

Modeling of hiking trail degradation using machine-learning techniques to find optimized recreational trails in an arid environment

Hamed Adab

h.adab@hsu.ac.ir

Hakim Sabzevari University

Zahra Ghelichipour

Hakim Sabzevari University

Azadeh Atabati

Hakim Sabzevari University

Research Article

Keywords: Trail Planning, Least Cost Path Analysis, Machine-Learning, Sarigol National Park and Protected Area (SNPP)

Posted Date: October 4th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-5019250/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

The present study aims to use different machine-learning algorithms to map trail susceptibility and use it to find the best hiking trail between specified locations across the Sarigol National Park and Protected Area (SNPP), Iran based on the least cost path analysis. Furthermore, the study compares the predictive performance of Artificial Neural Network, Support Vector Regression, and Gene expression programming model for trail susceptibility mapping. We have considered nine trail susceptibility conditioning factors as model input, namely Land use coverages, Landform classes, Annual precipitation, NDVI, Soil types, LS-factor, Wind explosion index, Topographic witness index, and Elevation. The study concluded that ANN gives better performance in overall accuracy assessment as compared to GEP and SVM models. The importance of predictor variables as identified by the ANN model indicated that the LS factor, Soil types, NDVI, and Landform classes represented the highest level of significance attributed to the model. The study found that LCPA is an efficient tool to find the “lowest land degradation” to connect two locations of hiking trails. This suggested that park planners should consider potential land degradation locations to identify suitable hiking trails.

1. Introduction

The demand for ecological tourism and intensive use of outdoor recreational areas have grown fast due to increasing visitors, leisure time, increased income, facilitated mobility, and consumer demands (Sutherland et al., 2001). Thus, nature tourism activities such as walking and mountain biking which are associated with trails, have been growing faster in developing regions (Eagles, 2014; Newsome et al., 2012). Recreational trails are increasingly created in many areas, including national parks and protected areas (Cole, 2004; Eagles, 2014), as well as urban and semi-urban natural areas (Ballantyne and Pickering, 2015). People of all ages and abilities can use trails with different functions such as access to attractive areas, recreation and wildlife watching (Grimwade et al., 2009; Santarém et al., 2015), sports activities, relaxation, and travel to the destination (Grimwade et al., 2009). Trails are designed to avoid the uncontrolled dispersal of visitors (Olive and Marion, 2009) and provide more infrastructure to allow access to natural areas (Ballantyne et al., 2014). As national Parks and protected areas are major destinations for outdoor recreation, the ecosystems of those areas are being extensively affected by human activities (Chatterjea, 2007). This may result in increased environmental degradation and wildlife habitat disruption by concentrating visitor activities into specified areas like trails and recreation sites (Chatterjea, 2007; Clius et al., 2012; Cole, 2004; Dixon et al., 2004; Farrell and Marion, 2001a; Olive and Marion, 2009).

Trail degradation is one of the most evident consequences of expanding visitor numbers in national parks. The magnitude of trail impacts is contingent on the intensity, frequency, timing, and type of use, as well as environmental conditions (Törn et al., 2009). The environmental impacts of recreational trails have been comprehensively presented worldwide. Trampling on trails changes soil surface compaction, mechanical properties, and hydrophysical behavior of watersheds which leads to greater on-trail erosion and changes in the micro-climate conditions (Chatterjea, 2007). The impacts of trampling activity on

water resources, plant communities, and wildlife species are observed in areas where visitor use is intensified, particularly along trails and campsites (Leung and Marion, 2000; Matulewski et al., 2021; Olive and Marion, 2009). It is well observed that new paths generated by visitors and trampling can significantly influence the inhibition of radial growth of trees, increasing trail widening, trail muddiness, soil organic matter removal, and soil particles compacting (Cole, 2004; Hill and Pickering, 2006; Leung and Marion, 2000; Matulewski et al., 2021). The decline in vegetation cover, changes in vegetation height, and introduction of weeds are the other negative impacts of recreation trails on the environment (Barros and Marina Pickering, 2017; Dolan et al., 2006). Therefore, trail development and use are often a major concern of natural area managers and visitors (Leung and Marion, 1996a).

Trail width is used as an indicator of degradation of the recreational trails (Tomczyk and Ewertowski, 2011). Generally, environmental predictor variables influence trail width because they influence human behavior (Wimpey and Marion, 2010). The predictor variables for estimating trail degradation can be referenced to environmental and managerial variables (e.g. elevation, slope, aspect, soil type, soil texture, vegetation type, etc (Nepal and Way, 2007; Spornbauer et al., 2023; Tomczyk and Ewertowski, 2011, 2013). The relationship between predictor variables and trail degradation and their negative impacts on the environment is not linear (Hammitt et al., 2015; Tomczyk and Ewertowski, 2013). Therefore, getting a reliable model for understanding the relationship between the trail width and predictor variables is required for effective tourism management (Farrell and Marion, 2001b; Leung and Marion, 2000). According to various studies in recent years, different predictive machine learning (ML) and deep learning (DL) algorithms were widely used in landslide susceptibility (Orland et al., 2020), soil moisture estimation (Adab et al., 2020) and forest fire vulnerability prediction (Adab, 2017). However, a comparative analysis of different predictive models using GIS techniques is very rare for the estimation of trail width (Nepal and Way, 2007; Spornbauer et al., 2023; Tomczyk and Ewertowski, 2013; Wimpey and Marion, 2010) and the prediction of recreational trail susceptibility (Sahani and Ghosh, 2021). In the context of trail width estimation, a comparative analysis of different predictive techniques has not been considered in the literature. Also, quality research is scarce, and a lack of scientific studies on the estimation of trail width in Iran. The accuracy of the algorithms may differ for the mapping of trail width, and the selection of the best model is important for decision-makers for maneuvering the natural stability of the trail. On the other hand, considerable research has been devoted to the context of local attractiveness for tourism and recreation (Khazaei Fadafan et al., 2022; Meshram et al., 2022) and optimal-designed road network by MCDM-GIS approach in Iran (Talebi et al., 2019) rather less attention has been paid to synthesizing knowledge on spatial optimization of recreational trails and role of human mobility patterns on the degradation (Talebi et al., 2022).

A research gap is found in the context of trail width estimation, as no comparisons have been made between multivariate machine learning models. The present study aims to assess the predictive performance of three multivariate machine learning predictive models for the estimation of trail width along the recreational trail of the Sarigol National Park and Protected Area (SNPP) and its optimization of recreational trails. The findings of this study will provide a foundation of comparative analysis of

machine learning algorithms for tourism researchers, and government officials to enhance land management and tourism management.

2. Materials and Methods

2.1. Study Area

The present study was conducted in the the Sarigol National Park and Protected Area (SNPP) with an area of 28,000 hectares which is divided into the national park and the protected area. This area is located in northeastern Iran (Fig. 1) which is the part of the eastern extreme of the Irano-Anatolian Biodiversity Hotspot (Mohammadi et al., 2021). Sarigol National Park was established under the Wildlife Protection Act of 1974, by the government of IRAN under the name of “Shah Jahan Protected Region”. After 1978, this area was renamed “Sarigol Protected Region”. Sarigol National Park covers an area of 6,000 hectares and falls under the International Union for Conservation of Nature (IUCN) category II. On the other side, the Sarigol protected area covers around 22,000 hectares area and belongs to IUCN category IV protected area. In 1979, 20% of the area of the protected region (about 8000 hectares) was separated for the protection of ecological value, which is completely preserved as “Sarigol safe region” (Rahchamani et al., 2014). These semi-arid areas receive 273 mm of annual participation and the mean annual temperature of the study area is 14°C. The elevation of the forest site ranges between 1234 and 3017 m above sea level (a.s.l.). Almost all of the area is covered with mountains, valleys, hilly, and a small area of the plain. The biodiversity of the area is considered rich due to the full conservation of the region (Rahchamani et al., 2014) and also specific climatic features corresponding to the Mediterranean or Irano-Turanian xeric-continental climate (Djamali et al., 2012). The Sarigol area has the potential to develop an outdoor recreation economy. The public has the opportunity to come into the Sarigol area to contact wildlife species through recreational observation.

2.2. Sampling and measurement procedures

Walking or hiking has been known as one of the most important recreational activities in the study area. In this study, the measurement of the width of individual hiking trails was conducted on both the national park and protected area by the researcher with two assistants from the Department of Environment of IRAN along the popular tracks. Traditionally, measurement of the width of hiking trails is used by field-based data sampling from different landforms. The large hiking trail sample size is ideal for statistical analysis. However, this is not always doable because these field surveys can be time-consuming and expensive. On the other hand, inadequate sample sizes influence the quality and accuracy of the model (Kotrlik and Higgins, 2001). In this study, the Cochran formula is used to find the most suitable sample size (Cochran, 1977). Based on the Cochran formula, the number of needed sample points was calculated at 115 points at 0.09 acceptable margin of error. However, it is not possible to measure the width for all 115 points due to time, budget, and other constraints. Therefore, three main hiking trails were selected for width sampling which are located in the south of the national park, on the border between the protected area and the national park, and in the north of the protected area (Fig. 1).

A total of 20 sample points were selected among 115 samples based on time and budget and the width of hiking trails was surveyed in the field. Locations selected along the hiking trails were sampled based on the topographic and land-cover characteristics. The field survey made use of hand-held eTrex Garmin GPS receivers with reasonable spatial accuracy of between 4 and 6 meters. The data were collected from June to September 2015. The field survey produced a total of 50 valid observations.

The Google Earth app is one of the online mapping applications that provide users with interactive mapping capabilities. This program is often used by academic users as an accurate and reliable geo data source for referrals or basic maps, as well as easy access and a free-cost source of image information (Ragheb and Ragab, 2015). Google Earth's imagery for the study area is available more the recent years, having been used in January 2013. The width of the rest trail samples ($n = 95$) was measured by the ruler tool of Google Earth Pro. These locations in a variety of geographic contexts were selected from Google Earth satellite imagery. Figure 1 shows hiking trails produced by Google Earth. To assess the efficacy of Google Earth's imagery-derived hiking trail widths, 20 hand-measured widths were used in the field along the study area.

2.3. Model Input Variables and Descriptions

As the aim of this study is to use different data-driven methods to estimate hiking trail degradation, we decided to not consider a large number of trail degradation predisposing factors, to minimize the possible combination of information resulting from the factors overlay. By reviewing previous literature, the most effective width of hiking trails predictors, a total of 9 independent variables including elevation, Normalized Difference Vegetation Index (NDVI), soil type, slope length and slope steepness factor (LS-factor), Channels, topographic position index (TPI), topographic wetness index (TWI), wind exposition, land use, and elevation were selected in this study to construct a spatial database (Bryan, 1977; Marion and Wimpey, 2007; Marion and Olive, 2006; Ólafsdóttir and Runnström, 2013; Randall and Newsome, 2008; Sahani and Ghosh, 2021). The data sources and summaries of the predisposing factors considered for the width of hiking trails assessment are shown in Table 1, and the spatial distribution patterns of the influencing factors are represented in Fig. 2. Additionally, the predisposing raster factors having different grid sizes were sampled at a 231 m resolution using the nearest-neighbor interpolation technique to match the resolution with the NDVI data acquired from MODIS. Because gridded climate data such as WorldClim lacks spatial detail, it is often needed to produce a new high resolution of climatic variables that can be used at the regional scale (Kusch and Davy, 2022). In this study, Kriging is used to interpolate data to coarse-scale precipitation data which can produce more accurate and reliable estimates at local scales (Kusch and Davy, 2022).

Concerning duplicate sampling in one pixel of a base resolution (231 m) due to different scales of sample and independent variables, it is expected that duplicate points in one pixel affect the result of the statistical analysis. To overcome this problem, it has been a common way to calculate the width mean of hiking trails of duplicate sampling in one pixel. From these procedures, 79 points of hiking trails, out of 115, were rarefied for the training set in the model (Fig. 1). This means 79 measurements are used to have a confidence level of 95% that the real value is within $\pm 11\%$ of the measured value. A total of 5317

(98 rows by 87 columns) cells were obtained for the application of the three different statistical approaches including Artificial Neural Networks, Support Vector Regression, and Gene expression programming.

Table 1

Brief descriptions of the eight influencing factors considered in this study for the prediction of the width of hiking trails

Independent variables	Description of data	Data Source
Elevation	The elevation is a key variable influencing the erosion rate of hiking trails fact fewer visitors can hike the trails at higher altitudes as it is more physically challenging and instead, they choose an enjoyable hike at lower elevations which is the easier and more accessible trails (Ólafsdóttir and Runnström, 2013).	Global Data Explorer (GDEx), United States Geological Survey (USGS) (Tachikawa et al., 2011). The LS-factor is derived from the Digital Elevation Model (DEM), at 30-m Resolution ASTER Global Digital Elevation Model (GDEM).
LS-factor	The L-factor measures the effect of slope steepness, and the S-factor represents the impact of slope length. the LS-factor shows the effect of topography on soil erosion (Panagos et al., 2015).	LS-factor calculated from GDEM, V2, SAGA software, Terrain Analysis, and Hydrology module. Gridded continuous value at 30-m.
TWI	The TWI represents the capacity of the landscape to accumulate water and soil moisture (Raduła et al., 2018). The trail becomes much more susceptible to degradation in the areas where the soil is wet or with a higher value of TWI (Sahani and Ghosh, 2021).	The TWI factor derived from DEM, at 30-m Resolution ASTER Global Digital Elevation Model (GDEM), SAGA software, Terrain Analysis, Hydrology, Topographic Indices module. Gridded continuous value at 30-m.
TAP	Total annual precipitation interacts with topography, vegetation, rainfall, and other environmental attributes to influence trail conditions (Ng et al., 2018).	WorldClim database of global climate layers (version 2), (Fick and Hijmans, 2017). Gridded continuous value at 834-m.
Wind exposition	Higher erosion rates on steep trail slopes, and therefore increased exposure to wind erosion (Leung and Marion, 1996b).	Calculated from DEM, at 30-m Resolution ASTER Global Digital Elevation Model (GDEM), SAGA software, Terrain Analysis, Climate, and Weather. Module Wind Exposition Index. Gridded continuous value at 30-m.
Topographic Position Index (TPI)	Land cover features such as valley bottom and open vistas are preferred recreational trails with visitors, however, they are susceptible to degradation and small perturbations of the natural biogeochemical cycles	Calculated from DEM, at 30-m Resolution ASTER Global Digital Elevation Model (GDEM), SAGA software, Terrain Analysis. Terrain Classification. Module TPI-Based Landform Classification. Gridded continuous value at 30-m.

Independent variables	Description of data	Data Source
	through human activities (Santiago et al., 2008).	
Soil type	Some soil types are most prone to particle erosion from recreational trails. Higher visitor rates readily collapse soil structures and lead to trail erosion (Olive and Marion, 2009).	Soil type produced from maximum likelihood classification of ASTER data image dated 2014 from digital soil map, National Synthesis Plan, Natural Resources and Watershed Management Organization, Iran. Kappa coefficient 0.56. Gridded categorical value at 12-m.
NDVI	Vegetation is the key variable influencing the erosion rate of trails (Magdalena Warter et al., 2017). Recreational trails cause more substantial abrasion destruction to vegetation, and roots (Olive and Marion, 2009). Some plant species such as dry grasslands are highly resistant to impact, which can minimize impacts like trail spread and trail widening. Conversely, trail degradation occurs more rapidly in open areas with less resistant vegetation types (Farrell and Marion, 2001b).	20 years of MODIS-NDVI products were retrieved by the AppEEARS portal (https://lpdaacsvc.cr.usgs.gov/appeears/). Gridded continuous value at 231-m.
Land use coverages	Land use has a direct effect on soil degradation through intense recreational use of trail networks, compaction, and trampling (Sutherland et al., 2001; Tomczyk, 2011).	Land use type produced from maximum likelihood classification of ASTER data image dated 2014 from digital Land use map, National Synthesis Plan, Natural Resources and Watershed Management Organization, Iran. Kappa coefficient 0.82.
Dependent variable		
Width of hiking trails	A total of 431 sampling sites were collected from a variety of geographic contexts. Point vector format	The geographic coordinates of each well have been located by GPS.

2.4. Machine Learning Algorithms

2.4.1. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is a subset of the machine learning approach that is applied for solving both regression and classification problems to find complex relations between the dependent and independent variables for the best result by a nonlinear function approximation algorithm (Jacinth Jennifer and Saravanan, 2022). In the field of mathematical modeling, many forms of artificial neural networks (ANN) such as radial basis neural networks (RBNN), multilayer perceptron neural networks

(MLPNN), and generalized regression neural networks (GRNN) have been applied in several diverse fields, including hydrology, agriculture and meteorology (Adab et al., 2020; Kişi, 2009). A MLPNN belongs to a general type structure of ANN called a feedforward neural network. The ANN is usually implemented by a set of processing units that are arranged in a layered feed-forward topology. MLP consists of two layers (input/output), and the complexity of the network increases with the addition of hidden interconnection layers (Ermini et al., 2005). The hidden layer shows the intermediary nodes that can help the network learn more complex multivariate nonlinear relationships and improve its performance. A set of nodes (neurons) is present in each layer, and each node is connected to nodes in the next layer with parameters called 'weights'. During the training phase, the neurons in the MLP learn the optimal weights for these connections using a learning process called the backpropagation algorithm, which minimizes the error between the predicted data and the actual data by repeatedly adjusting the weight values of the connections in the network. Once the learning algorithm is optimally adjusted, it can be used to make predictions on new data. In this study, a three-layered, multiplayer perceptron ANN model, with backpropagation algorithm was constructed by the Neural Network Toolbox of JMP V 10 software. The input layer consists of nine input parameters (Fig. 2) and is fed to one hidden layer. The hyperbolic tangent (sigmoid) function is applied to generate the transformation of a linear combination of the X variables (Mohanty et al., 2015).

The boosting method was utilized to build the best-performing ANN model. The boosting method of the Neural Network Toolbox of JMP is a set of boosting models and learning rates (Ling et al., 2021). The model is trained by robust fit using the least absolute deviations to minimize the impact of outliers which can be used for continuous responses (Ling et al., 2021). A squared penalty is considered to mitigate the over-fitting tendency of the neural networks. The number of tours is set to different times to find a proper fit for the data. Here, all of the tours converged by generating the model many different times for the best fit to the data, representing there is no problem with local optima.

2.4.2. Support Vector Regression (SVR)

SVR is a kernel-based method of supervised learning that is a promising approach for both linear and non-linear regression. The main idea of SVR is mapping input training variables data X_i from low dimensional into the high dimensional feature space by the kernel functions. Constructing several hyperplanes and a hyperplane with a maximum distance to the nearest training data point (called support vectors) was selected as the optimal hyperplane. Maximizing the margin of separation between the hyperplane and the support vectors can improve the generalization ability of the regressor. Kernels help SVR to implicitly create the optimal hyperplane in the feature space (Shin et al., 2011). This feature space can be formulated as a quadratic programming (QP) to find an optimized hyperplane that also shows the non-linear relationship between input (independent variables) and output data (dependent variables). Unlike neural networks, SVR creates global solutions, because SVR solves quadratic programming (QP) problems and focuses on finding the optimum hyperplane and minimizing the training error between training data and the e-insensitive loss function (Chen et al., 2017). In this study, the

sequential minimal optimization (SMO) algorithm is used for solving the quadratic programming (QP) problem.

An effective SVR model can be obtained by setting kernel-specific hyperparameters which are defined by the user to get optimal regression accuracy and generalization performance. Common kernel functions include linear kernel, and radial basis function (RBF). In this study, SVR was developed based on two kernel functions to get the optimal kernel function. The radial basis function kernel requires two hyperparameters to create a nonlinear hyperplane which includes the Cost penalty parameter and Gamma curvature parameter. The linear kernel function used only the Cost penalty parameter to create a linear hyperplane. The number of runs and the minimum and maximum values for the parameters of the kernel were applied for the tuning of SVR by grid search method by SVM Toolbox of JMP V 10 software.

2.4.3. Gene expression programming (GEP)

Genetic expression programming (GEP) as a conventional statistical method can be classified as a 'grey-box model' because it allows the model designer to control model structure and provides a better conceptualization of physical phenomena in a symbolic and uncomplicated way (Jalal et al., 2021). A GEP approach is one of the most intellectually compelling computational intelligence techniques based on principles from Darwin's evolution theory and Mendel's genetics. A main advantage of the GEP algorithm is their prediction using simple mathematical expressions, which yield reliability and improved prediction efficiency (Iqbal et al., 2022). In the present work, the GeneXproTools program (Version 5.0.3919) is applied to implement GEP models. The basic working mechanism of the GEP involves the following steps (Jalal et al., 2021).

1. GEP modeling begins with a random generation of chromosomes of the initial population and the Karva programming language which is a symbolic language for chromosome introduction is used for introducing the chromosomes. The chromosomes and genes are composed of a head and a tail. The head contains symbols that show both functions and terminals node, while the tail contains only terminals node. The size of the head of the chromosome depends on the complexity of each parameter, whereas the number of genes determines the number of sub-expression trees (sub-ETs) of the model
2. In 2nd step, this developed length of chromosomes is fixed, so the chromosomes are simply transformed into an algebraic expression. Each chromosome (gene) of GEP includes a fixed list of terms, consequently, each term is represented by a function, including basic arithmetic operations (+, -, ×, /), Boolean logic functions (AND, OR, NOT, etc.), mathematical functions (cos, sin, ln), conditional functions (IF, THEN, ELSE), or other function categories to get the optimum GEP model.
3. The chromosomes are represented by ETs, which have a variety of shapes and sizes. Afterward, the primary genetic operators are applied to each chromosome which consists of crossover, mutation, transposition, and gene recombination (1-point, 2-point, and gene recombination) according to their ratios. The process is completed when the stopping condition (highest number of generations or best solution) is reached (Ferreira, 2002).

4. The roulette wheel method (as well as greedy over the selection, ranking selection, tournament selection, and elite strategy) is employed when the maximum number of iterations or favorite fitness value is not met. Roulette wheel method and such that other methods select the viable chromosomes of the first generation and continue to the next generation. This process would be incrementally rewound for a specific number of generations or until finding the selection of the optimal solution. There are three methods to represent GEP output, including the Karva language (the gene language), Expression Tree (ET), and mathematical functions.

2.5. Tuning of hyper-parameters

Before the training phase, it is crucial to find the optimal set of hyper-parameters for machine learning algorithms, which directly affects an ML's model performance. ML algorithms are tunable by recognizing patterns in the training dataset and then predicting the output of the new dataset based on these patterns (Pannakkong et al., 2022). Picking the right hyper-parameter values is unknown and tuning may become a challenging problem and time-consuming to identify the right combination of hyper-parameter. Manual hyper-parameters tuning involves experimenting with different sets of hyper-parameters manually and then the value of the hyper-parameters is configured in the algorithm, the model executed and the accuracy assessed. This trial and error process is repeated to compare the accuracy by statistical evaluation metrics until the hyper-parameters that represent the optimal accuracy are chosen (Palani et al., 2008). In this study, different hyper-parameter sets for ANN, SVM, and GEP and methods were tried to make the models more accurate by trial and error process where several ANN, SVM, and GEP models were developed by changing the parameter values, as shown in Table 2. In the case of ANN, SVM, and GEP, the performance of these models is verified during the training phase by hold-back cross-validation and statistical checks. K-fold is usually used for size-limited datasets because it employs limited amounts of data to evaluate the performance of a predictive model efficiently (Ling et al., 2021). Hold-back cross-validation was chosen because comparing the performances of models can be obtained from the same validating data set. In this study, holdback randomly divides the data into the training and validation (holdback portion) sets by shuffling method. Here, 30% of the original data were held-back to prevent model overfitting from the ANN, SVM, and GEP models during holdback validation (Beucher et al., 2017).

A wide range of metrics are available to compare the performance of the developed models quantitatively; In this study, the Taylor diagram, Index of Agreement (D-index), and Kling-Gupta efficiency (KGE) were applied to evaluate the accuracy of predictive models. Taylor diagram was used to graphically see the precision of the modeled outcomes. In the Taylor diagram, a single point represents simultaneously three statistical parameters of model output on a polar plot, namely, correlation coefficient, root mean square error, and standard deviations. The point nearest to the observed data as a reference point indicates the perfect predictive model (Kardani et al., 2022). The index of agreement statistic criterion does not measure the correlation between two variables. Instead, it is designed to measure the agreement to which the model predictions match exactly the observed data with no proportionality (Aarnio et al., 2018); but, it is influenced by extreme values due to the squared

differences. KGE is a reliable goodness-of-fit statistic to overcome the limitations of the Nash-Sutcliffe efficiency (NSE). The KGE varies on the range $-\infty$ to 1 and higher values of KGE show a better agreement between predicted values and observations.

Table 2

Machine learning models and hyperparameter combinations are used in the manual hyperparameter search.

Model	Hyperparameters
Artificial Neural Network	
Boosting algorithm	Number of models = 10 Learning rate = 0.1
Activation function	hyperbolic tangent (sigmoid)
Configuration of the network	3 neurons in 1 hidden layer
Fitting method	Robust fit, Least Absolute Deviations (LAD)
Penalty method	Squared
Number of tours (local optima)	200; All 200 of the tours converged on the best fit to the data, showing there was no problem with local optima.
Support Vector Regression	
kernel function	Radial basis function (RBF), Linear
Cost penalty parameter, Gamma curvature parameter for RBF	Min = 0.01 ; Max = 5; Best Cost = 4.92 ; Min = 0.001 ; Max = 0.5 ; Best Gamma = 0.27
Cost penalty parameter for Linear	Min = 0.01 ; Max = 5 ; Best Cost = 0.0145
Number of Runs	40
Gene Expression Programming	
Chromosomes	40
Genes	10
Head size	14
Genetic operators	Optimal evolution
Linking function	Avg2
Function set	+, −, x, ÷, Exp, Ln, Inv,X2, 3Rt, Tanh, Atan
Mutation rate	0.00138
Inversion rate	0.00546
Constants per gene	20
Data type	Floating type
Lower bound	-10
Upper bound	10

2.6. Least Cost Path Analysis

The Least Cost Path Analysis (LCPA) is a GIS-based spatial optimization technique that allows environmental designers to find the “lowest cost” path to connect the source and the destination within a cost surface by Dijkstra Algorithm (Yu et al., 2003). Selecting an optimal path based on the width of hiking trails as a cost surface is the first step in path design which is described in Table 1. The cost surface is shown by a grid map in which each pixel involves the width of trails as a cost that represents how “expensive” it is to pass through that pixel. The accumulated cost surface is calculated from the cost surface, by computing the cumulative cost of each pixel from the source point. An optimal path between a source point and the destination point is traced down on the accumulated cost surface, by searching among the starting point's neighboring pixels, which has the lowest accumulated cost value (Albert and Sárközy, 2021) towards the destination point. In this study, the accumulated cost surface and the least-cost path are implemented using the Optimal Path as Line (OPL) tool in the Arc GIS Pro platform. The details of the proceedings are presented by de Lima et al. (2016). In this study, the start and endpoints of the existing trails within the study area were considered as source and destination points with start points are rural areas (Eyzi) and two endpoints for camping sports in mountain areas and Bedvaz Dam (Fig. 7). The least-cost path from the start point to the destination has been calculated for the optimal hiking trail after considering hiking trail width as cost surface.

3. Results

3.1. Results from Input Variables

Table 2 presents the summary statistics for each of the input parameters including the normalized interquartile range (nIQR), Shapiro-Wilk normality test, and Grubbs outlier test for the training datasets as well as the validating datasets. nIQR is a robust method for estimating symmetric and skewed distributions (Huang et al., 2022). The results showed that most of the data sets represented a slight asymmetric distribution. From the results, about 60% of data presented a normal distribution and, approximately, 14% (LS- data sets) indicated outliers, according to Jarque-Bera and Grubbs tests at a significance level of $\alpha = 0.05$, respectively. Moreover, the mean values of training and validating datasets show these tend to be close to each other. These results support the use of ANN, SVR, and GEP for the width of hiking trails analysis over the traditional statistical methods since neural networks are free of any statistical assumptions.

Table 1
Descriptive statistics (Quantitative data) for the hiking trail width and potential predictors.

Statistic	Min	Max	Mean	nIQR	Jarque-Bera	Grubbs test
WHT-T	45.0	165.0	101.7	51.0	0.068	1.00
WHT-V	60.5	165.0	117.4	43.8	0.487	1.00
WEI-T	0.85	1.22	1.01	0.08	0.285	0.67
WEI-V	0.88	1.12	1.01	0.06	0.753	0.73
TWI-T	4.4	15.9	7.8	2.1	0.008	0.09
TWI-V	4.5	13.5	7.2	1.3	0.032	0.04
NDVI-T	0.00	0.26	0.17	0.06	0.116	0.14
NDVI-V	0.01	0.31	0.17	0.05	0.843	0.29
LS-T	0.0	276.5	25.7	14.3	< 0.0001	< 0.0001
LS-V	0.3	83.0	22.2	16.3	0.007	0.01
AP-T	321	368	344	11	0.822	0.59
AP-V	324	365	345	8	0.840	1.00
DEM-T	1257	2854	1845	353	0.005	0.84
DEM-V	1253	2759	1810	224	0.027	0.11

As the computed p-value is greater than the significance level $\alpha = 0.05$, one cannot reject the null hypothesis H_0 . H_0 : The variable from which the sample was extracted follows a Normal distribution and there is no outlier in the data.

3.2. Implementing Machine Learning Models

The training set constructed was used for training the models. Then, the whole predictor dataset of the study area is applied as inputs for the deployed model to generate the estimation of width hiking as trail degradation index for the study area, as shown in Fig. 9. The estimation of width hiking was classified by the natural break into five classes (Very low, Low, Moderate, High, and Very high). The natural break method detects breakpoints statistically by picking the class breaks with relatively large jumps in the data values that best group similar values and maximize the differences between classes. The “Very low” trail degradation class area is represented by a dark blue color, and the “Very high” trail degradation class area is characterized by a crimson color. The colors changing from cold to warm represent that the trail degradation increases gradually.

Moreover, a histogram was drawn to provide details on the frequency of hiking pixels within each model in Fig. 4. It is observed that the majority of the grid cells for all models are located between 60 and 140,

respectively. It can be also seen that the spatial distribution of both GEP and SVM-RBF are the same and concentrated in the middle of the histograms. It is clear that the grid cell distributions of the GEP model are much wider and dispersive than the ANN and SVM models. By comparing the histogram between the four distributions (Fig. 4), we can see that the distribution of the SVM-linear model is much narrower than the others.

3.3. Model Validation and Comparison

The performance of models for the training and validating dataset is represented in Table 2. From the validating datasets, it showed that the ANN model has achieved the lowest values of MAE (17.64), and MAPE (14.1%) in predicting width hiking. It is important to note that, although the SVM-RBF model using the validating dataset has the best performance in terms of the logarithmic error (RMSLE) AUC (0.06), its performance in terms of MAE and MAPE is greatly inferior to that of the ANN selected model. Combined performance results represent that the ANN exhibited the highest performance rate value than the other machine learning models in predicting width hiking.

Table 3
Accuracy assessment of hiking trail width estimated by ANN, SVM, and GEP models

MLs	Dataset	RMSLE	MAE	MAPE	D-index	KGE
ANN	Training	0.26	18.25	21%	0.74	0.67
	Validation	0.08	17.64	14.1%	0.67	0.6
SVM-RBF	Training	0.25	3.90	4.5%	0.94	0.92
	Validation	0.06	21.13	19.4%	0.61	0.42
SVM-Linear	Training	0.32	25.85	28.3%	0.63	0.32
	Validation	0.11	28.95	23.7%	0.47	0.27
GEP	Training	0.04	18.95	20.8%	0.73	0.61
	Validation	0.17	32.24	27.8%	0.41	0.35

In addition, Fig. 3 shows a more quantitative graph of the agreement between model predictions against observed (reference) values which is obtained from the Taylor diagram using three different statistical parameters for validating the dataset. From Fig. 3, the statistical parameters used in the Taylor diagram are Pearson correlation coefficient (r), RMSE, and standard deviation. The observed hiking trail width is shown as the reference red point where the correlation is 1 and the centered root mean square error (CRMSE) is 0. The red line shows the standard deviation of the reference point. The SVM-linear, SVM-RBF, ANN, and GEP are located on the left of the red line. The result indicated that the hiking trail width has a high variation from the mean, while the ANN near the line shows a similar variation with the observed data. For more information, the Taylor diagram is the correlation between the hiking trail width predicted values and observation values. The ANN method is located between 0.6 and 0.65 sectors and

has a higher correlation than other models. The GEP is in the sector of 0.38 and has the lowest correlation. The SVM-linear and SVM-RBF are located in the sector of 0.43 and 0.54, respectively. The diagram represents information on the reliability of the modeling by attaining RMSE, which was expressed with color contours. The ANN is in the 29.8 cyan contour, which shows that the ANN method has the most reliable modeling process. The GEP is located on the yellow contour with an RMSE of 35, which indicates that a large average error is generated during the modeling process. Overall, the results showed that the ANN method had a higher accuracy for the validation dataset in predicting and spatially mapping hiking trail width for the study area.

3.4. Relative significance of predictors

A sensitivity analysis was performed to measure the importance of predictor variables in the ANN model (Fig. 6). Total effect as an importance index was used which represents the relative contribution of the predictor variables both alone and in combination with other predictor variables. Among all predictor variables, Slope Length Factor, Soil Types, and NDVI were the strongest predictors of hiking trail width in the ANN model, with relative importance values ranging from > 20 to 27%. The important predictor variables represent the variation in the factor that causes high variability in the hiking trail width, then that effect is important relative to the model (Fig. 6). The total effect of the rest parameters is less than 15 percent in the ANN model.

3.5. Least-Cost-Path analysis for the selected area

Figure 7 provides a close-up view of the existing hiking trail and the least-cost path as the hiking trail alternative on the hillshade map to reveal its real structure. The cost surface pixel values ranging from 40cm to 155cm were calculated according to the ANN model. The hiking trail alternative has the lowest cost in terms of trail width that connects three locations using accumulated cost surface that is determined by considering ANN estimation of hiking trail width as degradation index. The suitable hiking trails are shown in green and blue, and the existing hiking trails in red. Overall, the newly generated ANN-based hiking trail, and the existing hiking trail show a similar trail in some areas. The least cost path distance was 14.6km with an elevation gain of 1596m while the existing hiking trail distance was 16.4km with an elevation gain of 1445m. The difference of 151m between the existing hiking trail and the alternative was noticed and this would also go a long way in reducing the degradation of the hiking trail to be used if the alternative hiking trail as found by the least cost path method is to be followed.

4. Discussions

The result of this study shows that hiking trail width predictive machine learning techniques were a valuable tool for identifying susceptible zones in terms of erosion and deposition of soil and increasing human trail degradation. As described in the methodology section, three types of machine learning algorithms are applied to databases separately. Among all three machine learning techniques, ANN is found to be the most promising model to predict hiking trail width because of its ability to learn enormous volumes of data without depending on the scale of measurement and also the way data are

arranged (Taye, 2023; Yesilnacar and Topal, 2005). The superior performance of the ANN is linked to the complex computing capabilities of the algorithm structure through model training (Khan et al., 2022). Theoretically, modern machine learning techniques could be considered a technique to make highly predictive accuracy, however is not grunted that these techniques always offer the best result because the performance of each algorithm largely depends upon the variety of hyper-parameters, settings, and configurations which are typical for their general applications are chosen (Merghadi et al., 2020). Sahani and Ghosh (2021) used logistic regression (LR), decision tree (DT), and random forest (RF) for spatial prediction of recreational trail susceptibility. Comparing their results showed that LR indicates better overall performance as compared to DT and RF. As seen from Table 3, the SVM-RBF model has higher agreement than ANN and GEP in model calibration, however in the model test phase, the agreement of ANN is much higher than SVM-RBF, and GEP indicating ANN model is often superior in generalization ability. It was discussed that ANNs have self-learning non-linear abilities that allow them to produce better performances as more data becomes available (Kumar et al., 2021). Theoretically, the GEP model could be considered a technique to make highly predictive accuracy, but it has the highest capability to overfit which represented in Table 3.

Based on the findings of important hiking trail width factors, it was found that slope length factor, soil type, and NDVI are consistently recognized as the most important factors for the ANN model that influence the trail degradation occurrences in the Sarigol Protected Region. This finding demonstrates that the slope length factor was an important factor that influenced hiking trail width, which is in line with the findings of Wimpey and Marion (2010), Olive and Marion (2009). Among environmental factors, soil type was the most robust determinant of hiking trail width. This result is consistent with other studies that soil texture is the main factor that significantly affects the sustainability of a trail (Marion and Wimpey, 2017). The results of this study represent that NDVI is highly related to hiking trail width. Similar to the study by Sahani and Ghosh (2021) NDVI and vegetation type have the highest influence on the occurrence of susceptible trails. Additionally, TPI landform elements act as a softening mediator, which increases the probability of soil loss occurrences over the study area. Additionally, this finding found that TPI was an important factor in trail erosion with valley positions eroding more than mid-slope and ridge trails, which also agrees with Dixon et al. (2004) and Olive and Marion (2009). From the aforementioned analysis (Fig. 6), it can be noted that other predictors do not play a crucial role in hiking trail width, however, these factors are likely influenced on hiking trail width by a combination of flooding-related erosion periodically (Olive and Marion, 2009). The findings from Fig. 7 demonstrated that the identifying trail was well established by combining the least-cost path algorithm with GIS-based land suitability data. Moreover, trail design can assist us in identifying the most significant physical and natural features, which can help hikers in cognitive learning across the trail and increase their awareness of environmental problems. Concerning hiking trail width as a degradation index, the findings of this study emphasize the importance of trail design through the least cost path algorithm and replacing them with existing trail segments that could face substantial degradation routing when possible (Meadema et al., 2020; Sitzia et al., 2014). These finds could be indicative clues for the future trail design, which also agrees with Li et al. (2005) for tourism management.

5. Conclusions

Recreational trail susceptibility assessment is a key step in reducing trail degradation risk in an arid environment, especially for the restoration of natural protected areas which is achieving an adaptive balance between the conservation of nature and recreational opportunity. This paper represents a comparison of three machine learning algorithms for the estimation of hiking trail width as a trail degradation index based on the ANN, SVR, and GEP methods. Overall, the ANN model has the best performance, followed by SVM RBF. The GEP model is simple and easy to implement but does not have good generalization capabilities and tends to overestimate the hiking trail width for test sets to show a higher level. The highest errors, in both the SVM linear and the GEP models, were obtained for the hiking trail width where the GEP was predicted with the highest errors, representing more difficulty to apply it for very high-dimensional data. The results of this paper represent that four factors have better predictive ability, namely slope length, soil type, NDVI, and landform (topographic position). Other factors can be retained in the model because of the correlation, which can be formulated nonlinearly from way of the artificial neural network (ANN) model. For the study area, it can be seen from the hiking trail width map as a degradation index that the southwest part of the study area is highly prone to mine degradation due to slope, landform, climate, and vegetation conditions. Often, hiking trails are constructed to provide a path into natural areas that notably serve as a guide for hikers while also providing a sense of security and civilization. Surface erosion through trails that are already established can make a financial cost to operational organizations. Analysis of the results from the study area suggests that designing trails by Least Cost Path Analysis in GIS is beneficial and park managers can use it to minimize disruption of delicate ecosystems such as arid environments.

Declarations

Author Contribution

All authors contributed to the study conception and design. Hamed Adab and Zahra Ghelichipour carried out fieldwork. The manuscript was written and edited by Azadeh Atabati. All authors read and approved the final manuscript.

Data Availability

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

References

1. Adab, H., 2017. Landfire hazard assessment in the Caspian Hyrcanian forest ecoregion with the long-term MODIS active fire data. *Natural Hazards*: 1-19.

2. Aarnio MA, Kukkonen J, Kangas L, Kauhaniemi M, Kousa A, Hendriks C, Yli-Tuomi T, Lanki T, Hoek G, Brunekreef B, Elolähde T, Karppinen A (2018) A Model Evaluation Strategy Applied to Modelling of PM in the Helsinki Metropolitan Area. In: Mensink C, Kallos G (eds), *Air Pollution Modeling and its Application XXV*. Springer International Publishing, Cham. pp 103-109
3. Adab H (2017) Landfire hazard assessment in the Caspian Hyrcanian forest ecoregion with the long-term MODIS active fire data. *Natural Hazards* 87:1807-1825. doi:10.1007/s11069-017-2850-2
4. Adab H, Morbidelli R, Saltalippi C, Moradian M, Ghalhari GAF (2020) Machine Learning to Estimate Surface Soil Moisture from Remote Sensing Data. *Water* 12:3223.
5. Albert G, Sárközy Z (2021) Route planning on orienteering maps with least-cost path analysis. *Proc Int Cartogr Assoc* 4:4. doi:10.5194/ica-proc-4-4-2021
6. Ballantyne M, Gudes O, Pickering CM (2014) Recreational trails are an important cause of fragmentation in endangered urban forests: A case-study from Australia. *Landscape and Urban Planning* 130:112-124. doi:https://doi.org/10.1016/j.landurbplan.2014.07.004
7. Ballantyne M, Pickering CM (2015) Differences in the impacts of formal and informal recreational trails on urban forest loss and tree structure. *Journal of Environmental Management* 159:94-105. doi:https://doi.org/10.1016/j.jenvman.2015.05.007
8. Barros A, Marina Pickering CJEM (2017) How Networks of Informal Trails Cause Landscape Level Damage to Vegetation. 60:57-68. doi:10.1007/s00267-017-0865-9
9. Beucher A, Adhikari K, Breuning-Madsen H, Greve MB, Österholm P, Fröjdö S, Jensen NH, Greve MH (2017) Mapping potential acid sulfate soils in Denmark using legacy data and LiDAR-based derivatives. *Geoderma* 308:363-372. doi:https://doi.org/10.1016/j.geoderma.2016.06.001
10. Bryan RB (1977) The Influence of Soil Properties on Degradation of Mountain Hiking Trails at Grövelsjön. *Geografiska Annaler: Series A, Physical Geography* 59:49-65. doi:10.1080/04353676.1977.11879946
11. Chatterjea K (2007) Assessment and demarcation of trail degradation in a nature reserve, using GIS: case of Bukit Timah Nature Reserve. *Land Degradation & Development* 18:500-518. doi:https://doi.org/10.1002/ldr.793
12. Chen Y, Xu P, Chu Y, Li W, Wu Y, Ni L, Bao Y, Wang K (2017) Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. *Applied Energy* 195:659-670. doi:https://doi.org/10.1016/j.apenergy.2017.03.034
13. Clius M, Teleucă A, David O, Moroşanu A (2012) Trail accessibility as a tool for sustainable management of protected areas: Case study Ceahlău National Park, Romania. *Procedia Environmental Sciences* 14:267-278.
14. Cole DN (2004) Impacts of hiking and camping on soils and vegetation: a review. *Environmental impacts of ecotourism* 41:60.
15. de Lima RM, Osis R, de Queiroz AR, Santos AHM (2016) Least-cost path analysis and multi-criteria assessment for routing electricity transmission lines. *IET Generation, Transmission & Distribution* 10:4222-4230. doi:https://doi.org/10.1049/iet-gtd.2016.1119

16. Dixon G, Hawes M, McPherson G (2004) Monitoring and modelling walking track impacts in the Tasmanian Wilderness World Heritage Area, Australia. *Journal of Environmental Management* 71:305-320. doi:<https://doi.org/10.1016/j.jenvman.2004.02.006>
17. Djamali M, Brewer S, Breckle SW, Jackson ST (2012) Climatic determinism in phytogeographic regionalization: A test from the Irano-Turanian region, SW and Central Asia. *Flora - Morphology, Distribution, Functional Ecology of Plants* 207:237-249. doi:<https://doi.org/10.1016/j.flora.2012.01.009>
18. Dolan LMJ, et al., BOHEMEN H, WHELAN P, AKBAR KF, O'MALLEY V, O'LEARY G, KEIZER PJ (2006) Towards the sustainable development of modern road ecosystems. In: Davenport J, Davenport JL (eds). *The Ecology of Transportation: Managing Mobility for the Environment*. Springer Netherlands, Dordrecht. pp 275-331
19. Eagles PFJ (2014) Research priorities in park tourism. *Journal of Sustainable Tourism* 22:528-549. doi:10.1080/09669582.2013.785554
20. Ermini L, Catani F, Casagli N (2005) Artificial Neural Networks applied to landslide susceptibility assessment. *Geomorphology* 66:327-343. doi:<https://doi.org/10.1016/j.geomorph.2004.09.025>
21. Farrell TA, Marion JL (2001a) Identifying and assessing ecotourism visitor impacts at eight protected areas in Costa Rica and Belize. *Environmental Conservation* 28:215-225.
22. Farrell TA, Marion JL (2001b) Trail Impacts and Trail Impact Management Related to Visitation at Torres del Paine National Park, Chile. *Leisure/Loisir* 26:31-59. doi:10.1080/14927713.2001.9649928
23. Ferreira C (2002) Gene Expression Programming in Problem Solving. In: Roy R, Köppen M, Ovaska S, Furuhashi T, Hoffmann F (eds). *Soft Computing and Industry: Recent Applications*. Springer London, London. pp 635-653
24. Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*:n/a-n/a. doi:<http://dx.doi.org/10.1002/joc.5086>
25. Grimwade R, Horner B, Everhart GS (2009) Trail Design Guidelines for Portland's Park System, Portland Parks. *Portland Parks & Recreation*, Portland, Oregon. p 80
26. Hammitt WE, Cole DN, Monz CA (2015) *Wildland recreation: ecology and management*. John Wiley & Sons
27. Hill W, Pickering CM (2006) Vegetation associated with different walking track types in the Kosciuszko alpine area, Australia. *Journal of Environmental Management* 78:24-34. doi:<https://doi.org/10.1016/j.jenvman.2005.04.007>
28. Huang G, Wang D, Du Y, Zhang Q, Bai Z, Wang C (2022) Deformation Feature Extraction for GNSS Landslide Monitoring Series Based on Robust Adaptive Sliding-Window Algorithm. *Frontiers in Earth Science* 10. doi:10.3389/feart.2022.884500
29. Iqbal M, Elbaz K, Zhang D, Hu L, Jalal FE (2022) Prediction of residual tensile strength of glass fiber reinforced polymer bars in harsh alkaline concrete environment using fuzzy metaheuristic models. *Journal of Ocean Engineering and Science*. doi:<https://doi.org/10.1016/j.joes.2022.03.011>

30. Jacinth Jennifer J, Saravanan S (2022) Artificial neural network and sensitivity analysis in the landslide susceptibility mapping of Idukki district, India. *Geocarto International* 37:5693-5715. doi:10.1080/10106049.2021.1923831
31. Jalal FE, Xu Y, Iqbal M, Javed MF, Jamhiri B (2021) Predictive modeling of swell-strength of expansive soils using artificial intelligence approaches: ANN, ANFIS and GEP. *Journal of Environmental Management* 289:112420. doi:https://doi.org/10.1016/j.jenvman.2021.112420
32. Kardani N, Bardhan A, Samui P, Nazem M, Zhou A, Armaghani DJ (2022) A novel technique based on the improved firefly algorithm coupled with extreme learning machine (ELM-IFF) for predicting the thermal conductivity of soil. *Engineering with Computers* 38:3321-3340.
33. Khan K, Jalal FE, Khan MA, Salami BA, Amin MN, Alabdullah AA, Samiullah Q, Arab AM, Faraz MI, Iqbal M (2022) Prediction Models for Evaluating Resilient Modulus of Stabilized Aggregate Bases in Wet and Dry Alternating Environments: ANN and GEP Approaches. *Materials*,
34. Khazaee Fadafan F, Soffianian A, Pourmanafi S, Morgan M (2022) Assessing ecotourism in a mountainous landscape using GIS – MCDA approaches. *Applied Geography* 147:102743. doi:https://doi.org/10.1016/j.apgeog.2022.102743
35. Kişi Ö (2009) Daily pan evaporation modelling using multi-layer perceptrons and radial basis neural networks. *Hydrological Processes* 23:213-223. doi:https://doi.org/10.1002/hyp.7126
36. Kotrlik J, Higgins C (2001) Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research. *Information technology, learning, and performance journal* 19:43.
37. Kumar P, Lai SH, Mohd NS, Kamal MR, Ahmed AN, Sherif M, Sefelnasr A, El-shafie A (2021) Enhancement of nitrogen prediction accuracy through a new hybrid model using ant colony optimization and an Elman neural network. *Engineering Applications of Computational Fluid Mechanics* 15:1843-1867. doi:10.1080/19942060.2021.1990134
38. Kusch E, Davy R (2022) KrigR—a tool for downloading and statistically downscaling climate reanalysis data. *Environmental Research Letters* 17:024005. doi:10.1088/1748-9326/ac48b3
39. Leung Y-F, Marion J (1996a) Trail degradation as influenced by environmental factors: A state-of-knowledge review. *Journal of Soil and Water Conservation* 51:130-136.
40. Leung Y-F, Marion JL (1996b) Trail degradation as influenced by environmental factors: A state-of-the-knowledge review. *Journal of Soil and Water Conservation* 51:130-136.
41. Leung Y-F, Marion JL (2000) Recreation impacts and management in wilderness: a state-of-knowledge review. *Wilderness science in a time of change conference, USA*. pp 23-48
42. Li W, Ge X, Liu C (2005) Hiking Trails and Tourism Impact Assessment In Protected Area: Jiuzhaigou Biosphere Reserve, China. *Environmental Monitoring and Assessment* 108:279-293. doi:10.1007/s10661-005-4327-0
43. Ling Y, Wang K, Wang X, Li W (2021) Prediction of engineering properties of fly ash-based geopolymer using artificial neural networks. *Neural Computing and Applications* 33:85-105. doi:10.1007/s00521-019-04662-3

44. Magdalena Warter M, Peeters M, Kuppen E, Blok K, Dilly L (2017) Estimating soil erosion on hiking trails in the Sierra Mariola Natural Park in southern Spain. EGU General Assembly Conference Abstracts. p 81
45. Marion J, Wimpey J (2007) Environmental impacts of mountain biking: science review and best practices. International Mountain Bicycling Association
46. Marion JL, Olive N (2006) Assessing and Understanding Trail Degradation: Results from Big South Fork National River and Recreational Area. Reston, VA. p 80
47. Marion JL, Wimpey J (2017) Assessing the influence of sustainable trail design and maintenance on soil loss. *Journal of Environmental Management* 189:46-57.
doi:<https://doi.org/10.1016/j.jenvman.2016.11.074>
48. Matulewski P, Buchwal A, Zielonka A, Wrońska-Walach D, Čufar K, Gärtner H (2021) Trampling as a major ecological factor affecting the radial growth and wood anatomy of Scots pine (*Pinus sylvestris* L.) roots on a hiking trail. *Ecological Indicators* 121:107095.
doi:<https://doi.org/10.1016/j.ecolind.2020.107095>
49. Meadema F, Marion JL, Arredondo J, Wimpey J (2020) The influence of layout on Appalachian Trail soil loss, widening, and muddiness: Implications for sustainable trail design and management. *Journal of Environmental Management* 257:109986.
doi:<https://doi.org/10.1016/j.jenvman.2019.109986>
50. Merghadi A, Yunus AP, Dou J, Whiteley J, ThaiPham B, Bui DT, Avtar R, Abderrahmane B (2020) Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews* 207:103225.
doi:<https://doi.org/10.1016/j.earscirev.2020.103225>
51. Meshram SG, Abdollahi S, Ildoromi AR, Salmanmahini A, Fakheran S (2022) Assessing and mapping recreation value as an ecosystem service in central part of Esfahan Province, Iran. *Arabian Journal of Geosciences* 15:821. doi:10.1007/s12517-022-10039-z
52. Mohammadi A, Lunnon C, Moll RJ, Tan CKW, Hobeali K, Behnoud P, Moghadas P, Macdonald DW, Farhadinia MS (2021) Contrasting responses of large carnivores to land use management across an Asian montane landscape in Iran. *Biodiversity and Conservation* 30:4023-4037. doi:10.1007/s10531-021-02290-9
53. Mohanty M, Sinha NK, Painuli DK, Bandyopadhyay KK, Hati KM, Sammi Reddy K, Chaudhary RS (2015) Modelling Soil Water Contents at Field Capacity and Permanent Wilting Point Using Artificial Neural Network for Indian Soils. *National Academy Science Letters* 38:373-377.
doi:10.1007/s40009-015-0358-4
54. Nepal SK, Way P (2007) Characterizing and comparing backcountry trail conditions in Mount Robson Provincial Park, Canada. *Ambio* 36:394-400. doi:10.1579/0044-7447(2007)36[394:cacbtc]2.0.co;2
55. Newsome D, Moore SA, Dowling RK (2012) Natural area tourism: Ecology, impacts and management. Channel view publications

56. Ng S-L, Leung Y-F, Cheung S-Y, Fang W (2018) Land degradation effects initiated by trail running events in an urban protected area of Hong Kong. *Land Degradation & Development* 29:422-432. doi:<https://doi.org/10.1002/ldr.2863>
57. Ólafsdóttir R, Runnström MC (2013) Assessing hiking trails condition in two popular tourist destinations in the Icelandic highlands. *Journal of Outdoor Recreation and Tourism* 3-4:57-67. doi:<https://doi.org/10.1016/j.jort.2013.09.004>
58. Olive ND, Marion JL (2009) The influence of use-related, environmental, and managerial factors on soil loss from recreational trails. *Journal of Environmental Management* 90:1483-1493. doi:<https://doi.org/10.1016/j.jenvman.2008.10.004>
59. Orland E, Roering JJ, Thomas MA, Mirus BB (2020) Deep Learning as a Tool to Forecast Hydrologic Response for Landslide-Prone Hillslopes. *Geophysical Research Letters* 47:e2020GL088731. doi:<https://doi.org/10.1029/2020GL088731>
60. Palani S, Liong S-Y, Tkalic P (2008) An ANN application for water quality forecasting. *Marine Pollution Bulletin* 56:1586-1597. doi:<https://doi.org/10.1016/j.marpolbul.2008.05.021>
61. Panagos P, Borrelli P, Meusburger K (2015) A New European Slope Length and Steepness Factor (LS-Factor) for Modeling Soil Erosion by Water. *Geosciences* 5:117-126.
62. Pannakkong W, Thiwa-Anont K, Singthong K, Parthanadee P, Buddhakulsomsiri J (2022) Hyperparameter Tuning of Machine Learning Algorithms Using Response Surface Methodology: A Case Study of ANN, SVM, and DBN. *Mathematical Problems in Engineering* 2022:8513719. doi:10.1155/2022/8513719
63. Raduła MW, Szymura TH, Szymura M (2018) Topographic wetness index explains soil moisture better than bioindication with Ellenberg's indicator values. *Ecological indicators* 85:172-179.
64. Rahchamani N, Ezazi A, Ghahremaninejad F (2014) The flora of Sarigol National Park, Northern Khorassan Province, Iran. *Journal of Biodiversity and Environmental Sciences (JBES)* 4:278-307.
65. Randall M, Newsome D (2008) Assessment, evaluation and a comparison of planned and unplanned walk trails in coastal south-western Australia. *Conservation Science Western Australia* 7:19-34.
66. Sahani N, Ghosh T (2021) GIS-based spatial prediction of recreational trail susceptibility in protected area of Sikkim Himalaya using logistic regression, decision tree and random forest model. *Ecological Informatics* 64:101352. doi:<https://doi.org/10.1016/j.ecoinf.2021.101352>
67. Santarém F, Silva R, Santos P (2015) Assessing ecotourism potential of hiking trails: A framework to incorporate ecological and cultural features and seasonality. *Tourism Management Perspectives* 16:190-206. doi:<https://doi.org/10.1016/j.tmp.2015.07.019>
68. Santiago LE, Gonzalez-Caban A, Loomis J (2008) A Model for Predicting Daily Peak Visitation and Implications for Recreation Management and Water Quality: Evidence from Two Rivers in Puerto Rico. *Environmental Management* 41:904-914. doi:10.1007/s00267-008-9079-5
69. Shin J, Kim HJ, Kim Y (2011) Adaptive support vector regression for UAV flight control. *Neural Netw* 24:109–120. doi:10.1016/j.neunet.2010.09.011

70. Sitzia T, Rizzi A, Cattaneo D, Semenzato P (2014) Designing recreational trails in a forest dune habitat using least-cost path analysis at the resolution of visitor sight distance. *Urban Forestry & Urban Greening* 13:861-868. doi:<https://doi.org/10.1016/j.ufug.2014.09.011>
71. Spornbauer BS, Monz C, D'Antonio A, Smith JW (2023) Factors influencing informal trail conditions: Implications for management and research in Urban-Proximate parks and protected areas. *Landscape and Urban Planning* 231:104661. doi:<https://doi.org/10.1016/j.landurbplan.2022.104661>
72. Sutherland RA, Bussen JO, Plondke DL, Evans BM, Ziegler AD (2001) Hydrophysical degradation associated with hiking-trail use: a case study of Hawai'i Iloa Ridge Trail, O'ahu, Hawai'i. *Land Degradation & Development* 12:71-86. doi:<https://doi.org/10.1002/ldr.425>
73. Tachikawa T, Hato M, Kaku M, Iwasaki A (2011) Characteristics of ASTER GDEM version 2. *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*. IEEE. pp 3657-3660
74. Talebi M, Majnounian B, Makhdoum M, Abdi E, Omid M, Marchi E, Laschi A (2019) A GIS-MCDM-based road network planning for tourism development and management in Arasbaran forest, Iran. *Environmental Monitoring and Assessment* 191:647. doi:10.1007/s10661-019-7831-3
75. Talebi M, Nickabadi A, Majnounian B, Abdi E, Safabakhsh R, Izadyar N, Laschi A (2022) Forest road planning to improve tourism accessibility: a comparison of different methods applied in a real case study. *Geocarto International*:1-20. doi:10.1080/10106049.2022.2032389
76. Taye MM (2023) Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers*,
77. Tomczyk AM (2011) A GIS assessment and modelling of environmental sensitivity of recreational trails: The case of Gorce National Park, Poland. *Applied Geography* 31:339-351. doi:<https://doi.org/10.1016/j.apgeog.2010.07.006>
78. Tomczyk AM, Ewertowski M (2011) Degradation of recreational trails, Gorce National Park, Poland. *Journal of Maps* 7:507-518. doi:10.4113/jom.2011.1195
79. Tomczyk AM, Ewertowski M (2013) Planning of recreational trails in protected areas: Application of regression tree analysis and geographic information systems. *Applied Geography* 40:129-139. doi:<https://doi.org/10.1016/j.apgeog.2013.02.004>
80. Törn A, Tolvanen A, Norokorpi Y, Tervo R, Siikamäki P (2009) Comparing the impacts of hiking, skiing and horse riding on trail and vegetation in different types of forest. *Journal of Environmental Management* 90:1427-1434. doi:<https://doi.org/10.1016/j.jenvman.2008.08.014>
81. Wimpey JF, Marion JL (2010) The influence of use, environmental and managerial factors on the width of recreational trails. *Journal of Environmental Management* 91:2028-2037. doi:<https://doi.org/10.1016/j.jenvman.2010.05.017>
82. Yesilnacar E, Topal T (2005) Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology* 79:251-266. doi:<https://doi.org/10.1016/j.enggeo.2005.02.002>
83. Yu C, Lee JAY, Munro-Stasiuk MJ (2003) Research Article: Extensions to least-cost path algorithms for roadway planning. *International Journal of Geographical Information Science* 17:361-376.

Figures

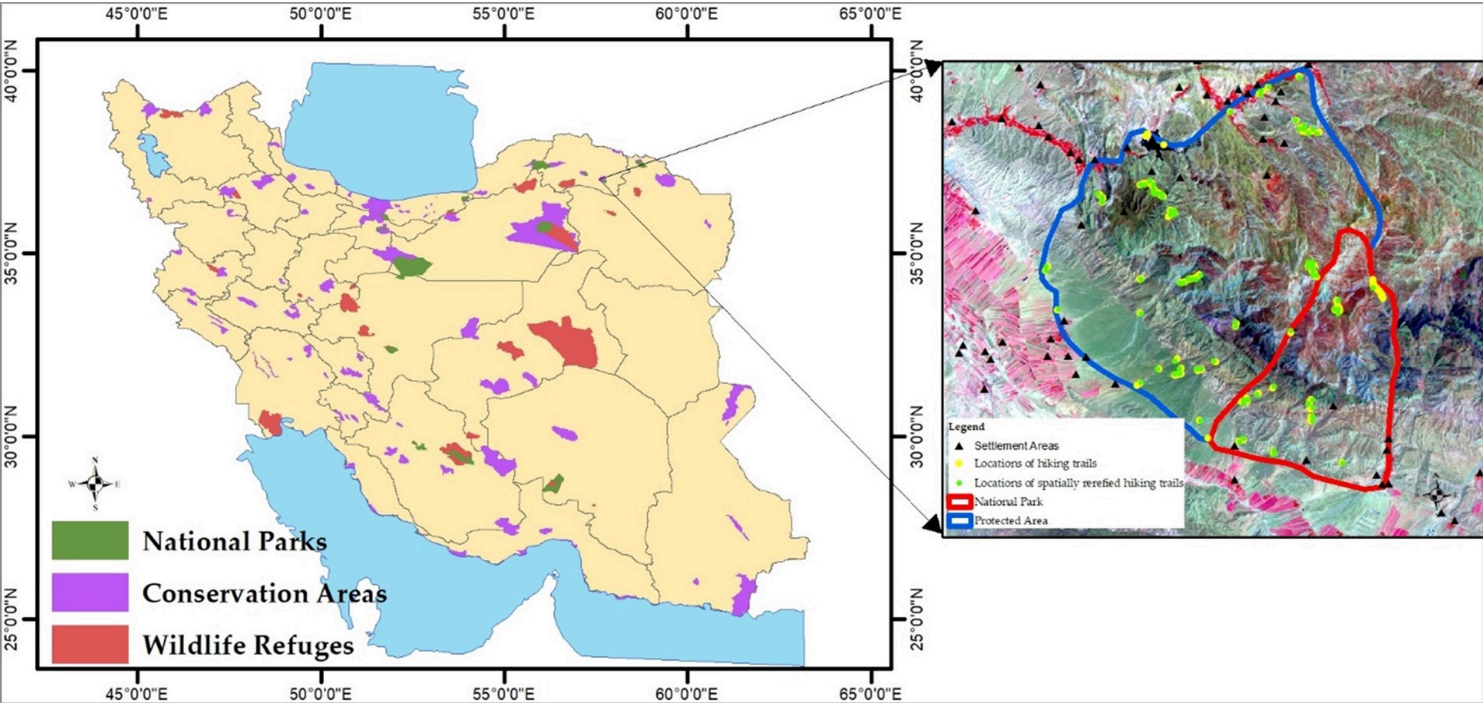


Figure 1

a) The study area in Northeast Iran, the Sarigol National Park and Protected Area (SNPP) b) hiking trails provided by Google Earth

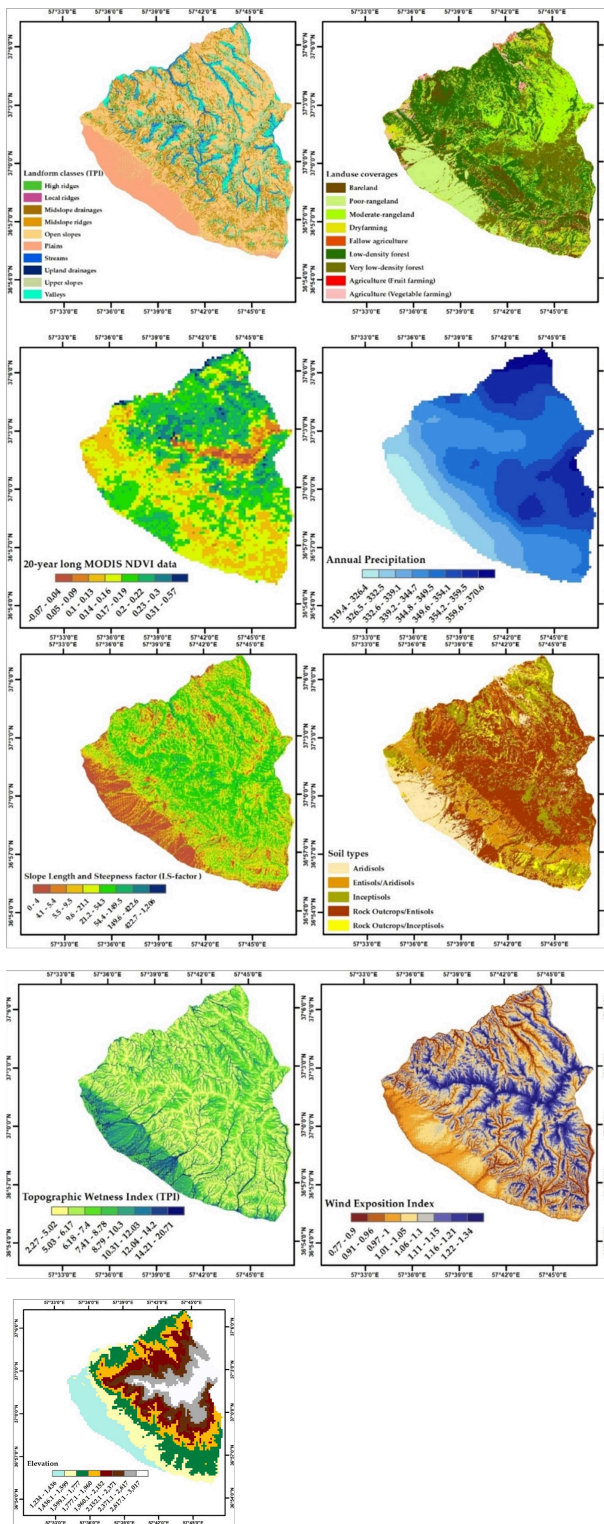


Figure 2

The spatial patterns of influencing factors on the width of hiking trails. (a): Land use coverages; (b): Landform classes (TPI); (c): Annual precipitation; (d): MODIS NDVI data; (e): Soil types; (f): LS-factor (The pixel size of data is same as original data). The spatial patterns of influencing factors on the width of hiking trails. (g): Wind explosion index; (h): Topographic witness index, (i): Elevation (The pixel size of data is same as original data).

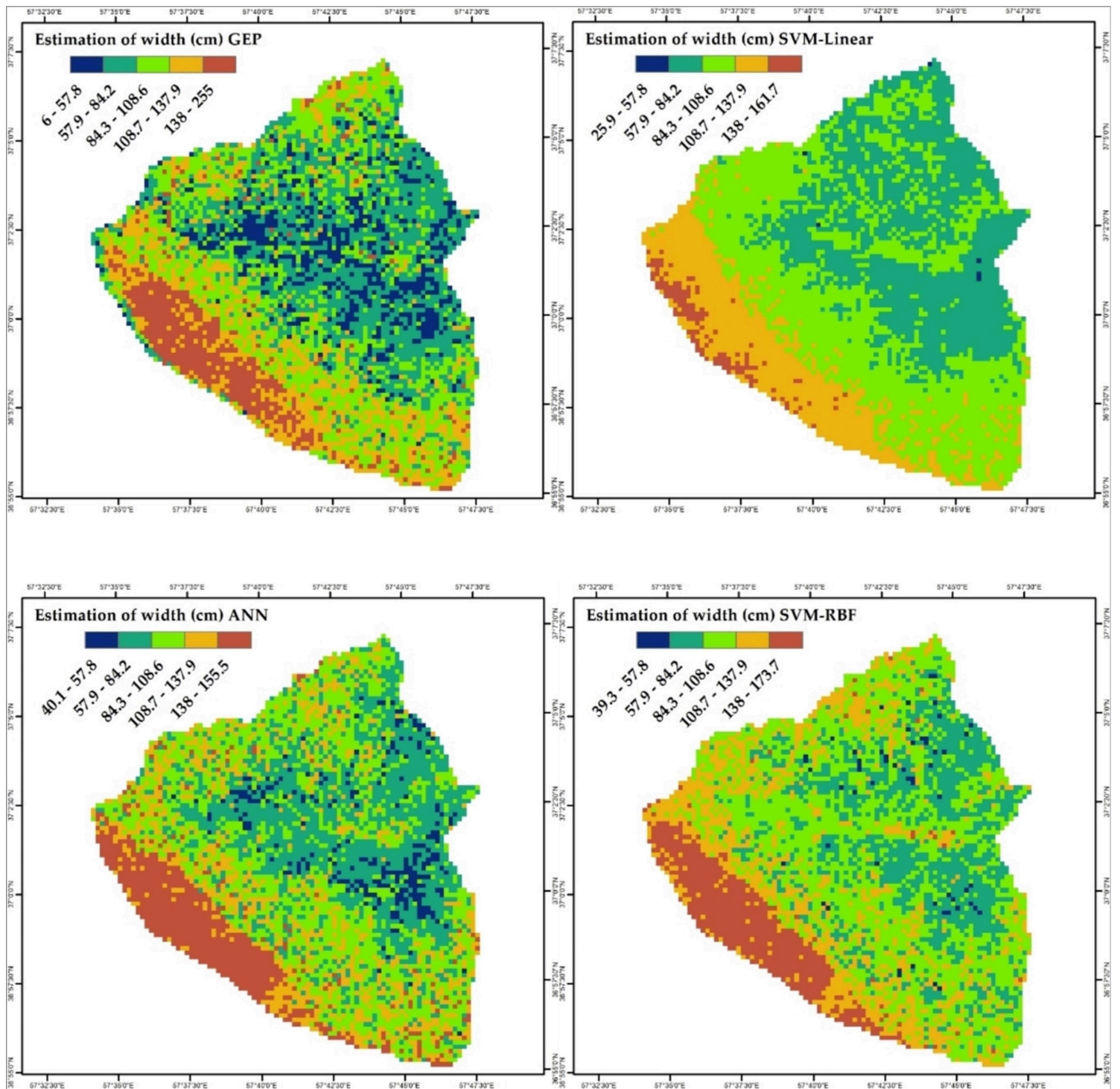


Figure 3

The estimation of hiking trail width as degradation index of SNPP derived through machine learning models

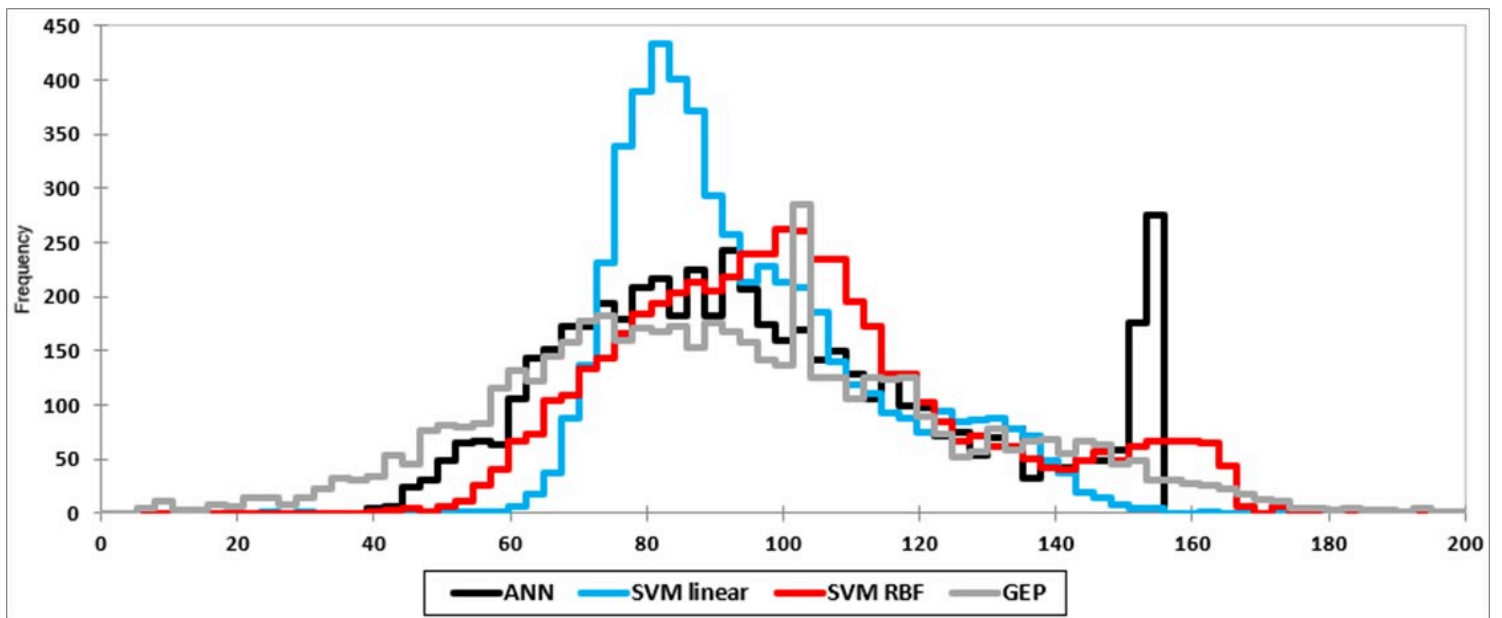


Figure 4

Histograms of the spatial distribution of the estimation of hiking trail width pixels as trail degradation index of SNPP derived through machine learning models

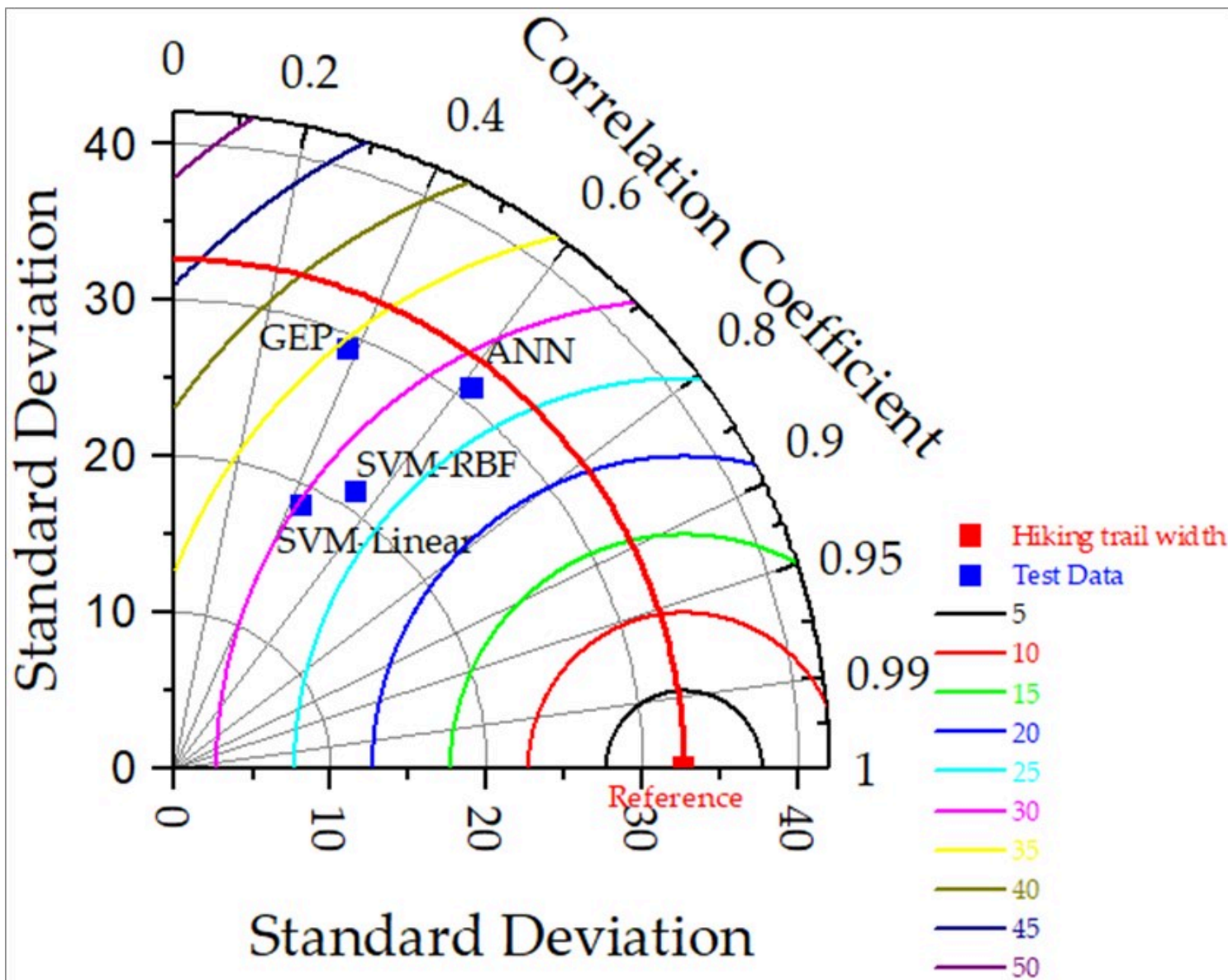


Figure 5

Performances of different models with Taylor diagram. The color line shows RMS

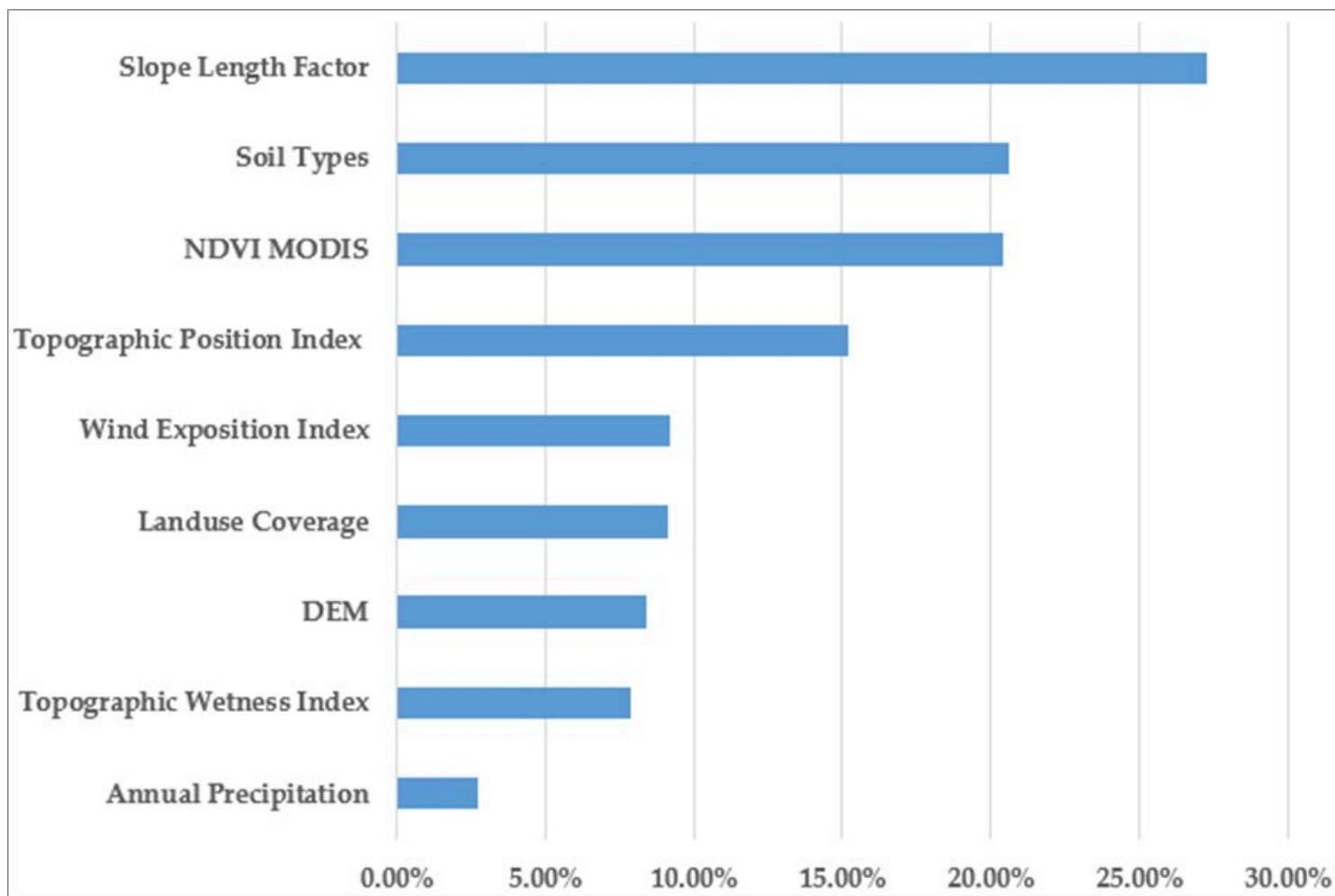


Figure 6

Importance of predictor variables as identified by the ANN model. The total effect of the predicting factors, with a higher value showing a greater significance attributed to the model.



Figure 7

The existing hiking trail was used to identify the start point and destination point of the study area