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/

https://www.Snap4City.org Powered by

Paolo Nesi, paolo.nesi@unifi.it
https://www.Km4City.org
https://www.disit.org

Landslide Prediction

Rainfall induced landslide is one of the main geological hazard in Italy and in the world.

- Worldwide based on the study [1] of Natural Hazards and Earth System Sciences
  - from 2004 to 2016, 55997 people were killed in 4862 non seismic landslide events worldwide
  - The same authors identified rainfall as the main the triggering factor of 79% of non-seismic landslides.
- In Italy based on the ISPRA report:
  - 19.9% of the Italian territory is at risk of landslides (59981km^2)
  - Tuscany is among the regions with the largest areas at risk (26%)

Accurate short-term **PREDICTIONS** (1 day in advance) 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.

[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

Scenario

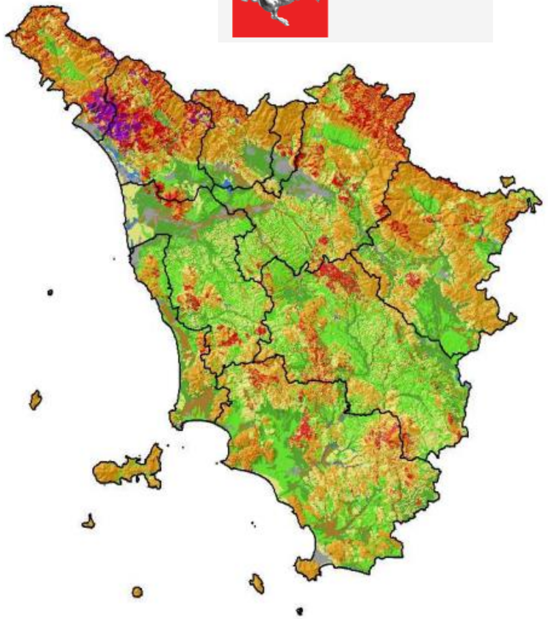
- The solution and its validation have been performed by using data collected in in the area of the **Metropolitan City of Florence** with
  - 41 Municipalities
  - 3514 Km^2 of Surface Area
  - altitude between 100-1000 above the sea level
  - land predominantly of deciduous forests and cultivated areas
  - 1.5 M inhabitants
- The data history covers the years 2013-2019 with a total of **341** landslide events

Prediction | Susceptibility

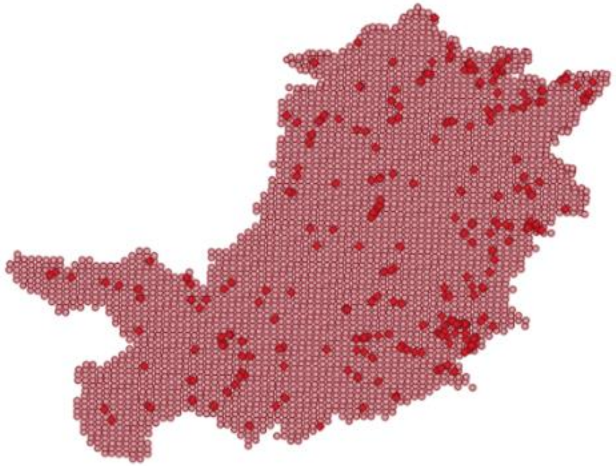
per municipality

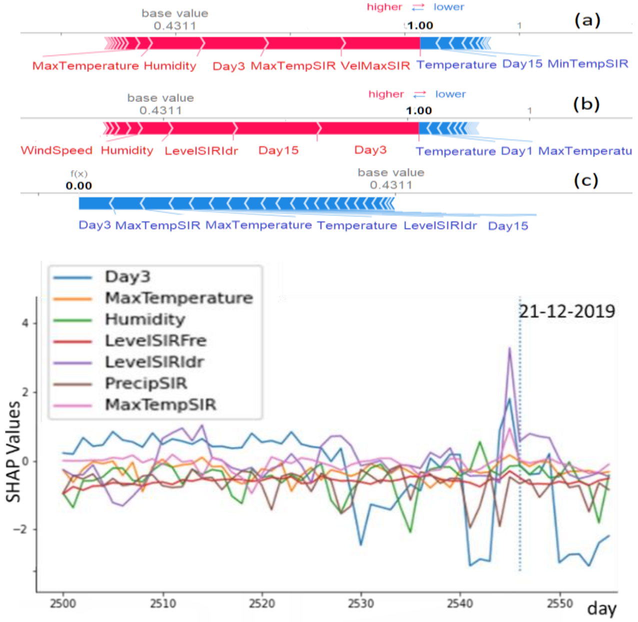
dynamic hazard heatmaps

Prediction	Susceptibility
Useful for early warning systems	Useful for long term land usage planning
static + dynamic features	static features based
Can be computed daily	1 or 2 times per year



Predicting Land slides





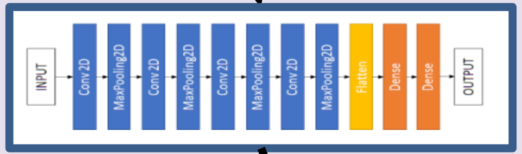
Dataset Construction

Dataset Construction

Previsional Model

Model training And validation

Model



Data

SNAP4City Advanced APIs

landslide DB

Real Time data from Field: rain, weather, etc.

Ingestion Processes

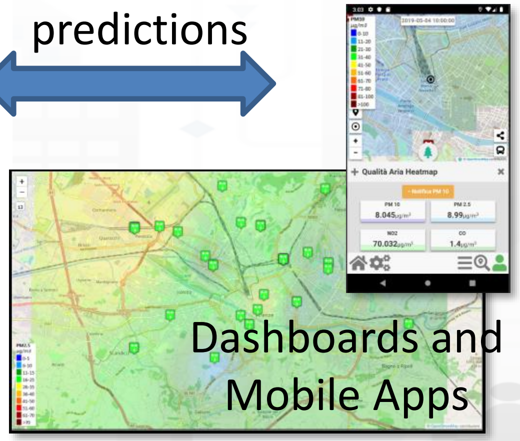
Big Data Storage and KB

Predictions

Model execution Shap Assessment

Data Analytics IOT App Management

predictions



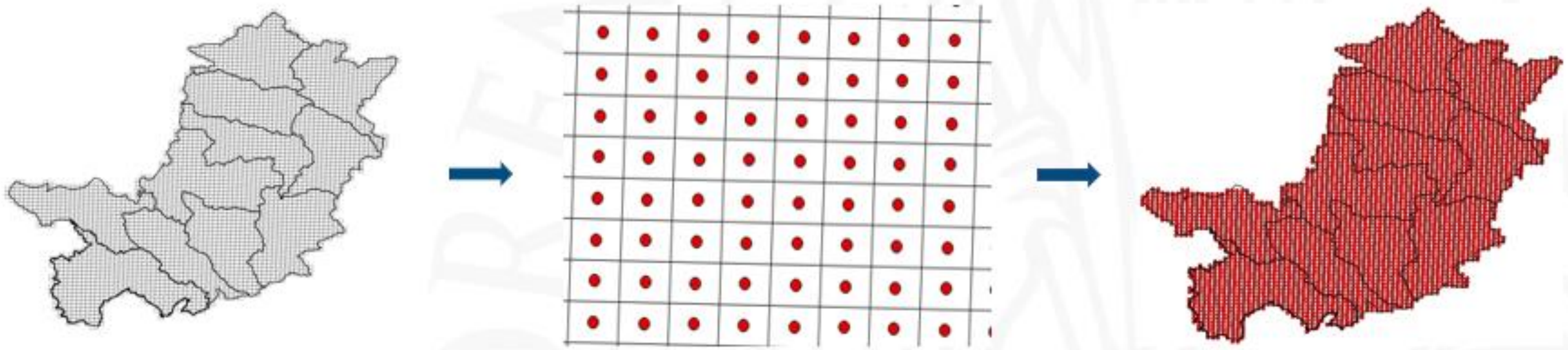
Dashboards and Mobile Apps

Snap4City Servers and Tools: Dashboard manager, Heatmap manager, GeoServer, Smart City API.

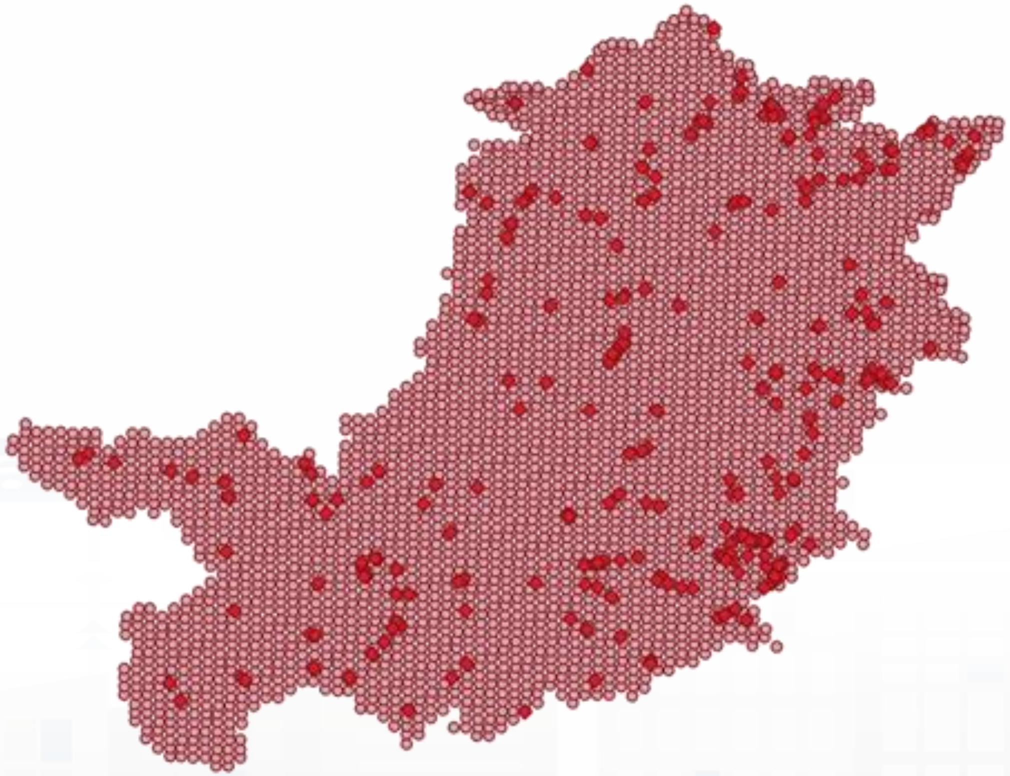
E. Collini, L. A. I. Palesi, P. Nesi, G. Pantaleo, N. Nocentini and A. Rosi, "Predicting and Understanding Landslide Events with Explainable AI," in *IEEE Access*, doi: 10.1109/ACCESS.2022.3158328.
https://ieeexplore.ieee.org/abstract/document/9732490

Dataset Details

- 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 grid was composed of 1000x1000 mt squares obtaining 3582 areas covering the whole Florence Metro area



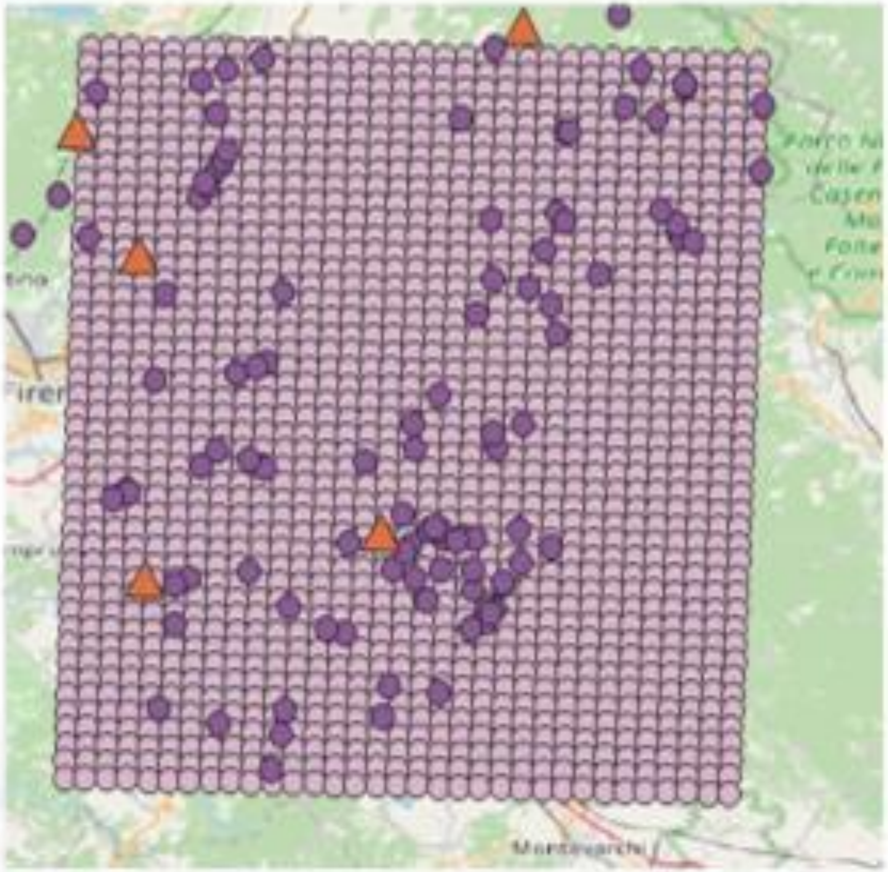
Dataset Details



- Grid and landslide events in the Florence Metro Area (Tuscany, Italy) from 2013 to 2019.
- the RED dots are the events of landslide registered

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

Static and Dynamic Features



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 area 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 on the day before the observation	mm	12.453
Day3	Rainfall on the 3 days preceding the observation	mm	15.072
Day15	Rainfall on the 15 days preceding the observation	mm	16.160
Day30	Rainfall on 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

The values of sensors have been interpolated by using **IDW** (Inverse Distance Weighting) algorithm

Data Analytic Solutions

- 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.

- On the basis of the above-described dataset, a number of techniques to predict landslide events has been tested:
  - Random Forest, **RF**
  - eXtreme Gradient Boosting, **XGBoost**
  - Convolutional Neural Network, **CNN**
  - Autoencoders, **AE**
  - decisional algorithm **SIGMA**

Comparing Predictive Model Architectures

- The considered dataset is composed of about 9 million estimations, among which 2342 positive events (labeled with Value = 1)
- The dataset was divided into two groups: training set (80%) and test set (20%)

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

Global Explainable AI Feature Relevance

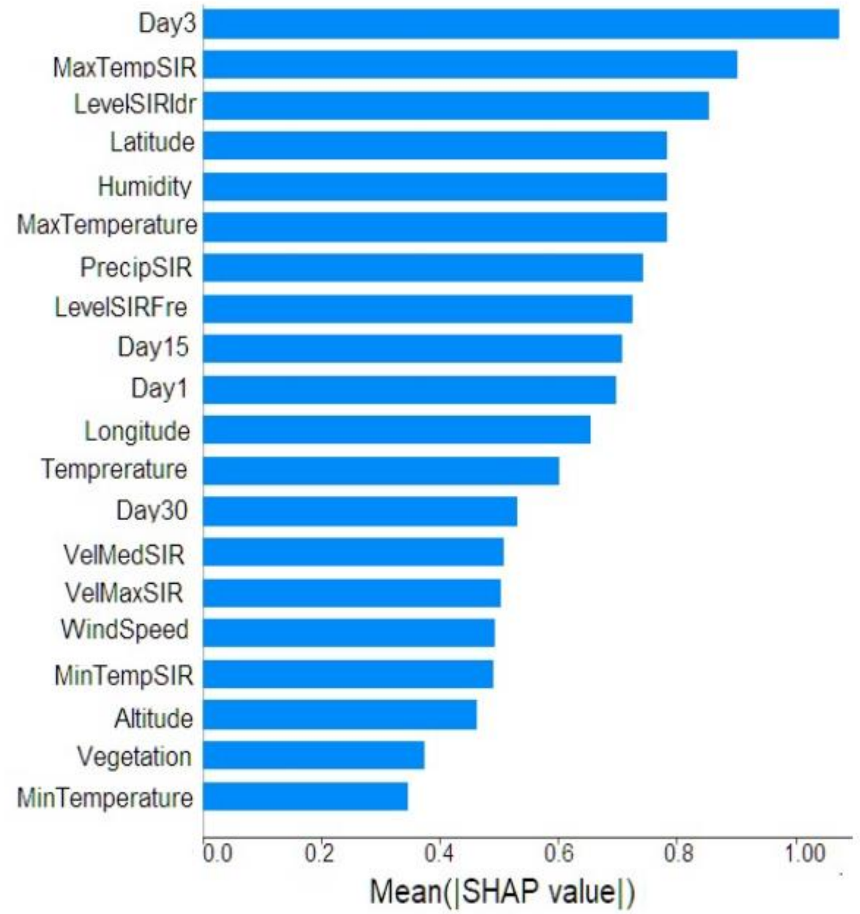


FIGURE 8. Global feature relevance as mean of the absolute SHAP global features importance for XGBoost (only the first 20).

- 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.
- 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

Global Explainable AI Feature Relevance

- The graph describes the distribution of SHAP values for each feature, sorted by relevance.
- 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 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.

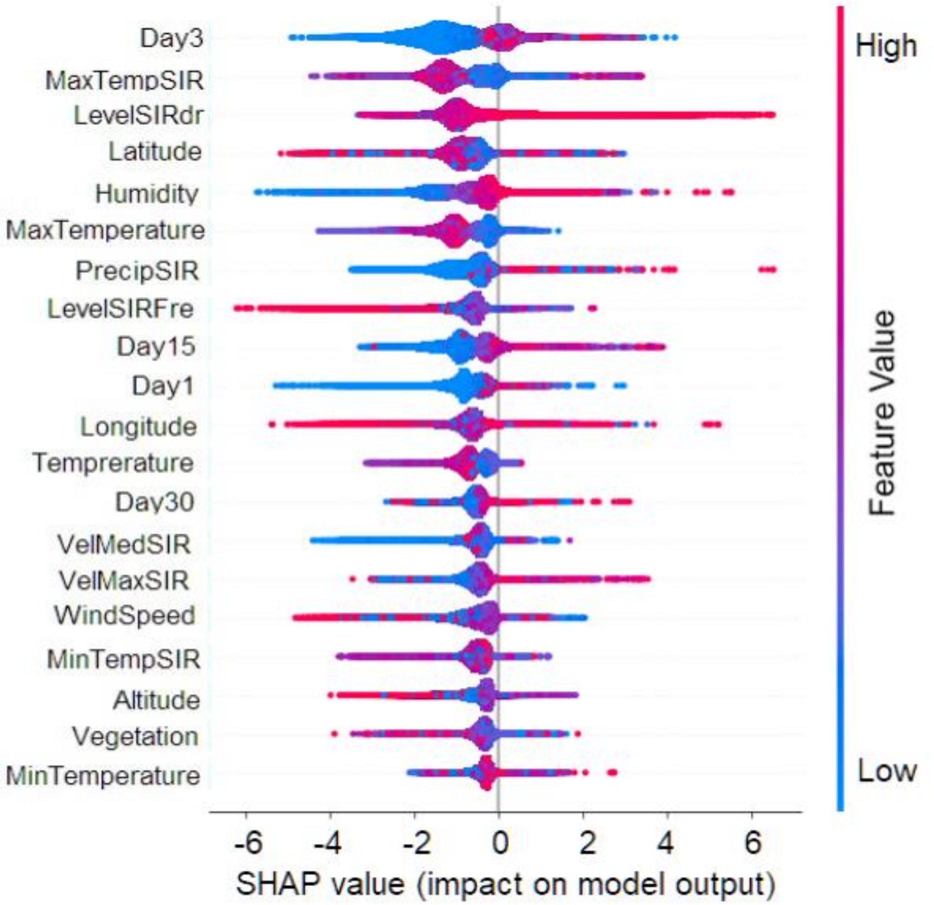


FIGURE 9. SHAP summary plot for XGBoost. x-axis reports the SHAP value of the feature, while on y-axis the features. The color codes the magnitude of the value, and the size the density of values.

Local Explainable AI - understanding the single event

- The local explanation puts in evidence the features which provided major contribution to the prediction
- For example considering Figure10a, the value of VelMaxSIR, MaxTempSIR, Day3 and Humidity contributed significantly to the classification of the observation as a l**andslide 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.

Local Explainable AI - understanding the single event

The trends of the SHAP values of the most relevant features 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**.

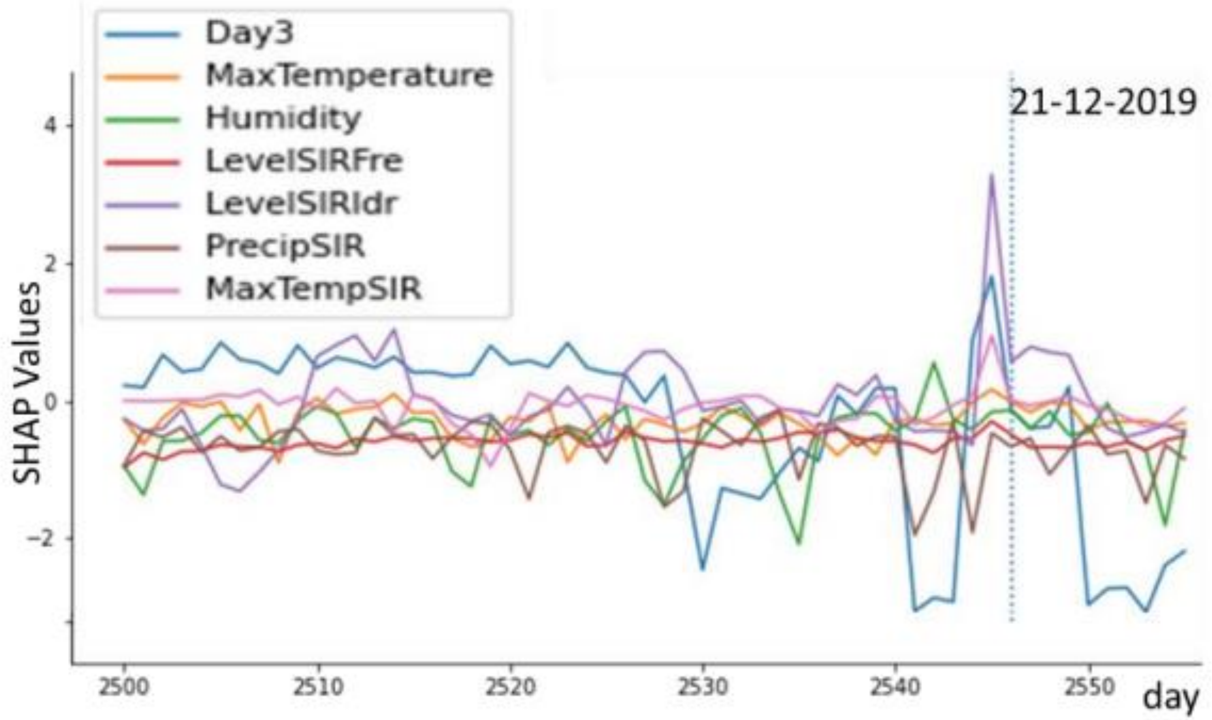
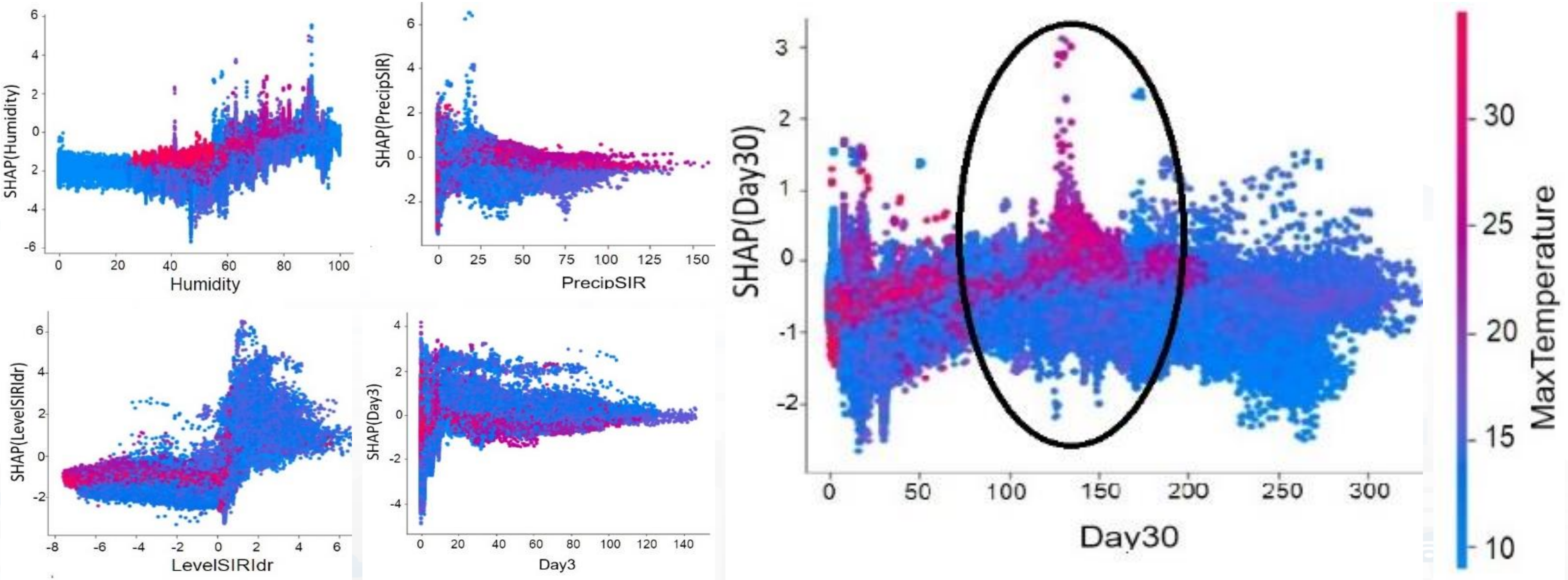


FIGURE 11. Time trend of SHAP values of most relevant features around the landslide event of 21-12-2019: values estimated by using data collected in the neighboring area of the event.

Impact of Features on corresp. SHAP Values vs MaxTemp

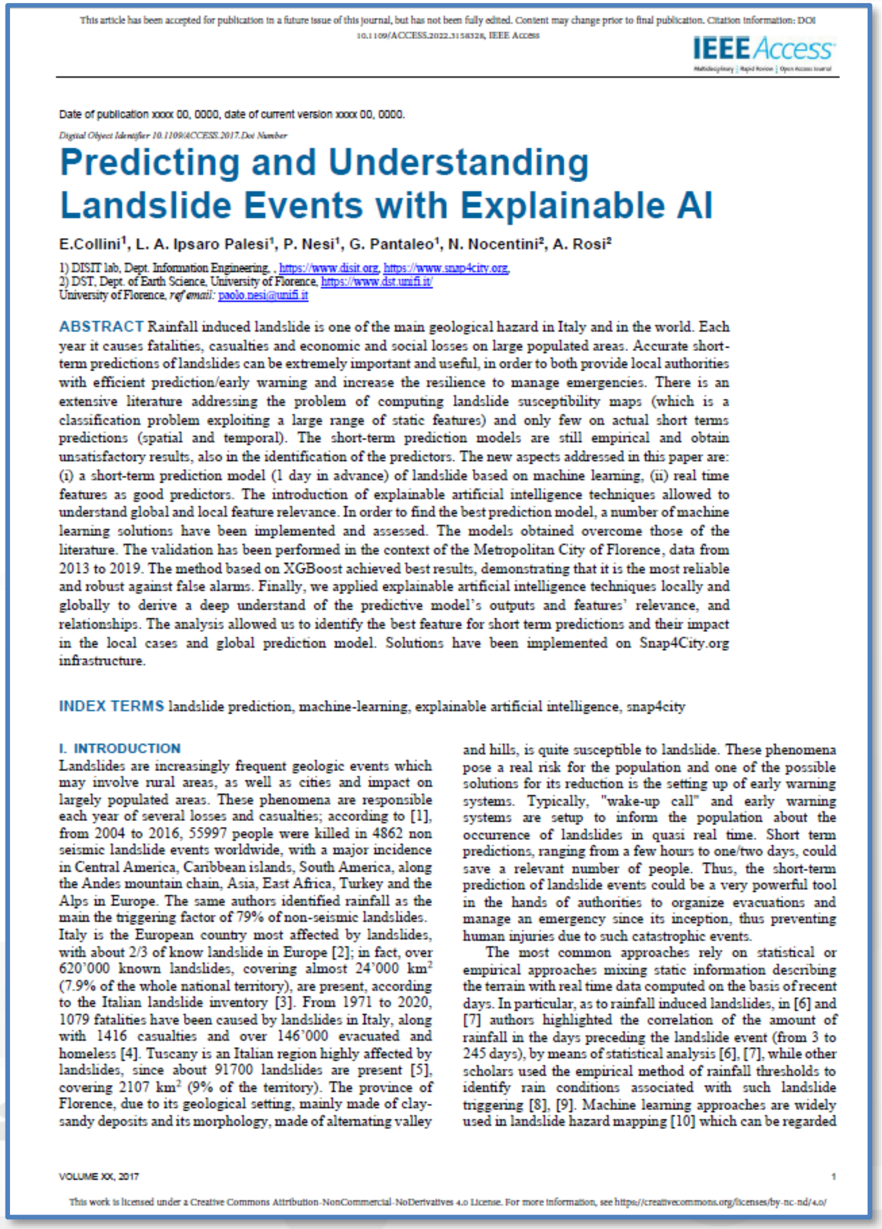


Conclusions

- **The problem of landslide event prediction** has been addressed, for early warning specific to the case study in the Metropolitan City of Florence, using
  - static land description,
  - dynamic features as rain fall, temperature, wind, etc.
- **Numerous AI solutions has been compared**
  - **t**he best performing architecture has been XGBOOST
- **XAI: based on Shapley** additive explanation (SHAP), global and local, derived relevance:
  - rain the last 3 days, max temperature in the previous day, lever of water in the river
  - land static features are preconditions for landslide, while they are not efficient in creating an early warning system.
- **Computationally**: predictions can be assess every day,
  - susceptibility map usually are computed 1 or two times per year.
- **Prediction** models can prevent disaster
  - susceptibility map are mainly used for taking decision on planning.

IEEE Access

- E. Collini, L. A. I. Palesi, P. Nesi, G. Pantaleo, N. Nocentini and A. Rosi, "Predicting and Understanding Landslide Events with Explainable AI," in *IEEE Access*, doi: 10.1109/ACCESS.2022.3158328.
- https://ieeexplore.ieee.org/abstract/document/9732490



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

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

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

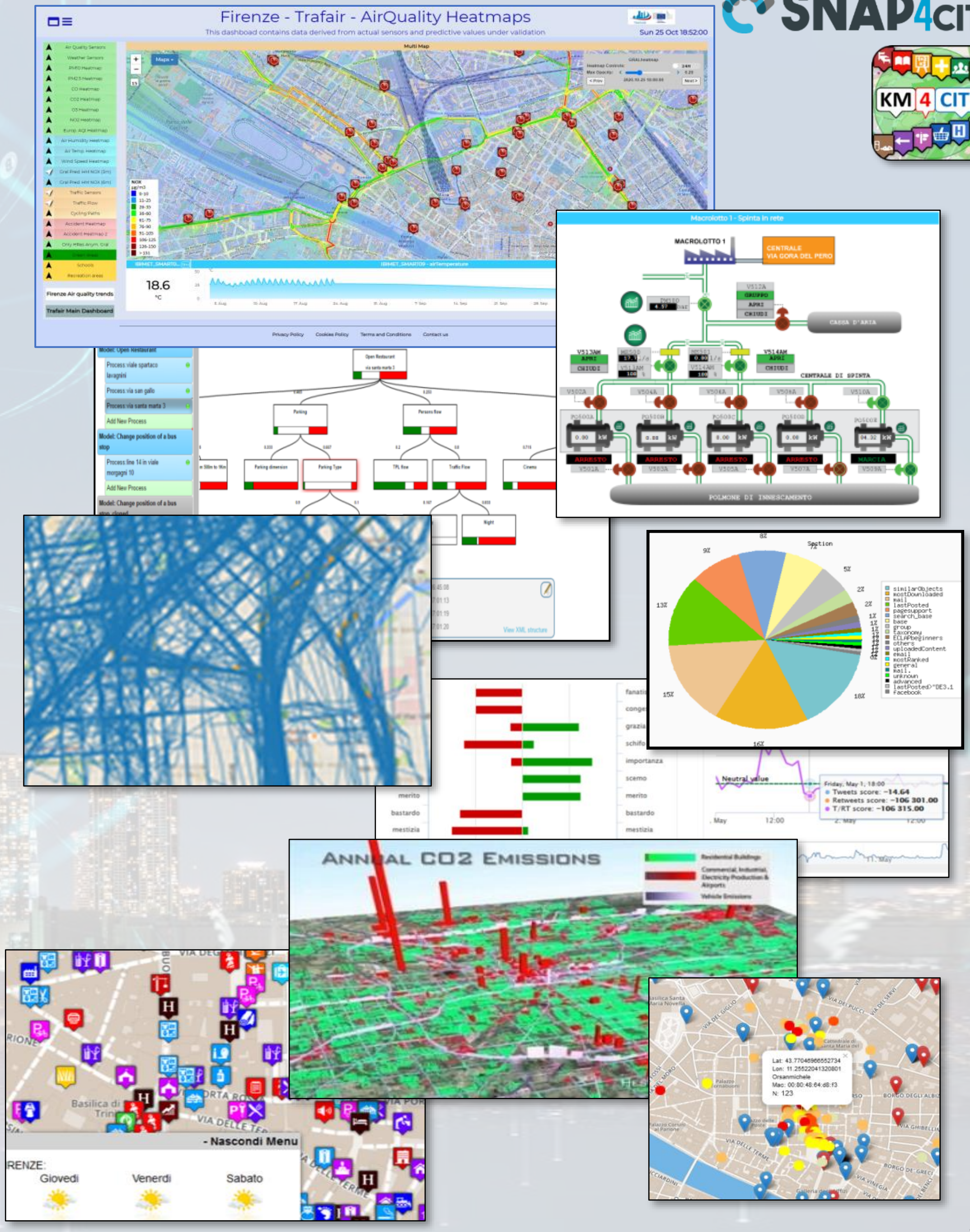
I. INTRODUCTION

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

Data Driven Decision Support

- Decision Support system
- Assessment / Strategies
- Data Rendering, visual analytics
- Data Analytics
- Data aggregation, Storage, indexing
- Data Ingestion



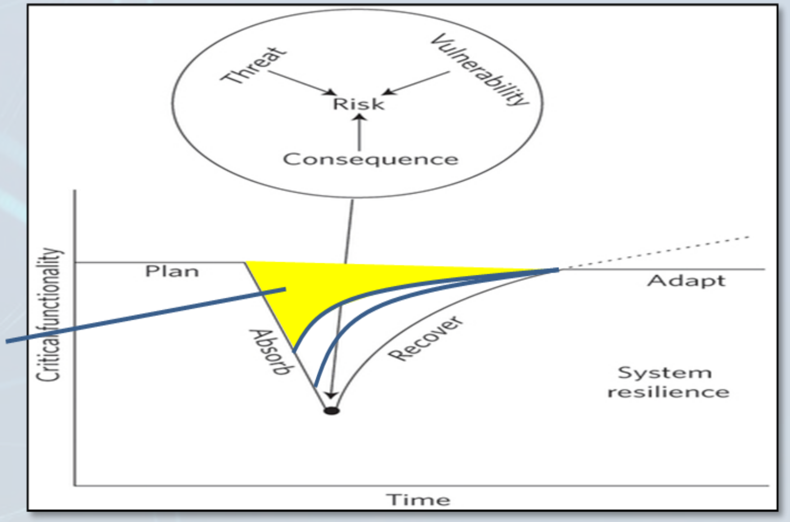
UNIVERSITÀ DEGLI STUDI FIRENZE | DINFO DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE | DISIT DISTRIBUTED SYSTEMS AND INTERNET TECHNOLOGIES LAB

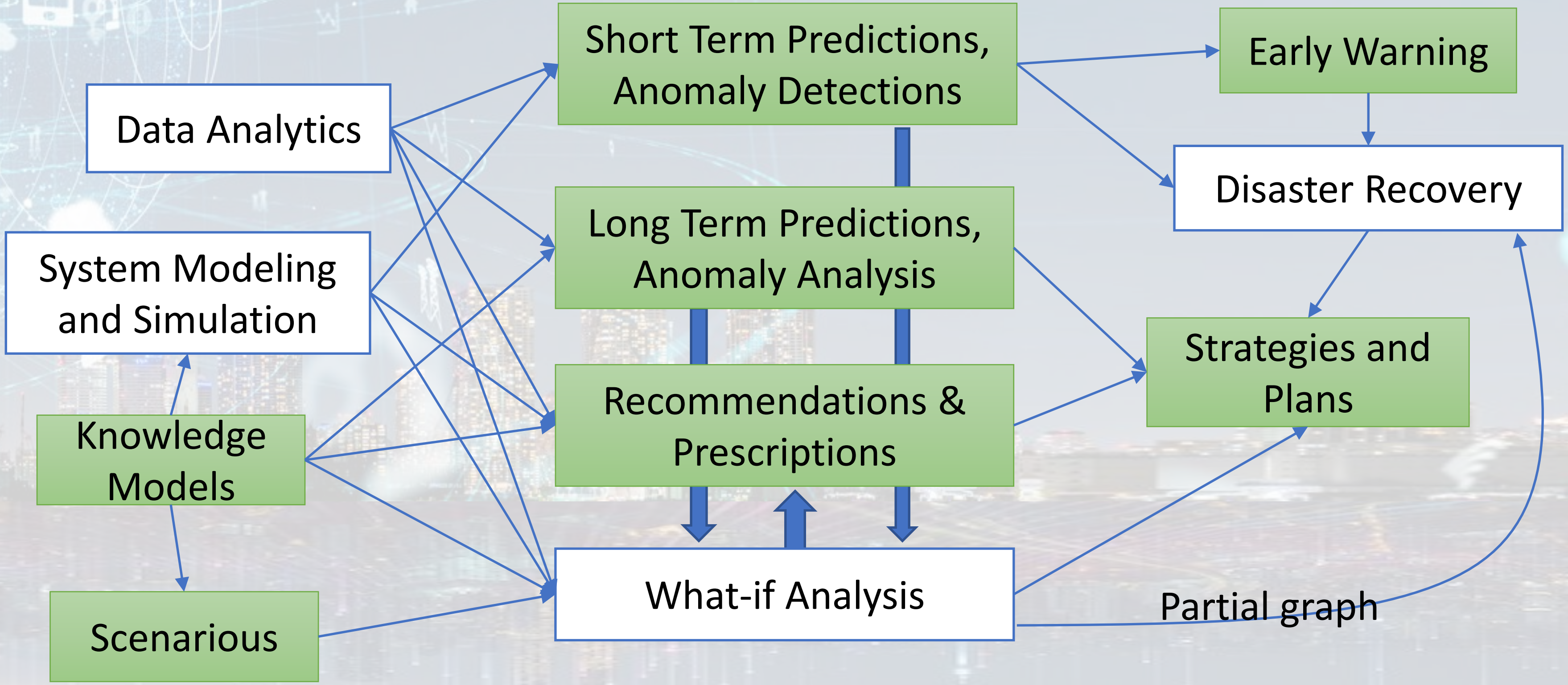
Snap4City Analytics

Prepare
Absorb
Recover
Adapt

- Decision support systems
- Improvement of life quality
- Sustainable Solutions
- Reduction of costs
- Risk Assessment
- Resilience

damage





Decision Support System, targeting: Quality of Life, KPI, SDG, 15MinIndex,…

100% OPEN SOURCE

Powered by FIWARE

FREE TRIAL

PEN Test Passed

EU GDPR COMPLIANT

SNAP4 Appliances and Dockers Installations

EUROPEAN OPEN SCIENCE CLOUD

Node-RED

JS Foundation

E015 digital ecosystem

NVIDIA

SNAP4CITY KM4CITY

SMART SOLUTIONS AND DECISION SUPPORT SYSTEMS

CONTROL ROOMS - DECISION SUPPORT SYSTEMS - WHAT-IF ANALYSIS - BUSINESS INTELLIGENCE - SIMULATIONS - SMART APPLICATIONS

DASHBOARDS - VISUAL ANALYTICS - SYNOPTICS - DIGITAL TWIN - GRAPHICAL WIDGETS - ANALYTICS - GUI CUSTOM STYLES - VISUAL PROGRAMMING

DASHBOARDS, WIDGETS TEMPLATES

PREDICTION - ANOMALY DETECTION - CLUSTERING - ROUTING - SENTIMENT NLP - TRAFFIC FLOW
PEOPLE FLOWS - SDG - 15 MIN CITY INDEX - KPI - HEATMAPS - ORIGIN DESTINATION - ETC...

API - MICROSERVICES - GIS - BPM
VIDEO - REPORTS - MAPS - 3D ...

ANY: DATA, BROKER, NETWORK AND VERTICAL

EXPERT SYSTEM, KNOWLEDGE BASE
SEMANTIC REASONING
SMART DATA MODEL
IOT DEVICE MODELS, STORAGE

BIG DATA ANALYTICS, ARTIFICIAL INTELLIGENCE
EXPLAINABLE AI, MACHINE LEARNING
OPERATIVE RESEARCH, STATISTICS

VISUAL PROGRAMMING, ADAPTERS
DATA FLOWS, WORKFLOWS
PARALLEL DISTRIBUTED PROCESSING
DATA DRIVEN

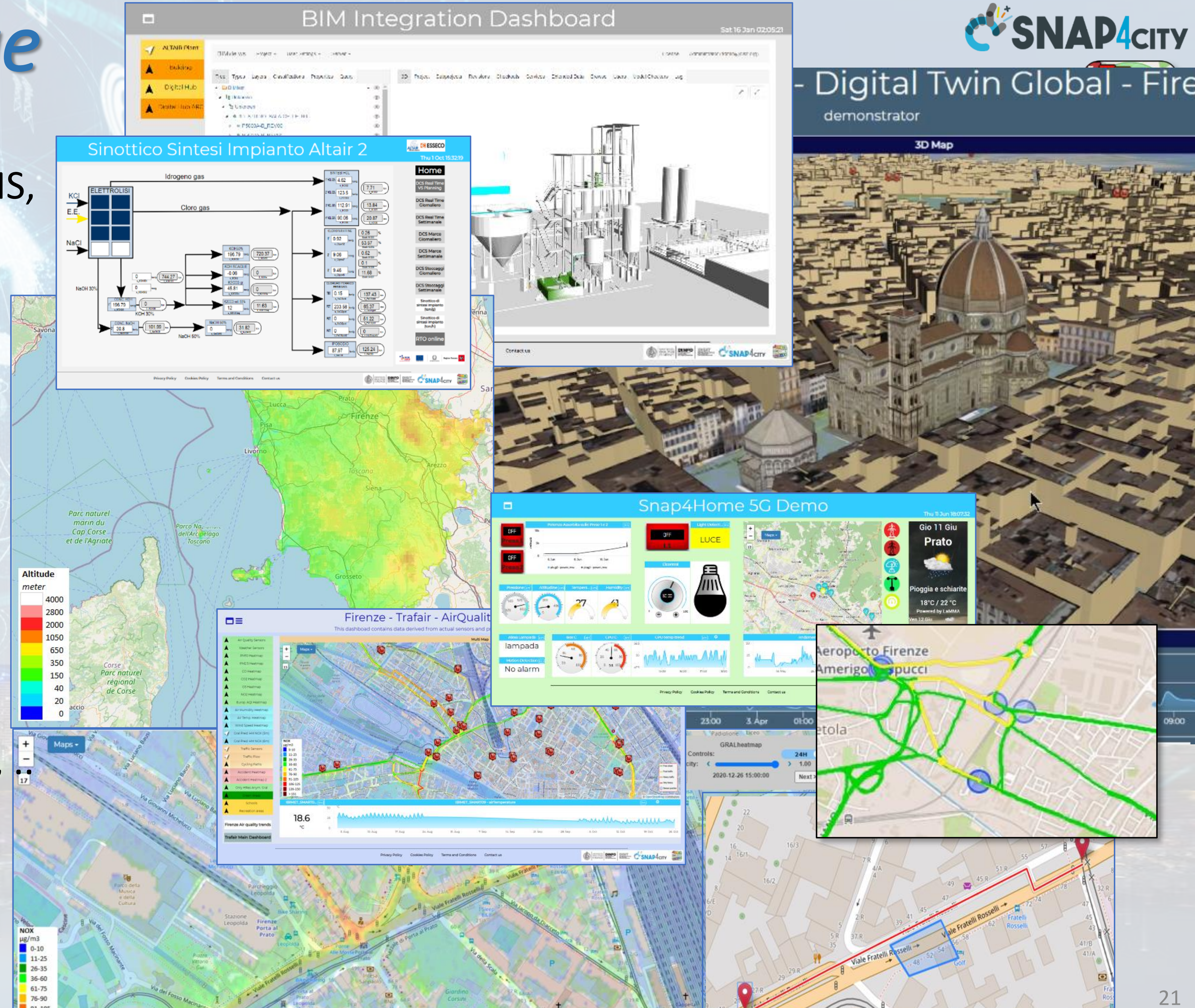
Native and External Applications

- Smart Parking
- Smart Light
- Smart Waste
- Smart Energy
- Social Media Analysis
- ...

METHODOLOGIES
LIVING LABS
COURSES AND COMMUNITY
DEVELOPMENT TOOLS

Data Type Coverage

- POI, IOT Devices, shapes,..
- GIS, maps, orthomaps, WFS/WMS, GeoTiff, calibrated heatmaps, ..
- Satellite data, ..
- traffic flow, typical trends, ..
- trajectories, events, Workflow, ..
- 3D, BIM, ..
- Dynamic icons/pins, ..
- OD Matrices of several kinds, ..
- Synoptics, animations, ..
- KPI, personal KPI,..
- social media data, TV Stream,
- routing, multimodal, constraints, ..
- decision scenarios, ….
- prediction models, ….
- etc.



UNIVERSITÀ DEGLI STUDI FIRENZE
DINFO DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
DISIT DISTRIBUTED SYSTEMS AND INTERNET TECHNOLOGIES LAB

https://www.snap4city.org/577

On Line Training Material (free of charge)

	1st part (*)	2nd part (*)	3rd part (*)	4th part (*)	5th part (*)	6th part (*)	7th part (*)
what	General	Dashboards	IOT App, IOT Network	Data Analytics	Data Ingestion processes	System and Deploy Install	Smart City API: Web & Mob. App
PDF							
Inter active							
Video1							
Video2							
Video3							
Video4				none		none	none
duration	2:55	3:16	3:41	2:00	2:48	2:35	1:47

2022 booklets

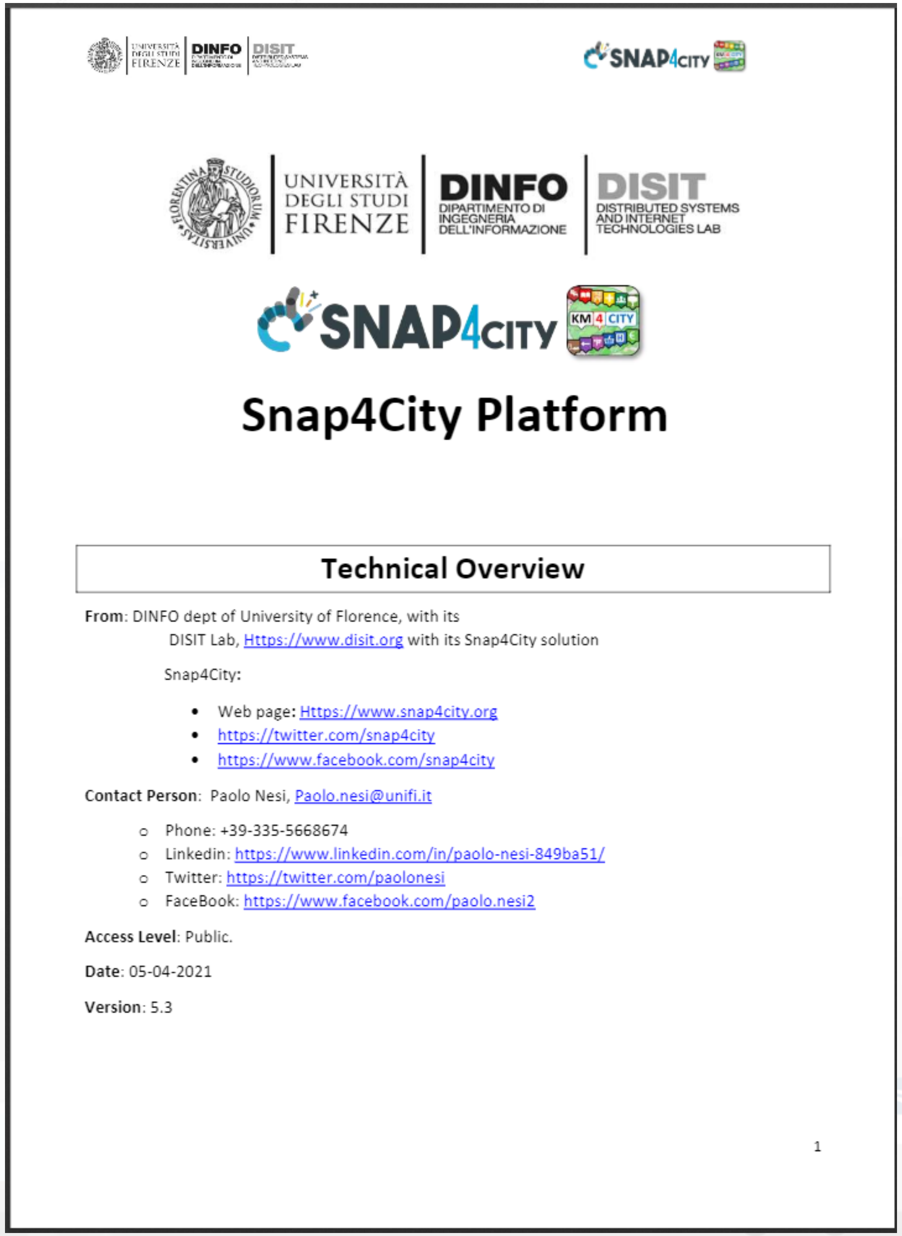
- Snap4City

https://www.snap4city.org/download/video/DPL_SNAP4CITY_2022-v02.pdf

- Snap4Industry

https://www.snap4city.org/download/video/DPL_SNAP4INDUSTRY_2022-v03.pdf

Overview



UNIVERSITÀ DEGLI STUDI FIRENZE | DINFO DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE | DISIT DISTRIBUTED SYSTEMS AND INTERNET TECHNOLOGIES LAB

SNAP4CITY

Snap4City Platform

Technical Overview

From: DINFO dept of University of Florence, with its DISIT Lab, Https://www.disit.org with its Snap4City solution

Snap4City:

- Web page: Https://www.snap4city.org
- https://twitter.com/snap4city
- https://www.facebook.com/snap4city

Contact Person: Paolo Nesi, Paolo.nesi@unifi.it

- Phone: +39-335-5668674
- Linkedin: https://www.linkedin.com/in/paolo-nesi-849ba51/
- Twitter: https://twitter.com/paolonesi
- FaceBook: https://www.facebook.com/paolo.nesi2

Access Level: Public.

Date: 05-04-2021

Version: 5.3

1

- https://www.snap4city.org/drupal/sites/default/files/files/Snap4City-PlatformOverview.pdf

TOP

Be smart in a SNAP!

SNAP4 Appliances and Dockers **Installations**

CONTACT

www.snap4city.org

DISIT Lab, DINFO: Department of Information Engineering
Università degli Studi di Firenze - School of Engineering

Via S. Marta, 3 - 50139 Firenze, ITALY
https://www.disit.org

Email: snap4city@disit.org

Office: +39-055-2758-515 / 517
Cell: +39-335-566-86-74
Fax.: +39-055-2758570