**Stefania Magrì · Monica Solimano · Fabio Delogu · Tania Del Giudice ·
Mauro Quagliati · Michele Cicoria · Francesco Silvestro**

Modelling rainfall-induced landslides at a regional scale, a machine learning based approach

Abstract In Italy, rainfall represents the most common triggering factor for landslides; thus, many Italian Regional Departments of Civil Protection are setting up warning systems based on rainfall recordings. Common methods are mainly based on empirical relationships that provide the rainfall thresholds above which the occurrence of landslide phenomena is likely to be expected. In recent years, the use of machine learning approaches has gained popularity in landslide susceptibility analysis and prediction. To support the operational early warning system of Liguria Civil Protection Department for landslides hazard, we propose the implementation of a polynomial Kernel regularized least squares regression (KRLS) algorithm, for predicting the daily occurrence of shallow landslides in the five Alert Zones in Liguria (North Western Italy). The model provides an estimate of the number of landslides associated with the set of three different hydrological features, also used for the Hydrological Assessment procedure: the soil moisture, the accumulated precipitation over 12 h and the precipitation peak over 3 h. Results of the model are converted to an Alert Scenario of landslide occurrence, based on the magnitude of the expected event and identified according to the National and Regional legislation (Regional Civil Protection guidelines D.G.R. n. 1116, 23/12/2020). The performance of the predictive model (e.g. accuracy of 93%) is deemed satisfactory and the methodology is considered a valuable support to the operational early warning system of Liguria Civil Protection Department. The choice of predictive variables allows, in future development, the values obtained from historical data to be replaced by those obtained from meteorological forecast models, introducing the use of the developed model in the operational forecasting chain.

Keywords Shallow landslides · Early warning system · Regression curve · Machine learning

Introduction

In many Italian regions, rainfall is the primary trigger of shallow landslides that can cause fatalities, widespread damages and economic losses (Guzzetti 2000). Due to its geographical, geological and geomorphological settings, Liguria region is prone to floods and landslides. In the last decades, Liguria Region has often experienced considerable Damaging geo-Hydrological Events, mainly related to intense and sometimes very localized rainfall (Silvestro et al. 2012) leading to important environmental, social and economic consequences (Cevasco et al. 2015).

For this reason, the mechanisms of triggering surface landslides by rainfall have been a topic of interest for a vast production of studies in Italy, that apply different methods and that are driven by different purposes. The ability to forecast rainfall-induced shallow landslide spatial and temporal occurrence is of primary interest, among others, to the Italian National Department of Civil Protection (DPCN) who would benefit from effective warning systems to mitigate the associated risk (Ratto et al. 2013). Physically based models, which aim at reproducing the nature of the phenomenon, coordinate a hydrologic model of the sub-surface flow, resulting from the meteoric event, with a geotechnical model for describing the stability conditions of the slope, requiring accurate knowledge and definition of the numerous physical parameters involved, which makes their application over large territories difficult (Segoni et al. 2009). On a regional scale, the most widely used methods are based on the definition of an empirical correlation between rainfall and landslides, obtained from historical data of landslide occurrence. Such models are generally obtained by deriving a mathematical equation that represents the threshold beyond which landslides have occurred in the past, and it is assumed they will occur in the future. In literature, different rainfall variables have been considered: the most common are intensity and duration of critical precipitation (Aleotti 2004; Guzzetti et al. 2008), accumulated antecedent precipitation from 1 day up to a few months (Cardinali et al. 2006; Martelloni et al. 2012), soil moisture (Segoni et al. 2018). Guzzetti et al. (2008) highlighted that a standardized procedure to define critical rainfall does not exist, as an example, the rainfall intensity (and associated duration) may refer to either the peak precipitation rate measured during the rainstorm or to an average estimate of the mean precipitation rate of the whole rainstorm. The definition of the best variables and adequate time periods for the rainfall cumulates are not straightforward (Martelloni et al. 2012). The choice of the right parameters to consider depends primarily on the landslide typology: there is a general agreement in recognizing that debris flow and shallow landslides are preferentially triggered by short and intense rainfalls, while deep-seated landslides are more commonly connected with prolonged and less intense rainfall events (Crosta 1998). In those areas affected by both shallow and deep-seated landslides, it is essential to define a methodology that could be flexible enough to encompass both of them. Some authors have tried to indirectly include other predisposing factors, such as lithological information or morphology, by applying calibration at a sub-regional scale (Gariano et al. 2015). As a result, existing and new thresholds are difficult to compare and evaluate quantitatively (Guzzetti et al. 2008; Brunetti et al. 2010). These methods focus on the temporal scale, considering individual landslides as discrete point events, and thus do not provide spatial information, while landslide hazard management would require the combined anticipation of "when", but also "where" and "how

large" a landslide is expected to be (Lari et al. 2014). Recent studies have worked in combining the spatial and temporal occurrence prediction of landslides, by applying statistical or machine learning algorithms (Lombardo et al. 2020).

Data-driven methods identify patterns or trends by analysing and interpreting the relationship between data. These methods have become increasingly popular over the last few years, thanks to the availability of big data and improvements in computational power. Machine learning models, such as random forest, artificial neural network, support vector machine and ensemble techniques, have been successfully applied in landslides detection based on images (Yu et al. 2017), landslide susceptibility analysis (Park 2019; Reichenbach et al. 2018; Lee et al. 2020; Tien et al. 2016; Kumar et al. 2017) and prediction (Huang and Xiang 2018).

The objective of the present work is to respond to the need for a Landslide Early Warning System (LEWS) for the Liguria Civil Protection Department, with a predictive model of rainfall-induced shallow landslides occurrence in the five Alert Zones (AZs) in which Liguria region is divided, that can be easily integrated in the existing operational early warning system for weather forecasting. Simplicity, low computation cost and explainability—that is, the ability to understand the patterns that the model uses to link to the training datasets and the factors that influence its outputs—were crucial factors for the selection of the model.

Material and methods

The study area

The Liguria Region lies between the Ligurian Alps and the Ligurian Apennines to the north and faces the Ligurian Sea to the south, with an unbroken chain that forms a veritable ridge.

From the line of the watershed, an asymmetrical arrangement of the slopes can be observed, involving an average higher energy value along the Tyrrhenian sector than in the Po valley. In particular, the maritime side, in the Apennine stretch, presents an arrangement of the valleys parallel to the coastline, unlike the Alpine area where the valley axis has an average submeridian structure. Geological information can be found on the

official regional cartographic website (http://www.banchedati. ambienteinliguria.it/).

Due to its geographical position and morphology, Liguria is prone to various types of weather scenarios, among them, two main types of precipitation can be distinguished: the thunderstorm, which is more sudden and intense, and the advective one, which is generally weaker in intensity but more persistent.

The analysis of precipitation fields shows that the eastern part of the region is rainier than the western part, not only in terms of accumulated precipitation (Fig. 1), but also in terms of the number of rainy days, daily precipitation values and consecutive dry and rainy days (Agrillo and Bonati 2013).

The operative system for flood and landslide risks in Liguria is based on the division of the regional area into five Alert Zones (AZs), defined for hydro-meteorological monitoring and early warning (Fig. 2). The subdivision is based on physiographic zoning that respects the integrity of catchment areas, municipal administrative areas, extension on spatial scales compatible with the limits of forecast reliability and the distinction into homogeneous climatic areas. The hydrogeological and hydraulic risk is thus determined for each AZ.

Rainfall data

The reference document that defines the hydrogeological/hydraulic warning procedure (D.G.R. 1116 of 23/12/2020) identifies two specific quantities to be evaluated and compared to a defined threshold for the Hydrological Assessment procedure: the 3-h cumulative precipitation height averaged over 100 km², which refers to the precipitation intensity, and the cumulative precipitation height in 12 h averaged over Alert Zone, which refers to precipitation quantity. These two rainfall-related variables were then considered as predictors for the developed model, as they are daily evaluated from the meteorological forecast models for the hydrogeological/hydraulic assessment procedure. The variables were estimated starting from the hourly measurements from the ground network of the Liguria Region, for the period 2014–2019. Based on 120 rain gauges distributed as shown in Fig. 3,
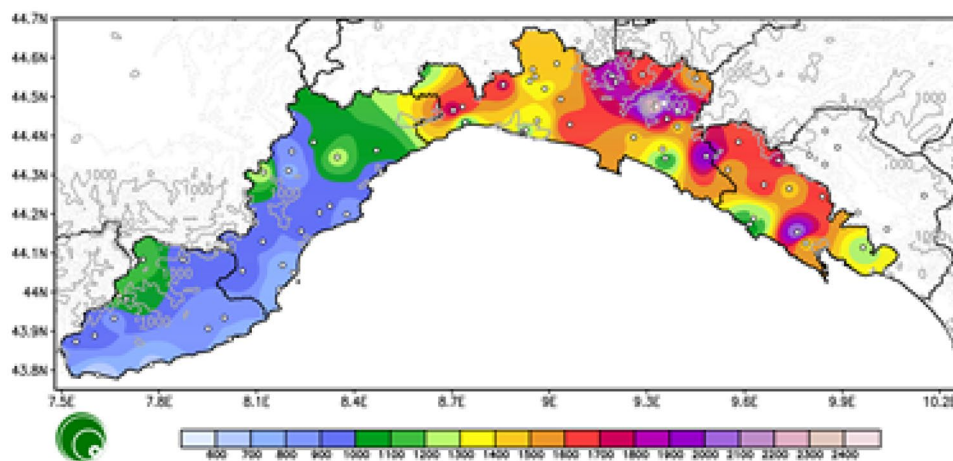

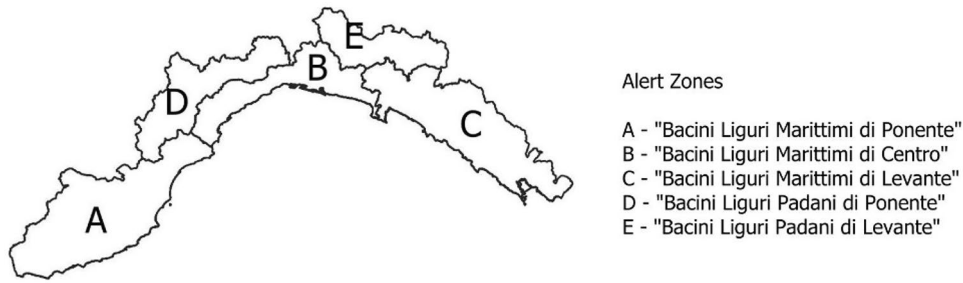**Fig. 1** Cumulative precipitation (mm)—annual average 1961–2010 (Agrillo and Bonati 2013)

Fig. 2 Identification of the Alert Zones

Alert Zones

A - "Bacini Liguri Marittimi di Ponente"
B - "Bacini Liguri Marittimi di Centro"
C - "Bacini Liguri Marittimi di Levante"
D - "Bacini Liguri Padani di Ponente"
E - "Bacini Liguri Padani di Levante"

the measurements were interpolated using the inverse distance weighting (idw) method to perform a spatial interpolation with a known scattered set of points. Using a weighting power equal to 2 and a radius of 0.2 degrees along the x and y coordinates, the application of the idw method has defined the 1-h interpolated rainfall map of the entire Liguria domain.

Based on the observed and interpolated 1-h rainfall measurements, the two different variables previously described were identified: (1) the 12-h rainfall accumulation map that represents the accumulated rainfall in the defined time period and (2) the 3-h rainfall peak over a domain of 100 km², that represents the maximum rainfall in the defined time period and domain.

The 12-h accumulated rainfall map was averaged over each AZ domains, while for the 3-h rainfall peaks the maximum value for each AZ domain was computed. The so obtained 12-h accumulated and the 3-h rainfall peak time series, saved with daily frequency, have been finally considered input features.

We point out that single rainfall events were not identified and thus single events that cover two different days are considered separately for the amount of rain measured for the single solar day.

Soil moisture data

Soil moisture was estimated through the continuous and distributed hydrological model continuum (Silvestro et al. 2013, 2015; Cenci et al. 2016). It schematizes the behaviour of the root zone with a modified version of the Horton method (Gabellani et al. 2008), which simulates the maximum water storage with a tank of maximum dimension $V_{max}(x,y)$, while the variable $V(t,x,y)$ describes the variation of the storage along time t, in the location x,y. It is possible to evaluate the relative soil moisture (SM) with the formulation:

$$SM = \frac{V(t,x,y)}{V(x,y)_{max}}$$

The model was used and implemented for many applications (Poletti et al. 2019; Silvestro et al. 2021; Lagasio et al. 2022), especially for flood forecast and water management purposes. In this case, an implementation quite similar in terms of calibration and parameterization strategy to the one found in various works (Davolio et al. 2017; Bruno et al. 2021) was carried out.
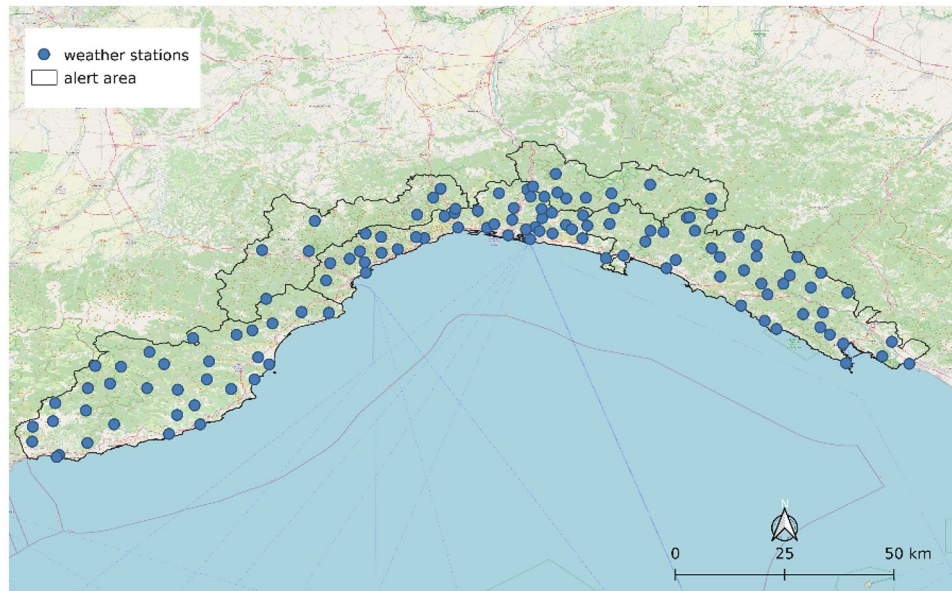

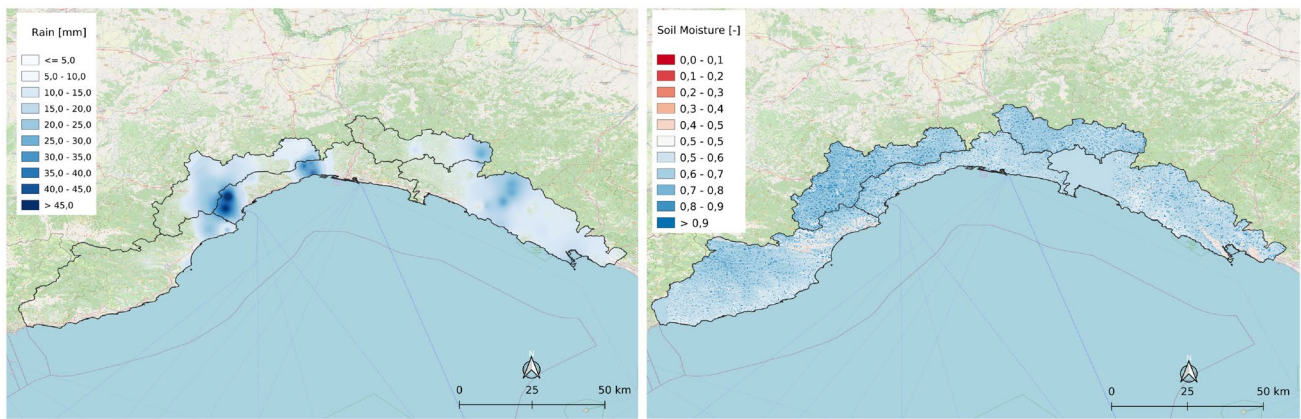**Fig. 3** Spatial distribution of weather stations in the Alert Zone in Liguria region

Fig. 4 Example of rainfall map derived by interpolating the observed measurements (left image) and soil moisture map derived by the hydrological model (right image)

The spatial resolution is set to 170 m, while the time resolution is 1 h. Since the model solves both mass and energy balances, continuous simulation allows the estimation of soil moisture based on a physical approach, the final product is a map that covers all Liguria Region at a high resolution (Fig. 4).

Soil moisture, obtained from hydrological reanalysis of the period considered, was then aggregated at the spatial scale of the Alert Zone, estimating the mean value in order to be used as an input feature in the proposed algorithm.

Landslide data and scenarios

For the analysed period 2014–2019, there were 359 days with conditions leading to 2191 shallow landslides (debris and earth slides, debris flows), while deep landslides and collapses due to gravity were excluded (Fig. 5). We obtained information on landslides occurrence from various sources like newspapers, technical reports provided by Liguria Region Civil Protection, municipal damage report sheets, Fire Brigade, Liguria Region Soil Defense Sector. Figure 6 shows the spatial distribution of landslides recorded per year. The fact that the number of landslides is higher in the latest years can be determined by the increased efficiency in identifying and recording the landslides. We acknowledge that, because of incompleteness inherent to the landslide information, the number of documented landslides is a subset of all rainfall-induced landslides that occurred in the 2014–2019 period.

Table 1 Scenarios and alert scale definition

Alert scenario	# landslides	Extension
Low	0–2	Low probability of phenomena
Ordinary	3–5	Localized phenomena
Moderate	6–13	Widespread phenomena
High	> 14	Phenomena extended over an AZ

Days with no occurrence of landslides represent about 95% of the whole dataset, and the maximum number of landslides occurred in one single day and AZ is 45 (Fig. 5).

In accordance with the Regional Civil Protection guidelines (D.G.R. n. 1116 /2020), the considered Alert Scenarios are:

- Low: absence or low probability of phenomena at the local level
- Ordinary: localized phenomena/effects on the ground
- Moderate: widespread phenomena (in single basins or portions of the alert area)
- High: numerous and/or typically extended phenomena over an entire alert area

Based on the study of the spatial distribution and the number of landslides within each AZ and their comparison, number of landslides was attributed to each scenario as shown in Table 1. Figure 7 shows the occurrence of the different scenarios for each AZ in the considered period.

Model implementation

Machine learning algorithms are used to model non-linear relationships between the input variables (also called *predictive* variables) and the output variable, and have been applied to "learn" the relationship among landslide occurrence and landslide-related predictors (Liu et al. 2021).

Here, we introduce a polynomial kernel regularized least squares regression (KRLS) model to predict the daily number of landslides, using five features as predictors, as previously described:

- The 12 h accumulated rainfall averaged over the AZ;
- The 3-h rainfall peak over an area of 100 km²;
- Soil moisture obtained at 11 p.m. of the previous day;
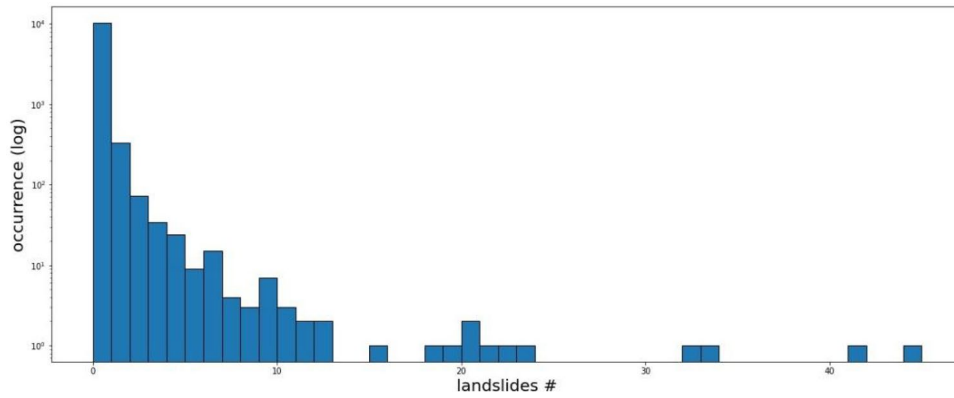- The Alert Zone;
- Day of the year.

Fig. 5 Histogram of the number of registered landslides per day

The soil moisture at 11 p.m. of the previous day represents the predisposing element, while the precipitation for the same day represents the triggering factor, for both types of precipitations: short and intense events (3 h peak over an area of 100 km²) and prolonged rainfall events (12 h accumulated rainfall over the AZ).

The Kernel method is based on the use of a kernel function K (x, x'), symmetric and positive definite, to implicitly have a non-linear function ϕ(x) mapping the vector of input variables x from the input space X to a new dot product space F, also called feature space. Through the kernel function, we can model our data by a linear function in the feature space, while the model is no longer linear in the original input space. In this way, we can apply non-linear model computations that are essentially the same of the linear case.

In particular, we refer to the regularized least squares (RLS), an algorithm defined by the square loss that consider linear functions as possible solutions:

$$f(x) = x^T w$$

where x is the vector of the input variables and w is the vector of weights. When introducing the feature map ϕ(x), described by the kernel matrix K (x, x'), we will have:

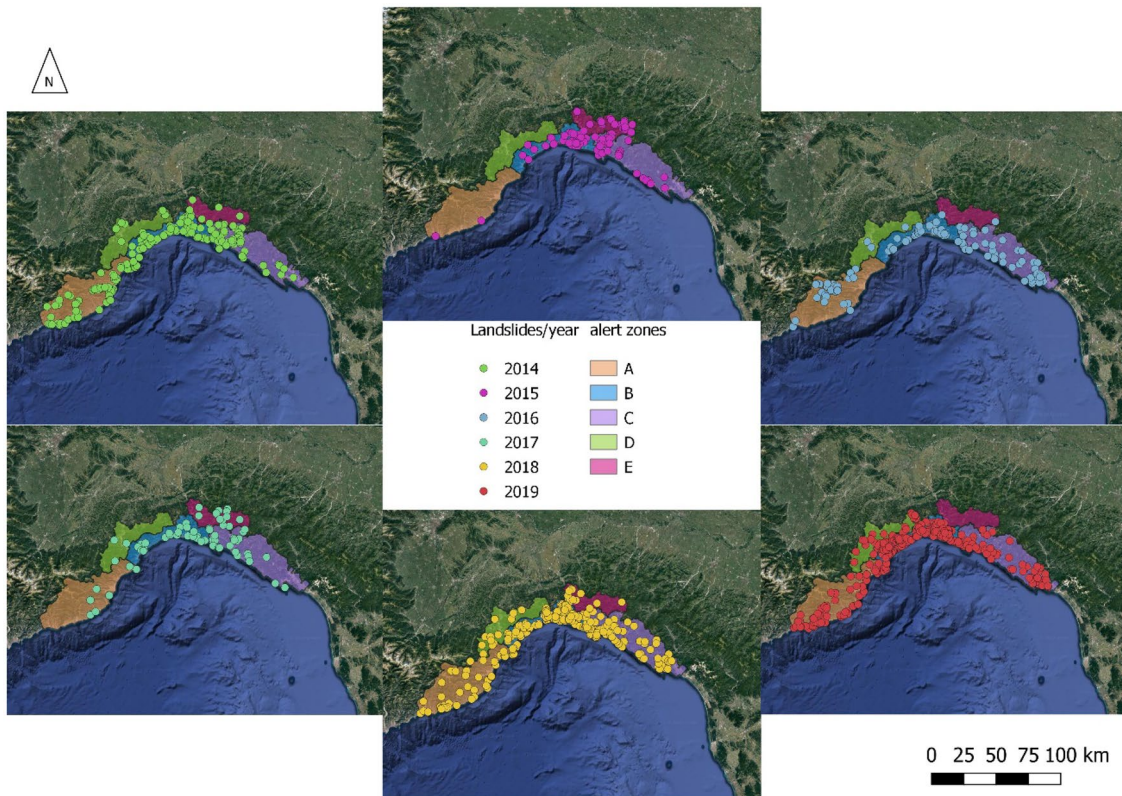$$f(x) = w^T \phi(x)$$

which further becomes:

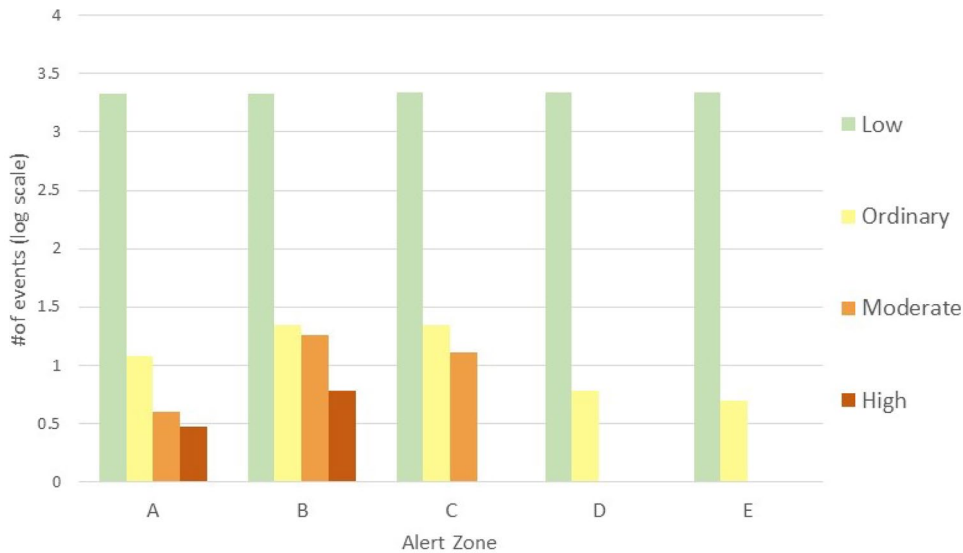**Fig. 6** Spatial distribution of landslides recorded in the database per year

Fig. 7 Occurrence (in log scale) of the different types of scenarios for each Alert Zone

$$f_w(x) = \sum_{j=1}^{n} K(x_j x)c_j$$

where $x_1, \ldots, x_n$ are the inputs in the training set, $c = (c_1, \ldots, c_n)$ a set of coefficients and n is the number of training data. The hyperparameters of the model are tuned by minimizing the squared loss over the available training samples. For the square loss function, the vector c of coefficients can be computed solving the following linear system:

$$Y_n = \left(K_n + \lambda nI\right)c$$

where Y_n is the vector of observed output variables and λ a regularisation term introduced to ensure that the minimization problem is well defined. There are many possible kernel functions that can be used, a polynomial kernel was chosen (Eq. (1)), where the degree of the polynomial d is a hyperparameter to be tuned.

$$K(x, x') = \left(x^T x' + 1\right)^d \ where \ d \in N \qquad (1)$$

The polynomial kernel of degree d thus computes a dot product in the space spanned by all monomials up to degree d in the input coordinates. The advantage of using such a kernel as a similarity measure is that it allows us to construct algorithms in dot product spaces (Hofmann et al. 2008).

The model is trained and validated with the database described in the "Landslides data and scenarios" section. The dataset of 2018 was used for validation, while the remaining dataset was used for calibration. A K-fold cross-validation was used, with $K = 10$. The choice of the calibration and validation datasets is crucial, and accurate analysis was performed to select the optimum datasets and consider the inhomogeneity of registered landslides through the different years (Fig. 6).

Root mean squared error (RMSE) and the mean absolute error (MAE) were used to determine model accuracy, as well as the normalized root mean square error (NRMSE), applied per class of values. A feature importance analysis for the proposed model was carried on through permutation test.

Next, the predicted number of landslides was converted to an alert level according to the described methodology and compared with the alert level for the same day corresponding to the registered landslides. Correct predictions were consequently defined through the use of a contingency table (Table 2) as described in Stehman (1997). Results were then analysed calculating correct predictions (true positives (TP) and true negatives (TN)) and incorrect predictions (missed alarms or false negatives FN and false alarms or false positives FP), based on the level of alert: an overestimated predicted alert is considered as FP while underestimated predicted alert is considered as FN.

Several skill scores, or performance indicators, are proposed in the literature for the calculation of the performance indicators of the confusion matrix. The indicators used for the performance evaluation of the ML method investigated herein: accuracy (ACC), probability of detection (POD), probability of false detection (POFD) and efficiency index (EI).

- Accuracy (ACC) is the most used indicator to analyse the confusion matrix. It gives an overall evaluation of the number of correct predictions (TN + TP) over the total number of predictions. The value ranges from 0 or no correct predictions, to 1 where 100% of the predictions are correct.

Table 2 Confusion matrix showing the four possible outcomes of a classifier

Forecasted events		Observed events	
		Yes	No
	Yes	TP	FP
	No	FN	TN

- Probability of detection (POD), also called sensitivity or hit rate or true positive rate, represents the rate of occurred events positively predicted, where 0 means none of the occurred events was predicted, and 1 where all the events having occurred were predicted correctly.
- Probability of false detection (POFD) is the counterpart of POD since it measures the ratio of forecasted events to have occurred when no events have been observed. It ranges between 0 and 1.
- Efficiency index (EI), also called threat score or critical success index, is the ratio of true positives to the sum of the true positives and the unsuccessful predictions. The indicator evaluates the performance without considering TN, which number often lies an order of magnitude above the other classifiers in the confusion matrix.

Results and discussion

Model results

A 3-degree polynomial kernel showed the best cross-validated fitting with measured data. Validation was conducted considering the year 2018, the model was run to determine the number of landslides predicted for each AZ and results were compared with registered landslides. In the validation year, 86 days reported at least one landslide event, corresponding to 333 landslides.

The root mean squared error (RMSE = 1.5) and mean absolute error (MAE = 1.2) are deemed satisfactory, although results show a tendency to overestimate the number of landslides. The RMSE and NRMSE applied per class of values are shown in Table 3. We can observe that for higher values of the number of landslides, the RMSE increases while NRMSE decreases.

The differences between prediction and observation were computed in terms of alert levels: as the final output of the procedure is to provide support to identify the alert level adopted in the civil protection procedures for the early warning system, the number of predicted landslides was converted to the corresponding four alert levels: "low", "ordinary", "moderate" and "high" (Table 1). Results were then analysed through the use of a contingency table or confusion matrix (Table 4). We can notice, for instance, that the two "high" criticality levels were correctly predicted.

Correct predictions and errors were analysed for each AZ (Table 5). We can notice that the AZs D and E have almost no landslide occurrences observed in the validation period and have a higher number of false alarms.

Table 4 Confusion matrix of results for the different classes of Alert Scenarios and for the whole validation dataset. The confusion matrix compares the predicted alert level (left) with the corresponding observed one (top)

	Low	Ordinary	Moderate	High
Low	1676	2	0	
Ordinary	102	6	4	
Moderate	17	8	3	
High	1	1	3	2

When analysing the performance and results through validation procedures, it is important to consider the uncertainty inherent in the incompleteness of information, especially in remote areas, where landslides may have occurred and gone unreported. This is coherent with the tendency of the model to overestimate the number of landslide events.

As an example, we report in Table 6 the comparison of predicted and registered number landslides for the same day in the different AZs, where the similar input conditions highlight the evidence that unreported landslides, especially in the remote areas of the AZ D and E, may be a cause of over estimation from the model.

Taking into account the confusion matrix (Table 4), the described skill scores were computed to quantitatively define the effectiveness of the model, as reported in Table 7.

Feature importance

The relative importance of the single variable used for the proposed predictive models is estimated through a permutation analysis: after random re-orderings (shuffling) of the predictive variable the test statistics is recalculated (Anderson and Robinson 2001). The statistic used is the RMSE and the number of runs used to achieve stable feature rankings was 50. Figure 8 shows the increase in the percentage of the RMSE after shuffling each of the features: the higher the RMSE value when a particular feature is shuffled, the more weight it has in the model. We can observe that the soil moisture ("sm") and the 12 h accumulated rainfall ("cum_12H") are the most important features, but all features contribute to the model prediction. The importance of soil saturation has been highlighted in other studies (Ratto et al. 2013).

Table 3 Root mean square error (RMSE) and normalized root mean square error (NRMSE) for different classes

Scenario	# landslides	# data	RMSE	NRMSE
Low	0–2	1796	1.5	4.6
Ordinary	3–5	17	5.3	1.5
Moderate	6–13	10	5.3	0.7
High	>14	2	12.8	0.5

Table 5 Results of validation through the use of confusion matrix for each Alert Zone (right)

AZ	A	B	C	D	E	TOT
TP	2	6	3	0	0	11
FP	14	26	24	33	32	129
TN	345	331	335	332	333	1676
FN	2	2	2	0	0	6

Table 6 Comparison of observed and predicted landslides occurrence for different Alert Zones for a same day (29/10/2008)

Day	Soil moisture	Rain cum. 12 h (mm)	Rain peak 3 h (mm)	AZ	# observed events	# predicted events
29/10/2018	0.82	48.7	33.4	A	0	12
29/10/2018	0.92	59.6	63.2	B	20	21
29/10/2018	0.80	67.7	62.4	C	9	14
29/10/2018	0.90	61.3	108.8	D	1	18
29/10/2018	0.96	68.1	63.2	E	3	11

Table 7 Skill scores based on the confusion matrix shown in Table 5 are used to perform the validation of the alert level

Skill score	Formula	
Accuracy (ACC)	(TP + TN)/Ntot	0.93
Probability of detection (POD)	TP/(TP + FP)	0.08
Sensitivity (POD/true positive rate)	TP/(TP + FN)	0.65
Probability of false detection (POFD)	FP/(FP + TN)	0.07
Probability of false alarm (POFA)	FP/(TP + FP)	0.92
Efficiency (EI)	TP/(FP + TP + FN)	0.08

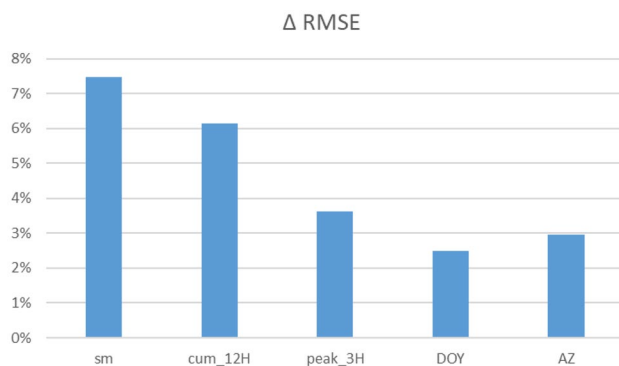*TN* true positives, *PF* false positives, *FN* false negatives, *TN* true negatives

Fig. 8 Feature importance for the proposed predictive model represented in terms of increase of RMSE (%) obtained by permutation test for each predictive feature (sm = soil moisture, cum_12H = 12 h accumulated rainfall over the AZ; peak_3H = 3-h rainfall peak over an area of 100 km.^{2}; DOY = day of the year; AZ = Alert Zone)

Conclusions

In order to model the occurrence of landslides at a regional scale and define a reliable alert system to support the Liguria Civil Protection Department, a machine learning-based algorithm was defined and applied based on a database of recorded landslides covering the period 2014–2019. The model is based on the application of a polynomial kernel regularized least squares regression (KRLS), based on a set of five predictive variables: the Alert Zone, the day of the year, soil moisture averaged over the AZ, rain accumulated over 12 h averaged over each AZ, the 3-h rain peak over an area of 100 km². The model provides, for each day and single AZ, the predicted number of landslides occurrence, which is then converted to the corresponding Alert Scenario, according to DGR n.1116/2020. Although the model has shown a tendency to overestimate the number of landslides, and uncertainties of the model cannot be easily quantified, also due to the uncertainty inherent in the incompleteness of observed landslide information, the performance of the model described can be considered satisfactory when considered in the context of the Hydrological Assessment procedure operative at the Civil Protection Department. The tested methodology allowed the validation of the use of the selected predictive features to predict the landslide occurrence and to identify the most significant ones. Future development of the model will consist of substituting the observed values of the hydrological variables with a forecasted value obtained from the predictive model operational at the Regional Agency for Environmental Protection of Liguria for the Civil Protection Department, moving toward an operative application, where the model will provide, at a daily basis, and separately for each AZ, the predicted level of criticality based on the weather forecasts and observed soil moisture.

The major advantages of the proposed methodology consist in fact in the input data, which have been selected to be substituted with values accessible in an operational procedure, to support in a simple and rapid way the identification of a warning level. The type of output of the model, that is the number of predicted landslides occurrence per AZ, is considered by the authors an easier parameter to be integrated into an early warning system, rather than a simple probability of exceeding a defined threshold, as it can be more easily interpreted by the operators to integrate the information into a "criticality bulletin".

The system can be improved by further subdividing the AZs according to lithological areas, obtaining a more detailed splitting up, and a better compromise between operational and scientific constraints.

Finally, it is important to point out that the continuous updating of the database, with the combined information on rainfall and records of landslides that have occurred, makes it possible to improve the calibration of the algorithm parameters.

Data availability

The landslide data used was provided by Istituto di Ricerca per la Protezione Idrogeologica CNR-IRPI Perugia (ITAlian rainfall-induced LandslIdes CAtalogue - ITALICA) and can be accessed at https://doi.org/10.5281/zenodo.8009366.

Declarations

Competing interests   The authors declare no competing interests.

References

Agrillo G, Bonati V (2013) Atlante Climatico della Liguria. ARPAL – Centro Funzionale della Regione Liguria -Progetto Resmar – Programma Marittimo Italia-Francia

Aleotti P (2004) A warning system for rainfall-induced shallow failures. Eng Geol 73:247–265. https://doi.org/10.1016/j.enggeo.2004.01.007

Anderson MJ, Robinson J (2001) Permutation tests for linear models. Aust N Z J Stat 43:75–88. https://doi.org/10.1111/1467-842X.00156

Brunetti MT, Peruccacci S, Rossi M, Luciani S, Valigi D, Guzzetti F (2010) Rainfall thresholds for the possible occurrence of landslides in Italy. Nat Hazards Earth Syst Sci 10:447–458. https://doi.org/10.5194/nhess-10-447-2010

Bruno G, Pignone F, Silvestro F, Gabellani S, Schiavi F, Rebora N, Giordano P, Falzacappa M (2021) Performing hydrological monitoring at a national scale by exploiting rain-gauge and radar networks: the Italian case. Atmosphere 12(6):771. https://doi.org/10.3390/atmos12060771

Cardinali M, Galli M, Guzzetti F, Ardizzone F, Reichenbach P, Bartoccini P (2006) Rainfall induced landslides in December 2004 in Southwestern Umbria, Central Italy. Nat Hazards Earth Syst Sci 6:237–260. https://doi.org/10.5194/nhess-6-237-2006

Cenci L, Laiolo P, Gabellani S, Campo L, Silvestro F, Delogu F, Bon G, Rudari R, Puca S (2016) Assimilation of H-SAF soil moisture products for flash flood early warning systems. Case Study: Mediterranean Catchments. IEEE J Select Top Appl Earth Observ Rem Sens 9(12):5634–5646. https://doi.org/10.1109/JSTARS.2016.2598475

Cevasco A, Diodato N, Revellino P, Fiorillo F, Grelle G, Guadagno FM (2015) Storminess and geo-hydrological events affecting small coastal basins in a terraced Mediterranean environment. Sci Total Environ 532:208–219. https://doi.org/10.1016/j.scitotenv.2015.06.017

Crosta G (1998) Regionalization of rainfall thresholds: an aid to landslide hazard evaluation. Environ Geol 35:131–145. https://doi.org/10.1007/s002540050300

Davolio S, Silvestro F, Gastaldo T (2017) Impact of rainfall assimilation on high-resolution hydro-meteorological forecasts over Liguria (Italy). J Hydrometeor 18:2659–2680. https://doi.org/10.1175/JHM-D-17-0073.1

Gabellani S, Silvestro F, Rudari R, Boni G (2008) General calibration methodology for a combined Horton-SCS infiltration scheme in flash flood modeling. Nat Hazards Earth Syst Sci 8:1317–1327. https://doi.org/10.5194/nhess-8-1317-2008

Gariano SL, Brunetti MT, Iovine G, Melillo M, Peruccacci S, Terranova O, Vennari C, Guzzetti F (2015) Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy. Geomorphology 228:653–665. https://doi.org/10.1016/j.geomorph.2014.10.019

Guzzetti F (2000) Landslides fatalities and the evaluation of landslide risk in Italy. Eng Geol 58:89–107. https://doi.org/10.1016/S0013-7952(00)00047-8

Guzzetti F, Peruccacci S, Rossi M, Stark CP (2008) The rainfall intensity–duration control of shallow landslides and debris flows: an update. Landslides 5:3–17. https://doi.org/10.1007/s10346-007-0112-1

Hofmann T, Schölkopf B, Smola AJ (2008) Kernel methods in machine learning. Ann Statist 36:1171–1220. https://doi.org/10.1214/009053607000000677

Huang L, Xiang L-y (2018) Method for meteorological early warning of precipitation-induced landslides based on deep neural network. Neural Process Lett 48:1243–1260. https://doi.org/10.1007/s11063-017-9778-0

Kumar D, Thakur M, Dubey CS, Shukla DP (2017) Landslide susceptibility mapping & prediction using support vector machine for Mandakini River Basin, Garhwal Himalaya, India. Geomorphology 295:115–125. https://doi.org/10.1016/j.geomorph.2017.06.013

Lagasio M, Fagugli G, Ferraris L, Fiori E, Gabellani S, Masi R, Mazzarella V, Milelli M, Parodi A, Pignone F, Puca S, Pulvirenti L, Silvestro F, Squicciarino G, Parodi A (2022) A complete meteo/hydro/hydraulic chain application to support early warning and monitoring systems: the Apollo Medicane use case. Remote Sens 14:6348. https://doi.org/10.3390/rs14246348

Lari S, Frattini P, Crosta GB (2014) A probabilistic approach for landslide hazard analysis. Eng Geo 182:3–14. https://doi.org/10.1016/j.enggeo.2014.07.015

Lee D-H, Kim Y-T, Lee S-R (2020) Shallow landslide susceptibility models based on artificial neural networks considering the factor selection method and various non-linear activation functions. Remote Sensing 12(7):1194. https://doi.org/10.3390/rs12071194

Liu Z, Gilbert G, Cepeda JM, Lysdahl AOK, Piciullo L, Hefre H, Lacasse S (2021) Modelling of shallow landslides with machine learning algorithms. Geosci Front 12(1):385–393. https://doi.org/10.1016/j.gsf.2020.04.014

Lombardo L, Opitz T, Ardizzone F, Guzzetti F, Huser R (2020) Space-time landslide predictive modelling. Earth Sci Rev 209:103318. https://doi.org/10.1016/j.earscirev.2020.103318

Martelloni G, Segoni S, Fanti R, Catani F (2012) Rainfall thresholds for the forecasting of landslide occurrence at regional scale. Landslides 9:485–495. https://doi.org/10.1007/s10346-011-0308-2

Park S (2019) Landslide susceptibility mapping based on random forest and boosted regression tree models, and a comparison of their performance. Appl Sci 9(5):942. https://doi.org/10.3390/app9050942

Poletti ML, Silvestro F, Davolio S, Pignone F, Rebora N (2019) Using nowcasting technique and data assimilation in a meteorological model to improve very short range hydrological forecasts. Hydrol Earth Syst Sci 23:3823–3841. https://doi.org/10.5194/hess-23-3823-2019

Ratto SM, Stevenin H, Navillod E, Cauduro M (2013) Operational procedure for a hydrogeological warning system in Aosta Valley. In: Margottini, C., Canuti, P., Sassa, K. (eds) Landslide Science and Practice. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31445-2_84

Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F (2018) A review of statistically based landslide susceptibility models. Earth Sci Rev 180:60–91. https://doi.org/10.1016/j.earscirev.2018.03.001

Segoni S, Leoni L, Benedetti AI, Catani F, Righini G, Falorni G, Gabellani S, Rudari R, Silvestro F, Rebora N (2009) Towards a definition of a real-time forecasting network for rainfall induced shallow landslides. Nat Hazards Earth Syst Sci 9:2119–2133. https://doi.org/10.5194/nhess-9-2119-2009

Segoni S, Rosi A, Lagomarsino D, Fanti R, Casagli N (2018) Brief communication: using averaged soil moisture estimates to improve the performances of a regional-scale landslide early warning system. Nat Hazards Earth Syst Sci 18:807–812. https://doi.org/10.5194/nhess-18-807-2018

Silvestro F, Gabellani S, Giannoni F, Parodi A, Rebora N, Rudari R, Siccardi F (2012) A hydrological analysis of the 4 November 2011 event in Genoa. Nat Hazards Earth Syst Sci 12:2743–2752. https://doi.org/10.5194/nhess-12-2743-2012

Silvestro F, Gabellani S, Delogu F, Rudari R, Boni G (2013) Exploiting remote sensing land surface temperature in distributed hydrological modelling: the example of the Continuum model. Hydrol Earth Syst Sci 17:39–62. https://doi.org/10.5194/hess-17-39-2013

Silvestro F, Gabellani S, Rudari R, Delogu F, Laiolo P, Boni G (2015) Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data. Hydrol Earth Syst Sci 19:1727–1751. https://doi.org/10.5194/hess-19-1727-2015

Silvestro F, Ercolani G, Gabellani S, Giordano P, Falzacappa M (2021) Improving real-time operational streamflow simulations using discharge data to update state variables of a distributed hydrological model. Hydrol Res 52(6):1239–1260. https://doi.org/10.2166/nh.2021.162

Stehman SV (1997) Selecting and interpreting measures of thematic classification accuracy. Remote Sens Environ 62(1):77–89. https://doi.org/10.1016/S0034-4257(97)00083-7

Tien BD, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016) 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 13:361–378. https://doi.org/10.1007/s10346-015-0557-6

Yu H, Ma Y, Wang L, Zhai Y, Wang X (2017) A landslide intelligent detection method based on CNN and RSGR. IEEE International Conference on Mechatronics and Automation (ICMA), Takamatsu, Japan, pp. 40–44. https://doi.org/10.1109/ICMA.2017.8015785

Stefania Magrì · Monica Solimano · Tania Del Giudice · Mauro Quagliati

Regional Agency for the Environmental Protection of Liguria (ARPAL), Via Bombrini 8, 16149 Genova, Italy

Fabio Delogu · Francesco Silvestro (✉)

CIMA Foundation, Via Armando Magliotto, 2, 17100 Savona, Italy

Francesco Silvestro

Email: francesco.silvestro@cimafoundation.org

Michele Cicoria

Regional Agency for the Environmental Protection of Lombardia (ARPALombardia), Via Rosellini 17, 20124 Milano, Italy