7*(1), 33–44. https://doi.org/10.1016/j.gsf.2014.10.002

Sun, D., Wen, H., Wang, D., & Xu, J. (2020). A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology*, *362*, 107201. https://doi.org/10.1016/j.geomorph.2020.107201

Szatmári, G., Bakacsi, Z., Laborczi, A., Petrik, O., Pataki, R., Tóth, T., & Pásztor, L. (2020). Elaborating Hungarian segment of the global map of salt-affected soils (GSSmap): National contribution to an international initiative. *Remote Sensing*, *12*(24), 4073. https://doi.org/10.3390/rs12244073

Van Leeuwen, C. C., Cammeraat, E. L., de Vente, J., & Boix-Fayos, C. (2019). The evolution of soil conservation policies targeting land abandonment and soil erosion in Spain: A review. *Land Use Policy*, *83*, 174–186. https://doi.org/10.1016/j.landusepol.2019.01.018

Wang, F., Shi, Z., Biswas, A., Yang, S., & Ding, J. (2020). Multi-algorithm comparison for predicting soil salinity. *Geoderma*, *365*, 114211. https://doi.org/10.1016/j.geoderma.2020.114211

Zeraatpisheh, M., Ayoubi, S., Mirbagheri, Z., Mosaddeghi, M. R., & Xu, M. (2021). Spatial prediction of soil aggregate stability and soil organic carbon in aggregate fractions using machine learning algorithms and environmental variables. *Geoderma Regional*, *27*, e00440. https://doi.org/10.1016/j.geodrs.2021.e00440

Zeraatpisheh, M., Garosi, Y., Reza, H., Ayoubi, S., Taghizadeh-mehrjardi, R., Scholten, T., & Xu, M. (2022). Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. *Catena*, *208*, 105723. https://doi.org/10.1016/j.catena.2021.105723

Zingg, A. W. (1940). Degree and length of land slope as it affects soil loss in run-off. *Agricultural Engineering*, *21*, 59–64.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

How to cite this article: Mohammed, S., Jouhra, A., Enaruvbe, G. O., Bashir, B., Barakat, M., Alsilibe, F., Cimusa Kulimushi, L., Alsalman, A., & Szabó, S. (2023). Performance evaluation of machine learning algorithms to assess soil erosion in Mediterranean farmland: A case-study in Syria. *Land Degradation & Development*, *34*(10), 2896–2911. https://doi.org/10.1002/ldr.4655