Math Geosci (2012) 44:47–70
DOI 10.1007/s11004-011-9379-9

Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy

Cristiano Ballabio · Simone Sterlacchini

Received: 14 December 2009 / Accepted: 29 November 2011 / Published online: 3 January 2012

Abstract The aim of this study is the application of support vector machines (SVM) to landslide susceptibility mapping. SVM are a set of machine learning methods in which model capacity matches data complexity. The research is based on a conceptual framework targeted to apply and test all the procedural steps for landslide susceptibility modeling from model selection, to investigation of predictive variables, from empirical cross-validation of results, to analysis of predicted patterns. SVM were successfully applied and the final susceptibility map was interpreted via success and prediction rate curves and receiver operating characteristic (ROC) curves, to support the modeling results and assess the robustness of the model. SVM appeared to be very specific learners, able to discriminate between the informative input and random noise. About 78% of occurrences was identified within the 20% of the most susceptible study area for the cross-validation set. Then the final susceptibility map was compared with other maps, addressed by different statistical approaches, commonly used in susceptibility mapping, such as logistic regression, linear discriminant analysis, and naive Bayes classifier. The SVM procedure was found feasible and able to outperform other techniques in terms of accuracy and generalization capacity. The over-performance of SVM against the other techniques was around 18% for the cross-validation set, considering the 20% of the most susceptible area. Moreover, by analyzing receiver operating characteristic (ROC) curves, SVM appeared to be less prone to false positives than the other models. The study was applied in the Staffora river basin (Lombardy, Northern Italy), an area of about 275 km^2 characterized by a

C. Ballabio (✉)
Environmental and Land Sciences Dept., Università degli Studi di Milano-Bicocca, Milano 20126, Italy
e-mail: cristiano.ballabio@unimib.it

S. Sterlacchini
Institute for the Dynamic of Environmental Processes, National Research Council (CNR-IDPA), Milano 20126, Italy

very high density of landslides, mainly superficial slope failures triggered by intense rainfall events.

Keywords Support Vector Machines · Landslide susceptibility mapping · Spatial prediction · Cross-validation

1 Introduction

The main aim of landslide susceptibility modeling is to identify areas prone to future mass movements, on the basis of previous knowledge about the spatial distribution of past occurrences. The basic idea behind this approach is to identify areas where a particular combination of physical properties could indicate the predisposition towards similar events in the future. This goal is usually achieved by relating the spatial distribution of past landslides with the spatial distribution of supporting morphometric, geological, geomorphological, and land use properties to produce a susceptibility map. As the relationship between these properties and the spatial probability of occurrence is usually unknown, susceptibility prediction is achieved through the use of statistical modeling or machine learning techniques. Landslide susceptibility assessment can be seen as a typical classification problem whose main aim is to separate instances belonging to two classes: (i) areas where landslides occurred or are likely to occur in the future; and (ii) areas where landslides did not occur or are not likely to occur in the future.

Several researchers applied different classification techniques to landslide susceptibility prediction and many of these are reviewed in Carrara et al. (1995), Guzzetti et al. (1999), Aleotti and Chowdhury (1999), Brenning (2005). Most of these techniques fall into two main categories: the former comprises simple ranking functions based on Bayes' theorem, like naive Bayes (NB), or the weights of evidence (WofE) (Goodacre et al. 1993; Bonham-Carter 1994) classifiers; the latter uses parametric statistical techniques such as linear discriminant analysis (LDA) (Carrara et al. 2003; Guzzetti et al. 2006) or logistic regression (LR) (Ayalew and Yamagishi 2005; Eeckhaut et al. 2006; Carrara et al. 2008). The former category uses non-parametric techniques and is thus not bounded by a particular distribution. However, these approaches suffer from conceptual limitations related to predictors independence assumption, which is usually violated in practice. Moreover, predictors classification is required by these models, a practical issue which reduces the informative content of the input and limits model accuracy. On the contrary, parametric models can deal with continuous variables; however collinearity remains a limit for these models and linearity is frequently not met in practice. Further, these models are strictly parametric and must satisfy several restrictive assumptions on data distribution. In this study, we propose the application of support vector machines (SVM) (Vapnik 1995, 1998; Cherkassky and Mullier 2007) to landslide susceptibility mapping. SVM are based on a non-linear transformation of the covariates in a high-dimensional space, where different classes are linearly separable. Recently, several authors applied two SVM-related methods, namely support vector classification and support vector regression, to a broad range of spatial problems (Kanevsky and Canu 2000; Smirnoff et al. 2008;

Ballabio 2009; Samui and Sitharam 2010; Pozdnoukhov et al. 2011). When used properly, SVM are able to outperform many classifiers (Meyer et al. 2003). To assess the performance of SVM, we chose to compare the performances of this technique with the ones of the techniques most commonly utilized by geomorphologists to produce maps of landslide susceptibility. To keep the outcome of the study coherent with the existing literature, we chose to use performance metrics of common use in this field of study.

1.1 Study Area

The study area is entirely within the Staffora river basin (Lombardy), an area of about 275 km^2 belonging to the northern sector of the Apennines, facing the Po Plain (Fig. 1). The area is characterized by outcropping formations mainly composed by clay and marl-rich terrigenous formations (Beatrizzotti et al. 1969; Braga et al. 1985) which are subject to frequent landsliding. A thoroughly geomorphological and historical description of landslides in the study area, along with a temporal analysis of their evolution, has been performed by Carrara et al. (2003), Meisina et al. (2006) from 1950 to 2000. The authors have shown that most of the slope failures were triggered in 1951, 1959, 1976–1977, 1993, 1997, 2000: all periods characterized by heavy rainfall events following dry periods. These landslides which can be classified in terms of types, materials involved, estimated volumes, velocities and degree of activity (Cruden and Varnes 1996), are mostly represented by active mud-earth flows, prediction of which is the aim of this study. Mud-earth flows are widespread all over the study area occupying over 36 km^2 with 1574 observed occurrences.

Some authors have already produced maps of landslide susceptibility for the study area, by applying different techniques. Carrara et al. (2003) performed a Linear Discriminant Analysis on 2245 morphological-lithological terrain units. Sterlacchini et al. (2004), Poli and Sterlacchini (2007) applied the weights of evidence modeling technique on a pixel base. In the first case, the terrain units used in the analysis may limit the spatial resolution of the resulting susceptibility map and, as these units could be further subdivided into smaller units or aggregated into larger ones, their choice is somehow ambiguous. Moreover, the linear discriminant analysis approach is strictly applicable only when the underlying variables are jointly normal with equal covariance matrices and devoid of outliers, a situation which does not commonly occur in practice, where class imbalance and sampling bias (Oommen et al. 2011) is a common issue even for less restrictive methods. Finally, the assumption of linearity between response and variables is usually not realistic. In the case of Poli and Sterlacchini (2007), the authors were obliged to classify the original continuous predictive variables (due to the model requirements) thus reducing the informative content of their input data and, in so doing, the predictive accuracy. Both cited methodologies also require the explanatory variables to be non-collinear. This is a concerning issue as most of the morphometric, geological and land-use variables are inherently correlated (i.e. slope and lithology, slope and land use, altitude and land cover, etc.). The use of SVM can help to overcome some assumptions and limitations implied in the models above mentioned, improving the accuracy of the final predictions. In particular, SVM are an inherently robust and collinearity resistant technique.

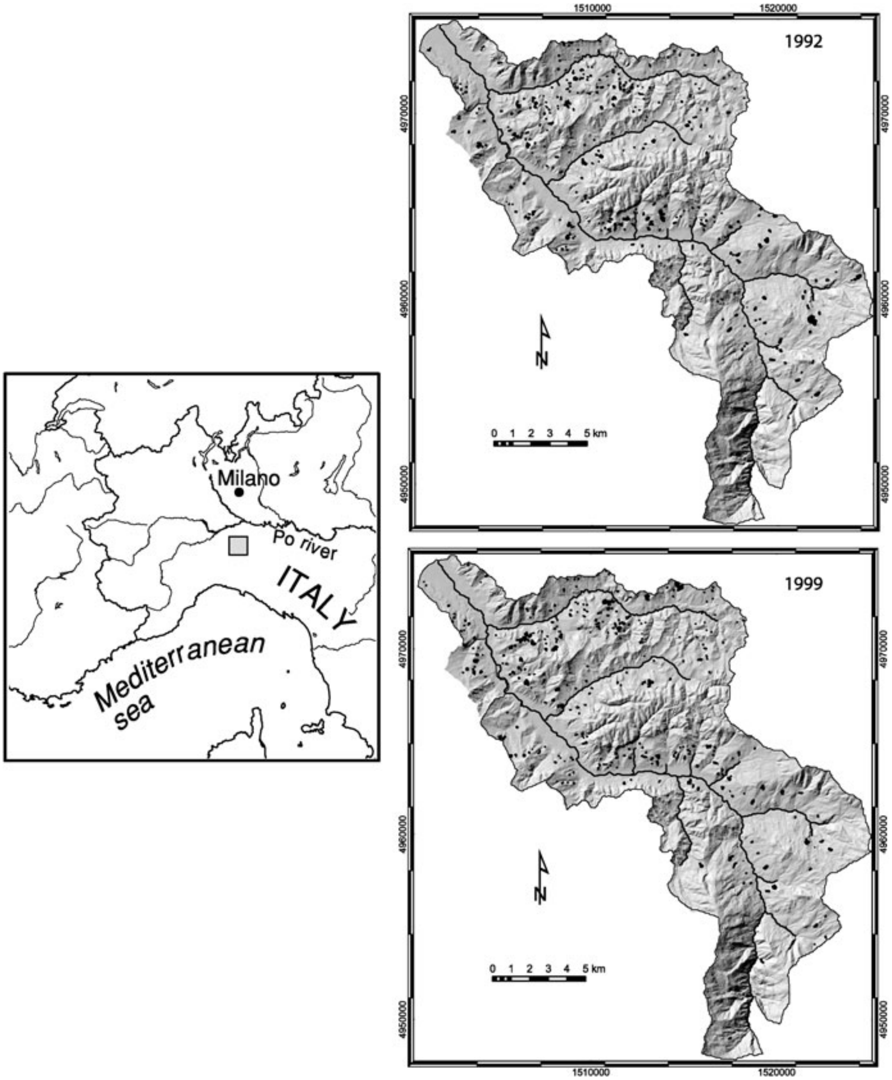


Fig. 1 On the *left*, the Staffora river basin location in northern Italy is indicated by the gray square. On the *right*, shaded relief maps of the study area with the two sets of occurrences; in the *upper-right* figure are the events that occurred before 1992, while in the *lower-right* figure are the events that occurred between 1992 and the end of 1999

1.2 Predictor Variables and Landslide Inventories

In order to predict landslide susceptibility, a set of spatially distributed predisposing variables was related with the spatial distribution of past landslide occurrences. More precisely, two inventories of active and quiescent landslides (Cruden and Varnes 1996), showing the geographical distribution of mass movements in two different time periods. 1992 and 1999, have been used in the model (Fig. 1). The events that occurred before 1992 will be named Set1; this set contains 378 distinct events with an average per-event area of 14400 m^2 and a total area of 5.5 km^2. Set2 contains the events which occurred between 1992 and 1999 and contains 336 events with an average per-event area of 12250 m^2 and a total area of 4.6 km^2. Since one goal of the study is to assess the forecasting capability of different models, we utilized only Set1 for model training and parameters fitting by randomly splitting the model into two subsets, a training set and a testing set. This subdivision was repeated in a k-fold procedure in order to avoid the possible influence of a particular data configuration. Set2 was utilized exclusively for the cross-validated performance evaluation of the fitted models.

All predictors used in the analysis are summarized in Table 1, and can be subdivided into two groups: morphometric terrain features and geo-lithological and land use features. Morphometric features were derived from the regional digital elevation model (DEM), available for the study area, with a ground resolution of 20 m. From the DEM, the following terrain features were derived: (i) altitude and slope (Zevenbergen and Thorne 1987), internal relief (expressed as the maximum elevation change per unit area), which are key-factors usually used in landslide susceptibility assessment studies; (ii) compound terrain index (CTI) (McKenzie and Ryan 1999) also known as the Wetness index is the logarithm of the ratio between upslope contributing area and slope gradient. The CTI models the soil water content and water saturation within landscapes, as flows need water saturation to occur; (iii) insolation (Dubayah and Rich 1995), expressed in $\mathrm{kW\,h\,yr^{-1}}$, calculated on a seasonal basis. Direct insolation is the amount of radiation reflected by the terrain: its maximum intensity is on south-facing slopes. This parameter plays different roles on landslide susceptibility: it directly influences the vegetation cover and the micro-climate; it indirectly controls water availability (and soil moisture, after all) by snow melting and evapotraspiration; (iv) distance from channel base level, which is the limiting level below which a stream cannot erode its channel. Local base levels occur where the stream meets a resistant rock body, or where the stream empties into a lake. This variable identifies areas where mud earth-flows triggering is less likely as these events cannot occur at the local channel base level; (v) convergence index, which characterizes soil and debris erosion and deposition within the landscape; (vi) morphological protection index (MPI) expresses the positive openness (Yokoyama et al. 2002) which expresses the degree of dominance or enclosure of a location within the landscape; (vii) downslope distance gradient (Hjerdt et al. 2004) which expresses how far a given amount of water must travel in the landscape to lose a certain amount of potential energy. This index differs from CTI because it takes into account the downslope topography, which may enhance or dampen local drainage; and (viii) stream power Index, defined as $SPI = \ln(C_a \cdot \tan(G))$ where C_a is the catchment area and G is slope steepness. This index is directly proportional to stream flow energy and as such expresses the erosive potential of overland flow.

Along with morphometric-related features, the spatial distribution of ten lithological units outcropping in the study area was derived from the regional geological map (Beatrizzotti et al. 1969; CARG project 1992). The presence of clay and marl-rich sedimentary formations clearly influences the overall susceptibility in the study area. Structural data (bedding planes, faults, thrusts, etc.) were derived from existing geological maps and added to the set of predictor variables. The influence of the attitude

Table 1 List of covariates included in the analysis. Continuous variables are summarized by their minimum, average and maximum value in the overall study area, in the occurrences of Set1 and in the occurrences of Set2 respectively. Discrete variables are summarized by the relative cover of each class within the study area, within Set1 and within Set2

Name	Continuous variables								
	Area			Set1			Set2		
	min	average	max	min	average	max	min	average	max
Channel base level	0	17.27	163.7	0	10.59	94.53	0	12.38	107.8
Convergence index	−100	0.2	100	−57.65	−3.32	92.4	−87.5	−3.6	71
CTI	1.29	6.69	21.9	3.7	6.7	17.2	3.1	6.6	17.6
Distance from faults	0	767	5916	0	566	1926	0	564.9	4587
Distance from thrust	0	1890	7851	0	1414	5341	0	1804	6310
Downslope distance gradient	0	0.325	1.466	0	0.27	0.89	0	0.312	1.056
Elevation	150	617	1690	200	546	1120	197	474	1246
Insolation	0.85	2.46	3.31	1.45	2.47	2.99	1.48	2.42	2.99
Internal relief	0	31.09	190	4	26.5	100	3	28	100
Morphological protection Idx	0	0.325	1.46	0	0.13	0.5	0	0.13	0.59
Slope	0	29.5	147	0	26	90	0	27.7	91
Stream power index	0.02	61.3	10^5	0.02	67.6	3272	0.02	73	4531

Discrete variables (relative surface)

Class name	Area	Set1	Set2
Land.1	3.1%	2.9%	1.5%
Land.2	34.1%	49.6%	52.2%
Land.3	13.0%	4.2%	1.7%
Land.4	2.8%	0.2%	0.2%
Land.5	12.0%	0.3%	0.3%
Land.6	15.1%	9.9%	8.0%
Land.7	9.2%	11.6%	5.4%
Land.8	6.3%	19.3%	23.9%
Land.9	1.0%	0.0%	0.6%
Land.10	3.3%	1.9%	6.1%
Geo.1	2.3%	1.9%	%2.5
Geo.2	7.1%	0.07%	0.4%
Geo.3	14.8%	6.3%	4.8%
Geo.4	23%	29.5%	36.1%
Geo.5	9.4%	28.2%	33.3%
Geo.6	8.9%	0.3%	0.3%
Geo.7	11.5%	10.3%	9.6%
Geo.8	1.6%	2.5%	2.5%
Geo.9	8.0%	20.8%	9.1%
Geo.10	13.3%	0.07%	1.2%
Str.1	23.9%	18%	19.3%
Str.2	21.4%	11.4%	13.2%
Str.3	31.7%	59.6%	50.2%
Str.4	23.0%	11.0%	17.3%

of the strata is rather evident: if bedding planes are oriented in the same direction of the slope, the slope is increasingly prone to failure. Thus, four classes describing different orientations and inclinations of rock strata with respect to slope were added among the predictors. Moreover, as geomechanical stress can result in rock fracture increasing landslide susceptibility, distance from faults and thrusts was also added to the set of predictors. Finally, as land cover and land use can influence hydrologic balance, soil erosion, and run-off, ten land use classes were derived from the DUSAF project (DUSAF project 2003) and added to the set of predictive variables. The data frame of Table 1 was then converted into a dummy matrix x of dimension $M \times N$, where $M = 36$ is the number of independent variables and N is the sample size.

In the present case, the classification task aims at discriminating between the landslide prone areas (class 1) and the rest of the study area (class -1). We can notice from Fig. 1 that the ratio between landslide area and the overall area is generally very low, thus the unaffected area can be resumed by a random sample of the total population, excluding landslide areas. It should be noted that class imbalance and sampling bias can influence the outcome of several models (Oommen et al. 2011; Lin et al. 2002), so in order to have comparable sample size, the dimension of the -1 sample has to be similar to the number of the pixels in the landslide areas (susceptible/not-susceptible). This sample was taken as to have a distribution which is a normal approximation of the distribution of the topographic and geo-environmental features (as summarized in Table 1) within the study area.

2 Methodology

2.1 Support Vector Machines

Support vector machines are a set of machine learning techniques based on the concept of optimal separating hyperplane, developed by Vapnik (1995). SVM can be thought as non-linear classifiers whose goal is to find the widest margin between two classes in feature space.

When dealing with landslide susceptibility mapping, by the use of statistical approaches, two major issues can be identified. The first resides in the strong assumptions of the classic parametric models, which usually assume data linearity. If the relation between dependent and independent variables is not linear, these models will perform poorly. The second issue is model overfitting, which implies that fitting a model with an arbitrary high number of degrees of freedom, aiming only at minimizing the error, will result in a model fitting not only the general trend, but also the random noise. Thus, the model will be poorly predictive on unobserved data, as it is not discriminating between the trend and the noise. This issue is related to the curse of dimensionality (Bellman 1961; Hastie et al. 2001; Evangelista et al. 2006) that can be explained as the effect of the exponential reduction of observational informative content with the increase of dimensions. This is due to the sampling density that is proportional to $N^{1/M}$ where M is the dimension of the input space and N is the sample size. This behavior is shown in Fig. 2 where prediction error is depicted as a function of the model complexity. Low complexity models, like linear models, may have higher error rates if the model fits the data

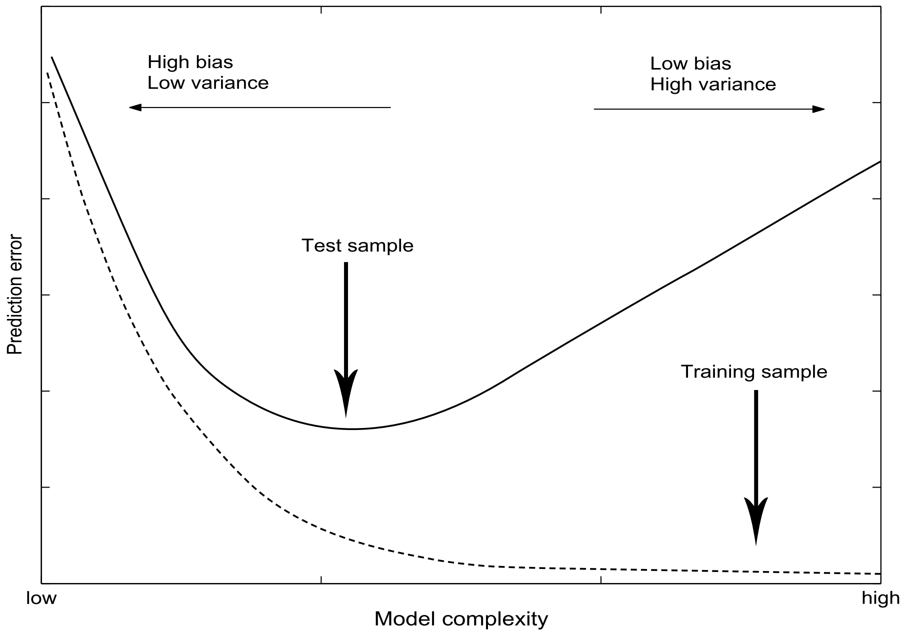


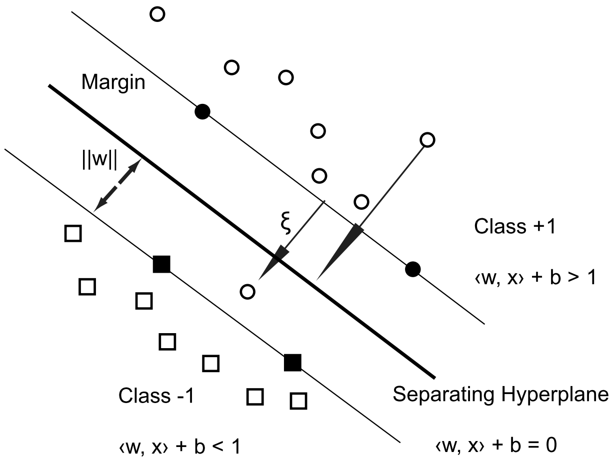
Fig. 2 Graphical representation of the relation between training error and testing error, as a function of model complexity. This is a general model in machine learning systems

poorly: in this case, error is due to model bias. On the other hand, increasing model complexity may arbitrarily reduce model bias and error for the training sample but the resulting model is over-fitted and has a higher variance. The optimal model complexity corresponds to the one which minimizes the error on unobserved test sample, which corresponds to the solid line of Fig. 2.

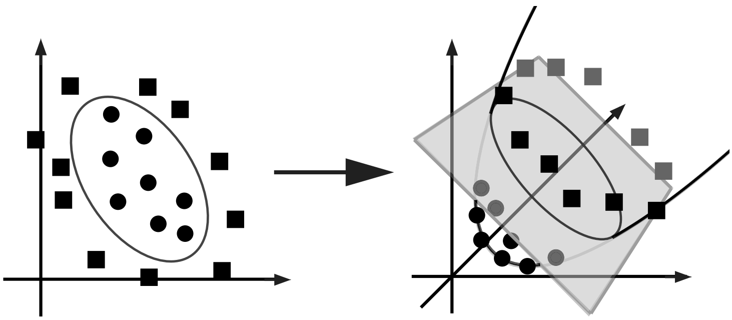
The SVM approach aims at minimizing both the error test and the model complexity. In other words, SVM are data-dependent models, which means that the model capacity is tuned to match data complexity. This paradigm, also termed: structural risk minimization (Vapnik 1995; Cherkassky and Mullier 2007), is the basis of the SVM learning algorithm. In this study, our aim is to discriminate between susceptible (1) and not susceptible (-1) pixels. In this case, SVM have to separate the N observations into different classes using a function f expressing a hyperplane defined as

$$f(x) = \langle \mathbf{w}, x \rangle + b = \sum_{i=1}^{N} (w_i x_i) + b = 0, \tag{1}$$

were, the vector x represents the independent variables (i.e. morphometric and geo-environmental predictors), while the vector w represents the vector of weights which has to be found by the model, and b is a constant. The concept behind (1), is rather simple. Suppose that we want to discriminate between two classes. The most intuitive approach is to find a function which acts as a threshold between the instances of the two different classes. As the hyperplane is defined by being equal to zero, pixels



(a)



(b)

Fig. 3 (**a**) Illustration of the Optimal Separating Hyperplane and Support Vectors. (**b**) Graphical depiction of the Kernel trick mapping data from two to three dimensions

classification is given by their sign. The simplest possible function is a straight line in two-dimensional space or a plane/hyperplane in more dimensions. This is exactly the approach of linear methods. The SVM approach attempts to find the maximum separating hyperplane, which is the plane passing through the widest gap between the instances of the two classes. Figure 3(a) illustrates this concept: the separating hyperplane is just one of the possible planes which separates the instances of the two classes with the same amount of misclassification error. This hyperplane is known as optimal separating hyperplane (Vapnik 1998). However, it is not always possible to define a separating hyperplane, because the two classes can, and usually do, overlap each other. This situation is illustrated in Fig. 3(a), where one of the instances lies within the margin. To manage this issue, it is possible to introduce slack variables (Cortes and Vapnik 1995; Vapnik 1998), defined as

$$\xi_i > 0 \quad (i = 1, \dots, N). \tag{2}$$

Slack variables are linear penalty terms aimed at minimizing the amount of the misclassified instances. Conceptually, the slack variables act in a way similar to the minimization of the root mean squared error (RMSE) in linear regression. The distance between a point x on the hyperplane, as in (1), and a point x' is found by minimizing $\|x - x'\|^2$. To achieve this goal, it is necessary to minimize both the complexity of the model, measured by $\|\mathbf{w}\|^2$ and its classification error $\max(\xi)$. A possible way to address this problem is the ν-SVM (Chen et al. 2005), in which the parameter ν expresses the lower and the upper boundary of the number of observations that could be taken as support vectors and the number of points allowed to lie on the wrong side of the hyperplane, respectively. Parameter ν is associated with variable ρ which is to be optimized during the training stage. Thus, the problem formulation could be written as

$$\begin{aligned} &\text{minimize } R_{w,\xi,\rho} = \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \frac{1}{N}\sum_{i=1}^{N}\xi_i \\ &\text{subject to} \quad \begin{cases} y_i(\langle \mathbf{w}, x_i\rangle) + b \geq \rho - \xi_i & (i = 1, \ldots, N) \\ xi \geq 0 & \rho \geq 0, \end{cases} \end{aligned} \tag{3}$$

where R is the structural risk, N is the number of observations, the term $\|\mathbf{w}\|^2$ is the Euclidean norm which expresses the distance between the hyperplane and the margins and $y_i = [-1, +1]$ are the classes of the observed pixels. It can be shown (Schölkopf 2001) that the effect of ρ can be explained considering that for $\xi = 0$ the constraint of (3) imposes that the two classes are separated by the margin $\frac{2\rho}{\|\mathbf{w}\|}$. This formulation is particularly attractive, as it can be shown that ν represents an upper limit for the margin error $\xi > 0$ and a lower limit for the number of SV (Schölkopf 2001). Thus, one can approximately define the percentage of allowed training misclassification beforehand, by changing the value of ν. The solution of $\mathbf{w}$ for (3) for the SVM can be found using the method of Lagrange multipliers expressing $\mathbf{w}$ as

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i x_i, \tag{4}$$

where α_i are the Lagrange multipliers.

Although SVM are hard $[+1, -1]$ classifiers, several approaches have been proposed to derive smooth probabilities from their outcome. The approach used in this work is based on the sigmoid transformation by Platt (1999) which has been successfully applied in other environmental related case studies (Kanevski et al. 2009; Pozdnoukhov et al. 2011)

$$p(y = 1|x) = \frac{1}{(1 + \exp(a \cdot f(x) + b))}, \tag{5}$$

where a and b are constants. In the present case, the parameters were chosen as to maximize the negative log-likelihood, which results in (5) providing estimates of the class-conditional posterior probabilities. Given its features, in the present study we applied probabilistic ν-SVM to predict mud earth-flow susceptibility.

2.2 Kernel Function

In susceptibility assessment, some of the most frequently used classification techniques are based on linear functions, as in Logistic Regression and Discriminant Analysis. However, the linear model could be considered at best as an approximation for most of the real world data. The SVM approach is rather different. As stated before, the hyperplane is derived from the values calculated through the dot product. However, it is possible to use a dot product which is able to project the data pairs into a space in which the classes are linearly separable. This procedure is known as the Kernel Trick (Schölkopf 2001). Given a projection $\Phi : X \rightarrow H$ from the original space X to a space H, the dot product $\langle \Phi(x_i), \Phi(x_j) \rangle$ can be represented by a kernel function k (Schölkopf 2001)

$$k(x_i, x_j) = \bigl\langle \Phi(x_i), \Phi(x_j) \bigr\rangle. \tag{6}$$

Since the weight vector can be expressed as a linear combination of the training points

$$\mathbf{w} = \sum_{i=1}^{N} \alpha \Phi(x_i), \tag{7}$$

Equation (1) can be written as

$$f(x) = \sum_{i=1}^{N} \alpha_i k(x_i, x_j). \tag{8}$$

Usually the kernel is chosen so that the f will be linear in feature space. This can be achieved by using the same criterion optimization as in (3). This situation is illustrated in Fig. 3(b): on the left of the figure, we have two classes in a two-dimensional space. In this space, it is not possible to separate these two classes using a linear function (a straight line). However, if we project the data into a three-dimensional space, we can successfully separate the two classes with a plane (which is a linear function in three dimensions), as shown in the right part of Fig. 3(b). Through the use of kernel mapping, SVM operating in an arbitrary number of dimensions can be built, thus making it possible to find a separating hyperplane even for highly complicated datasets. The radial basis function (RBF) Gaussian kernel

$$K(x_i, x_j) = \exp\bigl(-\sigma \|x_i - x_j\|^2\bigr) \tag{9}$$

is an example of mapping function (Aizerman et al. 1964; Nilsson 1965) and is the kernel utilized in this study due to its good generalizing properties. Given that the computational time required by a grid search is not excessive, both kernel σ and SVM ν parameters have been tuned using a fine grid search as suggested by (Hsu et al. 2007) using the testing set.

2.3 Variable Selection and Investigation

Variable selection is a crucial part of model optimization procedure, aimed at retaining only significant predictors in the analysis. This practice is widely used in statistics and its goal is to remove uninformative or noisy features from the final model. Variable selection in linear techniques is usually carried out to minimize the RMSE on the basis of some assumptions about data distribution. For instance, in landslide susceptibility assessment, Guzzetti et al. (2006) proposed the use of linear discriminant functions to rank possible predictors. A general way to deal with this problem is to maximize the separating margin between two classes (Fig. 3(a)). The aim is to select a set of features which maximizes the discriminant power and thus produces the widest margin between the two classes. Recursive feature elimination (RFE) (Guyon et al. 2002; Guyon and Elisseeff 2003) is a backward sequential feature elimination technique, which starts with all the features and discards one feature at a time. The squared coefficients of the weight vector w_j^2 ($j = 1, \dots, p$) are used as the feature ranking criteria as features with larger weights are more informative. The magnitude of w_j^2 corresponds to the change of the criterion R of (3) when the jth feature is removed

$$J = \frac{1}{2}\|\mathbf{w}\|^2 - \nu\rho + \sum_{i=1}^{N} \xi_i. \tag{10}$$

It can be shown (LeCun et al. 1990), that $\Delta J(j) \approx (\Delta w_j)^2$ where $J(j)$ is the value of J when the jth feature is removed, the relation can be rewritten as

$$J(j) \approx J + w_j^2. \tag{11}$$

Removing the feature with the smallest w^2 will produce the least increase of J; in other words RFE seeks to find the subset of variables which yields minimal J, thus increasing generalization (Zhou and Tuck 2007).

To test the reliability of the RFE procedure, a Wilcoxon rank sum test (Wilcoxon 1945; Mann and Whitney 1947) was performed. This procedure can rank input features of Table 1 on the basis of their p-values. The Wilcoxon rank-sum test is a non-parametric alternative to the two sample t-test which is based solely on the order in which the observations from the two samples fall. The implementation of the test follows the approach proposed by Boulesteix (2007), following the notation of the previous sections. The test assesses the equality of the median of the landslide prone and non-prone areas as two independent samples $y_i = -1$ and $y_i = 1$ in respect to a feature M_j. The test statistic is then given as

$$W = \sum_{i:y_i=-1} RK_i, \tag{12}$$

which is the sum of the ranks (RK_i) of the observations belonging to class -1. The p-value of the test is derived from the asymptotic result

$$\frac{W - N_{-1}(N+1)/2}{\sqrt{N_{-1}N_1(N+1)/12}} \sim \mathcal{N}(0, 1) \tag{13}$$

where N_{-1} and N_1 are the number of observations belonging to class -1 and 1, respectively. The p-value estimation is made more robust by the use of Monte-Carlo

Table 2 Collinearity check of input variables. The VIF values were calculated for all the candidate variables. The VIF values show that the effect of collinearity on the model should not be significant

	VIF	Df	$\mathrm{VIF}^{(\frac{1}{2}Df)}$
Elevation	1.75	1.00	1.32
Channel b l	4.50	1.00	2.12
Convergence index	1.69	1.00	1.30
Direct insolation	4.85	1.00	2.20
Distance from faults	1.27	1.00	1.13
Distance from thrust	1.89	1.00	1.37
Downslope gradient	1.80	1.00	1.34
Geology	24.95	9.00	1.20
Morphological protection idx	1.93	1.00	1.39
Slope	2.72	1.00	1.65
Stream power idx	4.64	1.00	2.15
Structure	3.26	3.00	1.22
Total insolation	1.85	1.00	1.36
Land use	2.87	9.00	1.06
LS factor	1.55	1.00	0.77

cross-validation (Boulesteix 2007). The p-values are then used to perform a univariate relevance ranking of environmental variables (Table 1) on the basis of their p-values which can be compared to the results of the RFE procedure.

Since collinearity among predictor variables can influence the model outcome, the predictor variables were checked for collinearity using variance inflation factors (VIF) (Marquardt 1970; Fox and Monette 1992) in the data preparation stage. In general, a VIF value of 10 or more is taken as proving high collinearity, although some caution should be taken before removing the variable from analysis (O'Brien 2007). The VIF values shown in Table 2 evidence that collinearity should not be a significant issue for this data set. The only variable which shows a significant (>10) degree of collinearity is the layer of the lithology units. However, this high value is due to the high number of degrees of freedom and the normalized VIF value is close to the one of the other variables. Given this result, no variable was discarded from the analysis due to collinearity.

2.4 Model Performance Evaluation

The classification capabilities of the SVM and other classifiers have been tested through the use of receiver operation characteristic (ROC) graphs and prediction rate curves. The ROC (Egan 1975; Fawcett 2006) is a graphical analysis of the success rate of a binary classification and it provides useful information about the proneness of a model to generate false positives errors. ROC graphs are built by plotting the ratio between the true positives and the false positives in a two dimensional plane; for discrete classifiers the resulting ROC is expressed by a point. Other classifiers yield a classification probability or score (such as SVM in the present formulation, naive Bayes, etc.) that represents the degree to which an instance is a member of a class. Such scoring classifier can be used with a threshold to produce a discrete classifier: if the classifier output is above a given threshold, the classifier will produce a positive or else a negative. Using a series of increasing threshold values will produce a series of point in ROC space and varying the thresholds by infinitesimal steps between $-\infty$ and $+\infty$ will draw a curve in ROC space. An interesting property of ROC is that the technique is insensitive to changes in class distribution, this means that while other performance metrics, such as accuracy and precision, can change in response to class skewness (Fawcett 2006), ROC curves remain the same. This is an attractive property as class skewness seems to be a common occurrence (Fawcett 2006). The result of the ROC analysis are summarized by the area under curve (AUC), which expresses a complete success in classification for an AUC $= 1$, and random classification for an AUC $= 0.5$.

The prediction rate curves method (Chung and Fabbri 1999; van Westen et al. 2003), have been widely applied for many years to assess the quality of susceptibility mapping, slicing the probability map into equal area classes, each ranked from a minimum to a maximum value. The curves identify the percentage of landslide area within a given probability class. Curve construction is performed by plotting the cumulative percentage of susceptible areas (starting from the highest probability values to the lowest ones) on the x axis and the cumulative percentage of events included in the cross-validation set on the y axis. The steeper the curve, the larger the number of events falling into the most susceptible classes.

3 Results

3.1 Variable Relevance and Probability Map Calculation

The RFE procedure selects 18 different features (out of 36 available) as the optimal subset of covariates. The initial set of variables is shown in Table 1. Figure 4 depicts the features with the highest overall influence on the model as estimated in the RFE procedure; the p-values were calculated using the Wilcoxon test; Fig. 4 also shows a substantial agreement between the ranking outcomes of both RFE and Wilcoxon test and cross-validation RMSE. The grid search approach for SVM parameters identifies an error minimum for $\nu = 0.17$ and $\sigma = 0.19$, a graphical depiction of the testing error surface is shown in Fig. 5. This surface is produced by the averaged testing error of the k testing sets used for parameters search. Once the model parameters have been optimally tuned we applied the trained SVM to the input variables to obtain the map of landslide susceptibility. As shown in Fig. 6, the output of the analysis is the landslide susceptibility map which portrays the spatial probability of event occurrence, the probability is shown by the color scale whose color variation is based on probability distribution quantiles. This probability scale is not the occurrence probability of an event in time, but the probability of a pixel to be classified in a given class by the SVM. So, a pixel with value equal to 0.8 has 80% chance to be classified as 1 and 20% to be classified as a -1.

3.2 SVM Classification Capabilities and Comparison with Other Techniques

To test the reliability of the SVM prediction map and to compare its performance against other techniques, cross-validations were performed on two different landslide

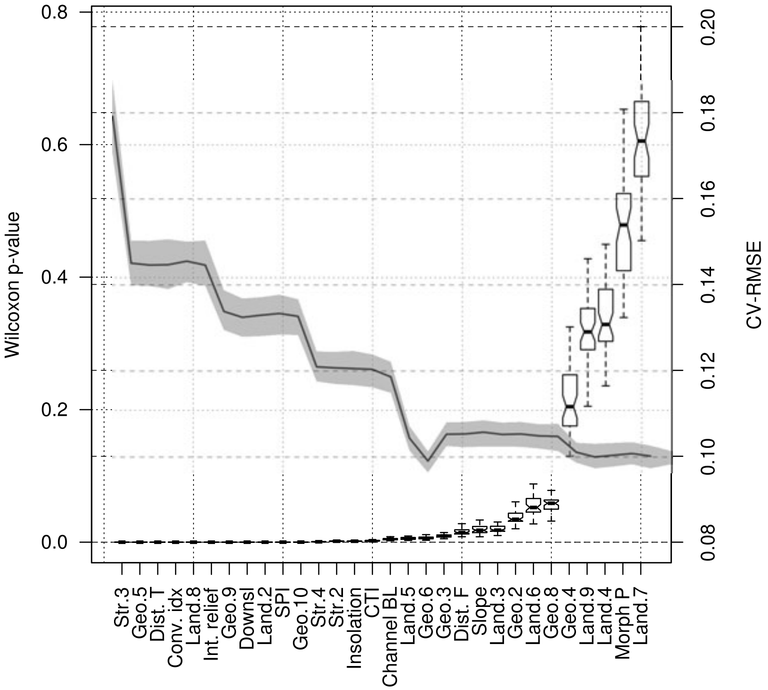


Fig. 4 Comparison of p-values obtained from MCMC Wilcoxon test and RFE ranking. The box plots show the calculated p-values from the Wilcoxon test whose scale is represented on the *left vertical axis*. RFE ranking is shown by the order of the variables on the horizontal axis. Cross-validation error is represented by the *solid line*, where the *gray* bands represent the 0.9 confidence intervals, CV-error scale is shown on the *right vertical axis*. Only the top 30 ranking variables are shown

inventories: the former (Set1, 1992) to verify how much the model fitted the occurrence of past landslides; the latter (Set2, 1999) to test the SVM predictive capability. Indeed, the degree of fit did not express how well the model predicted future landslides, because the events in Set1 were used to construct the prediction map. On the other hand, Set2 comprised only events of the same type but not used in the training stage. A good performance on Set2 would ensure a good predictive capability over events not yet observed which means a good landslide susceptibility prediction.

Success and prediction rate curves and ROC analysis were also used to investigate the outcomes of other three statistical techniques, commonly used in susceptibility mapping, namely, logistic regression, linear discriminant analysis, and the naive Bayes classifier weights of evidence. These techniques were applied using the best set of predictors. Variable selection procedure for logistic regression and discriminant analysis showed that the best set of covariates to be included in these models was equivalent to the one selected by recursive feature elimination, although the relative variable ranking is different; so that, the outcome of the different models could be compared. The susceptibility maps produced by each technique are shown in Fig. 7, also in this case color scale is based on probability distribution quantiles. The re-

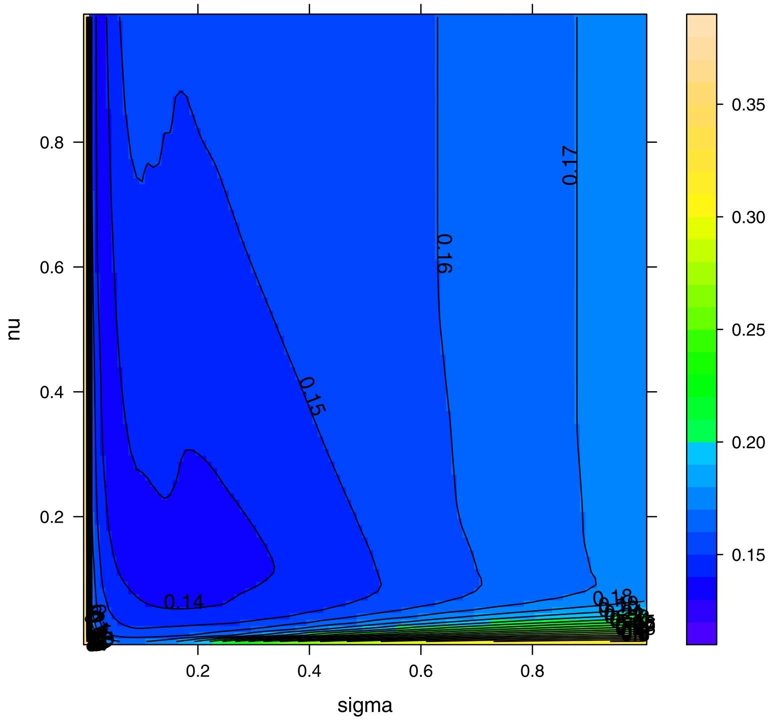


Fig. 5 Result of grid search for SVM ν and kernel σ, contour lines, and color scale show the mean square error of the testing set

lated ROC curves are plotted in Fig. 8(a) for the training set and in Fig. 8(b) for the cross-validation set. Table 3 compares the different approaches in terms of AUC.

Figure 8(b) shows the curves obtained from cross-validation over Set2; being based on a completely independent set of data, model performance over Set2 provides a reliable mean to compare different techniques. In this case, SVM clearly outperforms the other techniques, while logistic regression and discriminant analysis show substantially identical performances and Naive Bayes performs slightly worse than all the other approaches. Moreover, from the analysis of ROC-based curves of Fig. 8(b), SVM model appears to be less prone to false positives than the other models, as its curve starts with a very steep increase of the true positive rate which calls for a better discrimination capability between probable and not probable outcomes. While the exclusion of false positives from the plot can be justified by the real impossibility to observe future not-yet-occurred events, this is a somewhat ill-posed assumption. This consideration arises because we are not trying to discriminate between occurred and not-occurred events, but rather between the landslide distribution in variables space and a completely random distribution in variables space. The SVM ROC curve shown in Fig. 8(b) also proves that the model retains very good generalization properties on independent samples, thus reducing the likelihood of a possible over-fitting in the tuning stage.

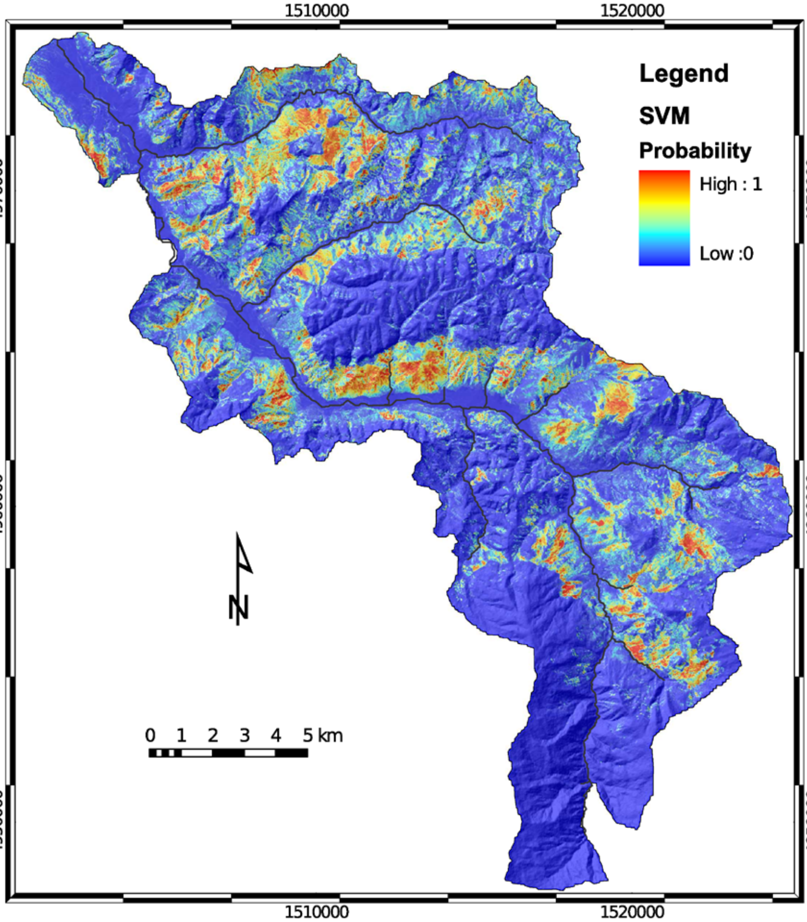


Fig. 6 Susceptibility map as predicted by the SVM model. The *color* scale represents the probability of a pixel being classified as landslide prone (spatial probability of occurrence)

Analyzing the results, it appears that SVM model outperforms the other classifiers and produces high probability values over a smaller area. This high specificity of the model confines approximately 78% of occurrences within the 20% of the most susceptible area, while the other models identify no more than 60%. This means an increased performance of 18% for the SVM.

Many authors (Chung and Fabbri 1999, 2003; Guzzetti et al. 2006) compute success and prediction rate curves as single lines for each model, by averaging the different susceptibility values falling within each occurrence. As each landslide scarp occupies a spatially defined area, the predicted probability values are not constant within the scarp area, but they range from a minimum to a maximum value. Thus, simple averaging can be reductive and lead to misinterpretation of the final result. For this reason, in this study we propose to compute different (maximum, minimum, and average) probability curves, as depicted in Fig. 9(a) for the training set (Set1). The wide separation between the predicted minimum and maximum values, and the

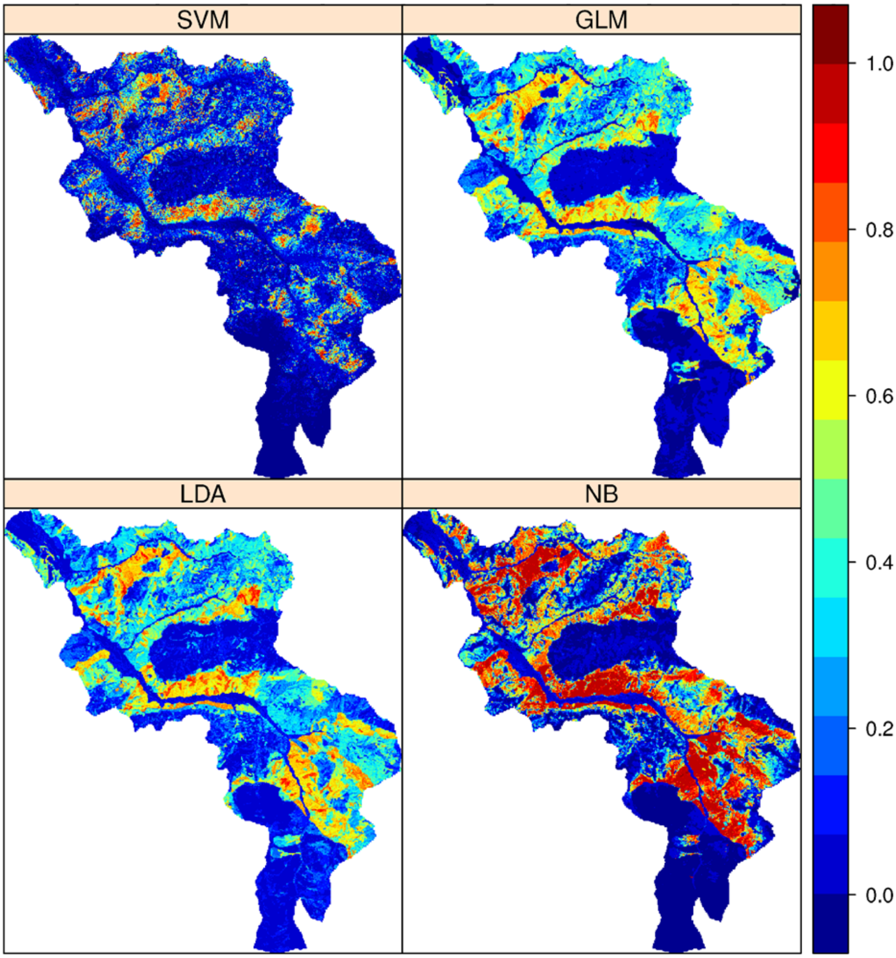
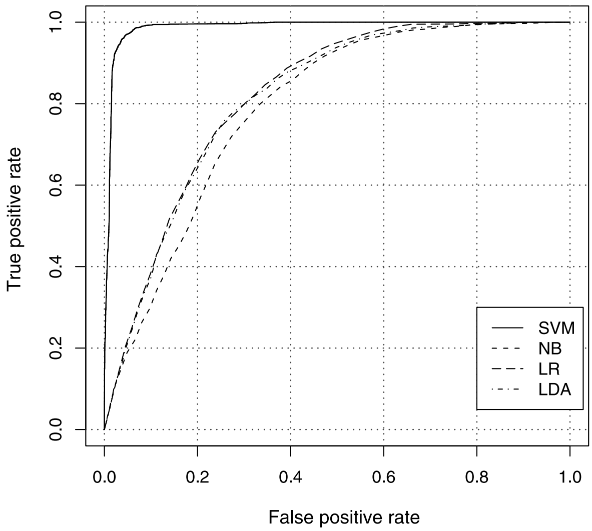


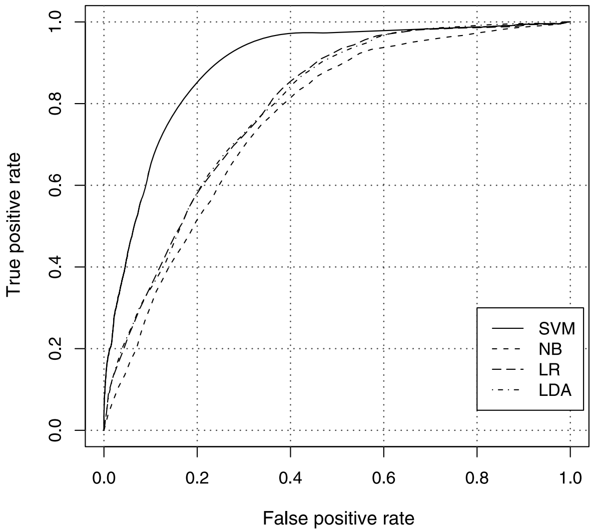
Fig. 7 Comparison between the SVM susceptibility map and the maps obtained by logistic regression, linear discriminant analysis, naive Bayes. The SVM model produces high probabilities less frequently than the other models, while preserving the predictivity on the observed occurrences. In this sense, the SVM model is more specific and less prone to type false positives errors. The *color* scale is the same in all the maps and indicates the spatial probability of occurrence

mean values very close to the maximum, is a strong evidence of SVM analysis specificity. The proximity between the mean and the maximum curves in Fig. 9(a) denotes that SVM are very specific learners, able to discriminate between the informative input and random noise. This behavior is also evident, albeit to a lesser degree, in Fig. 9(b) that shows the predictive curves for the cross-validation set (Set2). In conclusion, given the maximum precautionary principle, one should take into account the curve for the highest susceptibility rather than the average one. This implies a better definition of the most susceptible area and, consequently, a more precise financial allocation to mitigate slope instability problems.

Fig. 8 (**a**) ROC curves comparing classification performance of different techniques over the testing set (Set1): SVM; LR; LDA; NB. (**b**) ROC curves comparing classification performance for the cross-validation set (Set2)



(a)

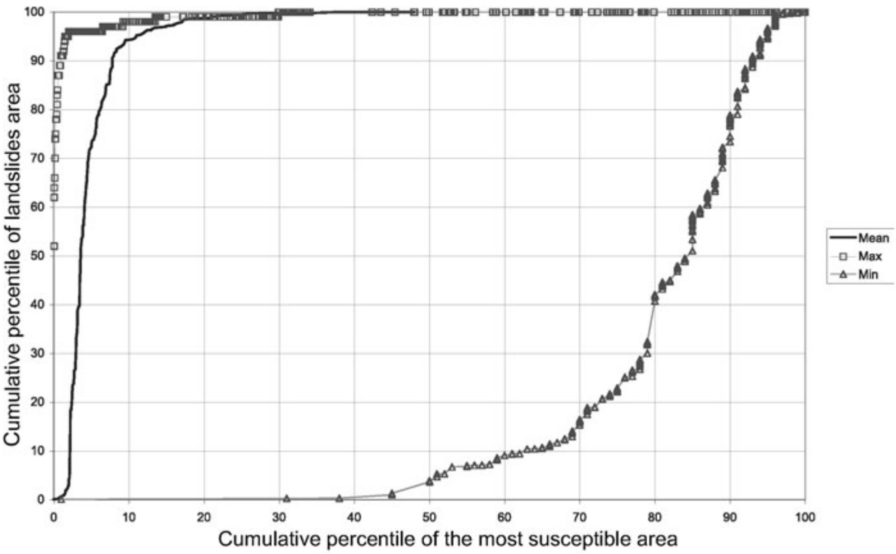


(b)

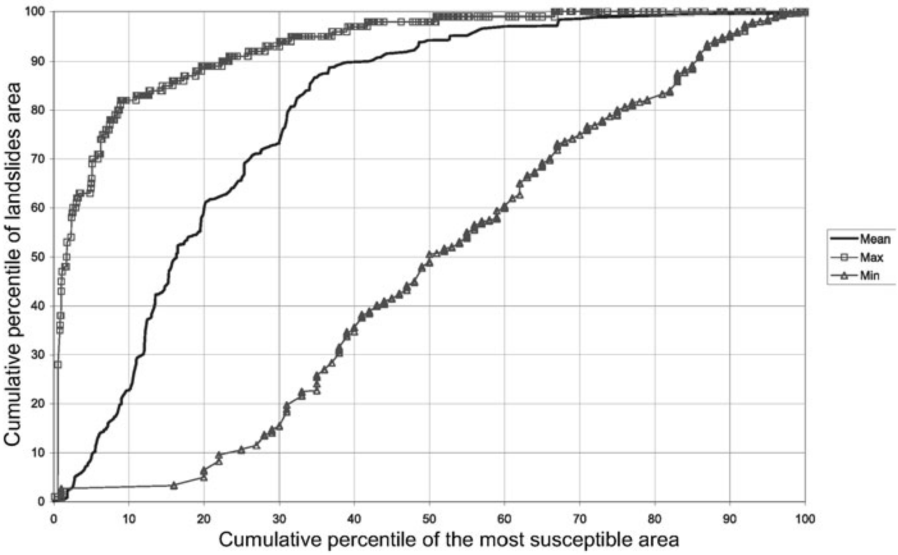
Table 3 AUC values for different techniques applied to the testing and to the cross-validation data sets

	SVM	LR	LDA	NB
Testing set (Set1)	0.98	0.82	0.81	0.79
CV set (Set2)	0.83	0.79	0.79	0.76

The last step of the study is aimed at verifying SVM prediction capability when only DEM derived continuous parameters are used in the analysis. As stated before,



(a)



(b)

Fig. 9 (**a**) Prediction rate curves for the testing set (Set1). The *solid curve* is calculated using the average probability within each mapped landslide. The "max" curve expresses the cumulative area of the highest probability class which falls within mapped landslide. The "min" curve expresses the cumulative area of the lowest probability class which falls within mapped landslide. (**b**) Prediction rate curves for the cross-validation set (Set2). *Curves* are calculated as in (**a**)

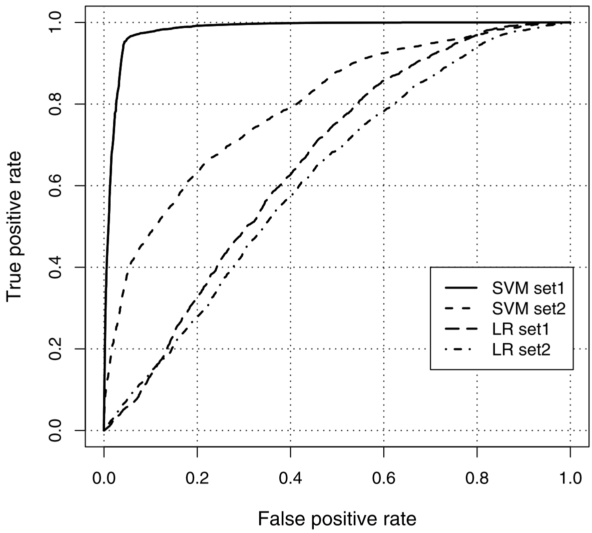


Fig. 10 ROC curves comparing the performance of SVM an LR when only morphometric covariates are utilized to fit the models. The models were fitted using only Set1 occurrences. The non-linear mapping of SVM produces a good fitting even with a reduced set of predictors, while LR suffers a greater performance reduction

one of the advantages of SVM over other techniques is the non-linear mapping. Linear models can cope with a certain amount of complexity and this usually requires the model to be stratified in multiple levels using discrete covariates. In this way, it is possible for a parametric model to have different intercepts and coefficient values for each class of the discrete covariates. However, if only DEM derived parameters are used, the model could suffer a significant performance reduction. This situation is depicted in Fig. 10, where ROC curves from SVM and logistic regression are compared. In this case, the models were not fitted on the best set of predictors, but only on the DEM-derived covariates. Logistic regression suffers a greater performance reduction, while SVM preserves most of its prediction capability. This feature is of great value considering the increasing availability of high resolution DEMs and the relative scarcity of high resolution geo-environmental features maps (e.g. geological maps). A technique which is able to produce reliable predictions using only morphometric descriptors, can be applied in areas where other covariates are not available.

4 Conclusions

This study provides a conceptual framework for the mapping of landslides susceptibility using support vector machines. This method is particularly attractive due to several peculiar features, which include: robustness to noise, non-linear decision boundaries, easily implementable probabilistic outcome and an inherent capability to deal with high dimensional classification problems. Several issues and common problems of SVM application in this field of study are discussed and suitable solutions are proposed. Particular attention has been dedicated to the description and application of feature selection procedures. In this study, we propose the adoption of the recursive feature selection algorithm (Guyon et al. 2002), which being based on the same principle of SVM, provides coherent feature ranking results. This ranking appears to be in agreement with the one produced by more traditional techniques as the Wilcoxon test. The evaluation of model performances has also been extensively addressed through the use of ROC and success rate curves. In particular, we strongly advocate the use of ROC curves given their attractive features, the most important one being their insensitivity to class distribution.

Finally, we dedicated part of the study to the comparison of SVM performances with the ones of more commonly used, at least in this field of study, techniques. We compared SVM with logistic regression, linear discriminant analysis and naive Bayes classifiers. Although this comparison might seem unfair, it is useful to advocate the use of non-linear classifiers and robust statistical techniques among geomorphologists and hazard experts. In this study, the SVM procedure was found feasible and able to outperform other techniques in terms of both accuracy and generalization capacity. The over-performance of SVM against the other techniques was around 18% for the cross-validation set, considering the 20% of the most susceptible area. This superior performance is retained even when the prediction is tested against a time independent set of events, thus proving the SVM ability to predict the location of possible future events. Another interesting aspect of SVM prediction is the reduced rate of false positives on the cross-validation set, as false positives contribute to increase the cost of remediation and mitigation. Moreover, SVM proved to perform particularly well when reduced data sets, comprising only morphometric covariates, are used in the analysis. This feature it is yet another effective advantage of SVM over other techniques as it makes their application more effective in areas where geo-environmental maps are missing or outdated, like in developing countries. The susceptibility map produced by SVM appears have a lower spatial variability when compared with the ones produced by other techniques, while retaining a superior prediction performance. If compared with other susceptibility maps previously produced for the study area (Carrara et al. 2003; Sterlacchini et al. 2004; Poli and Sterlacchini 2007) this study susceptibility map appears far more specific in its spatial delineation of potentially hazardous areas. Considering that hazard management and mitigation costs are directly related to the spatial extension of potentially hazardous areas, it is evident that more accurate and specific hazard maps will critically reduce both economic and social costs.

References

Aizerman MA, Braverman EM, Rozonoér LI (1964) Theoretical foundations of the potential function method in pattern recognition learning. Autom Remote Control 25:821–837

Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. Bull Eng Geol Environ 58:21–44

Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda–Yahiko Mountains, central Japan. Geomorphology 65:15–31

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

Beatrizzotti G, Bellinzona G, Beltrami G, Boni A et al. (1969) Geological map of Italy fg.71-Voghera (2nd ed). Tech rep, Servizio Geologico d'Italia, Roma

Bellman R (1961) Adaptive control processes. Princeton University Press, Princeton

Bonham-Carter G (1994) Geographic information systems for geoscientists: modelling with GIS. Elsevier, London

Boulesteix AL (2007) WilcoxCV: an R package for fast variable selection in cross-validation. Methods Biochem Anal 23:1702–1704

Braga G, Braschi G, Calculli S, Caucia F et al (1985) I fenomeni franosi nell'Oltrepò Pavese: tipologia e cause. Geol Appl Idrogeol 20:621–666

Brenning A (2005) Spatial prediction models for landslide hazard: review, comparison and evaluation. Nat Hazards Earth Syst Sci 5:853–862

CARG project (1992) The new Italian 1:50 000 geological map. Tech rep, National Geological Survey, Rome, Italy

Carrara A, Cardinali M, Guzzetti F, Reichenbach P (1995) GIS technology in mapping landslide hazard. In: Geographic information systems in assessing natural Hazards. Kluwer, Dordrecht, pp 125–175

Carrara A, Crosta G, Frattini P (2003) Geomorphological and historical data in assessing landslide hazard. Earth Surf Processes Landf 28:1125–1142

Carrara A, Crosta G, Frattini P (2008) Comparing models of debris-flow susceptibility in the alpine environment. Geomorphology 94:353–378

Chen P, Lin C, Scholkopf B (2005) A tutorial on ν-support vector machines. Appl Stoch Models Bus Ind 21:111–136

Cherkassky V, Mullier F (2007) Support vector regression. In: Learning from data: concepts, theory and methods, 2ed edn. Wiley, New York, pp 439–445

Chung CJ, Fabbri A (1999) Probabilistic prediction models for landslide hazard mapping. Photogramm Eng Remote Sens 65:1389–1399

Chung CJ, Fabbri A (2003) Validation of spatial prediction models for landslide hazard mapping. Photogramm Eng Remote Sens 65:451–472

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

Cruden D, Varnes D (1996) Landslides types and processes. In: Landslides: investigation and mitigation. Transportation Research Board Special Report, vol 247. National Academy Press, Washington, pp 36–75

Dubayah R, Rich P (1995) Topographic solar radiation models for GIS. Int J Geogr Inf Syst 9:405–419

DUSAF project (2003) Destinazione d'uso dei suoli agricoli e forestali. Tech rep, Lombardy Region, Milano, Italy

Eeckhaut MVD, Vanwalleghem T, Poesen J, Govers G et al (2006) Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes (Belgium). Geomorphology 76(3–4):392–410

Egan J (1975) Signal detection theory and ROC analysis. Academic Press, New York

Evangelista P, Embrechts M, Szymanski B (2006) Taming the curse of dimensionality in kernels and novelty detection. In: Applied soft computing technologies: the challenge of complexity. Advances in Soft Computing, vol 34. Springer, Berlin, pp 425–438

Fawcett T (2006) An introduction to ROC analysis. Pattern Recognit Lett 27:861–874

Fox J, Monette G (1992) Generalized collinearity diagnostics. J Am Stat Assoc 87:178–183

Goodacre C, Bonham-Carter G, Agterberg FP, Wright D (1993) A statistical analysis of spatial association of seismicity with drainage patterns and magnetic anomalies in western Quebec. Tectonophysics 217:285–305

Guyon I, Elisseeff A (2003) An introduction into variable and feature selection. J Mach Learn Res 3:1157–1182. Special Issue on Variable and Feature Selection

Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. Mach Learn 46:389–422

Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, central Italy. Geomorphology 31:181–216

Guzzetti F, Reichenbach P, Ardizzone F, Cardinali M et al (2006) Estimating the quality of landslide susceptibility models. Geomorphology 81:166–184

Hastie T, Tibshirani R, Friedman J (2001) Local methods in high dimensions. In: The elements of statistical learning: data mining, inference and prediction. Springer, New York, pp 22–27

Hjerdt K, McDonnel J, Seibert J, Rodhe A (2004) A new topographic index to quantify downslope controls on local drainage. Water Resour Res 40:6

Hsu CW, Chang CC, Lin CJ (2007) A practical guide to support vector classification. Department of Computer Science, National Taiwan University, Taipei 106, Taiwan

Kanevski M, Pozdnoukhov A, Timonin V (2009) Machine learning for spatial environmental data. EPFL-Press

Kanevsky M, Canu S (2000) Spatial data mapping with support vector regression. Tech rep, IDIAP

LeCun Y, Denker J, Solla S, Howard R et al (1990) Optimal brain damage. In: Advances in neural information processing systems 2. Morgan Kaufmann, Denver

Lin Y, Lee Y, Wahba G (2002) Support vector machines for classification in nonstandard situations. Mach Learn 46:191–202

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18(1):50–60

Marquardt D (1970) Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12:605–607

McKenzie N, Ryan P (1999) Spatial prediction of soil properties using environmental correlation. Geoderma 89:67–94

Meisina C, Zucca F, Fossati D, Ceriani M et al (2006) Ground deformation monitoring by using the permanent scatterers technique: the example of the Oltrepo Pavese (Lombardia, Italy). Eng Geol 86(3–4):240–249

Meyer D, Leisch F, Hornik K (2003) The support vector machines under test. Neurocomputing 55:169–186

Nilsson NJ (1965) Foundations of trainable pattern classifying systems. McGraw-Hill, New York

O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41:673–690

Oommen T, Laurie G, Vogel R (2011) Sampling bias and class imbalance in maximum-likelihood logistic regression. Math Geosci 43:99–120

Platt J (1999) Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In: Advances in large margin classifiers. MIT Press, Cambridge

Poli S, Sterlacchini S (2007) Landslide representation strategies in susceptibility studies using Weights-of-Evidence modeling technique. Nat Resour Res 16:121–134

Pozdnoukhov A, Matasci G, Kanevski M, Purves R (2011) Spatio-temporal avalanche forecasting with support vector machines. Nat Hazards Earth Syst Sci 11:367–382

Samui P, Sitharam T (2010) Applicability of statistical learning algorithms for spatial variability of rock depth. Math Geosci 42:433–446. doi:10.1007/s11004-010-9268-7

Schölkopf B (2001) The kernel trick for distances. In: Proceedings neural information processing systems, pp 301–307

Smirnoff A, Boisvert E, Paradis S (2008) Support vector machine for 3d modelling from sparse geological information of various origins. Comput Geosci 34(2):127–143

Sterlacchini S, Masetti M, Poli S (2004) Spatial integration of thematic data for predictive landslide mapping: a case study from Oltrepo Pavese area, Italy. In: Landslides: evaluation and stabilization. Taylor & Francis Group, London, pp 109–115

van Westen CJ, Rengers N, Soeters R (2003) Use of geomorphological information in indirect landslide susceptibility assessment. Nat Hazards 30:399–419. doi:10.1023/B:NHAZ.0000007097.42735.9e

Vapnik V (1995) The nature of statistical learning. Springer, New York

Vapnik V (1998) Statistical learning theory. Wiley, New York

Wilcoxon F (1945) Individual comparisons by ranking methods. Biom Bull 1(6):80–83

Yokoyama R, Shirasawa M, Pike R (2002) Visualizing topography by openness: A new application of image processing to digital elevation models. Photogramm Eng Remote Sens 68:257–265

Zevenbergen L, Thorne C (1987) Quantitative analysis of land surface topography. Earth Surf Processes Landf 12:47–56

Zhou X, Tuck D (2007) MSVM-RFE: extension of SVM-RFE for multiclass gene selection on DNA microarray data. Methods Biochem Anal 23:1106–1114