This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3158328, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Predicting and Understanding Landslide Events with Explainable AI

E.Collini[1], L. A. Ipsaro Palesi[1], P. Nesi[1], G. Pantaleo[1], N. Nocentini[2], A. Rosi[2]

1) DISIT lab, Dept. Information Engineering, , https://www.disit.org, https://www.snap4city.org,
2) DST, Dept. of Earth Science, University of Florence, https://www.dst.unifi.it/
University of Florence, *ref email:* paolo.nesi@unifi.it

ABSTRACT Rainfall induced landslide is one of the main geological hazard in Italy and in the world. Each year it causes fatalities, casualties and economic and social losses on large populated areas. Accurate short-term predictions of landslides can be extremely important and useful, in order to both provide local authorities with efficient prediction/early warning and increase the resilience to manage emergencies. There is an extensive literature addressing the problem of computing landslide susceptibility maps (which is a classification problem exploiting a large range of static features) and only few on actual short terms predictions (spatial and temporal). The short-term prediction models are still empirical and obtain unsatisfactory results, also in the identification of the predictors. The new aspects addressed in this paper are: (i) a short-term prediction model (1 day in advance) of landslide based on machine learning, (ii) real time features as good predictors. The introduction of explainable artificial intelligence techniques allowed to understand global and local feature relevance. In order to find the best prediction model, a number of machine learning solutions have been implemented and assessed. The models obtained overcome those of the literature. The validation has been performed in the context of the Metropolitan City of Florence, data from 2013 to 2019. The method based on XGBoost achieved best results, demonstrating that it is the most reliable and robust against false alarms. Finally, we applied explainable artificial intelligence techniques locally and globally to derive a deep understand of the predictive model's outputs and features' relevance, and relationships. The analysis allowed us to identify the best feature for short term predictions and their impact in the local cases and global prediction model. Solutions have been implemented on Snap4City.org infrastructure.

INDEX TERMS landslide prediction, machine-learning, explainable artificial intelligence, snap4city

I. INTRODUCTION

Landslides are increasingly frequent geologic events which may involve rural areas, as well as cities and impact on largely populated areas. These phenomena are responsible each year of several losses and casualties; according to [1], from 2004 to 2016, 55997 people were killed in 4862 non seismic landslide events worldwide, with a major incidence in Central America, Caribbean islands, South America, along the Andes mountain chain, Asia, East Africa, Turkey and the Alps in Europe. The same authors identified rainfall as the main the triggering factor of 79% of non-seismic landslides.

Italy is the European country most affected by landslides, with about 2/3 of know landslide in Europe [2]; in fact, over 620'000 known landslides, covering almost 24'000 km^2 (7.9% of the whole national territory), are present, according to the Italian landslide inventory [3]. From 1971 to 2020, 1079 fatalities have been caused by landslides in Italy, along with 1416 casualties and over 146'000 evacuated and homeless [4]. Tuscany is an Italian region highly affected by landslides, since about 91700 landslides are present [5], covering 2107 km^2 (9% of the territory). The province of Florence, due to its geological setting, mainly made of clay-sandy deposits and its morphology, made of alternating valley and hills, is quite susceptible to landslide. These phenomena pose a real risk for the population and one of the possible solutions for its reduction is the setting up of early warning systems. Typically, "wake-up call" and early warning systems are setup to inform the population about the occurrence of landslides in quasi real time. Short term predictions, ranging from a few hours to one/two days, could save a relevant number of people. Thus, the short-term prediction of landslide events could be a very powerful tool in the hands of authorities to organize evacuations and manage an emergency since its inception, thus preventing human injuries due to such catastrophic events.

The most common approaches rely on statistical or empirical approaches mixing static information describing the terrain with real time data computed on the basis of recent days. In particular, as to rainfall induced landslides, in [6] and [7] authors highlighted the correlation of the amount of rainfall in the days preceding the landslide event (from 3 to 245 days), by means of statistical analysis [6], [7], while other scholars used the empirical method of rainfall thresholds to identify rain conditions associated with such landslide triggering [8], [9]. Machine learning approaches are widely used in landslide hazard mapping [10] which can be regarded

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

as a classification rather than a prediction tools. Those approaches produce landslide *susceptibility* maps – to identify areas that could experience a landslide in the future. This kind of maps can be useful for long term land usage planning, but not for early warning purposes, since they do not give any information about the possible time of occurrence of the event.

At the state of the art, there are very few examples about the use of machine learning for short term forecast landslide occurrence [11]. In landslide events, the triggering is caused by the loss of cohesion in the soil, due to its saturation from rainwater or from the raising of groundwater level. This reduction of cohesion leads to the reduction of the shear stress of the slope, therefore restraining the factor of safety. In fact, in [11], the main factors analyzed were precipitation duration, mean intensity, and total volume (cumulated rainfall), obtaining with machine learning a TSS (true skill statistic) of 0.59. On such reasons the groundwater level (which is, in turn, influenced by rainfalls, in just previous days) is an important factor in the occurrence of landsides, as also remarked into the review of the state of the art reported [12].

Other relevant factors in terrain slope stability are the type of vegetation and soil, the slope of the topographic surface, profile curvature, distance from rivers, altitude, and soil landslide critic level (as assessed by experts). They are factors, change slowly over time, and may influence the stagnation level of rainwater in the soil [13]. Therefore, they influence somehow the consistency of the soil, and thus the groundwater level is an important factor correlated with the land instability and the occurrence of landslide events [9]. In many research works, field data have been the starting point for computing predictions, taking into account databases of registered geological and natural events (e.g., earthquakes, landslides, floods, river or lakes overflows) as reference event values. In most cases, events have been catalogued by experts according to their severity, depth, size, and persistence over time, and they are typically collected from blogs (RSS, etc.) or web pages [14], [15], recommendation systems of alert (e.g., like the ones from Civil Protection, national institutes of geophysics, etc.), sensor networks, statistical data and annual reports, etc., [16]. Moreover, current studies on landslide identification are based on optical images using pixel-based or object-oriented methods, and the digital terrain model (DTM) is combined with optical images and digital elevation model (DEM) derivatives to identify translational landslide scars using object-oriented methods [17], [18].

The creation of accurate forecasting models useful for early warning activities may be grounded on a wide range of data provided by different (static and real time) sources. Thus, taking into account recent events and the short-term conditions. This is one of the major differences with respect to the solutions which are computing susceptibility maps. The data aggregation implies to manage a variety of: licenses, protocols, standards, tools and formats. Thus, a multitude of historical and real-time data must be analyzed, so the data size and their processing speed are considerable. When it comes to the combination of these aspects, we can consider to be in the context of Big Data, for volume, variety, velocity, veracity, and value of data. Moreover, with the aim of producing predictions in a data driven approach, many different machine learning and deep learning algorithms have been applied in a variety of use cases: Logistic regression (LR), Support Vector Machine (SVM), Random forest (RF), Boosting, Convolutional Neural Network (CNN), as stated in [14], [17], [19], [20] [21]. The SIGMA algorithm, which was firstly developed in Emilia Romagna Region [6] and then tested in India [22], is a landslide early warning model based on the analysis of the probability related to exceedance of defined rainfall amounts. The latter has been also used and calibrated in our study area, which is the Province of Florence and then compared with some machine learning algorithms.

A. Related Words

The problem of computing landslide susceptibility and risk maps, displacements, and short-term predictions has been addressed through different approaches and this section presents the state-of-the-art of Artificial Intelligence solutions, as summarized in **Table I.** The main goal of the paper is on *short term prediction for early warning*, while the related works addressing similar problems could be useful to identify the features and context. In the context of this paper, the keyword susceptibility is used to describe the production of maps providing the proneness of the terrain to sliding which can be regarded as long-term prediction, which is a different goal.

Nam and Wang in [23] used Stacked Autoencoders combined with RF for the landslide *susceptibility* assessment. The areas of study were in Oda and Gotsu Cities in the Shimane Prefecture, in Japan, where 90 landslides occurred due to extreme precipitation from May to October 2013. The data used refer to the Digital Elevation Model (DEM), remote sensing and geological factors, all static variables. Researchers compared SVM, Stacked AutoEncoders (St-AE), Sparse Autoencoders (Sp-AE), and RF *classifiers*. As a result, they identified the best solution combining St-AE with RF, obtaining a True Positive Rate (TPR) of 0.93.

The Autoencoders have been also used by Huang et al., to predict the landslide *susceptibility* in the Sinan Country of Guizhou Province in China [24]. In that case, 306 landslide events were registered from the 1980s to 2010s. The data sources for the landslide predictions regarded 27 environmental static factors considering: topographic, geological, hydrological, and land covers features. The tested solutions were: SVN, Backpropagation Neural Network (BPNN), and Fully Connected Sparse Autoencoder (FC-SAE). The reported validation metrics have been the True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN), Positive Predictive (classification) Rate (PPR), Negative Predictive (classification) Rate (NPR), Accuracy. The FC-SAE architecture achieved its best results with an Accuracy of 85.2%, compared to 81.56% for the SVN and 80.86% for the BPNN.

Pham et al., in [25], used a Machine Learning (ML) technique for landslide *susceptibility* analysis, the CNN with a specific optimization algorithm for parameters selection. The study area of Lai Chau is a mountainous province of Vietnam, the dataset consisted in 2374 points of landslides and randomly selected non landslides with 12 area features (Elevation, Aspect, Slope, Stream Power Index (SPI),

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3158328, IEEE Access

TABLE I RELATED WORKS, DIFFERENT APPROACHES ANALYSING LANDSLIDES

Authors	Landslide Target	Features	Dataset	Model	Results
Nam and Wang, 2020 [23]	susceptibility	Altitude, Slope, Plan curvature, Distance to stream, SPI (Stream Power Index), TWI (Topographic Wetness Index), NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index), Rainfall, Distance to road, Geological age, Lithology	Oda City and Gotsu City in Shimane Prefecture Japan	Stacked AE combined with RF	St-AE + RF TPR: 93.2%
Huang et al., 2020 [24]	susceptibility	Elevation, Aspect, Plan Curvature, Surface roughness, Surface cutting depth, Slope Form, Geomorphic map, Total surface radiation, Surface temperature, Average annual rainfall, Topographic wetness index, Distance to river, MNDWI (modified normalized difference water index), Population density, Land use types, Distance to road, BSI (bare land soil index), NDBI (normalized difference building index)	Sinan Country of Guizhou Province in China	fully connected sparse AE	FC-SAE True pos: 6177 True neg: 5677 False pos: 1279 False neg: 780 PPR: 82.85% NPR: 87.92% Accuracy: 85.20%
Pham et al., 2020 [25]	susceptibility	DEM, Aspect, Slope, CTI , SPI , Curvature, NDVI, NDWI, NDBI, Distance to river, River density, cumulative rain for 4 months, Historical landslides occurrences,	Lai Chau province of Vietnam	CNN with Optim.Moth Flame Algorithm	CNN-FMO RMSE: 0.3685 MAE: 0.2888 AUC: 0.889 OA: 80.11%
Pei, Meng and Zhu, 2021 [26]	displacement	Water Level, Velocity of the water, Precipitation, Periodic Displacement	Three Gorges Reservoir area	CNN	CNN RMSE/mm: 9.97 MAE/mm: 8.29
Karunanayake et al., 2019 [27]	prediction of riskiness	Overburden Land use Slope Rainfall	Badulla and Nuwara Eliya districts, Sri Lanka	RF	RF TPR: 98.15%
Cheng et al., 2021 [28]	susceptibility	LULC (land-use/land-cover) types, Recharge of ground water, Distance to the bank of rivers, Distance to old landslides, Distance to dip slope, Geological line density, Distance to roads, River density, Aspect, Slope, NDVI, Wetness	Tsengwen River Watershed, Central Taiwan	RF	RF OA: 99.7 % Kappa Coefficient: 0.99
Wang et al., 2021 [29]	stability	Slope, Elevation, Curvature, Aspect	Santai County	XGBoost	Accuracy: 0.89 Recall: 0.94
Ngo et al., 2021 [30]	susceptibility	Altitude, slope degree, profile curvature, distance to river, aspect, plan curvature, distance to road, distance to fault, rainfall, geology and land-use	Iran	RNN, and (CNN)	AUC: 0.88 MSE: 0.007 RMSE: 0.083
Thai Pham, Binh, et al., 2019 [31]	susceptibility	LCF (Landslide Conditioning Factors), Overburden Depth, Land Cover, Geomorphology, Distance to Rivers, Distance to Roads, Curvature, Aspect, Slope, Valley Depth, SFM (Slope Forming Material)	Uttarakhand, India	ADtree, BAADT, RSADT, RFADT	Recall: 0.717 FPrate: 0.285 Precision: 0.771 Kappa: 0.433 RMSE: 0.397 AUC: 0.931
Tien Bui, Dieu, et al., 2019 [32]	Susceptibility	LS (European Slope length and Steepness factor), SPI (Stream power index), TPI (topographic position index), TWI (topographic water index), TRI (topographic roughness index), Land use, Lithology, Average Annual precipitation, Altitude, Slope, Aspect, General curvature, Plan curvature, Longitudinal curvature, Tangential curvature, distance to stream, distance to road, distance to fault	Sarkhoon watershed, Iran	ABSGD, SGD, LR, LMT, FT	ABSGD Accuracy: 0.776 Sensitivity: 0.833 Specificity: 0.835 RMSE: 0.411 RMSE: 0.861
Zhang, Tingyu, et al., 2018 [33]	susceptibility	Aspect, Slope, Altitude, Lithology, Mean Annual Precipitation, Distance to roads, Distance to rivers, Distance to faults, Land use, NDVI	Fugu County of Shaanxi Province, China	IOE (Index of Entropy method), (LR)-IOE, (SVM)-IOE	(LR)-IOE AUC: 0.8184
Abraham et al., 2021 [22]	prediction	Rainfall data	Idukki, India	SIGMA	Accuracy: 79.31% Sensitivity: 0.88 Specificity: 0.79 Likelihood Ratio: 5.62%

Compound Topographic Index (CTI), Curvature, NDWI, NDVI, Normalized difference build-up index NDBI, Distance to river, River Density, Precipitation in long term).

The exploited assessment metrics have been the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Area under Receiver Operating Characteristics (AUC), Overall Accuracy (OA). The proposed CNN architecture achieved better results compared to Random Subspace, RF and CNN using conventional Adagrad optimizer, with OA of 80.105%.

The CNN architecture has been also used by Pei et al., in [26]. Their study focused on the influence between time-varying trigger factors and the periodic landslide *displacement*. The specific area of this study is in Zigui, Hubei Province, China. In order to find the best solution for landslide displacement, researchers compared the 1-D CNN with the SVR. They stated that the 1-D CNN yields to more precise predictions, due to its feature extraction ability, and indeed the results in terms of RMSE/mm and MAE/mm are 9.97 and 8.29, respectively, compared to the 15.35 and 11.14 obtained by the SVR.

In the case study of Karunanayake et al., [27], for the *prediction of landslides riskiness*, the implemented ensemble learning techniques (RF) achieved better results compared to deep learning techniques. The work is based on the Badulla and the Nuwara Eliya districts in Sri Lanka. The dataset is made up of 81 landslides registered in each district, including measurements of the current weather and most significant and dynamic geographical conditions of that particular area. As evaluation metric researchers chose the TPR. The RF technique achieved better results compared to the DNN (Deep neural network) with a TPR on the test set of Badulla district of 96.29%, compared to the 92.59% obtained by the DNN, and a TPR of 100% for the Nuwara Eliya district, whereas the DNN correctly classified 26 out of 27 landslides. According to the above summarized TPR percentages, decision tree models outperformed the neural network models.

The ensemble learning techniques have been also used in the work of Chen et al., [28] for computing of landslide probability in long term (*susceptibility* map), for the area of Tsengwen River Watershed, Central Taiwan. Using optimal hydrological, geological, and topographical variables the RF technique achieved an overall Accuracy of 99,7%. Researchers stated that, despite different resolutions between ground reference and susceptibility maps that could determine an exaggeration in the landslide mapping accuracy, the used methods could provide reliable spatial and quantitative information on landslides.

Wang et al., in [29] compared the SVM Classifier, RF, and the XGBoost for the classification of landslide *stability*. Researchers used the topographic features extracted by the DEM elevation, slope, aspect, curvature and shape. The best classification technique turned out to be the XGBoost, providing an Accuracy of 89% and a Recall of 94%, outperforming the RF (which obtained an Accuracy of 88% and Recall of 91%) and the SVM (which achieved an Accuracy of 76% and Recall of 86%).

Thai Pham, Binh, et al., in their research study [31] assessed the problem of landslide *susceptibility* in Uttarakhand, India using a Hybrid Machine Learning Algorithm made of three meta-classifiers Bagging (BA), Random Subspace (RS) and RF combined with ADTree as a weak base classifier. The RF-ADtree was the best hybrid model based on the results of the paper achieving an AUC of 0.931.

Also Tien Bui, Dieu, et al. in [32] used a hybrid machine learning approach but applied to the problem of Shallow Landslides Prediction (which is a *susceptibility* map estimation as claimed by the authors in the conclusions). The method developed was a combination of a functional algorithm, stochastic gradient descent (SGD) and an AdaBoost Meta classifier. The researcher used 20 landslide conditioning factors to produce a reliable landslide susceptibility map for the Sarkhoon watershed in Iran.

The objective of producing reliable *susceptibility* maps for the Fugu County of Shaanxi Province, China was assessed by Zhang, Tingyu, et al., in [33]. They also used a hybrid integration approach but with the Index of Entropy (IOE), Logistic Regression (LR) and SVM. The LR-IOE model is the one with the highest accuracy of precisely 0.9011, it takes into account about rain by means of the annual average.

The SIGMA model has been used by Abraham et al., [22] in order to compute landslide *predictions* in the area of study of the Idukki district in India. Researchers used rainfall data and divided the district of study into 4 reference areas according to the topographic variability and location of rain gauges. The dataset used covers the years from 2009 to 2018 and the last one has been used to validate the SIGMA model. The model obtained a 79.31% mean Accuracy over the four areas.

B. Paper scope and structure

There is a large literature on landslide for susceptibility analysis, displacement, risk and prediction assessment, and most recent solutions are adopting machine learning and deep learning approaches. Susceptibility maps can be regarded as long terms predictions (proneness) providing a spatial map about probability of sliding. The estimation of susceptibility maps is performed on the basis of a number of static and quasi static variables describing the soil, terrain etc., and in some cases annual average for rain, etc. The identified features have been weather, rain, slope, vegetation, temperature, humidity, wind, soil kind, altitude, etc. The addressed machine learning techniques are: RF, XGBoost, CNN, and AE.

In this paper, the problem of computing short term predictions of landslide events has been addressed. The results can be used for immediate evacuation and early warning of population, rather than for planning, which is the main use of susceptibility map. The new aspects addressed in this paper are: (i) a short-term prediction model (1 day in advance) of landslides based on machine learning, which cab be used for early warning, (ii) a set of real time features as good predictors. Some of them have been also considered by the heuristics of the SIGMA predictive model [22]. The introduction of explainable artificial intelligence techniques allowed to understand and identify global and local feature relevance. In order to find the best prediction model a number of machine learning solutions has been implemented and assessed (e.g., RF as in [28] for susceptibility, XGBoost as in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3158328, IEEE Access

[29] for susceptibility maps, CNN as in [25], [30] for susceptibility maps, and AE). These models have been trained, validated and compared one another and with the SIGMA approach from the literature.

Solutions have been trained and validated by using data in the Metropolitan City of Florence from 2013 to 2019. The area is quite prone to landslide events. Thus, producing results to explain the approach and the phenomena. The research activity (named PC4City, civil protection for the city) has been partially funded by Foundation Cassa di Risparmio di Firenze and has been developed in collaboration with the Department of Earth Science of the University of Florence. The solution has been developed exploiting the data available in the area and the smart city infrastructure and living lab named Snap4City: https://www.snap4city.org

The structure of this paper is as follows. Section II describes the architecture of PC4City while stressing its relationships with the Snap4City framework adopted in the area. Section III describes the exploited data and the identified and computed features. In Section IV, both adopted machine learning techniques and SIGMA model have been presented along with their running parameters and metrics for result assessment. Section IV.B presents the results of the validation phase after training. Section V is focused on the local and global explanation of the best results obtained with the XGBoost method. Conclusions are drawn in Section VI.

II. P4CITY ARCHITECTURE

According to the above reported state of the art, some solutions aiming at computing some early warnings have been proposed. Early warning systems can be regarded as 24 hours predictors or early pattern detectors.

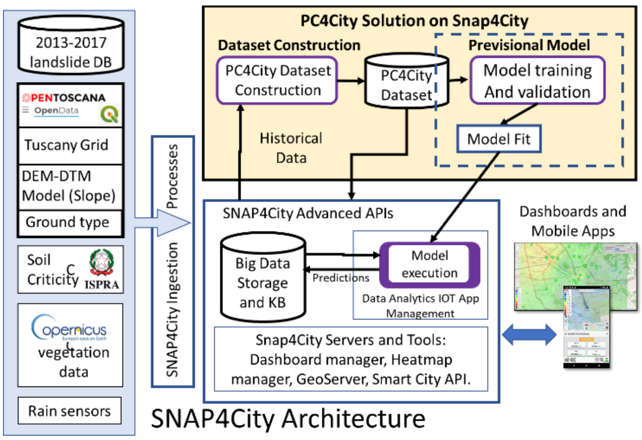


FIGURE 1. P4CITY Datasets and Solution in the context of Snap4City Architecture.

The complexity in this case is mainly due to the heterogeneity of data and the amount of data to be processed in short time. The solution presented in this paper is called PC4City, and it has been set up by exploiting the Snap4City architecture and service, which is in place in the Florence/Tuscany area, as well as on other regions in Europe [34], [35]. The Snap4City framework (briefly exploited in **Figure 1**, with its application within PC4City project) allows to collect data of any kind, to save them into a big data store where they can be queried for recovering specific historical data segments. The same storage can be used to collect data in real time and to save data analytic results.

The general workflow included activities of:

- **Data ingestion,** historical and real time data to be updated, for example rainfall, weather, data coming from satellites regarding vegetation, etc.
- **Dataset construction** for predictive model training and validation. This activity is preparing the dataset for the next step where the predictive model is produced and validated.
- **Predictive Model training and validation**. This activity is focused on producing the Predictive Model (Model Fit) (for example, based on machine learning or other solutions). The produced model is validated in other areas to assess its reliability, sensitivity and robustness.
- **Model execution**, takes in input both real time data and Model Fit to produce predictions which could be estimated 24 hours in advance and may be used to inform civil protection authorities, municipality, etc. The resulting model assesses in real time the probability of landslide events as early warning/prediction.
- **Publication of results** on specific Dashboards, Mobile Apps, etc.

In PC4City, data ingestion processes, as well as activation of data analytics, are performed by using Node-RED processes on docker containers. Node-RED flows can exploit the platform MicroServices with a specific library of node.js [35]. In addition, Data Analytics processes have been developed by using Python and/or Rstudio. In the case of PC4City, some Node-RED IoT Applications have been developed for data ingestion and specific Python processes have been developed for implementing the Predictive Model Training and validation and for the Model Execution. The IoT App in Node-RED governing the Python for Model Fit Execution may also decide to send alerts via Telegram, SMS, email. Finally, resulting data, as well as previous data, are visually presented by using a Dashboard exploiting the Dashboard Builder.

III. FEATURE AND DATA PREPARATION

In order to test and validate our approach we have collected a large dataset in Tuscany, in the Florence province (also called Metropolitan City Area) from 2013 to the 2019, with the aim of developing and validating a solution for the early warning and 1-day ahead prediction of landslide events. In the observation and analysis area, historical data regarding landslide events have registered 341 landslides from 2013 to 2019 [6]. To each and every landslide event we assigned an ID, the date when that landslide occurred, latitude and longitude expressed in EPSG:4326. Those points are located in their actual coordinates, and for each of them a given number of parameters is accessible such as: wideness, severity, duration, etc.

A. Grid definition

With the aim of computing a prediction / early warning in each point of the area, a dense grid of points was defined where the prediction could be estimated. The size of the grid is a critical aspect, since the prediction should be as much

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3158328, IEEE Access

precise as possible, while data would not be accessible with high precision and number of points would be prohibitive for computation.

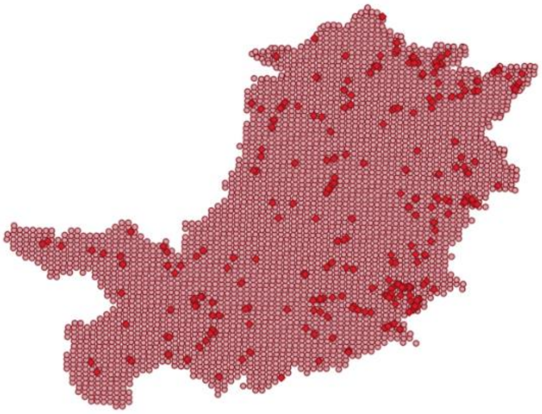


FIGURE 2. Grid and landslide events in the Florence Metro Area (Tuscany, Italy) from 2013 to 2019. An area in which live 1.5 M inhabitants.

So that a compromise is needed, the grid size has been defined according to the size of the landslide events, at least ½ to be sure to sample the event. For these reasons, the grid has been defined as a compromise (points distance of 1000 mt in both directions, obtaining 3582 areas, covering the whole Florence Metro area of 3514 Km^2, and a little more at the borders) as depicted in **Figure 2,** in which the RED dots are the events of landslide registered in 2013-2019.

The area presented a large number of landslide events having a relevant range of different features in terms of: criticism, altitude, slope, vegetation, cumulated rain, type of soil, etc. As a result, the set of points in the grid may have a set of associated data that would be taken from: sensors (for example: rain, temperature, humidity, etc.), geographical information systems of the territory, satellite services, and from the landslide occurred dataset, too.

B. Feature selection

The features in each area segment of the grid have been selected by analyzing the state of the art in studying landslides and from specific authoritative providers in the area. This allowed us to identify a number of possible features that may influence (and/or may be used as short-term predictors of) landslide events, and also taking into account terrain features as identified by long terms susceptibility analysis. One of the most relevant features that influences the landslides is the **soil water content**. These aspects can be directly measured with sensors in the soil, which is unfeasible for large areas and usually rain sensors on ground are not adopted. The same information could be indirectly measured based on the rainfall received in past days. The value of rain in each area of the grid cannot be estimated due to the lack of dense sensors, whereas data coming from satellite are very heavy to be processed and not precise, since also clouds contain water while covering the view of the terrain. On these reasons we decided to indirectly measure the amount of rainfall which reaches the ground from a number of sensors (so called SIR Sensors in Tuscany)**.** The values of sensors have been interpolated by using IDW (Inverse Distance Weighting) algorithm [36], which is also used in Snap4City to create Heatmaps. On the basis of such scattered data, we have estimated 4 derived features: *Day1*, *Day3*, *Day5*, *Day30*, which compute for each day the amount of rain in mm arriving on ground within a specific area in the last day, 3 days, 5 and 30 days, respectively, as performed in SIGMA model [22].

A second parameter which may be related to the landslide proneness may be the **geological nature** and the **terrain slope**. Geology is known to be a controlling factor when it comes to large and deep landslides, while small and shallow landslides (depth < 2 m) are somehow independent from the bedrock's geological nature, since they are usually located in more surficial layers of soil. On the other hand, the terrain slope, which is known as one of the main controlling factors of shallow landslides, may radically change in different parts of the same area. A Digital Terrain Model has been created by processing the Lidar survey carried out in 2017 and available among the Open Data of the Metropolitan City of Florence. The ***Slope*** feature has been associated to each area of the grid (as a percentage). Please note that these values change sporadically over time. Therefore, an update performed every month / year would be more than enough.

An additional aspect to consider is the **land usage** of the area. For this purpose, land use and land cover datasets of regional government, and in particular Tuscany Region geoserver, provided the data. This allowed to associate a value describing the type of ***Ground*** to each grid area in terms of identifiers referred to the CORINE Land Cover, CLC technical guidelines [37]. This work has been performed on a QGIS tool. Please note that these values change very slowly in time, and thus they have to be updated once a month or year.

A similar view but for a different purpose has been the identification of the vegetation which may also influence landslide events. **Vegetation** may keep the land connected to the ground. To this end, Copernicus satellite data have been collected exploiting the services of Snap4City Platform (https://www.snap4city.org/671 [38]) which automatically harvests, downloads and processes several different kinds of Copernicus data. The vegetation level may change over time, and thus the satellite data can give the precise, and almost real time information of the vegetation level. On the other hand, some processing has been made, since the satellite data may be influenced by clouds coverage, and they need also to be remapped from large to small grid areas.

Features have been enriched with some conditioning factors coming from the historical archives of the Regional Hydrological and Geological Sector (SIR). Tuscany region has a network in telemetry consisting of over 700 sensors for meteo-climatic data monitoring; such sensors are located in a homogeneous manner throughout the regional territory. 7 conditioning factors were obtained from these sensors involving wind speed, temperature, precipitation, daily hydrometric level and data providing information related to groundwater resources (water table data). Another feature

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3158328, IEEE Access

enrichment was made with data regarding temperature, moisture and average wind speed from the historical archive 'ilmeteo.it.' Compared to SIR data, ilmeteo.it could provide information associated with larger areas, such as cities (in our case the municipality of Florence).

Regarding the insertion of landslide data, 341 registered landslide events have been mapped over time to the grid, based on their positions and date of occurrence, and they have been labeled with the following criteria: value of 1 has been assigned to all grid cells included in an area of 1.5 km radius, centered on the coordinates of each landslide, in the previous day of its occurrence (for a total of 2342 areas impacted by landslide events); the value of 0 has been assigned to all other cells. The haversine formula has been used for distance evaluation. Please note that 7 years, multiplied by 365 daily values on 3582 areas compose a dataset of 9.153010 million elements, among which 2342 represent areas affected by landslide events.

At the end of the process, for each grid point, features composing the dataset have been the ones reported in Table II. Please note that most of them are new features that describe the short-term condition of the area, and thus they need to be actualized every day.

TABLE II FEATURES DETAILS

Feature	Description	Unit	Example
Date	Observation date, in the format YYYY-MM-DD	Day	2013-01-14
Latitude	Latitude of the area, EPSG:4326 format	Deg	43.86239
Longitude	Longitude of the are in the EPSG:4326 format	Deg	11.51586
Altitude	Altitude of the area	M	467.204
Slope	Acclivity of the area	%	45.942
Vegetation	Vegetation of the area	%	0.262
Ground	Soil type at the event site (class UCS)		223-Oliveti
Day1	Rainfall in the day before the observation	mm	12.453
Day3	Rainfall in the 3 days preceding the observation	mm	15.072
Day15	Rainfall in the 15 days preceding the observation	mm	16.160
Day30	Rainfall in the 30 days preceding the observation	mm	51.515
Temperature	Mean Temperature on the observation day (IlMeteo.it)	°C	6.965
MinTemperature	Minimum temperature on the observation day (IlMeteo.it)	°C	2.99
MaxTemperature	Maximum temperature on the observation day (IlMeteo.it)	°C	9.942
Humidity	Humidity (average) on the observation day (IlMeteo.it)	%	92.96
WindSpeed	Average wind speed on the observation day (IlMeteo.it)	Km/h	5.991
VelMedSIR	Average wind speed on the observation day (SIR)	m/s	0.9
VelMaxSIR	Maximum wind speed on the day of observation (SIR)	m/s	1.8
LevelSIRFre	phreatimetric data on the observation day (SIR)	m	-4.34
LevelSIRIdr	Water (river) level recorded on the observation day (SIR)	m	0.8
PrecipSIR	Precipitation on the observation day (SIR)	mm	0
MinTempSIR	Minimum temperature on the observation day (SIR)	°C	0.5
MaxTempSIR	Maximum temperature on the observation day (SIR)	°C	3.5

IV. DATA ANALYTIC SOLUTIONS

On the basis of the above-described dataset, a number of techniques to predict landslide events has been tested. Aiming at creating an early warning can be traced back to the estimation of areas presenting a high probability of landslide event occurrence in the next day, as in this case. Therefore, the dataset included several items representing non-slide events (referred hereafter as negative events) and items representing landslide cases (referred hereafter as positive events). As described in the previous section, the considered dataset is composed of about 9 million estimations, among which 2342 positive events (labeled with *Value* = 1). The input dataset was composed by the following variables:

- X = independent variables = {Latitude, Longitude, Altitude, Slope, Vegetation, Day1, Day3, Day15, Day30, Ground, Temperature, MinTemperature, MaxTemperature, Humidity, WindSpeed, VelMedSIR, VelMaxSIR, LevelSIRFre, LevelSIRIdr, PrecipSIR, MinTempSIR, MaxTempSIR}
- Y= dependent variable = *{Value of the day after}, 0 no sliding, 1 land sliding.*

In order to build the model, we have divided the dataset into two groups: training set (80%) and test set (20%). The selection of the data belonging to the two sets has been performed randomly but considering the same ratio of distributions for both positive and negative cases in training and test sets.

A. Machine Learning Models Adopted

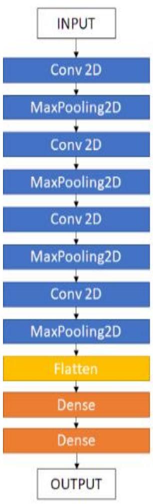
Most State-Of-The-Art works addressing the problem of landslide prediction are formulated as a classification problem. As a further development we investigated the possibility of predicting the occurrence of landslides 1-day in advance for this case study in the Florence Metropolitan Area. In this section, the considered machine learning techniques are compared with the aim of predicting landslide events. Therefore, each model is presented with a short overview and related information about how it has been used in this context.

Random Forest, RF, is a learning algorithm based on a set that includes n collections of uncorrelated decision trees. In our case, the model has been realized exploiting the *RandomForestClassifier* of the *sklearn* library. In order to classify the dataset, a high number of trees in the forest has been used (n_*estimators* = 100), each reaching a maximum depth given by: *max_depth* = 30. The criterion used to estimate the quality of each division is entropy. Since the input Dataset is unbalanced (in terms of negative and positive events), a weight to the classes in the dataset has been assigned, to give the right meaning to each value (through *class_weight*).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3158328, IEEE Access

eXtreme Gradient Boosting, XGBoost, is a specific implementation of the Gradient Boosting method using more accurate approximations to find the best tree model. A high number of trees in the forest has been used for classification (n_estimators=180), each reaching a maximum depth denoted by max_depth=40.

Convolutional Neural Network, CNN, is useful for learning spatial local features from input. It is a feedforward neural network using convolution instead of general matrix multiplication in at least one of its layers. It can capture global and local features with the aim of improving efficiency and accuracy. The model architecture is composed of four pairs of 2-dimensional convolutional layer, Conv2D, followed by a MaxPooling2D layer that down-samples the input along its spatial dimensions (height and width) by taking the maximum value over an input window for each input channel. Then, we added a flatten layer and finally we added 2 Dense layers, the former with 64 neurons and Relu activation function and the latter with a single neuron and a sigmoid activation function. An automated hyperparameters optimization was performed through a Randomized Search Cross-Validation. The best model resulting from the whole parameter optimization process and its related cross-validation is represented in **Figure 3**. The model is compiled to minimize the log loss (in our case, the binary_crossentropy metric) with an Adam optimizer.

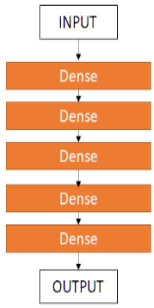


Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 1, 22, 16)	32
max_pooling2d (MaxPooling2D)	(None, 1, 22, 16)	0
conv2d_1 (Conv2D)	(None, 1, 22, 32)	544
max_pooling2d_1 (MaxPooling2D)	(None, 1, 22, 32)	0
conv2d_2 (Conv2D)	(None, 1, 22, 64)	2112
max_pooling2d_2 (MaxPooling2D)	(None, 1, 22, 64)	0
conv2d_3 (Conv2D)	(None, 1, 22, 64)	4160
max_pooling2d_3 (MaxPooling2D)	(None, 1, 22, 64)	0
flatten (Flatten)	(None, 1408)	0
dense (Dense)	(None, 64)	90176
dense_1 (Dense)	(None, 1)	65

Total params: 97,089
Trainable params: 97,089
Non-trainable params: 0

FIGURE 3. The adopted CNN model Architecture.

Autoencoders, AE, represents an unsupervised model generating an output by compressing the input in a space of latent variables. The model architecture is composed of five Dense layers, the first 4 with Relu activation function and the last one with linear activation function. The best model resulting from the whole parameter optimization process and its related cross-validation is represented in **Figure** 4. The model is compiled to minimize the log loss (in our case, the mean_squared_error metric) with an Adam optimizer. The training process of the Autoencoder has been made only on non-landslide data, as it occurs in anomaly detection the typical process is learnt. Then, whenever a landslide event is given in input to the trained model, the reconstructed output is likely not to follow the pattern of a typical process, and therefore it should be classified as an anomaly.



Layer (type)	Output Shape	Param #
dense (Dense)	(None, 32)	736
dense_1 (Dense)	(None, 16)	528
dense_2 (Dense)	(None, 16)	272
dense_3 (Dense)	(None, 32)	544
dense_4 (Dense)	(None, 22)	726

Total params: 2,806
Trainable params: 2,806
Non-trainable params: 0

FIGURE 4. Autoencoder Architecture Model.

The used Autoencoder reconstruction error has been the MSE and the threshold has been evaluated at 0.4 on the test set, based on the precision and recall curves reported in **Figure 5a**. If the reconstruction error is higher than the chosen threshold, it will be classified as landslide; this is visible in the reconstruction error for the validation set on **Figure 5b.**

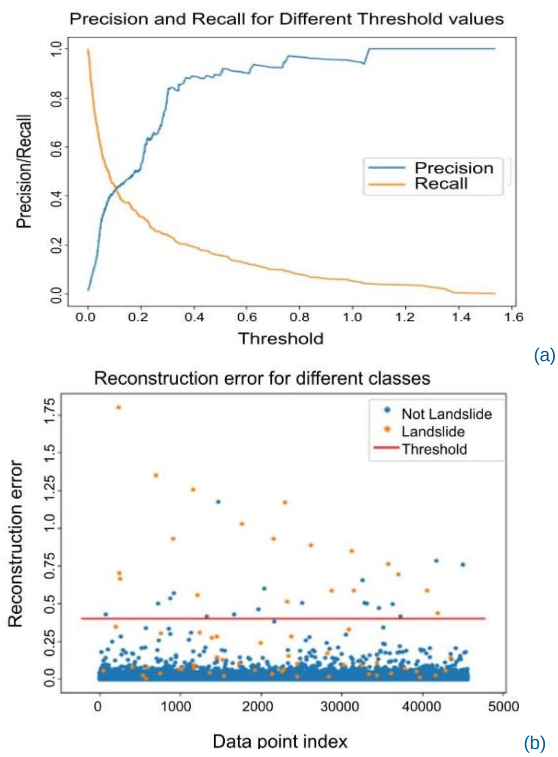
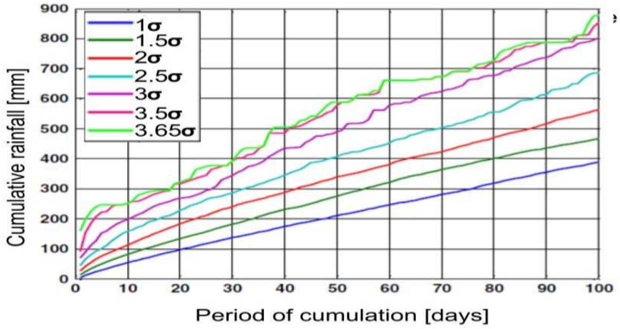
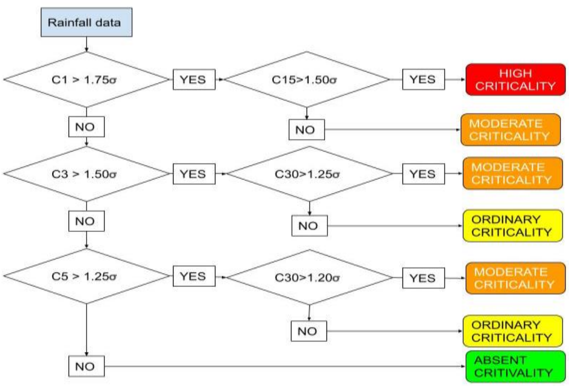


FIGURE 5. Autoencoder Model identified: (a) Precision and Recall Plot – (b) Reconstruction error plot for the validation set.

The decisional algorithm **SIGMA** has been taken into account, too (see **Figure 6**). The Sigma model has been calibrated for the city of Florence area according to the procedure described in [6], [22]. Since it is based on statistical analysis of rainfall data, rain gauges with at least 20 years of rainfall recordings have to be used and 9 rain gauges with the proper data series have been identified in the study area. For each rain station, the cumulative rainfall from 1 to n days is analyzed and the mean rain values and several standard deviation values (from 1 to 3, with steps of 0.5 standard deviation) are calculated. Then several Sigma curves, i.e., curves with the same standard deviation value for several time intervals, are defined **(Figure 6a). Figure 6b** reports the flow chart of the Sigma algorithm for early warning. Such scheme compares the cumulative rainfall in the days leading up to the event with a sigma coefficient. In order to make this sigma value more accurate, it was interpolated through the IDW algorithm (same methodology used previously to estimate Day_i cumulative rainfall and described in Section III), at each point in the dataset. In the schema reported in **Figure 6b**, values of *C1*, *C2*, *C3*, *C4*, and *C5* correspond to the *Day1*, *Day3*, *Day15*, and *Day30* values in the dataset, respectively, while the sigma symbols stand for standard deviation multiples (expressed in mm of rainfall) that must be exceeded to assign a level of criticality.



(a)



(b)

FIGURE 6. Sigma: (a) example of Sigma curves for duration from 1 to 100 days; from [1]; (b) flow chart of the algorithm: Cx represents the cumulative rainfall in x days, while the sigma symbol represents the standard deviation.

B. Method for Results' Assessment

The comparison with the results obtained at the state of the art works reported in Section 1.A, with respect to the solution proposed is discussed in this section. As stated above, most of the state of the art works for landslide analysis are focused on estimating susceptibility maps (which is a long term proneness of landslide), rather that computing predictions for early warning. For computing susceptibility map, mainly static or quasi static features metrics were used which do not depend on the specific short terms changes in land. On the contrary, predictive models such as that presented in [11] and SIGMA [6], [22] are based on rain fall data with some limited feature engineering, without the usage of explainable AI for features relevance assessment. For these reasons, in order to identify a more precise prediction model with respect to those at the state of the art we have applied a large number of machine learning approaches and SIGMA on the same area and data. To this end, we started from the same machine learning models adopted for susceptibility map and for prediction, and SIGMA.

Therefore, the results have been evaluated by using a large set of metrics defined as follows:

MAE, $MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$ (1)

Mean Squared Error (MSE),

$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ (2)

RMSE $= \sqrt{MSE}$ (3)

Accuracy $= \frac{TP+TN}{TN+FP+FN+TP}$ (4)

Sensitivity $= \frac{TP}{TP+FN}$ (5)

Specificity $= \frac{TN}{TN+FP}$ (6)

TSS = sensitivity + specificity - 1 (7)

probability of false alarm, $P.f.A = P(positive|negative)$

F1 score, $F1 - score = 2 * \frac{Recall*Precizion}{Recall+Precision}$ (8)

Matthews correlation coefficient (MCC),

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$ (9)

Overall Accuracy (OA),

OA = Accuracy + F1 + MCC (10)

Kappa index, $k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$ (11)

AUC: Area Under the Receiver Operating Characteristics (ROC) Curve.

C. Assessment of Results and Best Model Selection

In this work, we have compared the architectures used in the state of the art for susceptibility (RF as in [27], [28], CNN as in [26], XGBoost as in [29], etc.) with respect to their adoption for 1-day ahead landslide prediction in the area of Tuscany, Italy; with the adoption of different features since the feature used in susceptibility do not have short term predictive capabilities being in most cases static for the whole year or season. As a result, the XGBoost model achieved better results compared to the Autoencoders, CNN, RF, and SIGMA (which is a predictive model on [6], [22]) models. For SIGMA we assumed condition of *early warning* when the High Criticality is assessed.

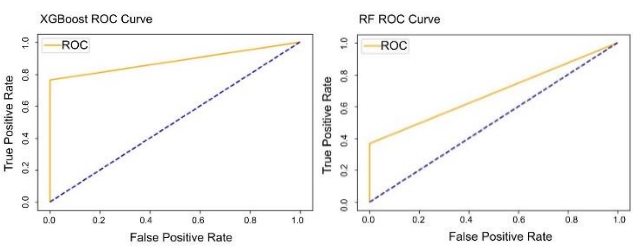
Table III shows the obtained results for landslide event predictions using machine learning models: RF, XGBoost, CNN, Autoencoders, and the SIGMA. In the machine learning model the features have been those reported in

Table II. They include a mist of features for land description (e.g., acclivity, slope, vegetation), and dynamic contextual data such as those describing rain fall, temperature, humidity, wind speed, water levels, etc. For the machine learning approaches, due to the unbalanced dataset, we have balanced the number of landslide cases in training dataset and test dataset in order to improve the RF, CNN and XGBoost performance. As to Autoencoder, all points located within a radius of less than 5 km of any landslide have been removed from dataset to prevent a non-landslide point located in the vicinity of a landslide, from presenting values of conditioning factors extremely similar to those associated with an actual landslide event.

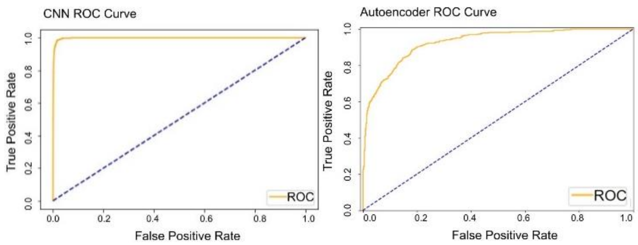
TABLE III **COMPARISON OF RESULTS OBTAINED USING MODELS FOR SHORT TERMS PREDICTION OF LANDSLIDES, BEST RESULTS IN BOLD.**

Model	XGBoost	RF	CNN	Auto encoder	SIGMA
MAE	**0.000173**	0.000334	0.000600	0.009218	0.004169
MSE	**0.000173**	0.000334	0.000259	0.009218	0.004169
RMSE	**0.0131**	0.0182	0.0160	0.0960	0.064572
Accuracy	0.99	0.99	0.99	0.99	0.99
Sensitivity	**0.79**	0.36	0.24	0.19	0.06
Specificity	0.99	0.99	0.99	0.99	0.99
TSS	**0.78**	0.35	0.23	0.18	0.05
PfA	**0.01%**	0.02%	**0.01%**	0.11%	0.39%
Precision	0.63	0.35	0.33	**0.64**	0.003
F1 score	**0.70**	0.36	0.27	0.29	0.007
MCC	**0.70**	0.36	0.28	0.35	0.01
OA	**2.40**	1.72	1.55	1.64	1.02
Kappa	**0.70**	0.36	0.27	0.29	0.01
AUC	0.89	0.68	**0.99**	0.92	0.53

According to **Table III**, on the basis of MAE metric, the best model resulted the XGBoost with a MAE of **0.000173**, compared to 0.000334 of the RF, 0.0006 of the CNN and 0.009218 of the Autoencoder. According to the results obtained in [11] as TSS=0.59, the solution proposed is better ranked reaching a TSS of =0.78. Please note that SIGMA [6], [22], provided an MAE of 0.0041, and a TSS of 0.05. A different comparative assess can be obtained by analyzing the ROC curves as reported in **Figure 7.** In this case, the CNN resulted to be the best (see AUC); while it present an unsatisfactory sensitivity and TSS. As a conclusion the best model for short term prediction of landslides which is one day in advance early warning resulted to be XGBoost (in terms of MAE, TSS, and OA).



(a) (b)



(c) (d)

FIGURE 7. ROC Curves for the 1-day ahead landslide prediction, for (a) XGBoost, (b) RF, (c) CNN, and (d) AE.

V. EXPLANATION OF THE PREDICTIVE MODEL

In order to better understand the relevance of features and their dependencies and correlation, we have applied technique for explainable AI, and interpreted the values predicted by the XGBoost model via SHAP (SHapley Additive exPlanation), both globally and locally. SHAP allows to understand the predictive model outputs and to explore relationships among features [39]. Theoretically, it is an approach from game theory explaining the output of machine learning models with respect to the values of features which act as score players in a coalition. In this case, the SHAP analysis allowed us to understand which factors are the most influential in the prediction of a landslide or not. To this end, we trained the SHAP explainer with the entire training dataset to estimate both global and local explanations as described in the following subsections.

A. Global XGBoost Model Explanation

In **Figure 8**, the graph describes the overall impact of features on predictions. The relevance of features is calculated as the average of the absolute Shapley values of the entire dataset. For example, features contributing most to the prediction of a landslide event or its absence are Day3, MaxTempSIR, and LevelSIRIdr. Therefore, we discovered that precipitation, temperature, water level in rivers, humidity are the main aspects for the prediction of landslide events. Regarding temperature, localized temperatures such as MaxTempSIR (from Regione Toscana) resulted to be more relevant that the generic area temperature: MaxTemperature which can be revered from generic services such as https://www.ilmeteo.it/. As expected, meteorological phenomena play an important role in the short-term prediction rather than other land location-related features, which can be valid for susceptibility analysis as vegetation, slope, ground kind results less relevant. Among variables concerning location, the most influential one is Latitude for the geological aspects of the territory, while this would change in different area.

Figure 9 shows the distribution of SHAP values for each feature, sorted by relevance. The x-axis represents the specific SHAP value while the y-axis represents features. Each dot/point represents the samples of our dataset, the color of the point stands for the value of a specific feature, with blue indicating a small value, while red large values for that feature. The horizontal position of the point denotes whether the feature value leads to a positive or negative

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3158328, IEEE Access

prediction. For example, as to feature LevelSIRIdr or Humidity or rain values (Day1, Day3, Day15, Day 30), high values (red dots) contribute positively to the prediction of a landslide. We can get a confirmation from the graph that high rainfall values associated with high temperatures and high levels of water within the soil have their main correlation with the prediction of landslide events.

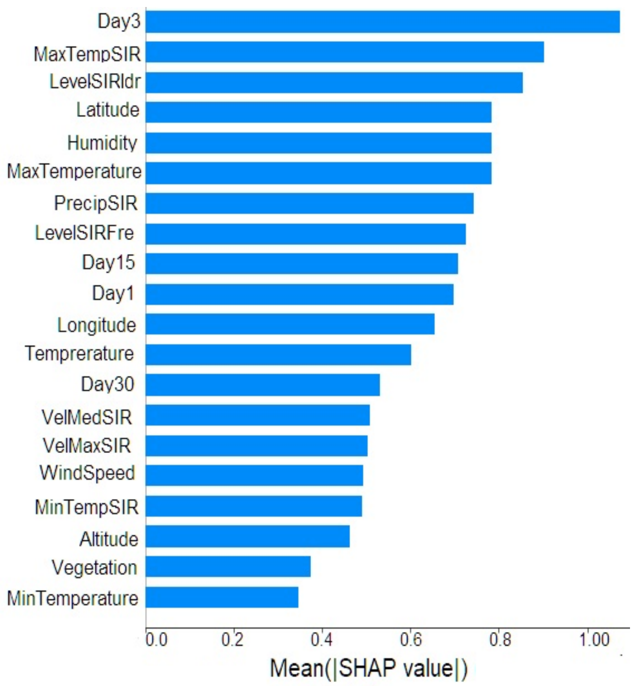


FIGURE 8. Global feature relevance as mean of the absolute SHAP global features importance for XGBoost (only the first 20).

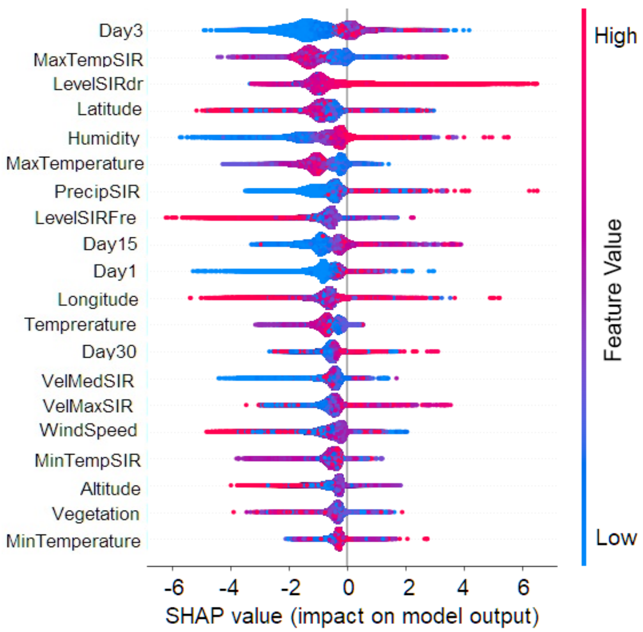


FIGURE 9. SHAP summary plot for XGBoost. x-axis reports the SHAP value of the feature, while on y-axis the features. The color codes the magnitude of the value, and the size the density of values.

B. Local XGBoost Model Interpretation

In addition to the global interpretation of the entire dataset on the XGBoust model, each single point, and thus the eventual landslide prediction, can be interpreted locally using SHAP. The local explanation puts in evidence the features which provided major contribution to the prediction. **Figure 10** illustrates 3 examples of the local interpretation of events: (a) and (b) as landslides, and (c) as a non-landslide. This SHAP plot decomposes final classification into the sum of contributions for input variables highlighting their contributions. The base value, in our case 0.4311, represents the value that would be predicted by the model if there were no knowledge of the features for current output. SHAP values are calculated in log odds. Features which increased prediction value towards a positive classification as landslide events are shown in red on the left, while features which lowered prediction value towards a negative classification are shown in blue. In our case, in **Figure 10**a the value of VelMaxSIR, MaxTempSIR, Day3 and Humidity contributed significantly to the classification of the observation as a landslide event. In **Figure 10b,** values related to rainfall in the last days, LevelSIRIdr and Humidity gave a relevant contribution to the landslide event prediction. While, in **Figure** 11c, values of features: Day3, MaxTempSIR, MaxTemperature, Temperature and LevelSIRdr have been determinant for the identification of the observation into a no landslide event.

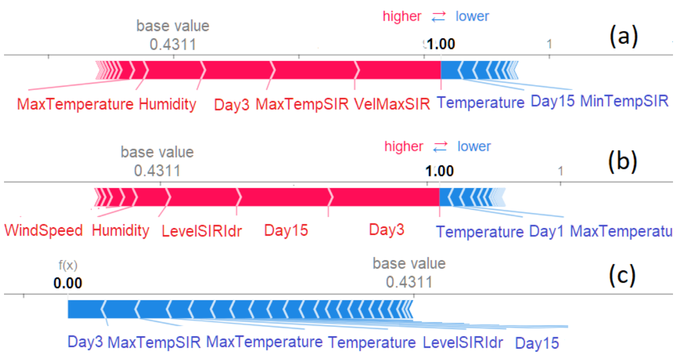


FIGURE 10. Local feature relevance via SHAP, as interpretation of events in terms of feature values: (a) and (b) are events with predictions of landslide, (c) a no landslide event.

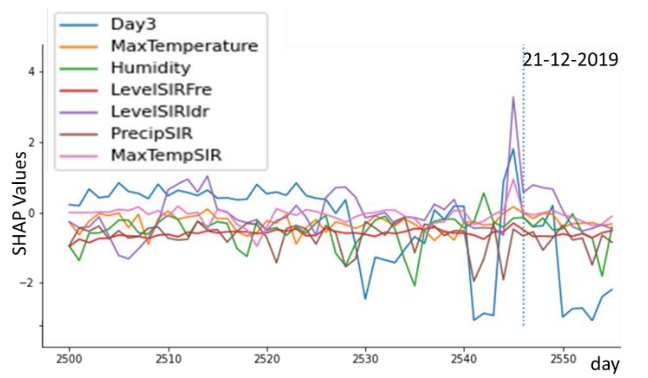


FIGURE 11. Time trend of SHAP values of most relevant features around the landslide event of 21-12-2019: values estimated by using data collected in the neighboring area of the event.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3158328, IEEE Access

A more detailed analysis of the landslide event of 21-12-2019 has been reported in Figure 11, where the trends of the SHAP values of the most relevant features according to Figure 8 have been plot with respect to the time/days. It can be noted that in coincidence of the day before the event, most of the SHAP values of the relevant features assumed a relevant value at the same time. And in particular for this event: LevelSIRIdr, Day3 and MaxTempSIR.

C. Features Dependency

In this section, some features that associate high SHAP values are furtherly analyzed. In order to understand the effect that a single feature has on the output of the model, the SHAP value of the features has been plot against the feature value for all instances of the dataset under consideration. The analysis reported in **Figure 12** presents the graphs for the most relevant features with respect to feature that has major influence or dynamic with them. Each point of the graphs in **Figure 12** represents an instance of the dataset. On the horizontal line we have the actual value of the selected feature, while the left Y axis presents the SHAP value associated with the feature. When a value along Y is positive the feature contributes positively to the occurrence of landslide event, if negative it favors the classification of the instance as a non-landslide event. The fact that the slope is upward, as in **Figure 12a,b (**where we have high values of variable with high value of SHAP), means that a higher value of the feature leads to a landslide event classification. Thus, high Humidity values or high-water levels (LevelSIRIdr) are associated with high SHAP values in predicting landslide events. Regarding the colored bar on the right, this is a reference scale for the values of a correlated second feature, the MaxTemp. In **Figure 12b,** we can see that high temperatures are typically associated with low SHAP values, thus no landslide. While in **Figure 12a,** it can be seen that high temperature with high level of humidity may lead to landslide. These graphs lead to immediate interpretation of the model. For example, similar values for a feature, as shown in **Figure 12c,** can lead to both positive and negative SHAP values to predict a landslide value. This means that the mean value of Day30 associated with high temperatures leads to higher SHAP values.

In **Figure 12e,** the high values of SHAP correspond to almost any kind of value for Day3. This means that having rain in the previous day is not enough to determine a landslide. While from **Figure 12d** we see high levels of SHAP with low levels of PrecipiSIR, which indicates the amount of rain of the day after. This may lead to confirm that the sliding may occur when the water had the time to penetrate and become stagnant.

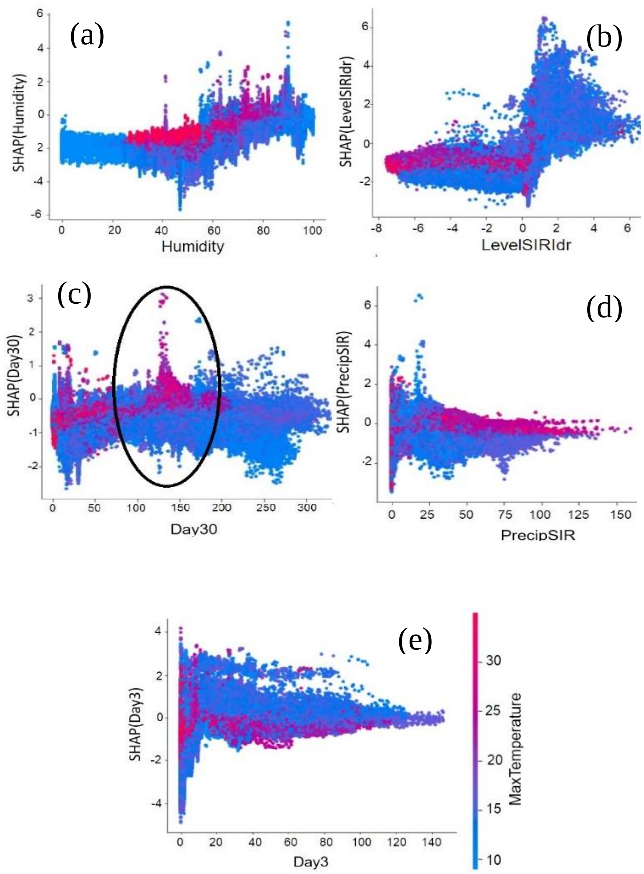


FIGURE 12. **Plots of SHAP value wrt to the feature value. Color of dots depends on MaxTemp feature: color legenda in (e) is valid for all the plots.**

VI. CONCLUSIONS

In this paper, the problem of landslide event prediction has been addressed, for early warning. A careful review of related works and solutions proposed in literature has been performed, making a comparative analysis of their results, where possible. Most of the work in the literature focused on computing susceptibility maps which is a sort of long terms estimation of landslide proneness, which are mainly based on static feature of the land. State of the art approaches for early warning (short term prediction) are empirical algorithms as SIGMA, while most recent state of the art solutions are based on machine learning. Their main limitations are represented by the fact that these systems have a low reliability (unsatisfactory TSS, OA and F1), and they are based on a limited number of features that have been supposed relevant a priori.

In this paper, we collected static and dynamic features addressing the land description but also the rain fall, temperature, wind, etc., in the previous days, in each point of a large territory and over several year. Then, a number of machine learning models have been tested to identify the best predictive model. To this purpose, this paper reports the implementation, tuning and testing of four machine learning methods, based on RF, XGBoost, CNN and AE. The models have been trained and validated by exploiting data collected in the context of the Metropolitan City of Florence since 2013 up to 2019; they have been compared with SIGMA decisional model, which is currently adopted in both Emilia Romagna and India. Comparative results showed that the method based on XGBoost achieved better results in terms of Sensitivity, MAE, MSE, TSS, OA and RMSE, with respect to SIGMA and to [6] which are the state of the art references on predictions. Moreover, a further analysis based on Shapley additive explanation (SHAP) has been carried out, globally and locally, for the XGBoost model which obtained best results. In this way, a deeper understanding of the predictive model outputs, as well as the relevance of features and their interdependency, has been provided. The results proved that features such as the amount of rain the last 3 days, the max temperature in the previous day, and the lever of water in the river are the most relevant predictors, and a number of other similar predictions may help, also on weather and water level of different kinds; also stressing that the land static feature are preconditions for landslide, while they are not efficient in creating an early warning system. From the computational point of view the short term prediction should be assess every day, while the susceptibility map usually are computed 1 or two times per year. On the other hand, the prediction models can prevent disaster while susceptibility map are mainly used for taking decision on planning.

ACKNOWLEDGMENT

Authors would like to thank Ente Cassa di Risparmio di Firenze and University of Florence, DISIT Lab of DINFO department and Earth Science Department which funded the research reported in this paper. Snap4City and Km4City are technologies and infrastructures of DISIT Lab of UNIFI. Authors would like to say thanks to LAMMA institution and Regione Toscana for data access and suggestions, and D. Cenni, L. Sbragi and D. Del Bimbo for some of the very early experiments, M. Paolucci for data ingestion.

REFERENCES

[1] Froude, M. J. and Petley, D. N.: Global fatal landslide occurrence from 2004 to 2016, Nat. Hazards Earth Syst. Sci., 18, 2161–2181, https://doi.org/10.5194/nhess-18-2161-2018, 2018.

[2] Herrera, G., Mateos, R.M., García-Davalillo, J.C. et al. (2018) Landslide databases in the Geological Surveys of Europe. Landslides, 15, 359-379.

[3] www.progettoiffi.isprambiente.it

[4] Bianchi e Salvati (2022) Rapporto Periodico sul Rischio posto alla Popolazione italiana da Frane e Inondazioni. Anno 2021 . DOI: 10.30437/report2021

[5] Rosi, A., Tofani, V., Tanteri, L., Tacconi Stefanelli, C., Agostini, A., Catani, F., & Casagli, N. (2018). The new landslide inventory of Tuscany (Italy) updated with PS-InSAR: geomorphological features and landslide distribution. Landslides, 15(1), 5-19.

[6] Martelloni, G., Segoni, S., Fanti, R., Catani, F. Rainfall thresholds for the forecasting of landslide occurrence at regional scale (2012) Landslides, 9 (4), pp.485-495. DOI: 10.1007/s10346-011-0308-2

[7] Lagomarsino, D., Segoni, S., Rosi, A., Rossi, G., Battistini, A., Catani, F., and Casagli, N.: Quantitative comparison between two different methodologies to define rainfall thresholds for landslide forecasting, Nat. Hazards Earth Syst. Sci., 15, pp.2413–2423, https://doi.org/10.5194/nhess-15-2413-2015 , 2015.

[8] Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorology and atmospheric physics, 98(3), 239-267.

[9] Abraham, M. T., Satyam, N., Rosi, A., Pradhan, B., & Segoni, S. (2020). The selection of rain gauges and rainfall parameters in estimating intensity-duration thresholds for landslide occurrence: case study from Wayanad (India). Water, 12(4), 1000.

[10] Goetz, J. N., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Computers & geosciences, 81, 1-11.

[11] Distefano, P., Peres, D. J., Scandura, P., & Cancelliere, A. (2021). Brief communication: Rainfall thresholds based on Artificial neural networks can improve landslide early warning. Natural Hazards and Earth System Sciences Discussions, 1-9.

[12] Chae B.G., Park H. J, Catani F., Simoni A., Berti M. Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. GeoScience Journal, Vol. 21, No. 6, pp.1033-1070, December 2017. http://dx.doi.org/10.1007/s12303-017-0034-4

[13] Kavzoglu, T., Sahin, E. K., & Colkesen, I. (2015). Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm. Engineering Geology, 192, 101-112.

[14] Battistini, A., Segoni, S., Manzo, G., Catani, F., Casagli, N. Web data mining for automatic inventory of geohazards at national scale (2013) Applied Geography, 43, pp.147-158. DOI: 10.1016/j.apgeog.2013.06.012

[15] Battistini, A., Rosi, A., Segoni, S., Lagomarsino, D., Catani, F., & Casagli, N. (2017). Validation of landslide hazard models using a semantic engine on online news. *Applied geography*, *82*, 59-65.

[16] Calvello, M., & Pecoraro, G. (2018). FraneItalia: a catalog of recent Italian landslides. Geoenvironmental Disasters, 5(1), 1-16.

[17] Wang, H., Zhang, L., Yin, K., Luo, H., Li, J. Landslide identification using machine learning. Geoscience Frontiers 12 (2021) pp.351–364. https://doi.org/10.1016/j.gsf.2020.02.012

[18] A. S. Santos et al., "Brazilian natural disasters integrated into cyber-physical systems: computational challenges for landslides and floods in urban ecosystems," 2020 IEEE International Smart Cities Conference (ISC2), Piscataway, NJ, USA, 2020, pp. 1-8, doi: 10.1109/ISC251055.2020.9239011.

[19] C. Lee, J. Huang, W. Ma, Y. Weng, Y. Lee and Z. Lee, "Analyze the rainfall of landslide on Apache Spark," 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI), Xiamen, China, 2018, pp.348-351, doi: 10.1109/ICACI.2018.8377482.

[20] S. T. Palliyaguru, L. C. Liyanage, O. S. Weerakoon and G. D. S. P. Wimalaratne, "Random Forest as a Novel Machine Learning Approach to Predict Landslide Susceptibility in Kalutara District, Sri Lanka," 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, 2020, pp. 262-267, doi: 10.1109/ICTer51097.2020.9325460.

[21] Korup, Oliver, and Amelie Stolle. "Landslide prediction from machine learning." Geology today 30.1 (2014): 26-33.

[22] Abraham, Minu Treesa, et al. "Forecasting landslides using SIGMA model: a case study from Idukki,

India." Geomatics, Natural Hazards and Risk 12.1 (2021): pp.540-559.

[23] Nam, K. and Wang, F. (2020). An extreme rainfall-induced landslide susceptibility assessment using autoencoder combined with random forest in Shimane Prefecture, Japan. *Geoenvironmental Disasters, 7(1), 6.*

[24] Huang, F., Zhang, J., Zhou, C. *et al.* A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction. *Landslides* **17,** pp.217–229 (2020). https://doi.org/10.1007/s10346-019-01274-9

[25] V. D. Pham, Q. Nguyen, H. Nguyen, V. Pham, V. M. Vu and Q. Bui, "Convolutional Neural Network—Optimized Moth Flame Algorithm for Shallow Landslide Susceptible Analysis," in *IEEE Access*, vol. 8, pp. 32727-32736, 2020, doi: 10.1109/ACCESS.2020.2973415.

[26] Pei, H., Meng, F. & Zhu, H. Landslide displacement prediction based on a novel hybrid model and convolutional neural network considering time-varying factors. *Bull Eng Geol Environ* **80,** pp.7403–7422 (2021). https://doi.org/10.1007/s10064-021-02424-x

[27] Karunanayake, K.B.A.A.M & Wijayanayake, Janaka. (2019). Application of Data Mining Technique to Predict Landslides in Sri Lanka. International Journal of Data Mining & Knowledge Management Process. 9. pp.53-67. 10.5121/ijdkp.2019.9404.

[28] Cheng, Y.S., Yu, T.T., & Son, N.T. (2021). Random Forests for Landslide Prediction in Tsengwen River Watershed, Central Taiwan. *Remote Sensing, 13(2).*

[29] C. Wang et al., "Classification of Landslide Stability Based on Fine Topographic Features," 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2020, pp.54-57, doi: 10.1109/ICCWAMTIP51612.2020.9317358.

[30] Ngo, Phuong Thao Thi, et al. "Evaluation of deep learning algorithms for national scale landslide susceptibility mapping of Iran." Geoscience Frontiers 12.2 (2021): pp.505-519.

[31] Thai Pham, B., Shirzadi, A., Shahabi, H., Omidvar, E., Singh, S., Sahana, M., Talebpour Asl, D., Bin Ahmad, B., Kim Quoc, N., & Lee, S. (2019). Landslide Susceptibility Assessment by Novel Hybrid Machine Learning Algorithms. Sustainability, 11(16).

[32] Tien Bui, D., Shahabi, H., Omidvar, E., Shirzadi, A., Geertsema, M., Clague, J., Khosravi, K., Pradhan, B., Pham, B., Chapi, K., Barati, Z., Bin Ahmad, B., Rahmani, H., Gróf, G., & Lee, S. (2019). Shallow Landslide Prediction Using a Novel Hybrid Functional Machine Learning Algorithm. Remote Sensing, 11(8).

[33] Zhang, T., Han, L., Chen, W., & Shahabi, H. (2018). Hybrid Integration Approach of Entropy with Logistic Regression and Support Vector Machine for Landslide Susceptibility Modeling. Entropy, 20(11).

[34] Badii, Claudio, et al. "Snap4city: A scalable iot/ioe platform for developing smart city applications." 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/ SCALCOM/UIC/ATC/ CBDCom/IOP/SCI). IEEE, 2018.

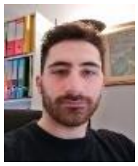
[35] Badii, C., Bellini, P., Difino, A., Nesi, P., Pantaleo, G., Paolucci, M. Microservices suite for smart city applications. Sensors (Switzerland). DOI: 10.3390/s19214798

[36] Hwang, Y.; Clark, M.; Rajagopalan, B.; Leavesley, G. Spatial interpolation schemes of daily precipitation for hydrologic modelling. Stoch. Environ. Res. Risk Assess. 2012, 26, 295–320.

[37] CLC2006 technical guidelines. European Environment Agency. EEA technical report, 2017. https://www.eea.europa.eu/publications/technical_report_2007_17

[38] P. Belllini, D. Cenni, N. Mitolo, P. Nesi, G. Pantaleo, "Exploiting Satellite Data in the Context of Smart City Applications", The 5th IEEE International Conference on Smart City Innovations - Track 1: Theory, Modeling and Methodologies, 2021.

[39] Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, pp.4768–4777, Long Beach, California, USA, 2017.

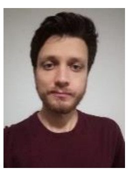


ENRICO COLLINI is an Eng. And candidate for Ph.D. of the University of Florence, DINFO Dept. His research interests include deep learning, mobility, data models.

L. A. Ipsaro Palesi is a Ph.D. student of the University of Florence, DINFO Dept. His research interests include deep learning, XAI, recommender system and predictive maintenance.

PAOLO NESI is a full professor at the University of Florence, DINFO Dept., chief of DISIT lab. His research interests include machine learning, massive parallel and distributed systems, physical models, IoT, mobility, big data analytic, semantic computing, formal model, machine learning. He has been chair of several international conferences. He is and has been the coordinator of several R&D multipartner international projects.

GIANNI PANTALEO is an aggregated professor at the University of Florence, DINFO Dept. His research interests include knowledge engineering, IoT, visual analytics, mobility, NLP, semantic computing. He has been coordinator of a number of WP in international R&D projects.



NICOLA NOCENTINI is a Ph.D Candidate in Geology at the University of Pisa and fellow at the Dept. of Earth Science of the University of Florence . His research activity is focused on the use of statistical models for landslide forecasting

ASCANIO ROSI is a fixed-term researcher at the Dept. of Earth Science of the University of Florence. His research interests include spatial and temporal landslide forecasting by means of statistical approaches or machine learning algorithms, as well as the use of remote sensing data for landslide mapping and monitoring.