JOURNAL OF DEGRADED AND MINING LANDS MANAGEMENT
Volume 11, Number 4 (July 2024):6471-6488, doi:10.15243/jdmlm.2024.114.6471
ISSN: 2339-076X (p); 2502-2458 (e), www.jdmlm.ub.ac.id

Review

A comprehensive survey exploring the application of machine learning algorithms in the detection of land degradation

Gangamma Hediyalad[1*], Ashoka K[2], Govardhan Hegade[3] , Pratibha Ganapati Gaonkar[4], Azizkhan F Pathan[5], Pratibhaa R Malagatti[6]

[1] Computer Science and Engineering Department, Bapuji Institute of Engineering and Technology, Davangere, Affiliated to Visvesvaraya Technological University, Belagavi-590018, India
[2] Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Affiliated to Visvesvaraya Technological University, Belagavi-590018, India
[3] Computer Science and Engineering, Manipal Institute of Technology, Manipal, India
[4] Artificial Intelligence and Machine Learning, St. Joseph Engineering College, Mangalore, An Autonomous Institution, 575001, India
[5] Information Science and Engineering, Jain Institute of Technology, Davangere, Affiliated to Visvesvaraya Technological University, Belagavi-590018, India
[6] Computer Science and Engineering Department, KLE Technological University, Hubli, 580020, India

[*]corresponding author: gangahediyalad@gmail.com

Article history:
Received 4 April 2024
Revised 28 May 2024
Accepted 10 June 2024

Keywords:
comprehensive survey
desertification
land degradation
machine learning

Abstract

Early and reliable detection of land degradation helps policymakers to take strict action in more vulnerable areas by making strong rules and regulations in order to achieve sustainable land management and conservation. The detection of land degradation is carried out to identify desertification processes using machine learning techniques in different geographical locations, which are always a challenging issue in the global field. Due to the significance of the detection of land degradation, this article provides an exhaustive review of the detection of land degradation using machine learning algorithms. Initially, the current status of land degradation in India is presented, along with a brief discussion on the overview of widely used factors, evaluation parameters, and algorithms used. Consequently, merits and demerits related to machine learning-based land degradation identification are presented. Additionally, solutions are prescribed in order to reduce existing problems in the detection of land degradation. Since one of the major objectives is to explore the future perspectives of machine learning-based land degradation detection, areas including the application of remote sensing, mapping, optimum features, and algorithms have been broadly discussed. Finally, based on a critical evaluation of existing related studies, the architecture of the machine learning-based desertification process has been proposed. This technology can fulfill the research challenges in the detection of land degradation and computation difficulties in the development of models for the detection of land degradation.

To cite this article: Hediyalad, G., Ashoka, K., Hegade, G., Gaonkar, P.G., Pathan, A.F. and Malagatti, P.R. 2024. A comprehensive survey exploring the application of machine learning algorithms in the detection of land degradation. Journal of Degraded and Mining Lands Management 11(4):6471-6488, doi:10.15243/jdmlm.2024.114.6471.

Introduction

Land is an important part of our ecosystem and a significant resource that we need to protect. Land degradation is a global issue that impacts many nations around the world. This denotes the combination of improper human actions and climatic variations that leads to a decline in the overall quality and fertility of the land, thereby reducing its productivity and impact on human lives, health, and food security, which is always a challenging issue in the global field. Desertification is the deterioration of land in arid, semiarid, and dry sub-humid regions due to weather variation and improper human activities that affect 8 billion people, especially farmers, poor people, and rural communities. Human activities such as afforestation, unsustainable agricultural practices, biomass burning, overgrazing, excessive groundwater extraction, rapid growth in industrialization and urbanization, inadequate irrigation methods, irrigation with poor water quality, and increasing population lead to an increase in land degradation. In addition to climatic variations such as drastic changes in climatic conditions such as high temperatures, high rain, and drought, which affect people to move from one place to another as a last livelihood strategy, extreme weather conditions can also lead to land degradation. Thus, through innovative and integrated approaches, policymakers can develop strong strategies to reduce or reverse land degradation. However, because of numerous complex factors, the detection of land degradation is always challenging. Because decertification is dynamic in nature, it is important to give continuous attention in order to achieve sustainable land management and conservation.

In recent years, different machine learning methods have assisted researchers in identifying desertification more accurately than traditional approaches because of their capacity to handle huge amounts of data, mainly non-linear, high-dimensional, and complicated interactions with missing values (Bhattacharya, 2013) and in finding the correlation between the target variable and other predictor variables from the training dataset using computer algorithms. Numerous mathematical models have been devised to outline regions susceptible to desertification (Salvati and Zitti, 2009; Dasgupta et al., 2013; Symeonakis, 2016). The Environmental Sensitivity Index is evaluated by computing the Soil Quality Index, Climate Quality Index, Vegetation Quality Index, and Management Quality Index within the MEDALUS methodology (Kosmas et al., 1999). Despite the model's limitation of assigning equal weight to all indices (Salvati et al., 2009), it serves as a valuable tool, especially in Mediterranean climates, due to its straightforward model construction and adaptability in indicator selection (Trott et al., 2015). Expert judgment is required for assigning scores to different classes within the model (Giordano et al., 2003). Salvati and Zitti (2009) introduced an index considering both biophysical and socioeconomic factors, yielding results comparable to MEDALUS. In another context, Karamesouti et al. (2015) utilized PESERA and TERON models to evaluate soil erosion rates, while the MEDALUS methodology assessed overall desertification risk in a traditional Mediterranean cropland ecosystem. Dasgupta et al. (2013) employed fuzzy inference rules for delineating environmentally vulnerable areas. Jafari and Bakhshandehmehr (2013) integrated fuzzy logic with the Environmental Sensitivity Index to map desertification sensitivity in Central Iran. Dutta and Chaudhuri (2015) investigated environmentally sensitive regions in the Jhunjhunun and Sikar districts of Rajasthan, India, revealing that 13% of the study area was highly susceptible to desertification**.** Machine learning approaches are successfully used in tackling earth-related issues because they significantly improve the accuracy of the model compared to traditional models. Random Forest (RF) is widely used for accurately predicting target variables (Devasena, 2014). Another significant machine learning algorithm is the Support Vector Machine (SVM), which has gained acceptance and usefulness with the advancement of artificial intelligence and Remote Sensing Geographic Information Systems (RS-GIS) techniques (Huang et al., 2018).

This is the first comprehensive review article that addresses the limitations of existing articles that identify land degradation using machine learning algorithms. Hence, the contributions of this paper are summarized as follows:

1. Providing the current status of land degradation
2. Describing the current aspects of the land degradation detection process
3. The benefits and challenges associated with features and machine-learning algorithms are investigated.
4. Expanding the areas of future research on machine learning-based land degradation identification.

A variety of machine-language techniques such as classification and regression tree (Lawrence et al., 2004), artificial neural network (Priori et al., 2014), K-nearest neighbor (Mansuy et al., 2014), multinomial logistic regression (Kempen et al., 2009), support vector machine (Kovačevic et al., 2010; Priori et al., 2014), and Random Forest Model (RFM) (Kim et al., 2012; Vågen et al., 2016) were tested by different authors in natural resources mapping. This paper mainly concentrates on the detection of land degradation using machine learning algorithms by gathering the required information from the most recent articles, as shown in Table 1, and provides future directions for research work. This paper takes a step forward by gathering a wide range of features and machine learning algorithms used in the assessment of diverse environmental issues in order to advance the adoption of machine learning algorithms in desertification. The rest of the paper is structured as follows: Section 2 explores the selection of articles based on inclusion and exclusion criteria. Section 3 presents the current status of land degradation in India. Section 4 provides the benefits and challenges associated with features, and machine learning algorithms are investigated. Section 5 recommends future directions for research challenges to improve the performance of the model for the detection of land degradation. Section 6 presents challenges and future directions for the detection of land degradation. Finally, the paper is concluded in Section 7.

Table 1. ML Techniques used in land degradation assessment.

Paper	Methods used	Outcome
Setargie et al. (2023)	RF	This study assessed gully erosion susceptibility areas with the finest resolution dataset by varying runoff curve numbers using Random Forest.
Das et al. (2023)	RF, GBM, and DL	An ensemble of RF and DL coefficients of determination (R2) of 0.89 and 0.55 and normalized root-mean-squared error of 0.15 and 0.16 in training and test datasets, respectively.
Bag et al. (2022)	SVM, CART, BRT, and RF	The author identified that elevation, drainage density (DD), and NDVI factors increase soil erosion hotspots using the best-performing RF model.
Chuma et al. (2023)	RF, Maximum of Entropy, ANN, BRT	The findings demonstrated that the RF and MaxEnt algorithms outperformed other techniques. The study highlights the relevance of elements that condition gully development.
Kulimushi et al. (2022)	SVM, BRT, KNN, RF	As per the findings, RF-BRT was the most accurate model (87.26%).
Abolhasani et al. (2022)	RF, BRT, SVM, CART	The author identified that RF showed the best performance among the remaining models and altitudes, and rainfall was the most contributing factor to LD.
Yulianto et al. (2022)	SVM, CART, GTB, NB, and RF	This author used Geo-AI model to map Land Degradation with best RF model compared to other models.
Huang et al. (2023)	RF, SVM, ANN, GLM	The RF approach predicted GES with the highest degree of accuracy among the four machine-learning methods.
Yan et al. (2022)	RF, neural network model	High prediction precision obtained with RF model compared to neural network model.
Saha A et al. (2022)	CV and REPTree, Boosting	LD Maps produced using ensemble boosting (REPTree) showed best model for prediction analysis.
Feng et al. (2022)	CART-DT, RF, CNN	The RF model worked well with remote sensing photos.
Yu and Deng (2022)	RF, ESAI	The North China Plain's sensitive zones were located using ESAI and RF models. As per the research results, the risk of Land Degradation showed a decline in 2015, primarily attributed to changes in socio-economic conditions.
Saha S et al. (2022)	SVM, DT, ANN, NB, RF	Compared to the other models, SVM demonstrated higher accuracy. This research is the inaugural examination of the likelihood of deforestation using high-precision machine learning techniques.
Ngo et al. (2020)	RNN, CNN	This research identified that the RNN approach outperforms the CNN method during both training and testing stages.
Wang et al. (2021)	DBSCAN, RF, SoftMax	Using Kriging interpolation and DBSCAN, the author examined the spatial distributions of soil characteristics and deterioration. Over 95% accuracy in clustering was suggested by validation using RF and SoftMax.
Sahour et al. (2021)	BRT, DL, and MLR	The analysis revealed that BRT performed superiorly than other algorithms in determining annual soil erosion, and the best algorithm was chosen to ascertain the soil erosion's spatial dispersion.
Meng et al. (2021)	DT, SVM, NB, MD, ME	The author showed that the maximum entropy method may produce an accurate degree of desertification.
Habibi et al. (2021)	PLSR, ANN	A map of the distribution of groundwater was created using the most efficient ANN results.
Bakhtiari et al. (2021)	RF	The results showed that apparent thermal inertia (ATI) was the most important variable and that considerable changes in land cover led to soil deterioration.
Singh et al. (2021)	Mnlogit	To verify the consistency of classification, research was created in India utilizing LANDSAT pictures from 2005, 2006, 2007, and 2016. With a Kappa statistic of 0.71–0.81, the results demonstrated good accuracy of 80–86% throughout the course of the four years.
Ahmadpour et al. (2021)	SVM, RF	Findings revealed that the ensemble model performed the best.
Chen et al. (2021)	BRT, GLM, and DB	The integrity of the DB model is noticeably higher than that of the other models.
Haghighi et al. (2021)	SVM, MARS, GLM, DA	The best learning and prediction capabilities were found in DA, together with the highest accuracy and efficiency.

Paper	Methods used	Outcome
Wanget al. (2020)	PLSR, CNN, SVM and RF	The RF model was the most successful predictive modeling approach utilized to map soil salinity.
Yousefi et al. (2020)	RF, CART, and SVM	The author discovered that RF was the most reliable method.
Chakrabortty et al. (2020)	ANN, GWR–ANN	The ensemble GWR-ANN is more effective in assessing vulnerability to water-induced soil erosion.
Grinand et al. (2019)	RF	According to the author, Random Forest has a marginally superior capacity for prediction than both maximal entropy and the generalized linear model.
Gayen et al. (2019)	MARS, RF SVM.	According to the AUC values, RF is the best model compared to the MARS, SVM, and FDA models.
Dharumarajan et al. (2017)	RF	To forecast desertification processes, the author employed the Random Forest model. The findings revealed that DVI, potential evapotranspiration, and NDVI were the three most crucial predictors.
Rahmati et al. (2017)	SVM, ANN, RF, and BRT	Accurate predictions were generated by the RF, the RBF-SVM, the BRT, and the P-SVM models because of their strong fitting and predictive abilities.

Remarks:

SVM	Support Vector Machine	K-NN	K-Nearest Neighbor
RF	Random Forest	DA	Dragonfly Algorithm
DT	Decision Tree	PSR	Partial Least Squares Regression
ANN	Artificial Neural Network	GBM	Gradient-Boosting Machines
CNN	Convolutional Neural Network	DL	Deep Learning
NB	Naïve Bayes	CV	K-Fold cross-validation
MARS	Multivariate Additive Regression Splines	REP	Reduced Error Pruning Tree
GSM	Generalized Linear Models	MD	Minimum Distance
ME	Maximum of Entropy	DB	Deep Boost
REPT	Reduced Error Pruning Tree	GWR	Geographically Weighted Regression
MLR	Multiple Linear Regression	DVI	Desertification Vulnerability Index
RNN	Recurrent Neural Network	NPP	Net primary Productivity
GARI	Green Atmospherically Resistant Vegetation Index	DEM	Digital Elevation Model
ME	Maximum Entropy	DBSCAN	Density-Based Spatial Clustering of Applications with Noise

Selection Criteria

The search s conducted by focusing on the important ideas that are relevant to the review's criteria. Machine learning is successfully applied in a diverse array of domains, which means that a lot of published articles fall outside the context of this review article. The following steps are used to select the most relevant papers:

Step 1: The related studies are selected from different databases by using "land degradation" and "machine learning" as the search inputs.

Step 2: During the screening process, articles are excluded or included based on predefined criteria. This involves removing duplicate articles and those without full text. The screening primarily focuses on the title, abstract, keywords, year of publication, and conclusion. After applying the inclusion and exclusion criteria, a total of 30 papers were chosen for the investigation. All relevant data is gathered from the chosen studies, and the data are used to tackle the study goals.

Table 2 shows a total of 30 papers selected from Science Direct, Wiley, MDPI, and Springer databases after applying inclusion and exclusion criteria. The number of publications of the selected documents by year for this review is depicted in Figure 1. It shows the highest number of publications over the last ten years (upon the omission of unrelated articles) for this particular review, suggesting the importance of this field of research.

Table 2. Percentage of papers published per database.

Database	After the exclusion criteria, select a number of articles	Percentage of papers
Science Direct	19	63.34
MDPI	4	13.33
Springer Link	3	10
Wiley	2	6.67
Scopus	1	3.33
SciELO Colombia	1	3.33
Total	30	100

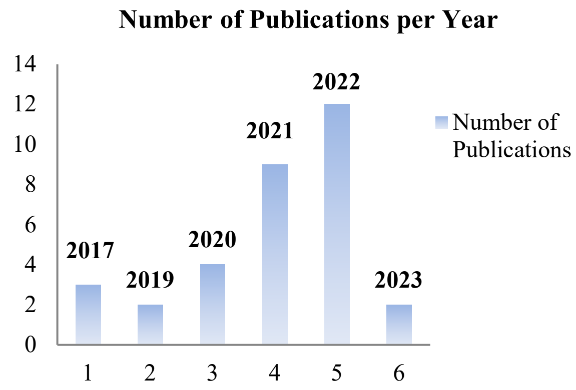

Figure 1. Number of publications per year.

Table 3 depicts that between 2013 and 2023, out of 30 selected papers, publications on the identification of land degradation using machine learning algorithms were conducted in different countries, such as eleven in Iran, seven in India, five in China, and two in Africa. One publication was done in Qinghai-Tibet, Mongolia, Germany, and the Mwenga countries. Lack of research in many countries means there needs to be continued research in the identification of land degradation so that policymakers can standardize the rules to reverse or combat it.

Table 3. Countries of selected papers

Country	Articles
Iran	11
India	7
China	5
Africa	2
Qinghai-Tibet	1
Mongolia	1
Madagascar	1
Germany	1
Mwenga	1

Current Status of Land Degradation in India

It is one of the major environmental issues that needs to be controlled because it affects farmers and forest dwellers the most. This leads to an increase in the change of climatic conditions, loss of biodiversity, food security, soil and water erosion, unsustainable land management and conservation, and the worst condition of human life. Most of the rain-fed farmland, which is responsible for food security in India, and forest land, which provides environmental benefits, have been degraded. As per the report provided by the Indian Space Research Organization (ISRO), almost 30% of the geographical area of India has already degraded over the past 15 years due to an increase in the loss of biodiversity, particularly in the northeastern states. Currently, 97.85 m ha of land has already been degraded between 2003-2005 and 2018-2019. With just 2.42% of the Earth's land area, India accommodates over 18% of the global population. According to Bhattacharya et al. (2013) there is a direct correlation between the fertility of land and the population it can support. On irrigated farms, there would always be crop cover, whereas in rain-fed areas, much erosion occurs because the topsoil remains exposed and is easily removed by rainfall. The issue was well known, but the solution was not as well discussed. In contrast to rain-fed farms, irrigated farms are better equipped to sustain resource-intensive, chemical-intensive, and energy-intensive cropping practices.

Deforestation, overgrazing, and/or other factors can cause vegetation to degrade, which is described as a decrease in biomass or a lack of ground cover vegetation, which contributes to soil erosion and the lack of organic soil matter. Deforestation is cutting the forested land in order to expand agriculture, obtain wood for fuel, construction, manufacturing purposes, and various developmental needs like commercial and industrial residential developments. In addition, excessive grazing and harvesting on short rotations are also causing the degradation. The process by which productive land turns into a desert or wasteland as a result of human improper activities and natural factors is referred to as desertification. Particularly in the states of Rajasthan, Gujarat, and Haryana, desertification is noticed in India. Rajasthan is largely surrounded by the Thar Desert, often called the Great Indian Desert, due to high temperatures, little rainfall, and wind erosion. Desertification is a problem in the Kutch region of Gujarat due to overgrazing, inadequate irrigation, and irresponsible land use practices that all contribute to the degradation of the soil's quality and decrease its vegetation cover. Erosion is especially noticeable in places with intense agriculture, like Punjab and Haryana, caused by inappropriate soil management practices, overuse of chemical fertilizers and pesticides, ineffective irrigation methods, and deteriorated soil quality.

To promote tree planting and stop the degradation of the land, programs like the National Afforestation Programme (NAP) and social forestry projects are being implemented. Utilize soil conservation methods, including contour flicking, terracing, and agroforestry, to decrease soil erosion and boost soil fertility. Promoting organic farming, efficient irrigation methods, crop rotation, and the use of organic fertilizers and bio-pesticides can minimize soil degradation and enhance agricultural sustainability. Public awareness initiatives, educational programs, and training efforts are undertaken to raise community awareness regarding the significance of implementing sustainable land management practices. Land degradation is a danger to agricultural productivity. It deteriorates soil health, which has an impact on the standard of living in rural communities. For instance, degraded soil is less able to absorb carbon dioxide (CO_2), the main greenhouse gas that intensifies global warming. The availability of surface and groundwater resources has decreased, and their quality has worsened. Under the best-case scenario of 1.5- °C warming, there will be 178 million more people who reside in dry regions by 2050 who are susceptible to water stress and drought intensity. Climate change is more challenging for communities and individuals who do not have adequate security measures in place.

Fundamental Aspects of Land Degradation Detection Process

The detection of land degradation using machine learning algorithms involves many stages, such as data collection in the first stage, data preprocessing in the second stage, and data partition and analysis in the next stage. Machine-learning-based regression and classification algorithms are used in data analysis.

Popular factors used in land degradation detection

The feature list is the most important component for the identification of land degradation. The investigation and analysis of features that directly or indirectly contribute to the degradation of the land are done in this study. According to the literature survey, the most important factors that lead to land degradation are climatic data, soil properties, human-induced factors, Geoenvironmental and topographic factors, and satellite data. Table 4 displays how frequently factors are applied to different geographical locations after the investigation of the selected papers. The most frequently used features are slope, land use, land cover, aspect, distance to river elevation, rainfall, soil type, temperature, soil salinity, urbanization, distance to roads, terrain, population, cloud cover, soil capacity, soil NPP, NDVI, and drainage density. Other factors also directly or indirectly lead to land degradation. The factor map that shows the significant features and sub-features is presented in Figure 2. Most of the factors have been categorized into five groups: soil information, climate data, human-induced data, geo-environmental data, and topographic data.

Table 4. Occurrence of factor usage.

Factors Used	# Times Used
Slope	12
Land Use/Land Cover	12
Aspect	10
Distance to Rivers	10
Elevation	9
Rainfall	9
Soil Type	8
Temperature	8
Soil Salinity	7
Urbanization	7
Distance to Roads	7
Terrain	7
Population	6
Pollution	6
Cloud Cover	6
Soil Capacity	5
Soil NPP	5
NDVI	5
Drainage Density	5
Soil Carbon Content	4
Altitude	4
Precipitation	3
Erosion	3
Overgrazing	2
Distance from Streams	2
Extreme Poverty	2
MODIS Images	2
Deforestation	1

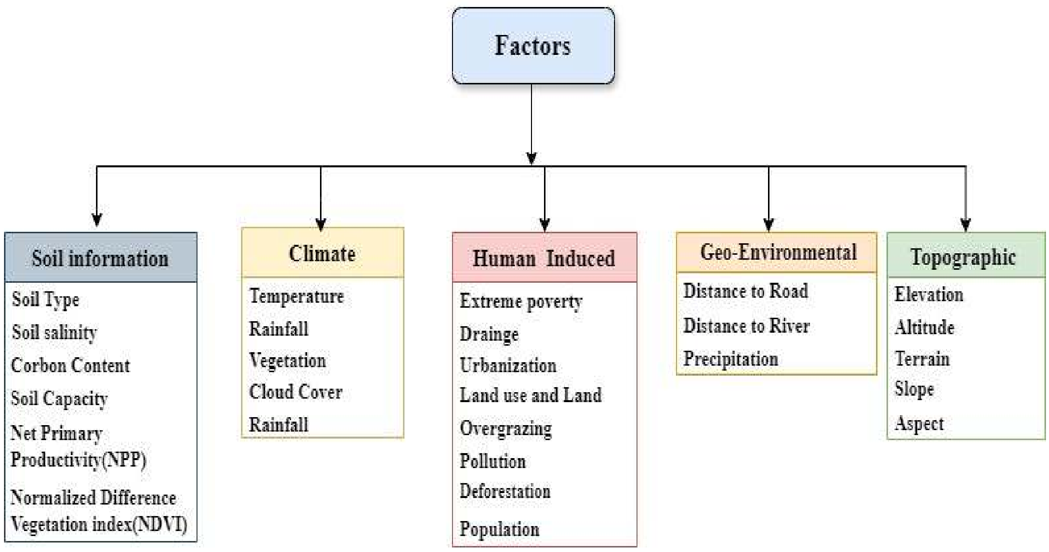

Figure 2. Factors diagram.

ML Algorithms

As per the literature survey shown in Table 5, various regression and classification-based ML algorithms have been used in the identification of land degradation in different geographical locations, such as Support Vector Machine, Variants of Boosting algorithms, Random Forest, Decision Tree, Classification and Regression Tree, Artificial Neural Network, Convolutional Neural Network, Naïve Bayes, and Multivariate Additive Regression Splines, which have been used at least twice. Other methods were used once in the previous work, as shown in Table 4. After conducting a comprehensive review of the existing literature for this study, which explored the effectiveness of various statistical and machine learning models in the domain of detection of land degradation, the SVM and variants of boosting methods emerged as the most suitable independent predictors for conducting assessments of land degradation in diverse geographical regions.

RF model

The Random Forest (RF) model is a supervised machine learning algorithm that serves both classification and regression tasks. It achieves this by building decision trees from a collection of independent and randomly sampled trees within the same distribution, as proposed by Breiman (2001). The model's performance relies on the quality of the individual trees and classifiers in the forest, as well as the relationships between them. Once a sufficient number of decision trees are generated, the model makes predictions by determining the majority vote for classification tasks and the average value for regression tasks. Due to the inherent randomness in constructing these trees, random forest models are considered reliable classifiers and regressors, according to Breiman (2001). The advantages of the Random Forest (RF) model are that it is a multivariate machine learning algorithm that doesn't make assumptions about statistical distributions. It exhibits strong resilience against overfitting when an adequate number of trees are grown and typically surpasses other classifiers in performance. The model's user-friendliness is enhanced by its reliance on fewer hyperparameters, such as the number of variables chosen randomly at each node, the number of trees, and tree depth within the forest. These hyperparameters are less sensitive to their values due to their random nature, as noted by Breiman (2001). Another benefit of the RFM model is its applicability to legacy data concerning desertification or land degradation processes. This model can be utilized to generate new maps or enhance existing mapping efforts Kidd et al. (2014).

SVM model

A widely employed machine learning technique known as Support Vector Machine (SVM) operates on the concept of minimizing risk, as introduced by Cortes and Vapnik (1995). This approach effectively divides the data into distinct classes using an optimal hyperplane, clearly showcasing the margin between these classes, as outlined by Abe. The training data points that are in close proximity to this hyperplane are referred to as support vectors, and the primary objective of this hyperplane is to accurately differentiate between various classes.

The effectiveness of SVM hinges on selecting appropriate kernel functions, such as the sigmoid kernel, Radial Basis Function (RBF), linear kernel, or polynomial kernel. Prior research (Choubin et al., 2018; Hong et al., 2018) suggests that the RBF kernel yields the most precise outcomes. Consequently, it was adopted in this study using R software ('e1071' package) (Meyer et al., 2019). The RBF kernel is widely employed in SVM classification across various kernelized learning algorithms. It is defined as follows (Vert et al., 2004; Cura, 2020).

$$K(x,x')=e^{-\gamma||x-x'||2}$$

where x and x′ represent two features for the RBF kernel, ||x−x′|| denotes the Euclidean distance between these features, and γ is a free parameter. The RBF kernel value diminishes as the distance increases, ranging between 0 and 1 when x=x′.

Decision Tree

The decision tree algorithm, as described by Yeon et al. (2010) is a commonly employed method characterized by the processes of tree expansion and tree refinement. It is a machine learning technique that partitions data into distinct subsets, with the tree's growth guided by the selection of attributes with the least entropy.

Classification and Regression Tree

CART is a data-centric machine learning technique applied in both classification and regression scenarios. Advantages are: easy to understand and interpret; identifying key features; handling nonlinear relationships and mixed data types; easy to apply; scalability; and data-driven decision-making. Nevertheless, it's crucial to recognize that CART models come with constraints, including their sensitivity to minor data fluctuations, the risk of overfitting, and challenges when dealing with complex relationships involving numerous variables.

Artificial Neural Network

An artificial neural network is a mathematical model inspired by the biological processes of neurons in the human brain. It was originally proposed by McCulloch and Pitts (1990). This model is capable of replicating complex, nonlinear relationships among variables. The most frequently employed type of artificial neural network is the Multi-Layer Perceptron (MLP), which comprises three distinct layers: the input layer, the hidden layer (which can consist of one or multiple layers), and the output layer. In artificial neural networks, the hyperbolic tangent, or sigmoid function, is commonly employed for mathematical convenience.

Convolutional Neural Network

CNNs, a type of deep learning model, have many benefits, such as being highly effective in processing and analyzing image data, excelling at identifying patterns and structures in the data, which is crucial for understanding the extent and nature of land degradation in different regions, recognizing different land cover types, soil erosion patterns, and vegetation changes, adapting to different spatial resolutions, and providing insights. Many open-source libraries and tools are available for developing and deploying CNN models.

Table 5. Used machine learning algorithms.

Applied ML algorithms	Frequency of usage
Support Vector Machine	14
Variants of Boosting Algorithms	14
Random Forest	13
Decision Tree	5
Classification and Regression Tree	5
Artificial Neural Network	4
Convolutional Neural Network	4
Naïve Bayes	2
Multivariate Additive Regression Splines	2
Generalized Linear Models	1
Maximum of Entropy	1
K-Fold cross-validation	1
Reduced Error Pruning Tree	1
Bagging	1
SoftMax	1
Multiple Linear Regression	1
Recurrent Neural Network	1
Geographically Weighted Regression	1
Maximum Entropy	1
Density-Based Spatial Clustering of Applications With Noise	1
K-Nearest Neighbor	1
Minimum Distance	1
Dragonfly Algorithm	1
Partial Least Squares Regression	1

Performance Evaluation Metrics

The evaluation metrics play a crucial role in determining the performance of a model, as they have the ability to distinguish between the results of various learning models (Elavarasan et al., 2020). As per the survey, several metrics, as shown in Table 6, are used to assess how well a regression and classification model performs. These include Area Under Curve, Area Under The (ROC) Curve, Overall Accuracy (OA) or Efficiency, Root Mean Square Error (RMSE), True Skill As Statics (TSS), Sensitivity (SST) or True Positive Rate, and Cohen's Kappa (K). The Area Under the Curve (AUC) is a frequently utilized metric in Receiver Operating Characteristic (ROC) analysis, commonly employed to evaluate the effectiveness of binary classification models. The ROC curve's area is a measure employed in the assessment of the model's capacity to differentiate between areas of land that are degraded and those that are not. Root Mean Square Error (RMSE) reflects the concentration of information on the optimal fit line and its utilization for estimating the standard deviation of residuals or forecasted errors.

Table 6. Number of times evaluation parameters were used.

Evaluation Parameter	# Of Times Used
Area Under Curve	11
Area Under The (ROC) Curve	11
Overall Accuracy (OA) or Efficiency	9
Root Mean Square Error (RMSE)	5
True Skill Statics (TSS)	4
Sensitivity (SST) or True Positive Rate	4
Cohen's Kappa (K)	4
Coefficient of Determination (R-Squared)	2
Normalized Root Mean Squared Error (NRMSE)	2
Specificity (SPF) or True Negative Rate	2
Specificity (SPF) or True Negative Rate	2
Mean Absolute Error (MAE)	1
Nash-Sutcliffe Efficiency (NSE)	1
Correlation Coefficient, Normalized Standard Deviation	1
Relative Error (RE)	1
NER Index and Softmax	1
Analytical Hierarchy Processes (AHP)	1
Topographic Threshold (TT)	1
Environmental Sensitivity Area Index (ESAI)	1

Critical Evaluation of Recent Studies Related to the Detection of Land Degradation

Soil Erosion Susceptibility

Soil erosion susceptibility assessment using ML algorithms has been investigated in the study (Chakrabortty et al., 2020; Wang et al., 2020; Sahour al., 2021; Wang et al., 2021; Bag et al., 2022; Kulimushi et al., 2022: Das et al., 2023). Das et al. (2023) proposed a soil salinity index using soil physical and chemical properties and detected salt-affected areas using ensemble methods of the RF and DL models with the best performance. Bag et al. (2022) showed that elevation, drainage density (DD), and NDVI factors contribute the most to soil erosion, and the RF model performed and predicted best compared to SVM, CART, and BRT. Kulimushi et al. (2022) used ensemble machine learning algorithms (ML-ALs) to evaluate the predictive power of combining algorithms like SVM, BRT, LB, and KNN with Random Forest (RF) as the base classifier for erosion susceptibility mapping (ESM) in the Elila catchment. Results showed that RF-BRT (87.26%) was the most accurate, followed by RF-SVM, RF-KNN, and RF-LB. Wang et al. (2021) analyzed spatial distributions of soil properties and degradation using Kriging interpolation and density-based spatial clustering of applications with noise (DBSCAN). Validation was done using random forest and SoftMax suggested that the accuracy of clustering was over 95%. Sahour al. (2021) showed that BRT outperformed DL and MLR in assessing annual soil erosion, and the optimal algorithm was selected for estimating soil erosion spatial distribution. Wang et al. (2020), nine predictive modeling approaches were used to model and estimate soil salinity, with the RF model being the most effective. Environmental factors and soil salinity indices, such as DEM, B10, and GARI, contributed the most to soil salinity estimation. Chakrabortty et al. (2020) proved the ensemble GWR-ANN is more optimal for determining water-induced soil erosion susceptibility.

Gully Erosion Susceptibility Assessment (GES)

ML-based gully erosion susceptibility assessment has been investigated in studies (Rahmati et al., 2017; Ahmadpour et al., 2021; Chenet al., 2021; Chumaet al., 2023; Huanget al., 2023; Setargie et al., 2023). Setargie et al. (2023) integrated detailed field investigations with high-resolution remote sensing products to assess gully erosion susceptibility and identify its controlling factors using the Random Forest (RF) model in six representative watersheds in the Upper Blue Nile basin of Ethiopia. Chuma et al. (2023) showed that RF and MaxEnt algorithms outperformed other methods, with higher prediction accuracies than BRT and ANN. The study highlights the importance of considering conditioning factors in gully occurrence, urging policymakers to adopt strategies that consider these factors to reduce gully risk and consequences. Huang et al. (2023) calibrated and validated four machine learning methods, such as RF, SVM, ANN, and GLM, with the RF method predicting GES with the greatest accuracy. Ahmadpour et al. (2021) showed that the ensemble model (SVM and RF) had the best performance (AUC = 0.982, TSS = 0.93) compared to the others. Chen et al. (2021) showed that DB machine learning methods have significantly higher accuracy than other approaches such as GLM and BRT. Rahmati et al. (2017) showed RF, RBF-SVM, BRT, and P-SVM models performed well in fitting and predictive performance, resulting in accurate predictions.

Mapping of Land Degradation

Abolhasani et al. (2022) used 15 geo-environmental factors, such as altitude, slope, land use, and temperature, that were considered LD predictive variables. The results showed that altitude was the most influential variable within RF, BRT, and SVM, while rainfall was the most important contribution in modeling based on the CART algorithm. Yulianto et al. (2022) evaluated the overall accuracy of the results of comparison and evaluation of machine learning-based predictions on the RF, CART, GTB, SVM, NB, and MD in the study area at 86.2%, 85.8%, 81.2%, 52.8%, 36.3%, and 34.5%, respectively. Therefore, the study concluded that the RF, CART, and GTB algorithms are proposed to be applied to produce land degradation maps in the study area. Saha et al. (2022) prepared susceptibility maps for undulating red and lateritic agro-climatic zones using hybrid techniques. The results show that Boosting-REPTree is the most optimal model for prediction analysis, with an AUC of 0.944 and 0.928. The ensemble of Boosting-REPTree can be applied as a new method for spatial land degradation prediction in future research. Yu and Deng (2022) used the Environmental Sensitivity Area Index (ESAI) and a random forest model to identify sensitive areas in the North China Plain. Results showed a decrease in land degradation risk in 2015, with socio-economic factors having the most significant impact. Habibi et al. (2021) observed that ANN has the highest efficiency, which agrees with other findings. The results of ANN have been used in the preparation of groundwater distribution maps. Using ANN, it is predicted that 100% of the area will be severely degraded by 2025. Haghighi et al. (2021) demonstrated that the primary factors influencing the landscape, environmental conditions, and human activities were incline, geological characteristics, and alterations in land use. DA had the highest accuracy and efficiency, with the greatest learning and prediction power. Yousefi et al. (2020) found that RF was most robust based on assessments of the trained and validated models. Grinand et al. (2019) showed Random Forest slightly better prediction ability than maximum entropy and a generalized linear model.

Challenges and Future Directions to the detection of Land Degradation

Lack of Optimal set of features

This review categorizes feature lists into 5 groups for easy understanding of the detection of land degradation. The most prevalent elements are slope, land use/cover, aspect, distance from rivers, elevation, rainfall, soil type, temperature, soil salinity, urbanization, distance from roads, topography, population, pollution, cloud cover, soil capacity, soil

NPP, and NDVI. Other characteristics are also employed, although only to a limited extent. Identifying the optimal sub-feature list for each set of features can pose a challenge, as further investigation is required. This is because previous studies lack a clearly defined optimal sub-feature set for each group. Explore optimum sub-features for different regions and record land degradation over several years to account for variations in factors like human activities, climatic factors, chemical, and physical variables. Identify vital features and remove redundant ones to improve predictive models' performance. Predicting land degradation in various regions poses a significant challenge due to the differences in their characteristics. The development of a universally accurate land degradation prediction model is difficult, as models tailored to one area may not apply to another. One key factor contributing to this challenge is the variation in predictive features across different sites. We suggest that identifying features that are consistently sensitive to land degradation across a diverse range of areas can enhance the accuracy of prediction models across various regions. Greater gaps were visible in coarser datasets compared to finer datasets, demonstrating that over-fitting did not affect the optimal goodness-of-fit on datasets with the highest resolution. Therefore, when multiple data repositories are available, consider taking accurate resolution settings into account in gully erosion susceptibility investigations.

Finding the optimal ML Algorithm

Various ML algorithms have been employed in earlier studies to detect land degradation. Although SVM is an extensively used prediction method, it has problems with large samples resembling training time, noise, underperformance, feature overlaps, and kernel selection difficulties. A well-liked ML technique called RF combines the predictive power of several decision trees. The advantages of random forests are high predictive accuracy, robustness to outliers and noisy data, handling large datasets, feature importance, and parallelization. The disadvantages of Random Forest are interpretability, computational expense, memory consumption, bias towards the majority class, and the fact that training a RF model can be slower relative to other algorithms. Derived from the specifications and traits of the issue, it is important to consider the pros and cons when deciding to select a machine learning algorithm. When the same types of land degradation detection models are judged, it is practically impossible to compare them when utilizing different performance metrics. Therefore, it should be advised that the land degradation detection model be quantified using a consistent, methodical technique or a single metric. A direct comparison between several land degradation detection models is achievable if the same performance metrics are utilized for each model. Future research should explore the latest geospatial data, use satellite imagery or other available data, and increase the scale of detailed research. Using medium- or high-resolution data, such as Sentinel-2, Landsat-8, and Landsat-9, can improve the accuracy of the model. Additionally, future research should use rank variables to test the importance of each variable in machine learning processing. While numerous machine learning studies have been conducted for the detection of land degradation, only a handful of them are directly relevant to the detection of land degradation. From these studies, it remains challenging to identify the optimal algorithm for the detection of land degradation. Nevertheless, certain machine learning regression algorithms, such as SVM, Variants of Boosting Algorithms, RF, DT, CART, ANN, and CNN, exhibit significant promise in the detection of land degradation. Furthermore, rather than relying on a single algorithm, it is advisable to explore an ensemble of multiple algorithms to enhance the robustness of the prediction model.

Application of remote sensing in the detection of land degradation

Remote sensing is a powerful technology that involves the collection of information about Earth's surface, typically using satellites, aircraft, drones, or ground-based sensors. It has a wide range of applications across various fields, such as monitoring climate-related changes, tracking deforestation and changes in land use, predicting natural disasters like hurricanes, wildfires, and floods, precision agriculture, crop health and yield prediction, monitoring urban growth, assessing infrastructure needs, planning for sustainable development, tracking changes in habitats and ecosystems, and aiding in wildlife conservation efforts. Wang et al. (2023) investigation indicates that both direct indicators (such as mineral composition, organic matter, surface roughness, and soil moisture content) and indirect measures (including vegetation health and changes in land use and land cover) prove to be effective in assessing soil degradation through RS methods.

The ability to monitor and assess different aspects of soil degradation can enhance our comprehension of the initiation and progression mechanisms. Additionally, this approach provides a promising perspective for refining our conceptual understanding and modeling of the underlying processes. We are now entering a new era of RS, characterized by the availability of vast amounts of data in terms of space, time, and spectral information. Furthermore, RS methods, whether based on the ground, drones, or satellites, offer unique insights and supportive evidence.

Prospective architecture of desertification process

In this study, a systematic review has been done for the identification of land degradation and suggested a future trend for the identification of land degradation framework, as shown in Figure 3. A wide range of data is gathered in the first step in order to detect land degradation, including information on soil properties, climate, vegetation indices, human induced, environmental and topographic data. Data must be pre-processed after collection in order to be used for additional analysis. The complete dataset is divided into a training set and a testing set after the data has undergone pre-processing. The prediction model is trained using the training dataset. In the model's training phase, various machine learning-based regression and classification techniques are subsequently utilized. The model's parameter is optimized if the trained model's performance is not satisfactory. The trained model is put to the test using the testing dataset after it reaches the threshold performance. If the performance of the model is good then recommend the model to map the desertification factors and identify the controlling factors.

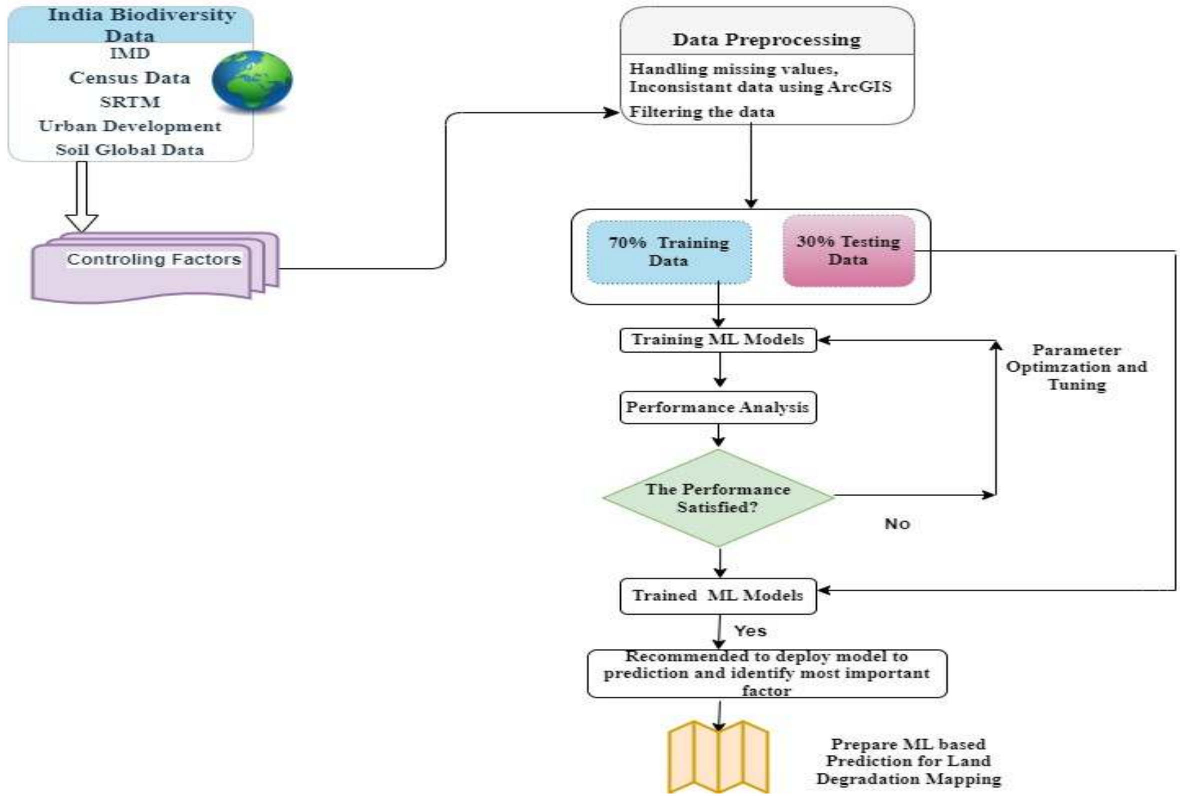

Figure 3. Prospective architecture for the detection of land degradation using ML.

Conclusion

In order to reverse land degradation, new technology needs to be implemented. Apart from this, policymakers need proper guidelines in time to allow them to forecast land degradation so that they can make strict rules and regulations to reduce land degradation. ML frameworks can provide clear insights by assessing huge amounts of data and interpreting the obtained information. ML models also establish the relationship between target variables and predictor variables. The current review demonstrates that the selected articles used a diverse set of factors, with a primary emphasis on data accessibility and research coverage. All the referenced articles detected land degradation using machine learning algorithms; the difference lies in the utilization of a wide array of features, methods, and different geographical locations. The selection of features depends on the availability of data and research objectives. The existing literature shows that a wide range of features may not give optimum results at all times in the detection of land degradation. Even though it is difficult to distinguish optimal ML models from the existing models, it's important to find the most frequently used ML models and their performance to get an overview. The most commonly used ML models are SVM, Boosti8/9ng algorithms, and RF. In addition to these algorithms, some deep learning models are also used, such as ANN, CNN, and RNN, in the detection of land degradation. To come to a specific conclusion about the best-performing model, it should be investigated with an existing, outperforming model. This paper is likely to lay the foundation for extensive research on the detection of land degradation using machine learning algorithms.

Acknowledgments

This article was written as part of the thesis research required to complete studies at BIET Institute, affiliated with VTU. The authors thank the reviewers for their valuable comments and suggestions, which helped improve the quality of the article.

References

Abolhasani, A., Zehtabian, G., Khosravi, H., Rahmati, O., Alamdarloo, E.H. and D'Odorico, P. 2022. A new conceptual framework for spatial predictive modeling of land degradation in a semiarid area. *Land Degradation and Development* 33(17):3358-3374, doi:10.1002/ldr.4391.

Ahmadpour, H., Bazrafshan, O., Rafiei-Sardooi, E., Zamani, H. and Panagopoulos T. 2021. Gully erosion susceptibility assessment in the Kondoran Watershed using machine learning algorithms and the Boruta feature selection. *Sustainability* 13(18):10110, doi:10.3390/su131810110.

Bag, R., Mondal, I., Dehbozorgi, M., Bank, S.P., Das, D.N., Bandyopadhyay, J., Pham, Q.B., Al-Quraishi, A.M.F., Nguyen, X.C. 2022. Modelling and mapping of soil erosion susceptibility using machine learning in a tropical hot sub-humid environment. *Journal of Cleaner Production* 36, doi:10.1016/j.jclepro.2022.132428.

Bakhtiari, M., Boloorani, A.D., Kakroodi, A.A., Rangzan, K. and Mousivand, A. 2021. Land degradation modeling of dust storm sources using MODIS and meteorological time series data. *Journal of Arid Environments* 190:104507, doi:10.1016/j.jaridenv.2021.104507.

Bhattacharya, M. 2013. Machine learning for bioclimatic modeling. *International Journal of Advanced Computer Science and Applications* 4(2):1-8.

Breiman, L. 2001. Random forests. *Machine Learning* 45:5-32.

Chakrabortty, R., Pal, S.C., Sahana, M., Mondal, A., Dou, J., Pham, B.T. and Yunus, A.P. 2020. Soil erosion potential hotspot zone identification using machine learning and statistical approaches in eastern India. *Natural Hazards* 104, 10.1007/s11069-020-04213-3.

Choubin, B., Darabi, H., Rahmati, O., Sajedi-Hosseini, F. and Kløve, B. 2018. River suspended sediment modelling using the CART model: a comparative study of machine learning techniques. *Science of the Total Environment* 615:272-281, doi:10.1016/j.scitotenv.2017.09.293.

Chuma, G.B., Mugumaarhahama, Y., Mond, J.M., Bagula, E.M., Ndeko, A.B., Lucungu, P.B., Karume, K., Mushagalusa, G.N. and Schmitz, S. 2023. Gully erosion susceptibility mapping using four machine learning methods in Luzinzi watershed, eastern Democratic Republic of Congo. *Physics and Chemistry of the Earth, Parts A/B/C* 29:103295, doi:10.1016/j.pce.2022.103295.

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

Cura, T. 2020. Use of support vector machines with a parallel local search algorithm for data classification and feature selection. *Expert Systems with Applications* 145:113133, doi:10.1016/j.eswa.2019.113133.

Das, A., Bhattacharya, B.K., Setia, R., Jayasree, G. and Das, B.S. 2023. A novel method for detecting soil salinity using AVIRIS-NG imaging spectroscopy and ensemble machine learning. *ISPRS Journal of Photogrammetry and Remote Sensing* 200:191-212, doi:10.1016/j.isprsjprs.2023.04.018.

Dasgupta, A., Sastry, K.L.N., Dhinwa, P.S., Rathore, V.S., and Nathawat, M.S. 2013. Identifying desertification risk areas using fuzzy membership and geospatial technique– A case study, Kota District, Rajasthan. *Journal of Earth System Science* 122:1107-1124.

Devasena, C.L. 2014. Comparative analysis of random forest, REP tree, and J48 classifiers for credit risk prediction. *International Journal of Computer Applications (0975-8887) International Conference on Communication, Computing and Information Technology (ICCCMIT-2014)* 30-36.

Dharumarajan, S., Bishop, T.F.A., Hegde, R. and Singh, S. 2017. Desertification vulnerability index - an effective approach to assess desertification processes: a case study in Anantapur District, Andhra Pradesh, India. *Land Degradation and Development* 29, doi:10.1002/ldr.2850.

Dutta, S. and Chaudhuri, G. 2015. Evaluating environmental sensitivity of arid and semiarid regions in northeastern Rajasthan, India. *Geographical Review* 105:441-461, doi:10.1111/j.1931-0846.2015. 12093.x.

Elavarasan, D., Vincent, P.M.D.R., Srinivasan, K. and Chang, C.Y. 2020. A hybrid CFS Filter and RF-RFE wrapper-based feature extraction for enhanced agricultural crop yield prediction modeling. *Agriculture* 10:400, doi:10.3390/agriculture10090400.

Feng, K., Wang, T., Liu, S., Kang, W., Chen, X., Guo, Z. and Zhi, Y. 2022. Monitoring desertification using machine-learning techniques with multiple indicators derived from MODIS images in Mu Us sandy land China. *Remote Sensing* 14:2663, doi:10.3390/rs14112663.

Gayen, A., Pourghasemi, H.R., Saha, S., Keesstra, S. and Bai, S. 2019. Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. *Science of the Total Environment* 668:124-138, doi:10.1016/j.scitotenv.2019.02.436.

Giordano, L., Giordano, F., Grauso, S., Iannetta, M., Sciortino, M., Rossi, L. and Bonati, G. 2003. Identification of areas sensitive to desertification in Sicily Region ENEA (eds). Italy: Rome.

Grinand, C., Vieilledent, G., Razafimbelo, T., Rakotoarijaona, J.R., Nourtier, M. and Bernoux, M. 2019. Landscape-scale spatial modelling of deforestation, land degradation and regeneration using machine learning tools. *Land Degradation and Development* 31, doi:10.1002/ldr.3526.

Habibi, V., Ahmadi, H., Jaffari, M. and Moeini, A. 2021. Prediction of Land Degradation by Machine Learning Methods: A Case Study from Sharifabad Watershed, Central Iran. *Earth Sciences Research Journal* 25(3):353-362, doi:10.15446/esrj.v25n3.75821.

Haghighi, A.T., and Darabi, H., Karimidastenaei, Z., Davudirad, A., Rouzbeh, S., Rahmati, O., Sajedi-Hosseini, F. and Klöve, B. 2021. Land degradation risk mapping using topographic, human-induced, and geo-environmental variables and machine learning algorithms, for the Pole-Doab watershed, Iran. *Environmental Earth Sciences* 80, doi:10.1007/s12665-020-09327-2.

Hong, H., Panahi, M., Shirzadi, A., Ma, T., Liu, J., Zhu, A.X. and Kazakis, N. 2018. Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Science of the Total Environment* 621:1124-1141, doi:10.1016/j.scitotenv.2017.10.114.

Huang, D., Su, L., Zhou, L., Tian, Y. and Fan. H. 2023. Assessment of gully erosion susceptibility using different DEM-derived topographic factors in the black soil region of Northeast China. *International Soil and Water Conservation Research* 11(1):97-111, doi:10.1016/j.iswcr.2022.04.001.

Huang, S., Cai, N., Pacheco, P., Narrandes, S., Wang, Y. and Xu, W. 2018. Applications of support vector machine

(SVM) learning in cancer genomics. *Cancer Genomics and Proteomics* 15(1):41-51, doi:10.21873/cgp.20063.

Jafari, R. and Bakhshandehmehr, L. 2013. Quantitative mapping and assessment of environmentally sensitive areas to desertification in central Iran. *Land Degradation and Development,* doi:10.1002/ ldr.2227.

Karamesouti, M., Detsis, V., Kounalaki, A., Vasiliou, P., Salvati, L. and Kosmas, C. 2015. Land-use and land degradation processes affecting soil resources: Evidence from a traditional Mediterranean cropland (Greece). *Catena* 132:45-55, doi:10.1016/j.catena.2015.04.010.

Kempen, B., Brus, D.J., Heuvelink, G.B.M. and Stoorvogel, J.J. 2009. Updating the 1:50,000 Dutch oil map using legacy soil data: A multinomial logistic regression approach. *Geoderma* 151:311-326, doi:10.1016/j.geoderma.2009.04.023.

Keshtkar, H., Voigt, W. and Esmaeill, A. 2017. Land-cover classification and analysis of change using machine-learning classifiers and multi-temporal remote sensing imagery. *Arabian Journal of Geosciences* 10, doi:10.1007/s12517-017-2899-y.

Kidd, D.B., Malone, B.P., McBratney, A.B., Minasny, B. and Webb, M.A. 2014. Digital mapping of a soil drainage index for irrigated enterprise suitability in Tasmania, Australia. *Soil Research* 52:107-119, doi:10.1071/SR13100.

Kim, J., Grunwald, S., Rivero, R.G. and Robbins, R. 2012. Multi-scale modeling of soil series using remote sensing in a wetland ecosystem. *Soil Science Society of America Journal* 76:2327, doi:10.2136/ sssaj2012.0043.

Kosmas, C., Ferrara, A., Briasouli, H. and Imeson, A. 1999. Methodology for mapping Environmentally Sensitive Areas (ESAs) to Desertification. In: Kosmas, C., Kirkby, M. and Geeson, N. (Eds.), *The Medalus Project: Mediterranean Desertification and Land Use Manual on Key Indicators of Desertification and Mapping Environmentally Sensitive Areas to Desertification* (pp. 31-47). Brussels: European Union 18882. ISBN: 92-828-6349-2.

Kovačevic, M., Bajat, B. and Gajić, B. 2010. Soil type classification and estimation of soil properties using support vector machines. *Geoderma* 154:340-347, doi:10.1016/j.geoderma.2009.11.005.

Kulimushi, L.C., Bashagaluke, J.B., Prasad, P., Heri-Kazi, A.B., Kushwah, N.L., Masroor, Md., Choudhary, P., Elbeltagi, A., Sajjad, H. and Mohammed, S. 2022. Soil erosion susceptibility mapping using ensemble machine learning models: A case study of upper Congo River sub-basin. *Catena* 222:106858, doi:10.1016/j.catena.2022.106858 .

Lawrence, R.L., Bunn, A., Powell, S. and Zambon, M. 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing and Environment* 90:331-336, doi:10.1016/j.rse.2004.01.007.

Mansuy, N., Thiffault, E., Paré, D., Bernier, P., Guindon, V. P., Poirier, V. and Beaudoin, A. 2014. Digital mapping of soil properties in Canadian managed forests at 250 m of resolution using the k-nearest neighbor method. *Geoderma* 35-236:59-73, doi:10.1016/j.geoderma.2014.06.032.

McCulloch, W.S. and Pitts, W. 1990. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology* 52(1-2):99-115, doi:10.1016/S0092-8240(05)80006-0.

Meng, X., Gao, X., Li, S., Li, S. and Lei, J. 2021. Monitoring desertification in Mongolia based on Landsat images and Google Earth Engine from 1990 to 2020. *Ecological Indicators* 129, doi:10.1016/j.ecolind.2021.107908.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.C., Lin, C.C. and Meyer, M.D. 2019. Package 'e1071'. The R Journal.

Ngo, P.T.T., Panahi, M., Khosravi, K., Ghorbanzadeh, O., Karimineja, N., Cerda, A. and Lee, S. 2020. Evaluation of deep learning algorithms for national-scale landslide susceptibility mapping in Iran. *Geoscience Frontiers* 12(2):505-519, doi:10.1016/j.gsf.2020.06.013.

Priori, S., Bianconi, N. and Constantini, E.A.C. 2014. Can γ-radiometrics predict soil textural data and stoniness in different parent materials? A comparison of two machine learning methods. *Geoderma* 226-227:354-364, doi:10.1016/j.geoderma.2014.03.012.

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

Saha, A., Pal, S.C., Chowdhuri, I., Islam, A.R.M.T., Roy, P. and Chakraborty, R. 2022. Land Degradation risk dynamics assessment in red and lateritic zones of eastern plateau, India: A combined approach of K-fold CV, data mining, and field validation. *Ecological Informatics* 69:10156, doi:10.1016/j.ecoinf.2022.101653.

Saha, S., Bhattacharjee, S., Shit, P.K., Sengupta, N. and Bera, B. 2022. Deforestation probability assessment using integrated machine learning algorithms of Eastern Himalayan foothills (India). *Resources Conservation & Recycling Advances* 14(2):200077, doi:10.1016/j.rcradv.2022.200077.

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

Salvati, L., Zitti, M., Ceccarelli, T. and Perini, L. 2009. Developing a synthetic index of land vulnerability to drought and desertification. *Geographical Research* 47:280-291, doi:10.1111/j.1745- 5871.2009.00590.x

Setargie, T.A., Tsunekawa, A., Haregeweyn, N., Tsubo, M., Fenta, A.A., Berihun, M.L., Sultan, D., Yibeltal, M., Ebabu, K., Nzioki, B. and Meshesha, T.M. 2023. Random forest-based gully erosion susceptibility assessment across different agro-ecologies of the Upper Blue Nile basin, Ethiopia. *Geomorphology* 431:108671, doi:10.1016/j.geomorph.2023.108671.

Singh, R.K., Singh, P., Drews, M., Kumar, P., Singh, H., Gupta, A.K., Govil, H., Kaur, A. and Kumar, M. 2021. A machine learning-based classification of LANDSAT images to map land use and land cover of India. *Remote Sensing Applications: Society and Environment*. 24:100624, doi:10.1016/j.rsase.2021.100624.

Symeonakis, E., Karathanasis, N., Koukoulas, S. and Panagopoulos, G. 2016. Monitoring sensitivity to land degradation and desertification with the environmentally sensitive area index: The case of Lesvos Island. *Land Degradation and Development* 27:1562-1573, doi:10.1002/ ldr.2285.

Trott, C., Menegoni, P., Frattarelli, F.H.M. and Iannetta, M. 2015. Assessing desertification vulnerability on a local scale: The Castelporziano study case (central Italy). *Rendiconti Lincei* 26:421, doi:10.1007/s12210-014-0362-5.

Vågen, T.G., Winowiecki, L.A., Tondoh, J.E., Desta, L.T. and Gumbricht, T. 2016. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. *Geoderma* 263:216-225, doi:10.1016/j.geoderma.2015.06.023.

Vert, J.P., Tsuda, K. and Schölkopf, B. 2004. A primer on Kernel Methods. In: *Kernel Methods to Computational Biology* 47:3570, doi:10.7551/mitpress/4057.003.0004.

Wang, J., Zhen, J., Hu, W., Chen, S., Lizaga, I., Zeraatpisheh, M. and Yang, X. 2023. Remote sensing of soil degradation: progress and perspective. *International Soil and Water Conservation Research* 11(3):429-454, doi:10.1016/j.iswcr.2023.03.002.

Wang, N., Xue, J., Peng, J., Biswas, A., He, Y. and Shi, Z. 2020. Integrating remote sensing and landscape characteristics to estimate soil salinity using machine learning methods: A case study from Southern Xinjiang, China. *Remote Sensing* 24, doi:10.3390/rs12244118.

Wang, Z., Wang, G., Ren. T., Wang, H., Xu, W. and Zhang, G. 2021. Assessment of soil fertility degradation affected by mining disturbance and land use in a coalfield via machine learning. *Ecological Indicators* 125:107608, doi:10.1016/j.ecolind.2021.107608.

Wei, C., Lei, X., Chakrabortty, R., Pal, S.C., Sahana, M. and Janizadeh, S. 2021. Evaluation of different boosting ensemble machine learning models and novel deep learning and boosting framework for head-cut gully erosion susceptibility. *Journal of Environmental Management* 284:112015, doi:10.1016/j.jenvman.2021.112015.

Yan, H., Ran, Q., Hu, R., Xue, K., Zhang, B., Zhou, S., Zhang, Z., Tang, L., Che, R., Pang, Z., Wang, F., Wang, D., Zhang, J., Jiang, L., Zhang, S., Qian, Z., Guo, T., Du, J., Hao, Y., Cui, X. and Wang, Y. 2022. Machine learning-based prediction for grassland degradation using geographic, meteorological, plant, and microbial data. *Ecological Indicators* 137:108738, doi:10.1016/j.ecolind.2022.108738.

Yeon, Y.K., Han, J.G. and Ryu, K.H. 2010. Landslide susceptibility mapping in Injae, Korea, using a decision tree. *Engineering Geology* 116(3-4):274-283, doi:10.1016/j.enggeo.2010.09.009.

Yousefi, S., Pourghasemi, H., Avand, M., Janizadeh, S., Tavangar, S. and Santosh, M. 2020. Assessment of land degradation using machine-learning techniques: A case of declining rangelands. *Land Degradation and Development* 32, doi:10.1002/ldr.3794.

Yu, Z. and Deng, X. 2022. Land Degradation in the North China Plain, driven by food security goals. *Ecological Engineering* 183:106766, doi:10.1016/j.ecoleng.2022.106766.

Yulianto, F., Raharjo, P.D., Setiawan, M.A., Sakti, A.D., Nugroho, S. and Budhiman S. 2022. Machine learning-based prediction for land degradation mapping using multi-source geospatial data in the Batanghari watershed, Sumatra, Indonesia. *Research Square*, doi:10.21203/rs.3.rs-2177125/v1.

Appendix A

Table A1. A summary of the features employed in the respective publications. The symbol "✔" in the box signifies the utilization of a specific feature

Paper	Elevation	soil salinity	soil capacity	soil type	soil carbon content	soil NPP	Drainage	Urbanization	Population	Temperature	overgrazing	pollution	altitude	deforestation	slope	aspect	rainfall	distance to roads	distance to rivers	precipitation	NDVI	distance from	drainage density	Cloud cover.	extremely	land use/land	erosion	terrain	MODIS images
Setargie et al. (2023)	✓			✓															✓			✓	✓			✓			
Das et al. (2023)		✓	✓	✓	✓	✓	✓																✓						
Bag et al. (2022)	✓						✓														✓								
Chuma et al. (2023)												✓			✓			✓	✓										
Kulimushi et al. (2022)									✓	✓				✓			✓							✓		✓			
Abolhasani et al. (2022)													✓		✓		✓												
Yulianto et al. (2022)	✓									✓					✓	✓				✓						✓			
Huang et al. (2023)	✓															✓													
Yan et al. (2022)		✓	✓	✓	✓	✓																				✓			
Saha A et al. (2022)		✓						✓	✓		✓	✓			✓	✓									✓		✓	✓	
Feng et al. (2022)																													✓
Yu and Deng (2022)									✓	✓							✓				✓		✓		✓	✓			
Saha et al. (2022)		✓	✓	✓	✓	✓		✓		✓		✓					✓	✓	✓										
Ngo et al. (2020)												✓			✓	✓	✓	✓	✓							✓			
Wang et al. (2021)		✓	✓	✓		✓																							
Sahour et al. (2021)			✓												✓						✓								
Meng et al. (2021)										✓		✓					✓							✓					
Habibi et al. (2021)	✓												✓		✓	✓	✓		✓									✓	
Bakhtiari et al. (2021)																													✓
Singh et al. (2021)																										✓			
Ahmadpour et al. (2021)	✓			✓			✓								✓	✓			✓							✓		✓	
Chen et al. (2021)										✓										✓				✓					
Haghighi et al. (2021)	✓							✓	✓						✓	✓		✓	✓		✓			✓		✓	✓	✓	
Wang et al. (2020)		✓																			✓							✓	
Yousefi et al. (2020)								✓				✓						✓											

Paper	Elevation	soil salinity	soil capacity	soil type	soil carbon content	soil NPP	Drainage	Urbanization	Population	Temperature	overgrazing	pollution	altitude	deforestation	slope	aspect	rainfall	distance to roads	distance to rivers	precipitation	NDVI	distance from	drainage density	Cloud cover.	extremely	land use/land	erosion	terrain	MODIS images
Chakrabortty et al. (2020																	✓												
Grinand et al. (2019)	✓							✓							✓	✓		✓	✓			✓							
Gayen et al. (2019)				✓									✓		✓	✓		✓	✓				✓					✓	
Dharumarajan et al. (2017)	✓	✓			✓	✓	✓	✓	✓	✓	✓		✓		✓	✓	✓						✓	✓		✓		✓	
Rahmati et al. (2017)				✓															✓	✓							✓		
Keshtkar et al. (2017)																										✓			

Table A2. A summary of the evaluation parameters employed in the respective publications. The symbol "✔" in the box signifies the utilization of evaluation parameters

Paper	RMSE	MAE	R-squared	NSE	NRMSE	NSD	RE	NER	AHP	TT	TSS	AUC	AROC	SST	SPF	OA	K	ESAI	RMSE	MAE	R^2	NSE	NRMSE	NSD	NER	AHP	TT	TSS	AUC	AROC	SST	SPF	OA	K	ESAI
Setargie et al. (2023)														✓	✓	✓	✓														✓	✓	✓	✓	
Das et al. (2023)	✓				✓														✓				✓												
Bag et al. (2022)									✓	✓																✓	✓								
Chuma et al. (2023)											✓		✓															✓		✓					
Kulimushi et al. (2022)													✓	✓	✓	✓	✓													✓	✓	✓	✓	✓	
Abolhasani et al. (2022)											✓	✓	✓															✓	✓	✓					
Yulianto et al. (2022)																✓																	✓		

Paper	RMSE	MAE	R-squared	NSE	NRMSE	NSD	RE	NER	AHP	TT	TSS	AUC	AROC	SST	SPF	OA	K	ESAI	RMSE	MAE	R^2	NSE	NRMSE	NSD	NER	AHP	TT	TSS	AUC	AROC	SST	SPF	OA	K	ESAI
Huang et al. (2023)	✓	✓											✓			✓			✓	✓										✓			✓		
Yan et al. (2022)							✓																		✓										
Saha A et al. (2022)												✓																	✓						
Feng et al. (2022)																✓																	✓		
Yu and Deng (2022)														✓				✓													✓				✓
Saha et al. (2022)												✓																	✓						
Ngo et al. (2020)												✓	✓																✓	✓					
Wang et al. (2021)								✓																	✓										
Sahour et al. (2021)			✓	✓	✓																✓	✓	✓												
Meng et al. (2021)																✓																	✓		
Habibi et al. (2021)	✓		✓																✓		✓														
Bakhtiari et al. (2021)														✓																	✓				
Singh et al. (2021)													✓																	✓					
Ahmadpour et al. (2021)											✓	✓	✓															✓	✓	✓					
Chen et al. (2021)												✓																	✓						
Haghighi et al. (2021)	✓					✓													✓					✓											
Wang et al. (2020)	✓																		✓																
Yousefi et al. (2020)												✓	✓																✓	✓					

Paper	RMSE	MAE	R-squared	NSE	NRMSE	NSD	RE	NER	AHP	TT	TSS	AUC	AROC	SST	SPF	OA	K	ESAI	RMSE	MAE	R^2	NSE	NRMSE	NSD	NER	AHP	TT	TSS	AUC	AROC	SST	SPF	OA	K	ESAI
Chakrabortty et al. (2020)												✓																	✓						
Grinand et al.(2019)												✓	✓																✓	✓					
Gayen et al. (2019)																✓	✓																✓	✓	
Dharumarajan et al. (2017)												✓	✓			✓													✓	✓			✓		
Keshtkar et al. (2017)																✓	✓																✓	✓	