DOI: https://doi.org/10.5592/CO/PhDSym.2026.01

Training AI algorithms for remote sensing monitoring of geotechnical structures using large-scale synthetic data

Lou Lerren Curacha[1], Irina Stipanović[2], Meho Saša Kovačević[3], Mario Bačić[4]

[1]Infra Plan Consulting Ltd., lou.curacha@infraplan.hr
[2]Infra Plan Consulting Ltd., Irina.stipanovic@infraplan.hr, University of Twente, i.stipanovic@utwente.nl
[3]University of Zagreb Faculty of Civil Engineering, meho.sasa.kovacevic@grad.unizg.hr
[4]University of Zagreb Faculty of Civil Engineering, mario.bacic@grad.unizg.hr

Abstract

Climate-related hazards threaten critical infrastructure, while monitoring remains limited by sparse and incomplete observations. This research proposes an AI-based workflow that integrates remote sensing data from stations, satellite, and drone through calibration, fusion, super-resolution, and forecasting to generate enhanced temporal datasets. Derived products are combined with other data to create synthetic training samples for damage detection and multi-hazard susceptibility mapping of geotechnical structures.

Key words: Remote Sensing, Machine Learning (ML), Super-Resolution, Generative Adversarial Networks (GAN), Damage Detection, Multi-Hazard Susceptibility Mapping, Structural Health Monitoring (SHM), Geotechnical Structures (GS)

Treniranje algoritama umjetne inteligencije za daljinsko praćenje stanja geotehničkih konstrukcija korištenjem velikih sintetičkih skupova podataka

Sažetak

Klimatske promjene sve više ugrožavaju kritičnu infrastrukturu, dok je praćenje njezina stanja ograničeno rijetkim i nepotpunim opažanjima. Rad istražuje primjenu umjetne inteligencije za integraciju podataka prikupljenih satelitima, dronovima i ugrađenim senzorima. Podaci će se objediniti kalibracijom, fuzijom, super-rezolucijom i prognoziranjem te koristiti za izradu sintetičkih skupova podataka za detekciju oštećenja i procjenu ranjivosti geotehničkih konstrukcija.

Ključne riječi: daljinska istraživanja, strojno učenje, super-rezolucija, generativne suparničke mreže, detekcija oštećenja, mapiranje podložnosti višestrukim opasnostima, monitoring, geotehničke konstrukcije

1 Introduction

Critical transportation infrastructures (CTI) such as roads, railways, bridges, and other structures are increasingly exposed to landslides, rockfalls, flooding, erosion, and land subsidence due to the growing impacts of climate change on geotechnical structures (GS). Recent studies show that global warming is intensifying extreme climate events and altering precipitation patterns, while transportation assets are increasingly recognized as highly vulnerable to climate-related hazards. These hazards can cause direct physical damage to infrastructure and disrupt mobility, accessibility, emergency response, and long-term recovery. Reliable monitoring and assessment therefore require continuous information on environmental, external, and infrastructure-related variables that influence hazard occurrence, spatial susceptibility, and infrastructure condition.

Across recent studies, precipitation, temperature, slope and other digital elevation model (DEM) - derived topographic factors, land use land cover (LULC), soil properties, and other factors are repeatedly identified as important conditioning or triggering variables for landslide, flood, rockfall, and land-subsidence analysis [1-3]. However, direct observations of these variables are often spatially sparse, temporally incomplete, inconsistent in resolution, or unavailable across large areas. To address these limitations, multi-source observations from station sensors, satellite data, LiDAR, and drones can be integrated to improve spatial coverage, temporal continuity, and data completeness. Station observations can be extended through interpolation and calibration with satellite products [4-6], while satellite-derived datasets can be enhanced through data fusion, downscaling, and super-resolution using LiDAR, drone, and other high-resolution reference data [7-9]. Together, these processes can support the generation of enhanced multi-source spatio-temporal (EMST) datasets with improved spatial coverage, temporal continuity, and data completeness on performance of GS to support forecasting of environmental loading conditions and estimate temporally incomplete observations.

Building on these capabilities, the objective of this study is to develop an AI-based workflow that integrates multiple observations to generate EMST and synthetic datasets for multi-hazard susceptibility analysis, damage detection, and monitoring of geotechnical structures along transport infrastructure networks. The workflow integrates calibration, interpolation, data fusion, super-resolution, forecasting, and synthetic data generation to support AI-based identification of hazard-susceptible areas, detection of affected CTI, and interpretation of hazard-related impacts within a GIS-based digital twin platform. Through this approach, the study addresses the need for more continuous, scalable, and data-driven methods for assessing CTI exposed to multiple hazards.

2 Research methodology

The research advances the current state of the art by moving beyond hazard susceptibility, damage detection, and CTI monitoring approaches toward an AI-based workflow for the integration and fusion of remote sensing-based observations into the monitoring of GS exposed to multiple hazards. As shown in Figure 1, the proposed methodology combines ground-based observations, satellite remote sensing, LiDAR, drone data, and other geospatial datasets to address limitations related to sparse field measurements, incomplete monitoring records, and inconsistent spatial and temporal resolutions for AI model training across large transport-infrastructure networks. These multi-source observations are processed through calibration, fusion, enhancement, forecasting, and synthetic data generation to produce EMST datasets representing external and environmental variables and damage extent. The resulting datasets are then integrated into AI models for multi-hazard susceptibility mapping, damage detection, and monitoring of GS. The overall workflow consists of four main components: multi-source observations, calibration and enhancement of spatial datasets, forecasting and synthetic data generation, and AI-based analysis of hazard susceptibility and infrastructure impacts. In this way, the study supports the European Union MSCA–FOURIER project vision by contributing to AI-based data fusion and interpretation, GIS-based digital twin visualization, and risk-informed early-warning and maintenance planning for exposed CTI.

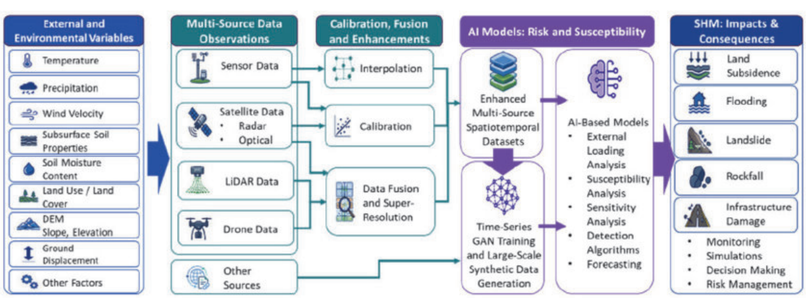

Figure 1. Scientific workflow for integrating multi-source observations for AI-based risk and susceptibility modelling

2.1 Multi-Source Observations

The methodology begins with the collection of multi-source observations representing the environmental and external variables that influence hazard occurrence, susceptibility, and infrastructure condition. These variables include precipitation, temperature, wind velocity, soil properties, LULC, topographic factors derived from DEM, and ground deformation. These variables were selected because they are directly related to the triggering, conditioning, and evolution of hazards related to GS [1-3].

	Temporal Frequency		Spatial Resolution		Spatial Coverage		Data Sources
Sensor Data	✓	Up to every 5 Minutes	X	Point Coordinates	X	Point Coordinates	DMHZ Stations FCEZG Stations
Satellite Data • Radar • Optical	–	Up to Hourly (Wind) Up to Daily (Temp.) Up to Weekly (DEM) Up to Weekly (Optical)	–	Up to 28km (Wind) Up to 1km (Temp.) Up to 100m (DEM) Up to 10m (Optical)	✓	Global Scale	Sentinel 1, MODIS Sentinel 2, ERA5 EGMS, Landsat 7/8/9 CHIRPS, SRTM
LiDAR Data	X	Upon Request	✓	15ppsm to 40ppsm	✓	Large Scale	State LiDAR Survey
Drone Data	X	Upon Request	✓	Up to 20cm Resolution	✓	Large Scale	FCEZG Drone Survey

Figure 2. Overview of multi-source observations used in the study: frequency, resolution, coverage

To capture these variables, the study integrates complementary observations from station sensors, satellite data, LiDAR, and drone platforms, as summarized in Figure 2. Station sensors provide localized and temporally frequent measurements, satellite datasets extend coverage across broader areas, and LiDAR and drone data improve spatial detail for selected locations. Together, these sources provide the basis for generating EMST datasets in areas where direct observations are sparse, incomplete, or unavailable. However, temporal gaps and data-quality limitations remain common, including missing station records, cloud-affected optical imagery, low radar coherence, and limited LiDAR or drone availability due to acquisition cost and temporal constraints. These limitations support the need for calibration, fusion, enhancement, and forecasting to improve data continuity and generate more reliable inputs for regional-scale hazard susceptibility analysis.

2.2 Calibration, fusion and enhancement

The limitations identified in the collected observations require processing procedures that improve data consistency, spatial detail, and temporal continuity. Therefore, calibration, fusion, enhancement, forecasting, and interpolation are applied to transform heterogeneous datasets into more reliable, higher-resolution, and analysis-ready inputs for hazard susceptibility analysis and monitoring of GS. In particular, point-based station observations are scaled to regional coverage through interpolation and calibration with satellite-derived variables, allowing sparse local measurements to contribute to continuous regional-scale datasets. Interpolation methods such as kernel interpolation with barriers will be tested because of their reported performance [5]. Equation for Linear Regression is:

$$y_i = B_0 + B_1 x_i + e_i \tag{1}$$

where:

y_i - dependent variable (ground measurement)
B_0 - intercept
B_1 - regression coefficient
x_i - independent variable(satellite observation)
e_i - residual error

Calibration is used to improve the consistency between satellite-derived measurements and local observations. In this study, linear regression, shown in Equation (1), is considered as one calibration approach for modeling the relationship between satellite variables and ground-based measurements. Regression-based calibration has been used in previous studies to improve or assess satellite products using reference observations. One sample study includes the MODIS land surface temperature with meteorological-station data [6] another is the MODIS precipitable water vapor products with ground-based GPS observations [4].

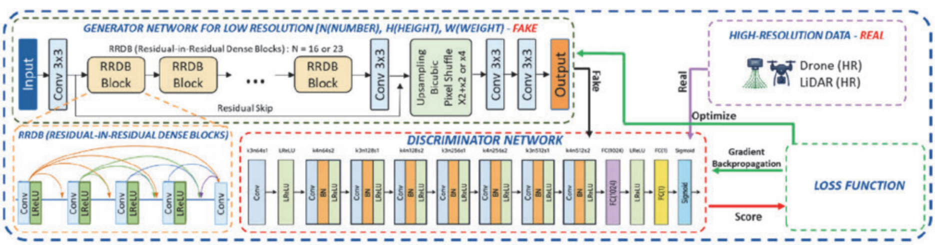

Figure 3. Proposed AI-based super-resolution framework for enhancing satellite-derived products

Satellite products are enhanced through data fusion and super-resolution such as enhanced super-resolution generative adversarial networks (ESRGAN) to improve spatial detail, terrain representation, and the delineation of surface and features. As shown in Figure 3, the pathway uses high-resolution data, such as LiDAR, drone products, or high-resolution optical imagery, to guide the enhancement of coarser satellite-derived products. When data are unavailable, an alternative is to use ESRGAN for single imagery approaches such as S2DR3 [7], although outputs should be used cautiously because performance may vary and visually enhanced textures may introduce hallucinated features [9][10]. This strategy follows UAV-satellite synergy principles, where drone observations provide high spatial detail and satellite observations provide broader spatial coverage and temporal continuity [8]. The ESRGAN framework proposed by Wang et al. (2018) [11] is considered as a reference because it can support both single-image enhancement and guided data-fusion approaches for optical and elevation data. Equation for Harmonic Regression with Linear Trend is:

$$y_t = B_0 + B_1 t + \sum_{k=1}^{K} (a_k \sin\left(\frac{2\pi kt}{P}\right) + b_k \cos\left(\frac{2\pi kt}{P}\right) + e_t \tag{2}$$

where:

y_t - dependent variable at time t
B_0 - intercept
k - harmonic index
B_1 - linear trend term
K - number of harmonics
P - period
a_k, b_k - sine and cosine coefficients
e_t - residual term
t - time

Temporal gaps in observational datasets can reduce data continuity and limit large-scale environmental monitoring and hazard analysis. To address this, forecasting methods are applied to EMST datasets to estimate missing observations and anticipate future environmental loading conditions such as temperature, precipitation, and other variables. For temperature, a harmonic regression model with a linear trend, shown in Equation (2), is proposed to capture both long-term and periodic variability using historical maximum, minimum, and mean values and the most recent seven-day observation window. Other approaches, such as long short-term memory networks may also be used for multivariate forecasting involving temperature and precipitation [12]. These forecast variables are then incorporated into the EMST for subsequent AI-based analysis of CTI exposure.

2.3 AI-based multi-hazard susceptibility mapping

The proposed workflow, shown in Figure 4, uses EMST datasets as the basis for extracting external and environmental variables for AI-based hazard analysis. These variables are derived from enhanced datasets and supplemented by geospatial, external and GS-related sources, such as DEM-based terrain factors, InSAR-derived deformation indicators, satellite-based vegetation and LULC information, and station observations, depending on the target hazard and data availability. Open geospatial databases may also provide exposure-related variables, such as traffic intensity and other infrastructure or socio-economic indicators, to represent conditioning factors, triggering factors, and exposed elements in hazard-prone areas.

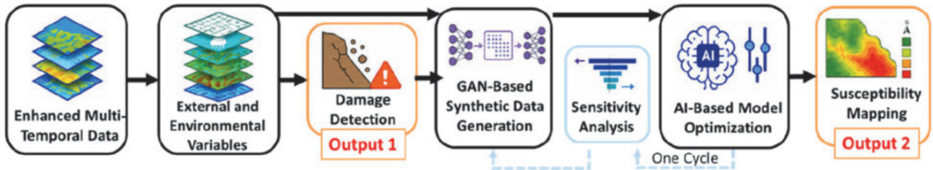

Figure 4. Proposed workflow of AI-Based Synthetic Data Generation and Multi-Hazard Susceptibility Mapping

The first output of the workflow is damage detection, which complements susceptibility mapping by identifying existing or visible impacts on CTI. Using EMST optical satellite imagery, deformation products, and other remote sensing datasets, AI-based models can be applied to detect potentially affected areas, including road damage, building damage, slope failures, and deformation-related effects. Previous studies have demonstrated the use of deep-learning, convolution neural network, InSAR-based, and vision transformer approaches for road-damage detection, building-damage assessment, and deformation-related infrastructure risk identification [13-14]. In this study, damage detection is used to provide direct evidence of observable CTI impacts and to support the interpretation of hazard-related consequences within an SHM context.
The second output is multi-hazard susceptibility mapping, where the extracted external and environmental variables are used with GAN - based synthetic data generation and AI-based model optimization. GAN-generated synthetic data are applied to produce additional training samples that support the learning process and improve model performance, which will be assessed before and after synthetic data augmentation, similar to the study by Al-Najjar and Pradhan (2020) for landslide susceptibility mapping [15]. Different AI-models or automated model-selection approaches may be evaluated under similar settings to identify suitable susceptibility models. After model optimization, a sensitivity analysis is conducted once to determine the most influential variables. These high-importance variables are then used in an additional cycle of GAN-based synthetic data generation and AI-based model optimization to refine the model and produce final multi-hazard susceptibility maps.

2.4 SHM impacts and consequences

The EMST datasets, damage-detection outputs, and multi-hazard susceptibility maps will be integrated into GIS-based digital twin platform to support life cycle management of GS exposed to climate change impacts which can cause landslides, rockfalls, flooding, erosion, land subsidence, and related hazards. Within this platform, external and environmental drivers, predicted hazard-prone areas, detected damage, and possible infrastructure consequences are linked to infrastructure elements, allowing the workflow to move beyond hazard mapping by providing stakeholders with a monitoring environment where exposed assets can be visualized, assessed, and prioritized for inspection, validation, or intervention. The susceptibility maps and damage-detection outputs function as dynamic decision-support layers that can be updated as new datasets become available, supporting continuous monitoring, early-warning applications, and SHM monitoring. When combined with exposure datasets such as buildings, road networks, population density, or other socio-economic indicators, the platform can also support broader consequence assessment and risk-informed decision-making.

3 Conclusion

The research presented in this paper aims to develop an AI-based workflow for generating enhanced EMST and synthetic datasets for AI-driven multi-hazard susceptibility analysis, detection, simulation, and monitoring applications for GS. The proposed research addresses the challenge of sparse, incomplete, and inconsistent observational data by integrating station measurements, satellite observations, LiDAR, drone data, and other geospatial sources into more complete and analysis-ready datasets. Through calibration, fusion, enhancement, forecasting, and synthetic data generation, the workflow is designed to produce enhanced datasets that are more suitable for AI-based hazard analysis and CTI monitoring.

The main benefit of the research is its potential to support scalable and data-driven assessment of CTI exposed to multiple hazards. By applying AI models to the EMST datasets, the research can support the identification of hazard-prone areas, detection of impacts, simulation of hazard scenarios, and interpretation of results within a GIS-based digital twin platform. Overall, the proposed research provides a flexible and scalable framework for improving multi-hazard susceptibility analysis, detection, and GS monitoring using EMST and synthetic datasets.

Acknowledgment

This research is performed within FOURIER project, GA Number 101169429, which received funding from the EU HORIZON-MSCA-2023-DN-01 program.

References

[1] Bajni, G., Camera, C. a. S., & Apuani, T. (2023). A novel dynamic rockfall susceptibility model including precipitation, temperature and snowmelt predictors: a case study in Aosta Valley (northern Italy). Landslides, 20(10), 2131–2154. https://doi.org/10.1007/s10346-023-02091-x

[2] Dai, X., Chen, J., Zhang, T., & Xue, C. (2025). Integrated landslide risk assessment via a landslide susceptibility model based on intelligent optimization algorithms. Remote Sensing, 17(3), 545. https://doi.org/10.3390/rs17030545

[3] Islam, T., Zeleke, E. B., Afroz, M., & Melesse, A. M. (2025). A Systematic review of urban flood susceptibility Mapping: remote sensing, machine learning, and other modeling approaches. Remote Sensing, 17(3), 524. https://doi.org/10.3390/rs17030524

[4] Xu, J., & Liu, Z. (2022). A linear regression of differential PWV calibration model to improve the accuracy of MODIS NIR All-Weather PWV products based on Ground-Based GPS PWV data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 7929–7951. https://doi.org/10.1109/jstars.2022.3204823

[5] Yang, R., & Xing, B. (2021). A Comparison of the Performance of Different Interpolation Methods in Replicating Rainfall Magnitudes under Different Climatic Conditions in Chongqing Province (China). Atmosphere, 12(10), 1318. https://doi.org/10.3390/atmos12101318

[6] Zhao, B., Mao, K., Cai, Y., Shi, J., Li, Z., Qin, Z., Meng, X., Shen, X., & Guo, Z. (2020). A combined Terra and Aqua MODIS land surface temperature and meteorological station data product for China from 2003 to 2017. Earth System Science Data, 12(4), 2555–2577. https://doi.org/10.5194/essd-12-2555-2020

[7] Akhtman, Y (2023). Sentinel-2 Deep Resolution. Available online: https://medium.com/@ya_71389/sentinel-2-deep-resolution-3-0-c71a601a2253 (accessed on 30 April 2026)

[8] Alvarez-Vanhard, E., Corpetti, T., & Houet, T. (2021). UAV & satellite synergies for optical remote sensing applications: A literature review. Science of Remote Sensing, 3, 100019. https://doi.org/10.1016/j.srs.2021.100019

[9] Chanev, M., Kamenova, I., Dimitrov, P., & Filchev, L. (2025). Evaluation of Sentinel-2 Deep Resolution 3.0 data for winter crop identification and organic barley yield prediction. Remote Sensing, 17(6), 957. https://doi.org/10.3390/rs17060957

[10] Massarelli, C. (2026). Super-Resolution of Sentinel-2 satellite images: A comparison of different interpolation methods for spatial knowledge extraction. Machine Learning and Knowledge Extraction, 8(1), 14. https://doi.org/10.3390/make8010014

[11] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C. C., Qiao, Y., & Tang, X. (2018). ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. arXiv. https://doi.org/10.48550/arxiv.1809.00219

[12] Hasan, K., Islam, M. S., Mitu, K. N., Salam, M. I., & Aghashahi, N. (2025). Evaluating the performance of CHIRPS and CPC precipitation data for streamflow forecasting using multiple linear regression and Long Short-Term Memory Neural Network model. MethodsX, 15, 103443. https://doi.org/10.1016/j.mex.2025.103443

[13] Cha, J., Lee, S., & Kim, H. (2025). Deep Learning-Based Detection and Assessment of Road Damage Caused by Disaster with Satellite Imagery. Applied Sciences, 15(14), 7669. https://doi.org/10.3390/app15147669

[14] Shahbazi, S., Barra, A., Gao, Q., & Crosetto, M. (2024). Detection of buildings with potential damage using differential deformation maps. ISPRS Journal of Photogrammetry and Remote Sensing, 218, 57–69. https://doi.org/10.1016/j.isprsjprs.2024.10.008

[15] Al-Najjar, H. A., & Pradhan, B. (2020). Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks. Geoscience Frontiers, 12(2), 625–637. https://doi.org/10.1016/j.gsf.2020.09.002