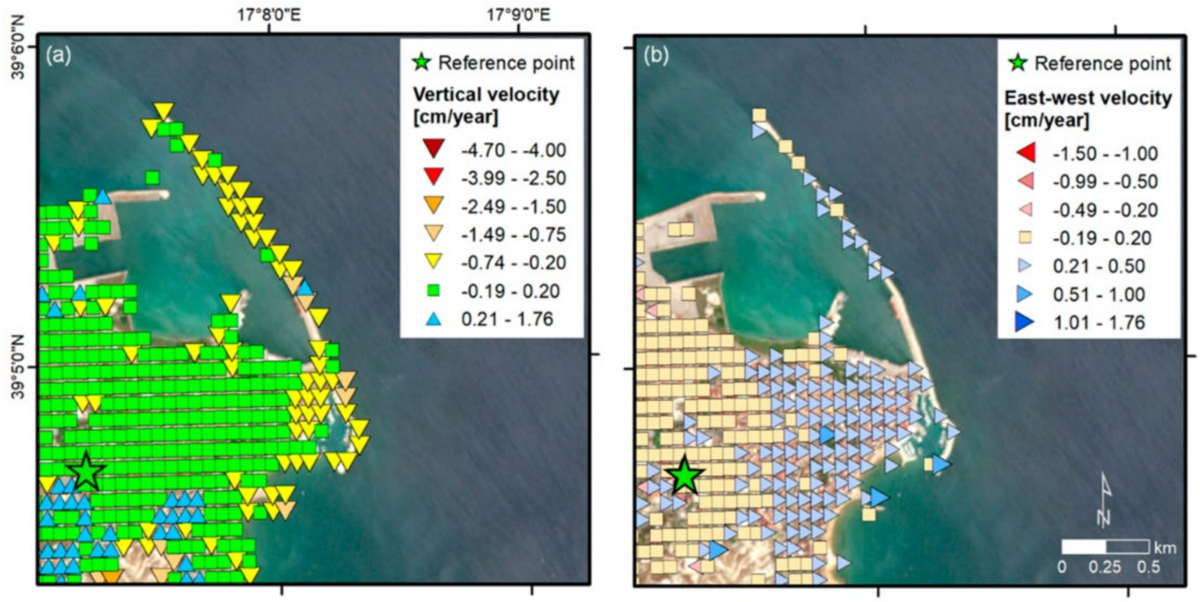
Opening the Black Box: Explainable AI for Slope Hazards Forecasting

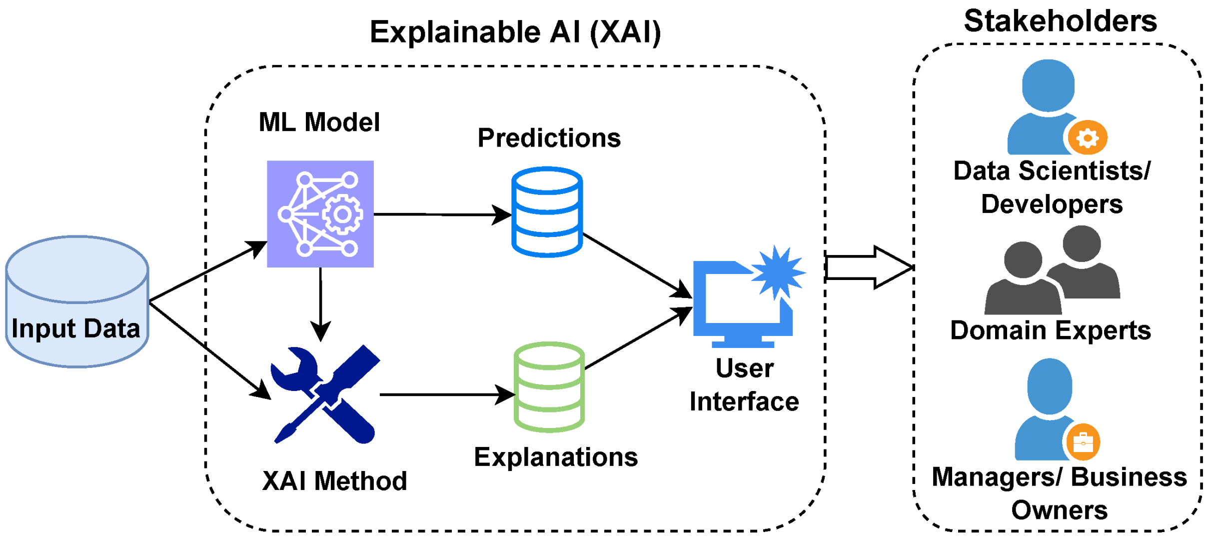
(Proposer: Prof. Filippo Catani)

The forecasting of slope and ground hazards such as mass movements, landslides, soil subsidence, and debris flows, is a challenging task that requires multi-disciplinary knowledge, advanced methodologies, and multi-scale information.
In the last decades, Artificial Intelligence, in its multiple forms, has become one of the most powerful tools available to scientists working on the topic to help them overcome limitations of traditional methods. Today, different typologies of AI contribute to almost all the phases of slope hazard forecasting, from data acquisition, to modelling and prediction, to risk assessment and early warning during and after emergencies.



However, much of the work performed by AI systems is still considered a black box, and researchers struggle to extract interpretable information from machine and deep learning algorithms—even though the predictive results and modeling scenarios are accurate. As a consequence, several research groups are pursuing new lines of investigation aimed at unveiling AI reasoning and outcomes, thereby enabling scientists to extract knowledge from models and understand the underlying causes of machine learning predictions. This would pave the way for an increased system understanding covering also complex systems such as those that characterize slope instability and mass wasting disasters, and to the possibility of improving predictive methods by eliminating errors and misinterpretations of data.

This doctoral project aims to contribute to the task of making AI more human understandable and usable and to the application of methodologies that exploit machine intelligence to improve rather than substitute human expertise in the realm of natural hazards prevention and mitigation.
In the specific context of landslide and mass wasting monitoring, mapping, and prediction, this project will seek to develop new XAI (eXplainable AI) tools that would improve interpretability of results and help unpack the "black box" nature of machine and deep learning in some of the most challenging tasks of hazard analysis, such as the rapid mapping and detection of disaster affected areas, the estimation of the type of geomorphic process acting on a given region, the measurement of the rates at which the slope dynamics acts, and the forecasting and early warning of hazards.



Methods of XAI that are foreseen for the PhD project include:

- **Feature Importance Metrics**, that help us understand which environmental variables most strongly influence predictions and results.
- **Explainable AI Tools**, incorporating existing techniques such as LIME (Local Interpretable Model-agnostic Explanations) or SHAP (SHapley Additive exPlanations) to generate local explanations for individual predictions, thereby providing insight into how the model weighs different factors in specific scenarios.
- **Hybrid Modelling Approaches**, that combine data-driven machine learning models with physics-based models. This not only leverages domain knowledge but also constrains the AI model with physical principles, making the outputs more interpretable and less prone to overfitting errors.

Those methods will be integrated into the existing AI and Remote Sensing methodologies already developed by the research group in the last five years.
The project will explore slope hazard predictions in regions with high slope dynamics such as the European Alps, the Tibetan Plateau, the Nepal Himalayas, and other areas recently affected by multi-hazard events triggered by climate extremes or earthquakes. Remote sensing data will be used as the main basis for detecting, analyzing, and forecasting slope displacements by fully integrating optical, SAR, and InSAR methods with the help of XAI techniques.

The following steps are envisaged for the development of the PhD:

- State of the art analysis, Data acquisition
- Initial experiments with baseline models
- Model development and application to selected test areas
- Field calibration and validation, model refinement
- Result interpretation with the help of XAI tools
- Development of XAI-supported prediction systems for use cases

Any detailed PhD proposals that partially or completely follow the lines provided are of potential interest, with special reference to those addressing important open science questions such as:

- What are the most important factors driving slope dynamics after large triggering events such as typhoons and earthquakes?

- What are the implications of projected climate scenarios on the predicted impact of mass wasting at large scales?
- How can we better integrate physics-based models and process understanding in artificial intelligence methods for the prediction of slope and ground disasters?

The project is integrated into current research lines of the Engineering Geology group at the Department of Geosciences and with the activities of the Machine Intelligence and Slope Stability Laboratory, in cooperation with international research institutions including the University of Cambridge (UK), the GFZ Potsdam (Germany), the CTTC Barcelona (Spain), the State Key Laboratory for Geohazard Prevention (SKLGP, Chengdu University of Technology, China), and the Tongji University (Shanghai, China).
Exchanges with one or more such institutions are expected, with study-abroad periods. The research will exploit the recently acquired GPU cluster NVIDIA DGX H100 and available research funds (SNAM Rete Gas, Autorità di Distretto del Po, ESA, National Civil Protection research contracts) to support field work, data acquisition, and travel to important scientific meetings such as EGU General Assembly, ICL World Landslide Forum, and AGU Assembly.