Geomechanics and Engineering, *Vol. 34, No. 3 (2023) 267-284*
https://doi.org/10.12989/gae.2023.34.3.267

Landslide risk zoning using support vector machine algorithm

Vahed Ghiasi*[1], Nur Irfah Mohd Pauzi[2a], Shahab Karimi[3b] and Mahyar Yousefi[4]

[1]Faculty of Civil Engineering, Malayer University, Malayer, Iran
[2]Department of Civil Engineering, College of Engineering, Institute of Energy Infrastructure, Universiti Tenaga Nasional, Malaysia
[3]Department of Civil Engineering, Faculty of Engineering, Malayer University, Malayer, Iran
[4]Faculty of Engineering, Malayer University, Malayer, Iran

(*Received August 8, 2022, Revised June 20, 2023, Accepted June 27, 2023*)

Abstract. Landslides are one of the most dangerous phenomena and natural disasters. Landslides cause many human and financial losses in most parts of the world, especially in mountainous areas. Due to the climatic conditions and topography, people in the northern and western regions of Iran live with the risk of landslides. One of the measures that can effectively reduce the possible risks of landslides and their crisis management is to identify potential areas prone to landslides through multi-criteria modeling approach. This research aims to model landslide potential area in the Oshvand watershed using a support vector machine algorithm. For this purpose, evidence maps of seven effective factors in the occurrence of landslides namely slope, slope direction, height, distance from the fault, the density of waterways, rainfall, and geology, were prepared. The maps were generated and weighted using the continuous fuzzification method and logistic functions, resulting values in zero and one range as weights. The weighted maps were then combined using the support vector machine algorithm. For the training and testing of the machine, 81 slippery ground points and 81 non-sliding points were used. Modeling procedure was done using four linear, polynomial, Gaussian, and sigmoid kernels. The efficiency of each model was compared using the area under the receiver operating characteristic curve; the root means square error, and the correlation coefficient . Finally, the landslide potential model that was obtained using Gaussian's kernel was selected as the best one for susceptibility of landslides in the Oshvand watershed.

Keywords: kernel functions; landslide; logistic function; mapping; prediction rate-area diagram; support vector machine

1. Introduction

Natural events are complex processes that affect all parts of the planet. In the meantime, landslides as one the natural hazards always occurring all over the world and are of great importance, so landslides have been introduced as one of the most important natural hazards in the world (Khalifa *et al*. 2020, Malamud *et al*. 2004). Landslide, as one of the mass movements, is the movement of a mass of rock and soil debris on the slope downwards under the influence of gravity. This phenomenon occurs when the shear stress on the slope exceeds the shear strength of the slope's materials (Cruden and Varnes 1996).

Mass movements and their consequences every year in most countries cause economic damage to roads, railway lines, power and communication lines, water and water supply channels, oil and gas extraction, and refining facilities and factories. Industrial centers and residential areas, natural and artificial dams and lakes, as well as the destruction of pastures and agricultural lands, accelerates erosion and the widespread transfer of sediments behind the dams. Considering these issues, it can be safely said that in construction projects such as choosing the route for the construction of highways and main and minor mountain roads, choosing the location for the construction of earthen and concrete dams, water transfer channels and tunnels, and projects such as development Forests and natural pastures and any mining development, the study of the stability of slopes is one of the most sensitive and important issues (Naderpour *et al*. 2011).

This phenomenon can be caused by many geological, geomorphological, hydrological, biological, and human factors. However, the most important factors driving landslides include heavy rainfall, rapid melting of snow, sudden changes in the underground water level, landslides and high-speed erosion (Sidle and Ochiai 2006), climatic conditions, geological features, topography, vegetation, or a combination of these. are factors (Ost *et al*. 2003). In active tectonic areas, the instability of slopes can be observed in various forms of landslides, landslides, falls, slope ruptures, creep, etc. Most of the large and catastrophic landslides on a global scale have occurred in mountainous areas and seismic belts with active faults (Abedini 2015).

In the literature related to assessing the earth's stability, it is often said that the past and the present are the keys to the future, which means that the future can be predicted based on the past and the present (Pavel *et al*. 2011). Landslide susceptibility is the tendency of soil or rock to produce many types of landslides. A landslide susceptibility map that combines some of the important factors that have contributed to the occurrence of landslides in the past shows the areas with the potential to occur in the future (Chalkias *et al.* 2014). Such maps are considered a useful and reliable

*Corresponding author, Assistant Professor
E-mail: v.ghiasi@malayeru.ac.ir
[a]Assistant Professor
[b]Master

Copyright © 2023 Techno-Press, Ltd.
http://www.techno-press.org/?journal=gae&subpage=7 ISSN: 2005-307X (Print), 2092-6219 (Online)

tool for assessing the potential of landslide risk and classifying an area according to its stability and instability potential and based on the number of preliminary factors such as lithology, geomorphology, and stimulating factors such as seismicity and rainfall. etc. are classified (Chauhan *et al*. 2010).

The purpose of zoning is to divide the land surface into homogeneous areas and classify them according to the amount of potential landslide risk. In other words, the earth's surface is divided into special and separate areas of actual or potential degrees of danger based on the effective factors in creating landslides (Varnes 1984).

Various approaches have been developed to study landslides, which can be divided into four categories: analysis of landslide inventory, exploratory or index method, statistical approach, and geotechnical method. Statistical approaches are usually considered for the zoning and mapping landslides in large and complex areas. For example, we can refer to multivariate regression analysis's defining analytical approach. Also, linking methods have been developed by combining two statistical methods and an artificial neural network (Varnes 1984).

Linked methods include artificial neural network method - fuzzy logic, Bayes analysis - artificial neural network, and decision tree analysis - Bayesian. But the method based on the artificial neural network cannot provide objective and fixed output because its results depend on the operator (Yao 2008). The logistic regression method has been developed (Ermini *et al*. 2005).

The problem that exists regarding landslides is how to identify areas with the potential risk of this phenomenon, and a method should be considered that can be used to study these areas with a logical and reasonable procedure. This regard introduces various factors such as slope, topography, soil type, the presence of fractures and faults, drainage and waterways, the amount of rainfall and density of waterways, vegetation, underground water level, and other conditions that are effective in causing landslides. In this regard, measuring the relative importance of each of the above factors and ranking the studied area, as well as simultaneously examining the impact of the above factors on the occurrence of landslides, is a fundamental issue. For this purpose, researchers have used geographic information system (GIS) capabilities. Therefore, various methods have been developed to produce natural resources and natural hazards, e.g., landslide, potential model (e.g., Yousefi and Carranza 2015a, b, Ghiasi *et al*. 2021, Yousefi *et al*. 2019, 2021), dividing into some general categories data-oriented (supervised) and knowledge-oriented (unsupervised) methods (Yousefi *et al*. 2021).

Iran has experienced numerous landslides that is due to its geological characteristics, seismicity, rainfall, weather conditions, amount of mountainous regions, and diverse topography.

On the other hand, the increasing growth of the country's population has led to increased construction, industrial and agricultural activities in steep areas. The encroachment of humans on the natural environment, the change of land use on slopes, and the development of residential areas and similar activities on steep slopes have increased the damage caused by this phenomenon. Considering that two-thirds of the area of our country consists of mountains and hills, having such natural conditions, and knowing that landslides occur in steep areas, it can be concluded that Iran is one of the countries prone to landslides. The distribution of landslides in the country shows that most landslides are located in the border areas of the Alborz and Zagros highlands. According to preliminary estimates, about 500 billion rials of financial damage are caused by landslides in Iran every year (Hossein Zadeh *et al.* 2010). Therefore, it is necessary to take necessary measures to reduce the damages caused by landslides, identify areas prone to landslides, determine the factors affecting the occurrence of landslides, and prepare a map of the susceptibility of landslides. Preparing a landslide susceptibility map is a basic tool for crisis management activities in mountainous areas (Dahal *et al.* 2010).

In Iran, many research has been done about landslides using models. Among these models, we can use the weighted linear combination model in zoning the potential of landslides in the region of Serkhon, (Karam 2004), fuzzy logic (Moradi *et al*. 2010), artificial neural network, the value of information and hierarchical analysis (Ghiasi *et al*. 2020) hierarchical analysis (Ghanavati 2011), logistic regression, correlation equation and hierarchical analysis (Mosfaei *et al*. 2001), multivariate regression and geographic information system, Fuzzy TOPSIS, and geographic information system (Hashemi *et al*. 2010).

Among the statistical methods that have attracted attention in the last few years is the Support Vector Machine (SVM) algorithm. Based on the statistical learning theory, this model is used to evaluate and test a set of data (Yamani *et al*. 2011). The use of this model will increase the accuracy of the prediction of landslides on a regional scale. This model has attracted much attention recently due to its good classification performance, error tolerance, and appropriate generalization (Ahmadabadi and Rahmati 2014).

Pourghasmi *et al.* prepared a landslide susceptibility map in the Golestan province by using six linear, second-order, third-order, fourth-order, radial (RBF), and circular functions and selecting 14 control factors. The radial function prediction rate was introduced as the best landslide susceptibility evaluation algorithm in Golestan province (Pourghasemi *et al*. 2013). Yamani et al. used the algorithm of support vector machines in zoning the risk of landslides in the Derkeh catchment area. They are evaluated using circular, polynomial, radial, and linear functions in SVM algorithm and effective criteria in identifying areas sensitive to landslides, including distance (from fault, network-drainage), lithology, and slope (value, angle), and elevation level. The possibility of land subsidence in the watershed of Derkeh in the north of Tehran was discussed. The research results showed that the ring function has the best performance and the linear function has the lowest accuracy in terms of performance due to its greater compliance with reality. Also, based on the implemented functions, the factor of distance from the fault and waterway and the lithological condition play the most important role in the sensitivity to landslides in the watershed (Yamani *et al*. 2011, Pourghasemi *et al*. 2013). Ghasemian *et al*. evaluated landslide sensitivity using a support vector machine algorithm in Kamiyaran city, located in Kurdistan province.

In this research, the training and validation of the RBF function of the SVM algorithm were investigated by a spatial database with a total of twelve landslide factors. The obtained results showed that the RBF function of the SVM model has a good performance in evaluating the landslide susceptibility in the study area(Pourghasemi *et al*. 2013). In 2014, Peng *et al*. prepared a landslide susceptibility map based on Rough Set theory and vector machine algorithm in three valley regions in China. The results showed that 19.7% of the region has a medium and high sensitivity to the occurrence of landslides, and the combined model, which includes RS theory and SVM model, to predict the risk of landslides, has better prediction skills, and Reliability is higher (Peng *et al*. 2014). In research, Hong et al. predicted the risk of landslides using Logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM) algorithms in the Yehang region of China. And the validation of the used models was evaluated using the (ROC) index, and the results showed that the value of the area under the curve of the ROC index was 92.5% for the logistic regression model and 88.8% for the support vector model. And for the decision tree model, it is 95.7%. Also, the predictability of landslide risk for the logistic regression model is 81.1%, for the support vector machine model is 84.2%, and for the decision tree model is 93.3% (Hong *et al*. 2015). Tian Bai *et al.* evaluated landslide susceptibility using support vector machine models, tree logistics, artificial neural networks, and nuclear logistic regression in Vietnam. Their results show the advantage of machine learning models in modeling the susceptibility of landslides in the study area (Tien *et al*. 2015).

Karmi et al. used the support vector machine algorithm to zonate landslide susceptibility in Aharchai catchment area. The results of this research showed that the radial basis function (RBF) has a better performance than circular, polynomial, and linear functions in the zoning of landslide susceptibility in the region (Karimi *et al*. 2018).

Ahmadabadi and Rahmati investigated the use of quantitative geomorphometric indices in identifying landslide-prone areas using SVM model on Khorram Abad-Zal bridge freeway as one of the important communication ways in the country. The research results showed that the use of polynomial and linear functions of the SVM model in the zoning of the studied area by involving the geomorphometric factors of the domains was able to identify the points with high landslide potential. Also, the evaluation of the accuracy of the modeling showed that the polynomial function with an overall accuracy of 89% shows the landslide-prone areas better than the linear function, which seems to be due to the complex and non-linear behavior of the variables involved in the occurrence of landslides in the study area (Ahmadabadi and Rahmati 2014). Abedini *et al.* identified and classified different landslides with the object-oriented approach in the distance between Nasir Abad village and Sattar Khan Ahar dam.

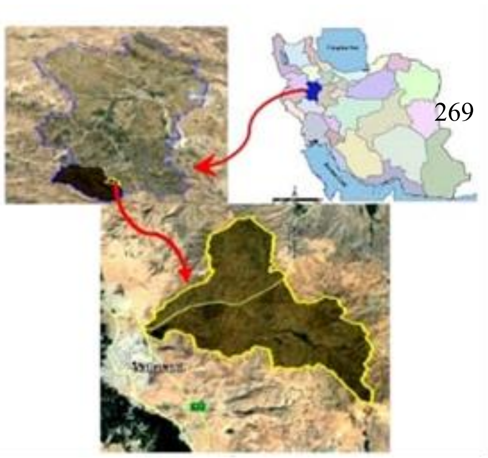

Fig. 1 Location of Oshvand watershed

This research showed that 34.02% of the area has a very high potential for landslide occurrence (Abedini *et al*. 2017). In a research, Lee *et al*. (2017) dealt with landslide susceptibility zoning using the radial function of the SVM model in two areas of Pyeongchang and Inje, located in Gangwon Province of Korea. The results of the studies showed the accuracy of the model implementation as 81.36% and 77.49%, respectively (Lee *et al*. 2017).

The purpose of this research is to identify the area with landslide potential in the study area (Oshvand watershed) and to collect data affecting landslides such as slope, slope direction, height, rainfall, the density of waterways, distance from the fault, the geology of the area, Analyzing the data and weighting them using the continuous fuzzy method, combining the weighted reference layers and producing the landslide potential model using the support vector machine and finally comparing the final obtained map with the real land maps. The landslide will be in the study area.

2. Materials and methods

2.1 Geographical scope of the region

Nahavand Plain, with a catchment area of 1902 square kilometers, is one of the plains of the upper basin of the Karkhe River. Oshvand watershed is located in Hamadan province and northeast of Nahavand city, which has an area of 47.16 square kilometers. The geographical coordinates are 48'51'22' to 48'53'33' east longitude and 34'51'1' to 15'34" 23" north latitude. The maximum and minimum height of this watershed is 2578.99 and 1709.65 meters from the sea level, respectively. In Fig. 1, the Oshvand watershed is shown (Hosseini 2015).

2.2 The climate of the region

The main source of atmospheric precipitation in the region is the Mediterranean air mass that causes precipitation when it hits the heights of the region. The minimum and maximum rainfall in the region are 431 and

482 mm, respectively, and the climate of the region is semi-humid, and the average temperature of the region is 8.8 degrees Celsius (regional water of Hamadan province).

2.3 Parameters affecting landslides

Many factors can play a role in the instability of the slopes, among them the slope of the slopes, the direction of the slope, the type, and composition of mineralogy and geological materials, the level of underground water, earthquakes, distance from the fault, land use, vibrations caused by the work of construction machinery or traffic, the presence of waterways, rain and snow, weathering cycle (wetting, drying, and dissolution), land subsidence, trenching and creation of pits (Kamranzad *et al*. 2013)

In this research, seven factors influencing the occurrence of landslides, including slope, height, rainfall, watercourse density, distance from the fault, the direction of slope, and geology in the region, have been selected. The geological map of the region was prepared on a scale of 1:100,000 by the Geological Organization of the country.

The map of the distance from the fault is made from the geological map. Also, the topographic map (DEM) of the area with a scale of 1:50000 has been used to prepare maps of height, slope, the direction of slope, and density of waterways. The rainfall map has also been received from the regional water department. A raster map of seven factors influencing the occurrence of landslides in Oshvand region was prepared in ArcGIS 10.4.1 software. After generating the raster maps of factors affecting landslides, the maps created are weighted by the logistic function. In this way, a weight between 0 and 1 is assigned to each of the criteria. And considering that this function has an increasing trend, the higher the numerical value of the criterion, the score is closer to (1), and the lower the numerical value of the criterion, the score is closer to (0).

In this research, the commands embedded in Matlab software are used to use the SVM classifier. In Matlab, in addition to providing a good guide to familiarizing with the concept of support vector machines, commands are also designed to work with SVM. Like any supervised learning model in Matlab, first, the support vector machine is trained by the training data, then the trained machine is used to classify (predict) the new data. In addition, in order to achieve proper prediction accuracy, different kernel functions can be used, and the parameters of these kernel functions can be adjusted.

2.3.1 Slope

The slope is one of the most important factors affecting the sliding of materials from the surface of the earth. Based on morphology, each study area may have different slopes [37]. As the slope increases, the gravity level also decreases. The shear stress caused by gravity in the alluvial and sedimentary soils increases and causes the domain to break more and more. Therefore, calmer slopes are less exposed to landslides (Dai *et al*. 2001). Although it is expected that the degree of instability in the soil will increase with the increase of the slope, the number of landslides will decrease. On very high slopes, due to the impossibility of soil formation and the instability of sediments, as well as rock outcrops and cliffs, the probability of landslides will decrease. Finds. Therefore, it is expected that the probability of landslides will increase in slopes between 10 and 40 degrees (Nojavan *et al*. 2018). The slope map of Oshvand watershed is according to Fig. 2.

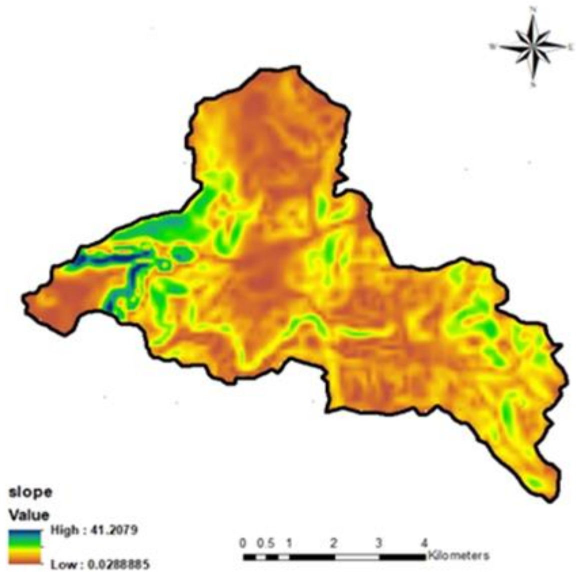

Fig. 2 Slope map

2.3.2 Slope direction

The direction of the slope shows the different effects of sunlight, hot and dry winds, and rainfall in different directions. Classification of the slope direction is done according to the presence of different factors in different directions of the slope of the domain and the difference in the process of expanding the domains. The southern slopes of the mountains receive more sunlight during the day than the northern slopes, and as a result, they have a drier climate, while the northern slopes are more humid and rainy. On the other hand, due to the general movement direction of winds from West to East, the amount of atmospheric precipitation on the western slopes is also higher than on the eastern slopes (Kamranzad *et al*. 2013). Also, in the areas where there is a big difference in the temperature of the slopes facing and facing the sun during different months of the year, this temperature fluctuation takes place in the range above and below the freezing point of water (especially in the snow-covered slopes). A significant relationship is established between the frequency distribution of domain instabilities and the direction of the domain (Nojavan *et al*. 2018).

Slope direction map for the studied area into nine classes North (0-5/22), Northeast (22/5-67), East (112-5/67), Southeast (112-5-157), South (5/202 - 5/157), Southwest (5/247 - 5/202), West (5/292 - 5/247), Northwest (5/337 - 5/292) and North (337-5-360) was classified. The map of the slope direction of Oshvand watershed is according to Fig. 3.

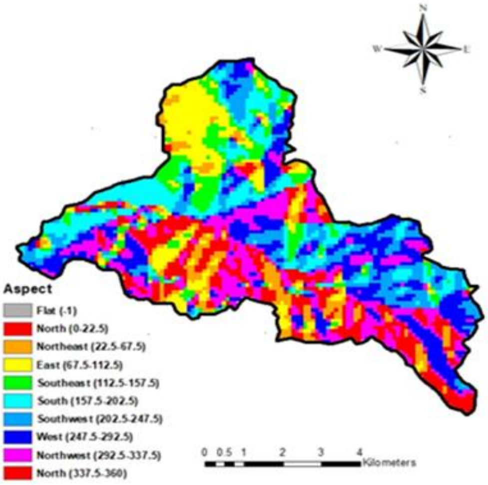

Fig. 3 Aspect map

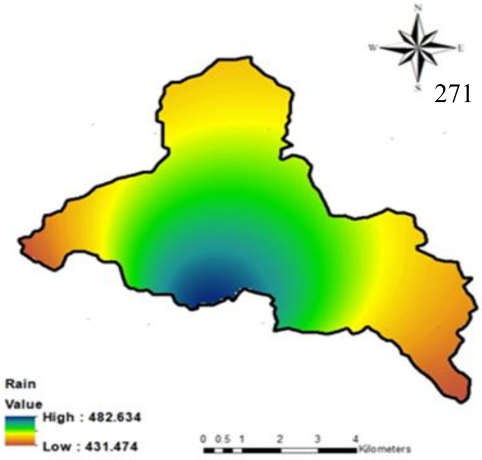

Fig. 5 Rain map

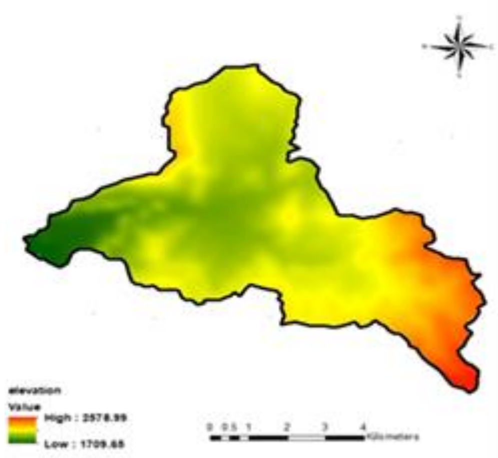

Fig. 4 Elevation map

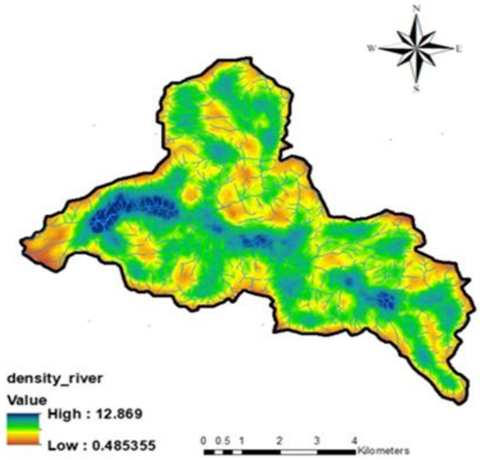

Fig. 6 Stream density map

2.3.3 Height

Some researchers use height as a controlling factor in the occurrence of landslides (Yilmaz 2010). This factor controls the direction of the waterways and the density of the drainage network and has a significant effect on the soil moisture and the slope of the slopes.

At higher altitudes, precipitation usually occurs in the form of snow, which has a lesser effect on increasing the potential for landslide occurrence compared to rain. Therefore, it can be expected that landslides are more likely to occur in areas with low altitudes. If we want to examine the two variables of altitude and precipitation in a combined form, it is also necessary to mention that the amount of precipitation is also lower at very low altitudes. Therefore, there is the highest probability of landslides in the middle latitudes. In addition, usually at altitudes higher than 3500 meters, there is less possibility of soil formation or placement of loose and fine-grained sediments, and landslides do not occur (Nojavan *et al.* 2018). The altitude range of the studied area is between 1709 and 2578 meters.

The elevation map is prepared from the Digital Elevation Model (DEM) of the area, which is according to Fig. 4.

2.3.4 Precipitation

Rainfall is one of the effective factors in creating instability of domains. The highest number of slope breaks occurs after heavy rains or snow melting in spring and due to water penetration in the cracks (Ghiasi *et al*. 2021). Also, atmospheric precipitations in the form of snow or rain will cause saturation of the range and change the level of underground and surface water, and as a result, mass movements will occur. Therefore, knowing the characteristics of rainfall is of particular importance. The existence of inherent conditions prone to the occurrence of landslides, such as the type of materials, increases the possibility of landslides with the increase in rainfall and humidity (Mafian *et al.* 2009). The amount of precipitation in the study area is 431 to 482 mm, which was prepared using interpolation of precipitation stations inside and outside the basin with a time base of 12 years (Ilderami *et al*. 2016). The precipitation map is according to Fig. 5.

2.3.5 Waterway density

Due to the presence of water drainage and steep walls, waterways usually have more slippage. Flowing waters can affect slope failures through erosion or saturation of

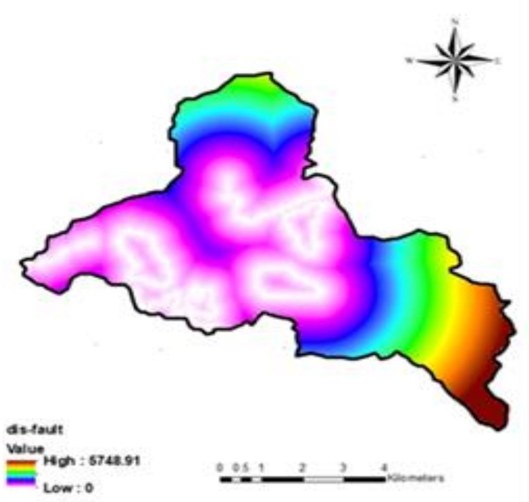

Fig. 7 Fault distance map

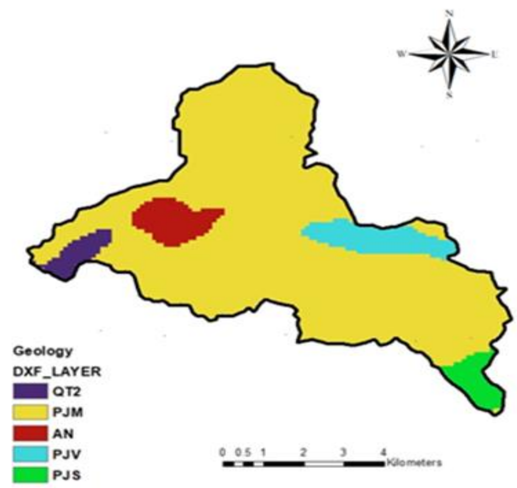

Fig. 8 Geology map

materials at the foot of the slope (Dehnavi *et al*. 2015). In fact, as a well-known principle, by moving away from waterways, the potential for landslides decreases. In other words, the higher the density of waterways in an area, the higher the probability of landslides (Dai *et al*. 2001). The topographic map of the region and Arc GIS software was used to prepare the waterway density map. The density map of the waterway is according to Fig. 6.

2.3.6 Distance from the fault

Faults play an effective role in creating or reactivating areas with sliding potential. Crushing and shearing in the fault areas, water penetration from these areas into the slopes, the emergence of discontinuity around the fault, erosion differences in the slopes, and crushing of the surrounding rocks are among the effects that can be mentioned. The movement of the fault can also be the initiator of the landslide in the desired area. The fault is considered a secondary factor and an independent variable in the occurrence of landslides . The risk of landslides is greater in areas that are less far from the fault, and the greater the distance from the fracture site, the less the impact of the fault on the occurrence of landslides. Therefore, the distance from the fault plays an important role in the occurrence of landslides.

To prepare the fault map, the geological map of the studied area was used on a scale of 1:100,000. For this purpose, Arc Gis software has been used. The distance map of the faults in the study area is according to Fig. 7.

2.3.7 Geology

The geological factor is one of the important parameters (independent variable) in landslide studies. Due to the difference in the type of constituent sediments and the conditions and period of formation, rocks show different resistances against external forces. In fact, the type of rock as one of the geological factors, controls the rate of change of rock facies and their transformation into rock debris, the occurrence of landslides and the increase of soil generation potential, and the creation of slopes with variable slopes and finally the occurrence of landslides. It has been a slip. Its diverse geology and structure cause differences in the stability and resistance of rocks, as well as the variety of soil types. The type of geological formation of each area plays an important role in the expansion of sliding zones. Even some researchers consider lithology as a fundamental factor in controlling landforms (Daei *et al.* 2001).

According to the geological map of the region on a scale of 1:100000 prepared by the Geological Organization of the country. The geological map is divided into five classes, anhydrite, and gabbro (AN), thick dark glossy gray bedrock with marble-shaped fossils (PJM), volcanic rocks with interlayers of tuff, and magma (PJV), Alluvial flat foothill deposits (QT2). And the shemshak formations of silites were classified together with dark and gray slates (PJS).

In this research, the geological map of the region with a scale of 1:100,000 was used to prepare the geological raster map. The geological map is according to Fig. 8.

2.3.8 Distance from the fault

Faults play an effective role in creating or reactivating areas with sliding potential.

2.4 Production of weighted control layers by a continuous fuzzy method using a logistic function

A circular logistic transformation can play an important role in many classification algorithms and detection patterns, such as statistical methods, neural networks, machine learning, and specialized systems (Yousefi and Carranza 2015a). For example, Nikkanen *et al*. used a logistic function for fuzzy scoring of continuous spatial witness values. Yousefi *et al.* applied linear and non-linear transformations (using the logistic function) to the exploratory witness data and compared the results, and showed that non-linear transformations are more suitable for weighting.

There is a group of logistic functions that can be used to transfer the data to the logistic space based on the minimum and maximum values of the data and the slope changes between them and to use the logistic function, and thus the total data to the range between zero and are transferred. In this research, the logistic function of Eq. (1) has been used to generate weighted control layers and assign fuzzy scores to continuous values. Continuous data such as distance from the fault, the density of waterways, slope, precipitation, and height are assigned a weight between a minimum of zero and a maximum of one without expert judgment. For this purpose, the logistic function stated in Eq. (1) has been used to fuzzify the witness maps (Yousefi and Nykänen 2016).

$$F_{EV} = \frac{1}{1 + e^{-s(EV-i)}} \quad (1)$$

F_{EV} :Fuzzy weight like E_V, i :Turning point, s :Slope of the function, E_V: The numerical value of each criterion.

The values of i and s are obtained from Eqs. (2) and (3).

$$i = \frac{E_{Vmax} + E_{Vmin}}{2} \quad (2)$$

$$S = \frac{9 \cdot 2}{E_{Vmax} - E_{Vmin}} \quad (3)$$

E_{Vmax}: The value of the largest pixel
E_{Vmin}: The value of the smallest pixel

The criteria of the slope, height, precipitation, watercourse density, and distance from the fault were weighted using the continuous fuzzy method and logistic function.

2.5 Production of weighted control layers by a continuous fuzzy method using a logistic function

For two geological criteria and slope direction, weighting is done using the classical fuzzy method. The approach of this fuzzy modeling is based on the opinion of experts and based on the unsupervised method, whereby fuzzy evidence points are assigned to classified (discrete) spatial data (weighting). In this research, according to experts, geological criteria were divided into five classes, slope direction criteria were divided into eight classes, and digital witness maps were considered for all criteria (Yousefi and Nykänen 2016).

2.6 Support vector machine algorithm

The support vector machine (SVM) algorithm is one of the newest supervised classification methods, which is based on learning theory and dimensional statistical theory and follows the minimum structural risk. This method, which has attracted the attention of a wide range of researchers in different fields in the last few years, was first presented by Vepnick in 1963. One of the features of this method is that it can be used jointly for two classifications (SVC) and regression (SVR) operations (Cristianini and Shawe-Taylor 2000). This method includes a set of classification functions, which has the ability to evaluate errors and generalize the information appropriately, and by using the information available in the layers of effective factors and high repetition of modeling, it reduces the complexity of the occurrence of landslides (Kornejadi and Pourqasmi 2015).

This model includes a training stage with input and output goal values . An algorithm (SVM) is used to evaluate and test a set of data. Also, in recent years, due to its good classification and regression performance, error tolerance, and appropriate generalization, it has attracted a lot of attention. Based on this theory, the error rate bound of the learning machine for unclassified data can be considered as the generalized error rate. These limits are a function of the total training error rate that shows the level of complexity of the classifiers.

Support vector machine models are divided into two main groups:

a) support vector machine classification model and **b)** support vector machine regression model are divided.

The support vector machine classification model is used to solve data classification problems that are placed in different classes, and the support vector machine regression model is used to solve prediction problems (Shokri *et al.* 2012). A support vector machine regression model is used in this research. Support Vector Machine Regression Model (SVR): As mentioned, Support Vector Machine is based on minimizing the risk structure, which is taken from the theory of statistical training (Vapnik 1995). For the application of support vector machines in regression problems, Vepnik used the error function called ε, which ignores the errors that are at a certain distance from the real values. This function is defined as Eq. (4) (Basak and Patranabis 2007).

$$\left(y\left(f(x.w)\right) = |y - f(x.w)|_\varepsilon = \begin{cases} 0 & for \quad |y - f(x.w)| \le \varepsilon \\ |y - f(x.w)| - \varepsilon & for \quad |y - f(x.w)| > \varepsilon \end{cases} \quad (4)$$

This error function does not consider error values less than ε. We consider the problem of approximating a set of the following data in Eq. (5) :

$$D = \{(x^1.y^1).\cdots.(x^l.y^l)\}. \quad (x \in R^m . y \in R^n) \quad (5)$$

The regression function is estimated by the following function in Eq. (6)

$$f(x) = \langle w.x \rangle + b \quad (6)$$

that in this regard:

x: Input vector ($x \in R^m$)

y: output value ($y \in R^n$)

w: Weight vector ($w \in R^m$)

b: Bios ($b \in R^n$) & $\langle \cdot \rangle$ It is an internal multiplication.

By using the Vepnik loss function, the controlling parameters of the optimal response function in SVM (that is, the weight and bias function) are obtained by minimizing the function in Eqs. (7) and (8) (Cristianini and Shawe-Taylor 2000).

$$\emptyset(w.\xi) = \frac{1}{2}\|w\|^2 + C\sum_{l=1}^{l}(\xi_i^- + \xi_i^+) \quad (7)$$

$$subject\ to \begin{cases} y_i - (\langle w.x_i \rangle + b \le \varepsilon + \xi_i \\ (\langle w.x_i \rangle + b) - b \le \varepsilon + \xi_i^* \\ \xi.\xi_i^* \ge 0 \end{cases} \quad (8)$$

where $C$is a predetermined value and ξ_i^- aξ_i^+re variables that determine the upper and lower limits of the system output.

ε: Acceptable error in the loss function is shown in Fig. 9.

$\|w\|^2$: Soft weight vector

As it was said in the optimal regression function of relation (7), εit is the acceptable error in the loss function and C is the regulating parameter ξ , and ξ^* the Slack variables, which are shown together with the loss function in Fig. 9, have been (Cristianini and Shawe-Taylor 2000).

Based on the Lagrange theory, the optimization problem of relation (7) can be converted into the Lagrange function in Eq. (9)

$$L(\alpha^*.\alpha) = -\varepsilon\sum_{i=1}^{1}(\alpha_i^* + \alpha_i) + \sum_{i=1}^{1} y_i(\alpha_i^* - \alpha_i) - \frac{1}{2}\sum_{i=1}^{1}\sum_{i=1}^{1}(\alpha_i^* - \alpha_i) - (\alpha_j^* - \alpha_j)(x_i x_j) \tag{9}$$

By maximizing the above function under the following constraints, the values of coefficients α^* a α re obtained. These coefficients are called Lagrange coefficients in Eq. (10).

$$\begin{cases} \sum \alpha_i^* = \sum \alpha_i \\ 0 \ll \alpha_i^* \ll C \\ 0 \ll \alpha_i^* \ll C \end{cases} \quad i = 1.2.\cdots.l \tag{10}$$

In the above relationships, L and C are penalty parameters or adjustment parameters.

It is noteworthy that the above optimization problem can be solved by nonlinear equation methods. As a result, reaching the general extremum will be certain, and there is no risk of getting trapped in the local extremum (Cristianini and Shawe-Taylor 2000)Thus, the final answer will be as follows in Eqs. (11)-(13) (Dibike *et al*. 2001).

$$w_0 = \sum_{sup\,portvectos} (\alpha_i^* - \alpha_i)x_i \tag{11}$$

$$b_0 = -(\frac{1}{2})w_0 \cdot [x_r + x_s] \tag{12}$$

$$f(x) = \sum_{sup}(\alpha_i^* - \alpha_i)(x_i \cdot x) + b_0 \tag{13}$$

In these equations:

x_i=The input vector by which the model is trained

x =Input vector

$x_s.x_r$ =Two support vectors

w_0 =Optimal weight vector

b_0 =Optimal bias vector

Data whose corresponding Lagrange coefficients are non-zero are known as support vectors. Geometrically, these data have a prediction error greater than $\pm\varepsilon$. Therefore, the support vectors are not included in the $\pm\varepsilon$ band, and the value of ε controls the number of support vectors; according to Eq. (11), it is observed that the data whose Lagrange coefficient is zero do not have a role in the final answer. In other words, they are the support vectors that determine the final regression function with the optimal response (Samui 2008). There are separate linear forms, an optimal level that separates the data without error and with the maximum distance between the plane and the nearest training points (support vectors) is trained. If we define the training points as $[x_i.y_i]$ the input vector$x_i \in R^m$, in the case that the data can be separated linearly, the Equation will be in the form of Eq. (14) (Shokri *et al*. 2012).

$$y = f(x) = sign\left[\sum_{i=1}^{N} y_i a_i \langle x_i.x\rangle + b\right] \tag{14}$$

where $y$is the output of the Equation andy_i is the class value of the experimental sample x_i . The vector $x = (x_1.x_2\cdots.x_n)$ indicates input data, and the vectors $i = 1.2.\cdots.Nx_i$ are the support vectors. If the data cannot be separated linearly, it is possible to move the samples to a higher space by applying pre-processing. In this case, relation (21-3) is written as Eq. (15) (Shokri *et al*. 2012).

$$y = f(x) = sign\left[\sum_{i=1}^{n} y_i a_i k(x.x_i) + b\right] \tag{15}$$

The function $k(x.x_i)$ is a kernel function that generates internal hits to create machines with different types of non-linear levels in the data space (Shokri *et al*. 2012).

To build the support vector machine model, parameters$C$ε are defined by the user. The parameter$C$ is a regulatory parameter and can accept values from zero to infinity. The role of this parameter is to create a balance between minimizing the experimental risk and maximizing the generalizability. Therefore, the effect of the principle of structural risk minimization can be seen (Dibike *et al*. 2001). In other words, it can be said that the Cvalue creates a balance between the performance of the model on the training data and the test data. When large values are assigned to this parameter, SVM does not allow errors to occur in the training data, and the result will be a complex model. On the other hand, when Cit tends to zero, the model can accept a large error because as the value of this parameter decreases, less attention is paid to the number of slack variables, and the model is more sensitive to the occurrence of errors on the training set. Shows less, and as a result, the complexity of the model will be less (Dibike *et al*. 2001).

The parameter ε can also accept values from zero to infinity. The value of this parameter is very effective in the state of the support vectors and, as a result, the efficiency of the model. Although choosing very large values εcauses a reduction in the number of support vectors and is desirable, it is wrong to achieve this goal by widening the bandε . On the other hand, very small values of this parameter cause a large number of support vectors to be selected, and the risk of overtraining increases (Dibike *et al*. 2001).

Table 1 Quantitative correlation of the area under the curve (AUC) in the ROC method (Yesilnacar and Topla 2005)

Qualitative weak	medium	very good	Good	Excellent
0.5 to 0.6	0.6 to 0.7	0.7 to 0.8	0.8 to 0.9	0.9 to 1

The problem of linear regression in SVM can be easily extended to non-linear regression. In this sense, kernel functions are used. Kernel functions map the data to a special space in which it is possible to use linear regression (Cristianini and Shawe-Taylor 2000). The reason for using kernel functions is to convert original non-linear data patterns into a format that includes a separate line in a high-dimensional special space (Yao *et al*. 2008). The selection of different kernel functions in SVM model is very important; although many kernel functions $K(X_i . X_j)$ have already been proposed and used, only some of these kernels are suitable for working in A wide variety of applications that are known to be useful. The most important kernel functions include the following functions(Yao *et al*. 2008).

Linear function $$K(X_i . X_j) = X_i^T \cdot X_j$$

Polynomial function $$K(X_i . X_j) = (\gamma \cdot X_i^T \cdot X_j + r)^d$$

Radial basis function $$K(X_i . X_j) = e^{-\gamma(x_i - x_j)^2}$$

circular function (sigmoid) $$K(X_i . X_j) = tanh(\gamma \cdot X_i^T \cdot X_j + r)$$

where$d\gamma$, and$r$ are parameters of kernel functions and are entered manually.

2.6.1 The process of evaluating the effectiveness of models

Validation is an essential part of landslide susceptibility, and landslide susceptibility maps are worthless without validation. In the current research, for each kernel used, a zoning map of landslides in the studied area has been prepared. 70% of the data was used for training, and 30% of the data was used for testing. In order to evaluate the support vector machine algorithms and the maps obtained from each of the kernels, the following parameters have been used:

A) ROC curve

To evaluate the performance of the functions in order to predict the sensitivity of landslides in the region, the criterion of the percentage of the area under the curve has been used. ROC curve is a graph whose horizontal axis includes (true positive or l-Specificity), and its vertical axis includes (false positive or Sensitivity). A positive true is a number of landslide pixels (value one) that are correctly classified as landslides. False positives are the number of landslide pixels (zero value) that are wrongly classified as landslides. The area under this curve is called AUC, and the model with the highest AUC value has a higher relative performance. Both training data and test data have been used to verify the accuracy of the location map of areas sensitive to landslides. This means that the ROC curve and the AUC value have been calculated for both training and test data. If a model cannot estimate the landslide event better from a probabilistic point of view, the AUC value will be equal to 0.5, and the closer this value is to one, the efficiency of the model will increase (Pontius and Schneider 2001). The higher the detection power of a test, the ROC curve will be above the square diameter and, therefore, will be closer to the ideal state (Ghasemian *et al*. 2016). Based on a conventional classification system, the surface under the curve can be classified and interpreted according to the Table 1 (Yesilnacar and Topla 2005).

B) Root mean square errors (RMSE)

The metric error value is in the same data as the original data and is a statistical index that is used to evaluate the efficiency of the model. RMSE measures the amount of error between two data sets. This parameter usually compares predicted values and measured values. Since, in the calculation of RMSE, larger errors are more important than small errors, this factor has become a common criterion for error measurement [53]. A lower RMSE value indicates a better performance of the sliding model (Ghasemian *et al*. 2016). Its value is obtained from the Eq. (16).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad (16)$$

In the above Equation, we have:
x_i: Predicted value
y_i: real amount
n: Number of data
In this research, RMSE is calculated for the training data as well as for the test data.

C) Correlation coefficient (R):

The correlation coefficient is a measure to measure the correlation between predicted and measured values, which is obtained from Eq. (17)

$$R = \frac{C_{y_j} d_j}{\sigma_{y_j} \sigma_{d_j}} \quad (17)$$

y_j: measured output value (desired)
d_j: model output value (predicted)
$C_{y_j} d_j$: covariance between model output and measured output
σ_{y_j}: standard deviation of the measured output
σ_{d_j}: standard deviation of model output

If the absolute value of the correlation coefficient is greater than 0.8, according to Smith, there is a strong correlation between the two data sets (Pontius and Schneider 2001). In this research, R has been calculated for both training data and test data. A higher value of R indicates a better performance of the sliding model.

3. Result and discussion

3.1 Fuzzification (weighting) of information layers

Five criteria affecting the occurrence of landslides

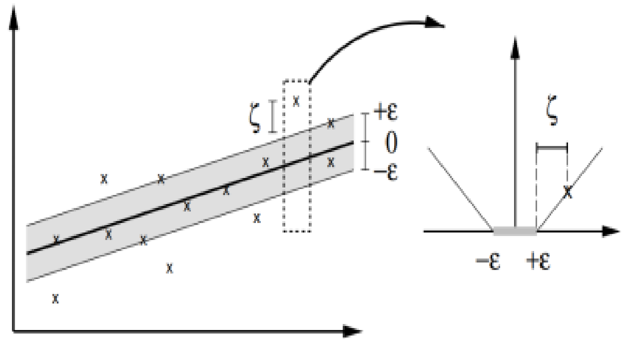

Fig. 9 Loss function and variables(Cristianini and Shawe-Taylor 2000)

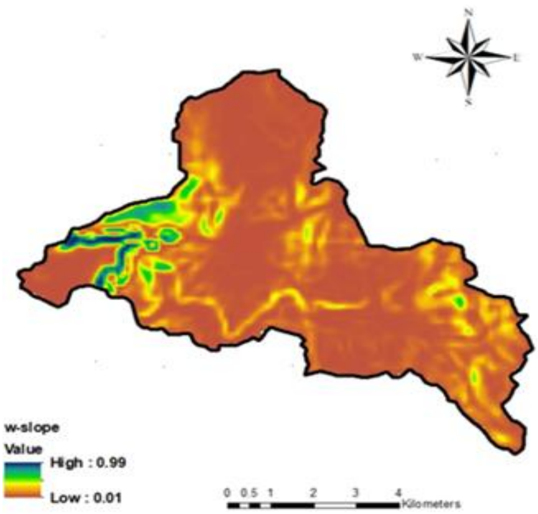

Fig. 10 Slope weighted map

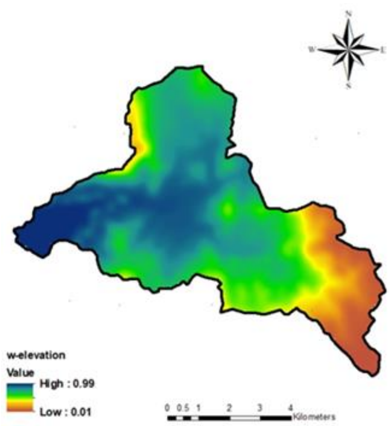

Fig. 11 Elevation weighted map

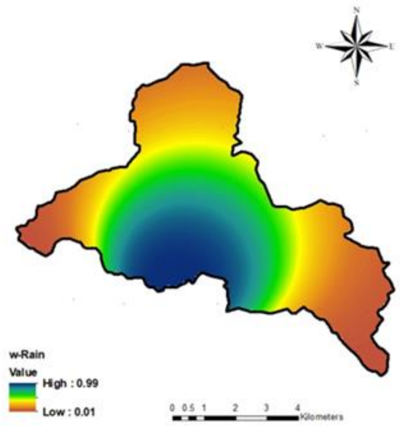

Fig. 12 Rain weighted map

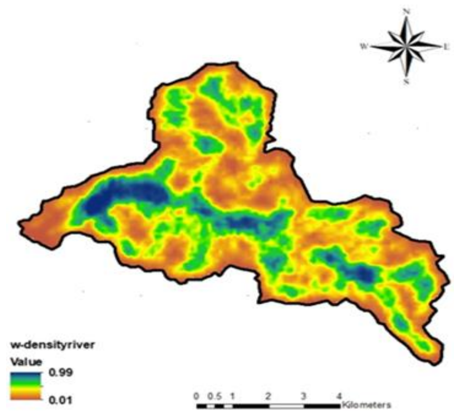

Fig. 13 Stream Density weighted map

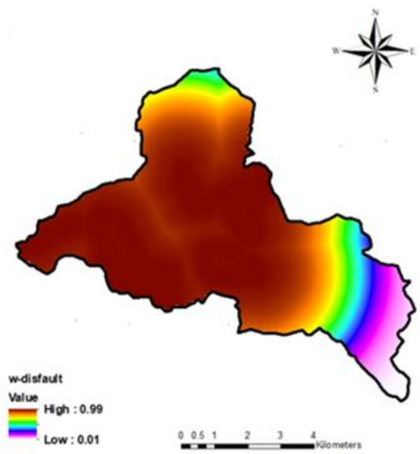

Fig. 14 Fault distance weighted map

include: slope, height, precipitation, watercourse density, and distance from the fault using the logistic function in ArcGIS 10.4.1 software according to Figs. 10 to 14 and two criteria for slope direction and geology with the help of ArcSDM tool in ArcGIS 10.4.1 software were weighted according to Figs. 15 and 16.

3.2 Combination of maps with support vector machine method

After preparing the weighted (fuzzy) maps of the effective factors in landslides, in the next step, these weighted maps are combined with each other using the Support Vector Machine (SVM) algorithm. For this purpose, the desired program (code) has been written

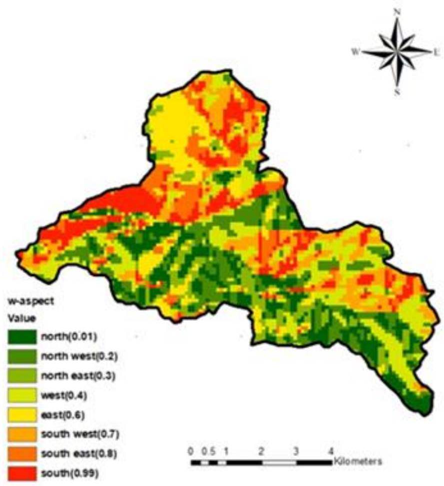

Fig. 15 Aspect weighted map

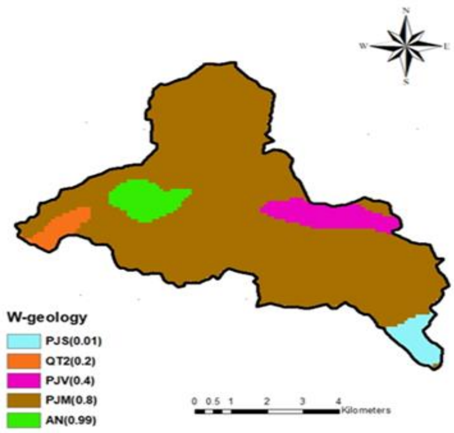

Fig. 16 Geology weighted map

using the SVM toolbox in the Matlab software, which will be explained below, but before that, the information obtained from the weighted layers should be in the form of Excel files. To be prepared as the output of the GIS software and the input of the backup vector machine.

In the first step, the map of the studied area contains 4799 square pixels with dimensions of 100 meters. By using GIS software, the desired pixels are converted into points. Each of the available 4799 pixels should have the characteristics of seven weighted information layers. Therefore, in the GIS software, by entering each layer of weighted information in the desired part, a table of 4799 desired points is obtained, where each point contains the information of seven weighted layers. Finally, the prepared table will be converted into an Excel file, which will be one of the inputs of the support vector machine.

3.3 Sliding points and non-sliding points

In the second step, the points where landslides have occurred have been identified using satellite images, as well as areas prone to landslides on the map, and the number of them is 81 points. These points, which are already marked on the map, should be converted from Raster mode into

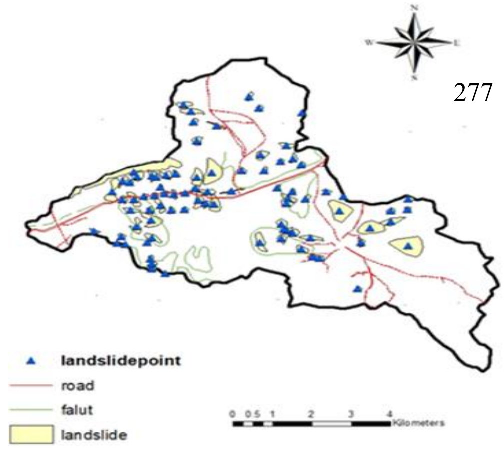

Fig. 17 Landslide distribution map in the study area

points that have the characteristics of seven weighted information layers. Therefore, for each pixel with the characteristics of all involved layers, a number of one has been considered due to the occurrence of landslides. A table in Excel format containing 81 definite landslide points will be prepared as the output of the GIS software and the input of the backup vector machine.

Landslides, the number of landslides that occurred in the studied area was already known, which was identified using satellite images and areas prone to and suspected of landslides in the region, and many landslides can be identified due to their small dimensions or appearance similar to the adjacent domain in satellite images. Therefore, to complete the information, all available landslides have been visited in the field, and 81 landslide points have been detected in the study area of Oshvand, according to Fig. 17.

In the third step, 81 non-sliding points (equal to sliding points) should be selected to have the characteristics of the seven layers of the weighted witness. Therefore, for each pixel with the characteristics of all involved layers, the number zero has been considered due to the absence of landslides. These points are usually chosen on low slopes and inside waterways, where the probability of landslides is negligible. The advantage of this method is the absence of any intervention on the part of the expert on the possible research results. For this purpose, we select the desired points manually with the help of GIS software. The excel file is formed with 81 non-landslide points, which will be the input of the support vector machine. Eighty-one landslide points and 81 non-landslide points will be used to train and test the support vector machine algorithm. Fig. 18 shows the selected non-slip points.

3.4 Support vector machine modeling

The training data are defined according to Eq. (18).

$$x_i = (i = 1.2. \dots . n) \tag{18}$$

where x_i are a set of training cells. In the current research, 81 points as landslide occurrence points $(y = 1)$and 81

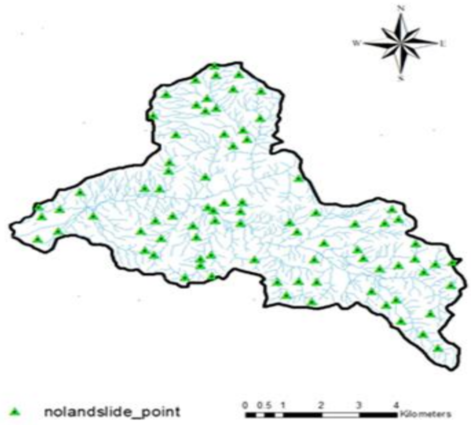

Fig. 18 Non Landslide distribution map in the study area

other points as non-sliding points $(y = 0)$, totaling 162 points, are used as inputs to the direction support vector machine. Training and testing have been introduced. Therefore, the input data will contain a matrix with 162 rows and eight columns (seven factors plus one or zero).

From this number, 70% of the data (113 pixels) is used as a matrix with 113 rows and eight columns for training, and 30% of the data (49 pixels) as a matrix with 49 rows and eight columns are used for testing. The main goal of building a model and testing it is to separate the set of training and test samples. If the random samples collected for training and testing are used simultaneously with the random samples based on sliding pixel points, the model's performance will increase (Kuang *et al*. 2020, Xinhua 2019, Taner San 2014) This research uses the support vector machine regression model for modeling. It was stated that the support vector machine is part of the category of supervised learning models (the model is trained with the help of a supervisor) along with the relevant training algorithms that are used in the analysis of data used in regression and classification. Its basis is based on structural minimization. Risk is taken from the theory of statistical training.

In general, linear data regression rarely happens, and in practice, the data placement is non-linear. Kernel functions generate internal hits to create machines with different types of non-linear levels in the data space. The kernel's task is to take data as input and transform it into the required form. Kernel functions are more useful in nonlinear separation problems, which means that in problems where the data are not linearly separable, the Equation of the separating plane for the nonlinear state with the intervention of the kernel function, which is responsible for mapping data from nonlinear to linear space, Achieved. In this situation, the non-linear method of support vector regression is used. In this case, the input patterns are mapped to a space with larger dimensions, so linear regression can be done in the mapped space. Usually, finding a mapping related to a specific kernel function or a kernel function related to a special mapping is difficult and even impossible. The selection of the kernel function is an important issue based on the test, knowledge of classification and regression issues, and theoretical considerations (Surmiri *et al*. 2020, Tsai and Liao 2019, Ghannadpour and Mehrparvar 2020, Feng *et al*. 2019, Erasto 2001).

In building an efficient support vector machine model, model parameters must be accurately calculated using an optimization method. These parameters are:

A- The type of kernel function used in modeling
B- Parameterε: acceptable error in the loss function
C- Parameter C : This is a penalty or regulator parameter that may be obtained by trial and error or through an optimization algorithm.
D- Parameters of kernel functions: such as$d\gamma r$, which are used in polynomial kernels, radial, and sigmoid basis functions.

The good performance of the regression function depends on the good choice of parameters ε ، C kernel function, and parameters of kernel functions. Due to the dependence of SVM model parameters, choosing the optimal parameter is complicated.

In this research, four linear, polynomial, radial basis function (RBF) (Gossin), and circular (sigmoid) kernels have been used to train the support vector machine. The linear kernel function is a special case of the polynomial kernel function, a common and widely used function in problems. The polynomial kernel function can be much more useful in complex problems. Gaussian and circular kernel functions are the most famous and widely used in support vector machine problems. They are used in problems with no information about the data type and nature.

3.4.1 Support vector machine modeling using a linear kernel function

Coding has been done in MATLAB using a linear function. The input data matrix (Excel file), which contains 81 sliding points and 81 non-sliding points (162 points in total), are each called individually by the software.

εC Parameters were optimized by trial and error. $\varepsilon = 0.1$ $C = 10000$. The best support vector machine model was generated using a linear kernel for values of and. For the obtained model, the values of AUC, RMSE, R, and the number of support vectors for the training and test data were obtained according to Table 2.

Finally, the model produced by the linear kernel of the support vector machine was applied and normalized on all data. The output data from MATLAB software was re-entered into GIS 10.4 software as an Excel file to visualize the landslide susceptibility map. 1. The landslide sensitivity map of Oshvand watershed is according to Fig. 19.

The area under the ROC curve can be used to evaluate the model. Also, the values of root mean square errors (RMSE) and correlation coefficient (R) have been calculated. According to the Table 2, it can be seen that for the training data, the values of RMSE and R are equal to 0.3887 and 0.6450, respectively. Also, the AUC value is equal to 0.8709, which according to the presented qualitative-quantitative correlation table, the efficiency of the linear kernel model for landslide susceptibility zoning is very good. Also, the ROC diagram obtained for training data and the level under the curve is according to Fig. 20.

Table 2 Final parameters of the model obtained by the linear kernel function

	Number of support vectors	R	RMSE	AUC	ε	C
Training data	99	0.6450	0.3887	0.8709	0.1	10000
Test data	94	0.6329	0.4046	0.8851	0.1	10000

Table 3 Final parameters of the model obtained by the polynomial kernel function

	Number of support vectors	R	RMSE	AUC	ε	C
Training data	90	0.6747	0.3872	0.9204	0.3	0.01
Test data	87	0.5921	0.4149	0.8256	0.3	0.01

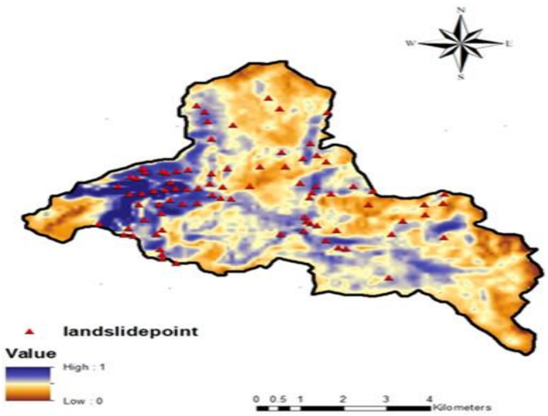

Fig. 19 The final map of landslide susceptibility in the area based on the linear kernel function

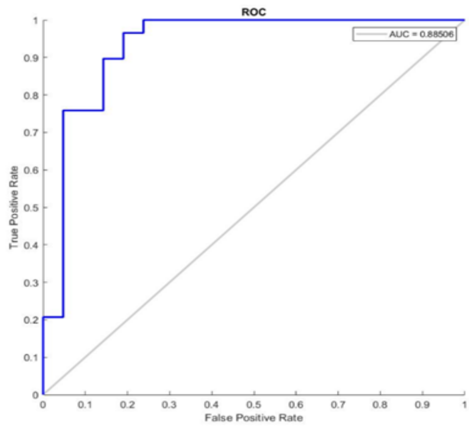

Fig. 21 ROC diagram for test data in linear kernel

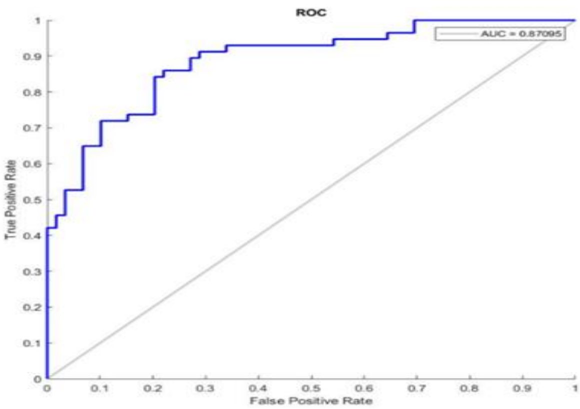

Fig. 20 ROC diagram for training data in linear kernel

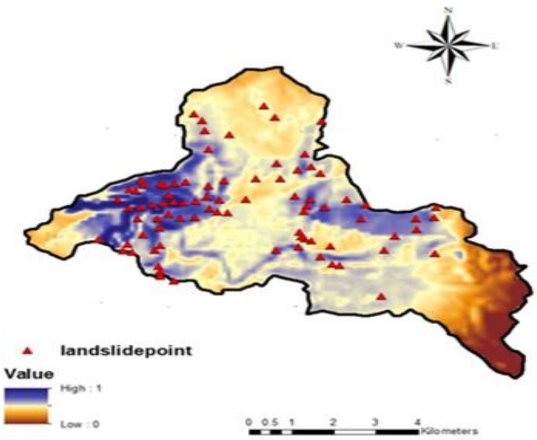

Fig. 22 The final map of the susceptibility of landslides in the region based on the polynomial kernel function

Regarding the test data, as it is clear from Table 2, the values of RMSE and R are equal to 0.4046 and 0.6329, respectively. Also, the AUC value is equal to 0.8851, which indicates the very good performance of the linear kernel function. Also, the ROC diagram obtained for the test data and the area under the curve are according to Fig. 21.

3.4.2 Support vector machine modeling using a polynomial kernel function

In this part, support vector machine modeling is done using the polynomial kernel function. As before, the input data matrix Excel file, including 81 sliding points and 81 non-sliding points (162 points in total), are called separately by MATLAB software. The process of machine learning is like linear mode. 70% of the input points are used for training, and 30% are used for testing.

εC Parameters were optimized by trial and error. For values of $\varepsilon = 0.3$ and $C = 0.01$ .The best support vector machine model was generated using a polynomial kernel. For the obtained model, the values of AUC, RMSE, R, and the number of support vectors for the training and test data were obtained according to Table 3.

Table 4 The final parameters of the model obtained by the kernel Gaussian function

	Number of support vectors	R	RMSE	AUC	ε	C
Training data	67	0.8647	0.2543	0.9830	0.3	10000
Test data	74	0.5933	0.4763	0.8309	0.3	10000

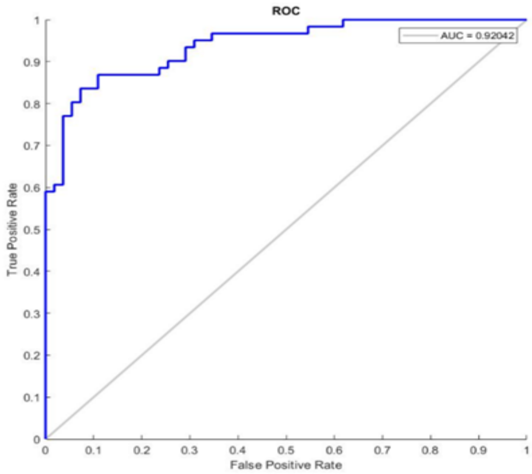

Fig. 23 ROC plot for training data in polynomial kernel

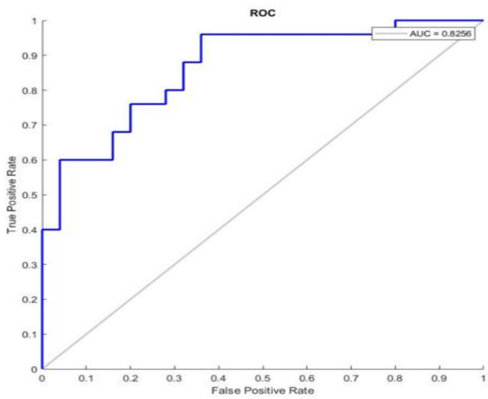

Fig. 24 ROC plot for test data in polynomial kernel

Landslide susceptibility map of Oshvand watershed prepared by the polynomial kernel of support vector machine is according to Fig. 22. According to the Table 3, it can be seen that the values of root mean square errors (RMSE) and correlation coefficient (R) for training data are 0.3872 and 0.6747, respectively. Also, the AUC value is equal to 0.9204, which according to the qualitative-quantitative correlation table, the effectiveness of the polynomial kernel model for landslide susceptibility zoning is excellent. Also, the ROC diagram obtained for the training data and the level under the curve is according to Fig. 23. Regarding the test data, as it is clear from the Table 3, the values of RMSE and R are equal to 0.4149 and 0.5921, respectively. Also, the AUC value is equal to 0.8256, which indicates the very good performance of the polynomial kernel function. Also, the ROC diagram obtained for the test data and the area under the curve are according to Fig. 24.

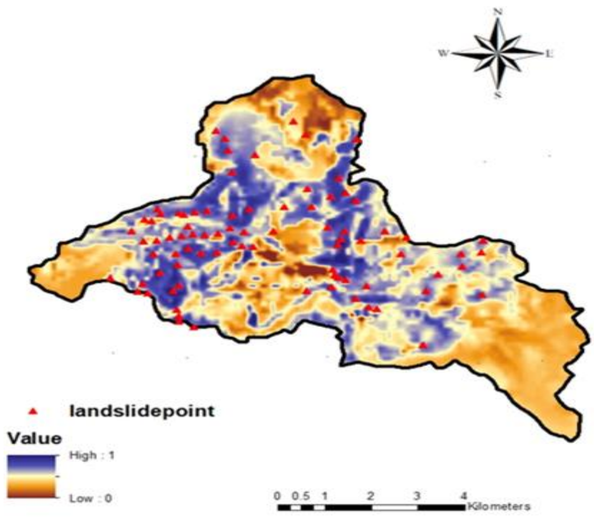

Fig. 25 The final map of the susceptibility of landslides in the area based on the kernel Gaussian function

3.4.3 Support vector machine modeling using radial or Gaussian basis kernel function

The relevant model is created using support vector machine regression, the Gaussian kernel function, and determining the relevant parameters. εC Parameters were optimized by trial and error. $\varepsilon = 0.3 C = 10000$. The best support vector machine model is generated using the Gaussian kernel function for values. For the obtained model, the values of AUC, RMSE, R, and the number of support vectors for the training and test data were obtained according to Table 4.

Landslide susceptibility map of Oshvand watershed prepared by Kernel Gossin vector machine is according to Fig. 25.

Table 4 shows that the values of root mean square errors (RMSE) and correlation coefficient (R) for training data are 0.2543 and 0.8647, respectively. Also, the value of AUC is equal to 0.9830, which according to the qualitative-quantitative correlation table, the efficiency of the Cornell Gossin model for landslide susceptibility zoning is excellent. Also, the ROC diagram obtained for the training data and the level under the curve is according to Fig. 26. As it is clear from Fig. 26, the ROC diagram has the greatest distance from the square diameter, which indicates the very high efficiency of the model.

Regarding the test data, as it is clear from Table 4, the values of RMSE and R are equal to 0.4763 and 0.5933, respectively. Also, the AUC value is equal to 0.8309, which indicates the very good performance of the Gaussian kernel function. Also, the ROC diagram obtained for the test data and the area under the curve are according to Fig. 27.

Table 5 The final parameters of the model obtained by the sigmoid kernel function

	Number of support vectors	R	RMSE	AUC	ε	C
Training data	102	0.9827	0.096	0.9833	0.1	10000
Test data	96	0.6074	0.5401	0.8161	0.1	10000

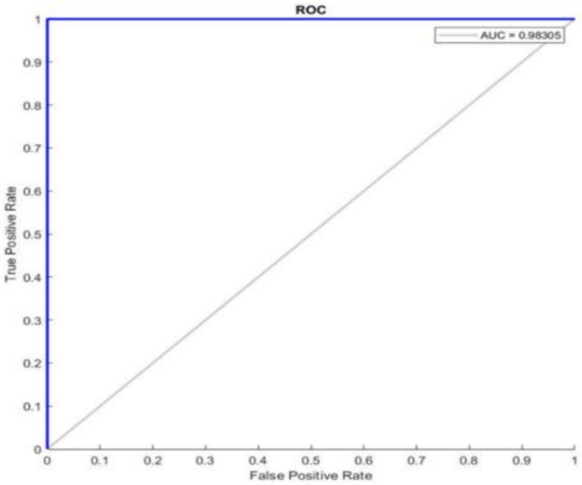

Fig. 26 ROC diagram for training data in Gaussian kernel

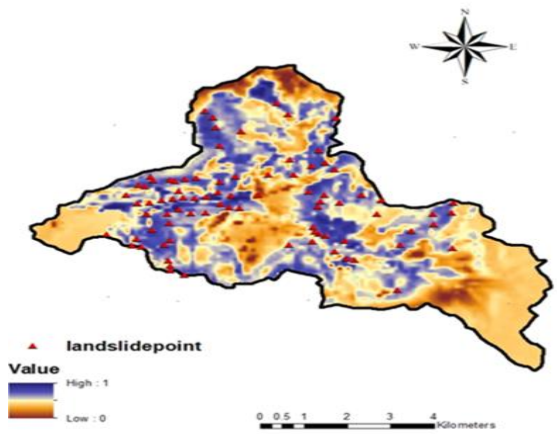

Fig. 28 The final map of landslide susceptibility in the region based on the sigmoid kernel function

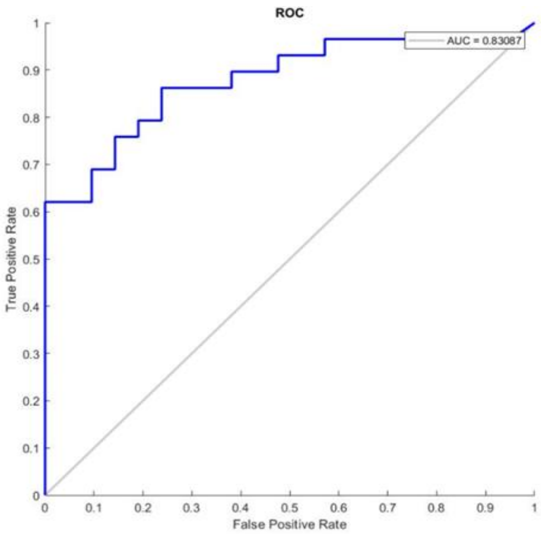

Fig. 27 ROC diagram for test data in Gaussian kernel

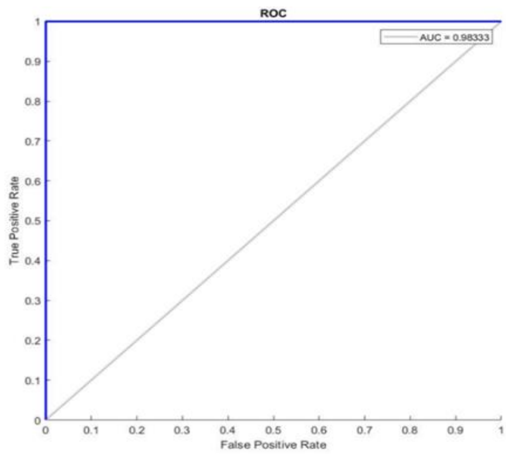

Fig. 29 ROC diagram for training data in sigmoid kernel

3.4.4 Support vector machine modeling using a circular or sigmoid basis kernel function

Modeling has been created using support vector machine regression, a circular kernel function (sigmoid), and determining the relevant parameters. εC Parameters were optimized by trial and error. $\varepsilon = 0.1 C = 10000$ The best support vector machine model is generated using the sigmoid kernel function for values of and. For the obtained model, the values of AUC, RMSE, R, and the number of support vectors for the training and test data were obtained according to Table 5.

Landslide susceptibility map of Oshvand watershed prepared by the sigmoid kernel of support vector machine is according to Fig. 28.

Table 5 shows that the values of root mean square errors (RMSE) and correlation coefficient (R) for training data are 0.096 and 0.9827, respectively. Also, the AUC value is equal to 0.9833, which according to the qualitative-quantitative correlation table, the efficiency of the sigmoid kernel model for landslide susceptibility zoning is excellent. Also, the ROC diagram obtained for the training data and the level under the curve is according to Fig. 29. As it is clear from Fig. 29, the ROC diagram has the greatest distance from the square diameter, indicating the model's very high efficiency.

Regarding the test data, as it is clear from Table 4, the values of RMSE and R are equal to 0.5401 and 0.6074, respectively. Also, the AUC value is equal to 0.8161, which indicates the very good performance of the sigmoid kernel function. Also, the ROC diagram obtained for the test data and the area under the curve are according to Fig. 30.

Table 6 Summary of the results obtained using four different kernels to prepare the zoning map

		AUC	RMSE	R
Linear kernel	Education	0.8709	0.3887	0.6450
	Test	0.8851	0.4046	0.6329
Polynomial kernel	Education	0.9204	0.3872	0.6747
	Test	0.8256	0.4149	0.5921
Colonel Gossin	Education	0.9830	0.2543	0.8647
	Education	0.8309	0.4763	0.5923
Sigmoid kernel	Education	0.9833	0.096	0.9827

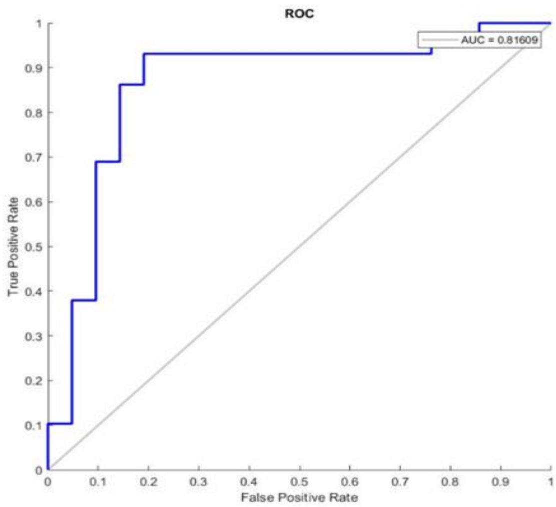

Fig. 30 ROC plot for test data in sigmoid kernel

4. Conclusions

According to the information, using a support vector machine algorithm, linear, polynomial, Gaussian, and sigmoid kernels were used to zoning the sensitivity of landslides in Oshvand region. According to the separate performance of each of the functions and the setting of parameters required for modeling, the zoning map of the area was obtained along with the parameters required to evaluate the models. Table 6 summarizes the results obtained using four different kernels.

As seen in the Table 6, among the four kernels used, the numerical value of AUC in Gaussian and Sigmoid kernel functions for training data is 0.9830 and 0.9833, respectively, which is higher than other functions. In the case of test data, this number has the highest values in the linear and Gaussian kernel functions, which are 0.8851 and 0.8309, respectively. It was previously stated that the larger the numerical value of AUC, the better the modeling has been done.

In the case of the RMSE parameter, values close to zero are more important. This parameter is equal to 0.096 and 0.2543, respectively, for the training data's sigmoid and Gaussian kernel functions, which have the lowest values. The lowest value of this parameter for test data is related to linear and polynomial kernel functions.

The R parameter is the opposite of the RMSE parameter; the closer its numerical value is to 1, the more important it is. As seen in Table 6, the value of this parameter in the case of the sigmoid kernel function for training data is 0.9827, which is the highest value, and then the Gaussian kernel function is in second place with a value of 0.8647. The highest value of this parameter in the case of test data is related to linear and sigmoid kernel functions, which are 0.6329 and 0.6074, respectively.

According to the mentioned materials and also the comparison of the landslide susceptibility maps obtained from four different support vector machine models, it can be concluded that the best model is the one obtained from the kernel Gaussian function and the susceptibility zoning map Landslide obtained from Kernel Gossin function is the best-prepared map which is according to Fig. 25.

Conflict of interest statement

There is no conflict of interest for the authors.

Ethics approval statement

Not applicable, because this article does not contain any studies with human or animal subjects.

Funding

The authors did not receive support from any organisation for the submitted work.

Credit author statement

Vahed Ghiasi: Nur Irfah. Mohd Pauzi. Conceptualization, Methodology, Investigation, Formal analysis, Roles/Writing – Review and Editing. Shahab Karimi: Conceptualization, Methodology, Investigation, Formal analysis, Roles/ Writing – Review and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The research described in this paper was financially supported by Malayer University, Iran.

References

Abedini, M. (2015), Tectonic geomorphology. The compilation, second volume, first edition, Mohaghegh Ardabili University Publications, 227.

Abedini, M., Raushit, S. and Fathi, M.H. (2017), "Identifying and classifying the types of landslides using spectral and spatial features with an object-oriented approach (Nasir Abad to Sattarkhan Ahar Dam)", *J. Geography Planning*, **22**(66). 178-205.

Basak, D., Pal, S. and Patranabis, D.C. (2007), "Support vector regression", *Neural Inf. Process*., **11**, 203-225.

Boulfoul, K., Hammoud, F. and Abbeche, K. (2020), "Numerical study on the optimal position of a pile for stabilization purpose of a slope", *Geomech. Eng*., **21**(5), 401-411 https://doi.org/10.12989/gae.2020.21.5.401.

Chalkias, C.H., Ferentinou, M. and Polykretis, C.H. (2014), "GIS-based landslide susceptibility mapping on the Peloponnese Peninsula, Greece", *Geosci*., **4**(3), 176-190. https://doi.org/10.3390/geosciences4030176.

Chauhan, S., Sharma, M., Arora, M.K. and Gupta, N.K. (2010), "Landslide susceptibility zonation through ratings derived from the artificial neural network", *Int. J. Appl. Earth Observation Geoinform.*, **12**, 340-350. https://doi.org/10.1016/j.jag.2010.04.006.

Cristianini, N. and Shawe-Taylor, J. (2000), "An introduction to support vector machines and other kernel-based learning methods", *Cambridge University Press*, 256. https://doi.org/10.1017/CBO9780511801389.013.

Cruden, D.M. and Varnes, D.J. (1996), Landslide types and processes. In: Turner A.K. and Schuster, R.L. Landslides, Investigation and Mitigation, Transportation Research Board Special Report, **247**, 36-75

Dahal, R.K., Hasegava, S.H., Nonoura, A., Yamanka, M., Dhakal, S. and Pauudyal, P. (2008), "Predictive modeling of rainfall-induced landslide hazard in the lesser Himalaya of Nepal based on weights of evidence", *Geomorphology*, **102**(3-4), 496-510. https://doi.org/10.1016/j.geomorph.2008.05.041.

Dai, F.C. and Lee, C.F. (2001), "Terrain-based mapping of landslide susceptibility using a geographical information system: a case study", *Can. Geotech. J.*, **38**, 911-923. https://doi.org/10.1139/t01-021.

Dehnavi, A., Aghdam, I.N., Pradhan, B. andV arzandeh, M.H.M. (2015), A new hybrid model usi the stepwise weight assessment ratio analysis (SWARA) technique and adaptive neuro-fuz inference system (ANFIS) for regional landslide hazard assessment in Iran Catena, **135**, 122-148. https://doi.org/10.1016/j.catena.2015.07.020.

Dibike, Y.B.,Velikov, S., Solomatine, D. and Abbot, M.B. (2001), "Model induction with support vector machines-introduction and applications", *J. Comput. Civ. Eng. ASCE*, **15**(3), 208-216. https://doi.org/10.1061/(ASCE)0887-3801(2001)15:3(208).

Erasto, P. (2001), Support vector machines-backgrounds and practice. BSc Thesis, University of Helsinki, Faculty of Science, Rolf Nevanlinna Institute, Helsinki, Finland. 78.

Ermini, L., Catani, F. and Casagli, N. (2005), "Artificial neural networks were applied to landslide susceptibility assessment", *Geomorphology*, **66**, 327-343. https://doi.org/10.1016/j.geomorph.2004.09.025.

Feng, Y., Jiang, L., Zhou, W., Lai, Z. and Chai, X. (2019), "An analytical solution to the mapping relationship between bridge structures vertical deformation and rail deformation of high-speed railway", *Steel Compos. Struct*., **33**(2), 209-224. https://doi.org/10.12989/scs.2019.33.2.209.

Ghanavati, E. (2011), "Landslide hazard zonation in Jajrud Basin Using Hierarchical Analysis Method", *J. Appl. Res. Geograph. Sci*., **20**, 51- 68.

Ghannadpour S.A.M. and Mehrparvar M. (2020), "Nonlinear and post-buckling responses of FGM plates with oblique elliptical cutouts using plate assembly technique", *Steel Compos. Struct*., **34**(2), 227-239. https://doi.org/10.12989/scs.2020.34.2.227.

Ghasemian, B., Abedini, M. and Raushi, S. (2016), "Landslide susceptibility evaluation using support vector machine algorithm. (Case study: Kamiyaran city, Kurdistan province)", *Quantitative Geomorph. Res.,* **6**(3), 15-36.

Ghiasi, V., Ghasemi, S.A.R. and Yousefi, M. (2021), "Landslide susceptibility mapping through continuous fuzzification and geometric average multi-criteria decision-making approaches", *Nat. Hazards*, **107**, 795-808. https://doi.org/10.1007/s11069-021-04606-y.

Ghiasi, V., Mirzaei, S. and Yousefi, M. (2020), "Development of Prediction - Area charts to improve the output of landslide potential models", *MCEJ*, **20**(6), 127-131. http://mcej.modares.ac.ir/article-16-39163-en.html.

Hashemi, S.H., Karami J., Amini J. and Alimohamadi, A. (2010), "Zoning areas susceptible to landslides using Fuzzy Topsis algorithms and GIS (Case Study: Lorestan Province)", *J. Remote Sens. GIS*, **8**, 23- 36.

Hong, H., Pradhan, B., Xu, C. and Bui, D.T. (2015), "Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression", *Alternating decision tree and support vector machines Catena*, **133**, 266-281. https://doi.org/10.1016/J.CATENA.2015.05.019.

Hossein Zadeh, M M., Tharvati, M.R., Mansouri, A., Mirbagheri, B. and Khazri, S. (2009), "Zoning the risk of mass movements using logistic regression model", Sanandaj-Dehgolan route **.** *Iranian Geology Quarterly*, **11**(3), 27-37.

Hosseini, N. (2015), Application of the data-driven index homopology method in landslide risk modeling. Master's thesis, Malayer University, Iran.

Ilderami, A., Nouri, H., Mohammadi, M. and Mousavi, M. (2016), "Investigating effective factors and landslide risk zoning using surface density model, hierarchical analysis (AHP), and logistic regression in Oshvand watershed", *Environ. Erosion Res.*, **7**, 1-23.

Kamranzad, F., Afshar, E., Mojerab, M. and Memarian, H. (2013), "Landslide risk zoning in Tehran province using data-driven methods and hierarchical analysis", *J. Earth Sci*., **25**(97). 101-114. https://doi.org/10.22071/gsj.2015.41372.

Karam, A. (2004), "Application of Linear mixed models in landslide occurrence potential zonation (Case Study: Sarkhon area in Charmahal and Bakhteyari Province)", *J. Geography Development*, **4**, 131-146. https://doi.org/10.22111/GDIJ.2004.3888.

Karami, F., Bayati Khatibi, M., Khairizadeh, M. and Mokhtari Asl, A. (2018), "Evaluating the effectiveness of the support vector machine algorithm in landslide susceptibility zoning of Aharchai catchment area", *Geography Environ. Hazards*, **32**, 1-17. https://doi.org/10.22067/GEO.V8I4.83263.

Kornejadi, I. and Pourqasmi, H. (2015), "Landslide susceptibility assessment using data-mining models (case study: Chelchai watershed)", *Watershed Eng. Management Scientific-Res. J.*, **11**(1), 28-42. https://doi.org/10.22092/ijwmse.2019.118436.

Lee, S., Hong, S.M. and Jung, H.S. (2017), "A support vector machine for landslide susceptibility mapping in Gangwon Provine, Korea", *Susceptibility*, **9**(48), 1-15. https://doi.org/10.3390/su9010048.

Lin, K.Y., Lin, T.K. and Lin, Y. (2020), "Real-time seismic structural response prediction system based on support vector machine", *Earthq. Struct*., **18**(2), 163-170. https://doi.org/10.12989/eas.2020.18.2.163

Mafian, S., Huat, B.B.K. and Ghiasi, V. (2009), "Evaluation on root theories and root strength properties in slope stability", *Eur.*

J. Scientific Res., **30**(4), 594-607.

Malamud, B.D., Turcotte, D.L., Guzzeti, F. and Reicenbach, P. (2004), "Landslide inventories and their statistical properties", *Earth Surf. Process*, **29**, 687-711. https://doi.org/10.1002/esp.1064.

Nojavan, M., Shahzeidi, S.S., davudi, M. and Aminroaya, H. (2019), "Zonation of landslide risk using fuzzy and AHP model (Case Study: Komeh Basin)", *Isfahan Province. Quantitative Geomorphological Research*, **7**(4), 142-159.

Ost, L., Van-Den, E.M., Poesen, J. and Vanmaercke-Gottigny, M. C. (2003), "Characteristics and spatial distribution of large landslides in the Flemish Ardennes (Belgium)", *Zeitschrift für Geomorphologie N.F.*, **47**(3), 329–350. https://doi.org/10.1127/zfg/47/2003/329.

Pavel, M., Nelson, J.D. and Fannin, R.J. (2011), "An analysis of landslide susceptibility zonation using a subjective geomorphic mapping and existing landslides", *Comput. Geosci*., **37**(4), 554-566. https://doi.org/10.1016/j.cageo.2010.10.006.

Peng, L., Ruiqing, N., Bo, H., Xueling, W., Yannan, Z. and Runqing, Y. (2014), "Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China", *Geomorphology*, **204**, 287-301. https://doi.org/10.1016/J.GEOMORPH.2013.08.013.

Pontius, R.G. and Schneider, L.C. (2001), "Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agriculture", *Ecosyst. Environ.*, **85**(1), 239-248. https://doi.org/10.1016/S0167-8809(01)00187-6.

Pourghasemi, H.R., Jirandeh, A.G., Pradhan, B., Xu, C. and Gokceoglu, C. (2013), "Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran", *J. Earth Syst. Sci.*, **122**(2), 349-369. https://doi.org/10.1007/s12040-013-0282-2.

Samui, P. (2008), "Support vector machine applied to the settlement of shallow foundations on cohesionless soils", *J. Comput. Geotech.*, **35**, 419-427. https://doi.org/10.1016/j.compgeo.2007.06.014.

Shokri, S., Mohammad Taqi., S. and Mehdi, A.M. (2012), "Presenting a combined method of data preprocessing in the regression vector machine to predict the quality of refined diesel", *Oil Res.*, **23**(75). 116-102.

Sidle, RC. and Ochiai, H. (2006), "Landslides: Processes, prediction, and land use", *Water Resour. Monograph*, **18**.

Surmiri, A., Nayebi, A., Rokhgireh, H. and Varvani-Farahani, A. (2020), "Anisotropic continuum damage analysis of thin-walled pressure vessels under cyclic thermo-mechanical loading", *Struct. Eng. Mech*., **75**(1), 101-108. https://doi.org/10.12989/sem.2020.75.1.101.

Taner San, B. (2014), "An evaluation of SVM using polygon-based random sampling in landslide Susceptibility mapping: The Candir catchment area (western Antalya, Turkey)", *Int. J. Appl. Earth Observation Geoinform.*, **26**, 399-412. https://doi.org/10.1016/j.jag.2013.09.010.

Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B. and Revhaug, I. (2015), Spatial Prediction Models For Shallow Landslide Hazards: A Comparative Assessment of The Efficacy of Support Vector Machines, Artificial Neural Networks, kernel Logistic Regression, and Logistic Model Tree, Landslides.

Tsai H.C. and Liao, M.C. (2019), "Knowledge-based learning for modeling concrete compressive strength using genetic programming", *Comput. Concrete*, **23**(4), 255-265. https://doi.org/10.12989/cac.2019.23.4.255.

Vapnik, V.N. (1995). The Nature of Statistical Learning Theory, *Springer Verlag*, New York.

Varnes, D.J. (1984), Landslide hazard zonation: A review of Principle and Practice, Unexco, Paries, 63.

Xue, X. (2019), "Application of a support vector machine for prediction of piping and internal stability of soils", *Geomech. Eng.*, **18**(5), 493-502. https://doi.org/10.12989/gae.2019.18.5.493.

Yamani, M., Ahmedabadi, A. and Zare, G. (2011), "Application of support vector machine algorithm in landslide risk zoning. (Case study: Derkeh watershed)", *Geography Environ. Hazards*, **3**, 125-142.

Yao, X., Tham, L.G. and Dai, F.C. (2008), "Landslide susceptibility mapping based on Support Vector Machine: a case study on natural slopes of Hong Kong. China", *Geomorphology*, **101**(4), 572-582. https://doi.org/10.1016/j.geomorph.2008.02.011.

Yesilnacar, E. and Topla, T. (2005), "Landslide susceptibility mapping a comparison of logistic regression and neural networks methods in a medium scale (Turkey)", *Eng. Geol.,* **79**, 251-266. https://doi.org/10.1016/j.enggeo.2005.02.002.

Yousefi, M. and Carranza, E.J.M. (2015a), "Fuzzification of continuous-value spatial evidence for mineral prospectivity mapping", *Comput. Geosci.*, **74**, 97-109. https://doi.org/10.1016/j.cageo.2014.10.014.

Yousefi, M. and Carranza, E.J.M. (2015b), "Geometric average of spatial evidence data layers: A GIS-based multi-criteria decision-making approach to mineral prospectivity mapping", *Comput. Geosci*., **83**, 72-79. https://doi.org/10.1016/j.cageo.2015.07.006.

Yousefi, M. and Nykänen, V. (2016), "Data-driven logistic-based weighting of geochemical and geological evidence layers in mineral prospectivity mapping", *J. Geoch. Exploration*, **164**, 94-106. https://doi.org/10.1016/j.gexplo.2015.10.008.

Yousefi, M., Carranza, E.J.M., Kreuzer, O.P., Nykänen, V., Hronsky, J.M.A. and Mihalasky, M.J. (2021), "Data analysis methods for prospectivity modelling as applied to mineral exploration targeting: State-of-the-Art and Outlook", *J. Geochem. Exploration*, **229**, 106839. https://doi.org/10.1016/j.gexplo.2021.106839.

Yousefi, M., Kreuzer, O.P., Nykänen, V. and Hronsky, J.M.A. (2019), "Exploration information systems—a proposal for the future use of GIS in mineral exploration targeting", *Ore Geol. Reviews*, **111**, 103005.

GC