# Machine-learning approaches to enhance tsunami damage modeling for coastal roads: lessons from the 2011 Great East Japan Event

**Anna Rita Scorzini,**[1] Mario Di Bacco, James H. Williams, Daisuke Sugawara

## Abstract

In the context of disaster risk assessment, the accurate prediction of tsunami-induced damage to coastal road infrastructure remains a critical challenge. Conventional approaches often rely on fragility functions, which, while valuable, might not fully capture the intricate non-linear interactions among diverse variables influencing the damage mechanisms. This study aims at addressing this limitation by employing advanced multi-variable machine learning models, based on the road damage dataset compiled by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) of Japan after the 2011 Great East Japan Event in the Tōhoku region.

The dataset comprises a line shapefile that represents the actual length of each damaged road segment, with an associated damage level, distinguished into three classes, as minor, moderate or severe damage. To address the absence of information regarding undamaged assets into the original MLIT dataset, all roads within the inundated area are incorporated into it. This integration involves sourcing road data from OpenStreetMap or digitizing them from aerial imagery. As a result of this comprehensive data compilation effort, the dataset finally encompasses approximately 4300 km of inundated roads, with roughly 20% of them associated with a specific damage level. This empirical dataset is one of the few globally available for tsunami-damaged transportation infrastructure, making it a valuable asset to be used in data-intensive machine-learning approaches. Moreover, to obtain insights into the multi-variable nature of tsunami damage mechanisms on roads, the dataset has been enriched by associating additional potential explanatory variables to each segment. These are not limited to the hydraulic features of the inundation, but they also encompass morphological features at the road location as well as proxy variables for accounting for shielding effects and debris impacts exerted by nearby structures, all of which can be computed through straightforward geospatial analysis.

The utilization of this extended dataset within a machine-learning framework allows for an enhanced understanding of road damage dynamics during tsunami events, overcoming the limitations of conventional fragility functions. The insights generated by multi-variable models, such as the outcomes from the feature importance analysis, can empower stakeholders with the essential knowledge for more informed and efficient decision-making in disaster risk management. This encompasses targeted and optimized resource allocation, both in the ex-ante risk assessment phases and during the post-event emergency response. As a result, the improved model outcomes, along with the explicit treatment of modeling uncertainties and efficient input data retrieval and computation, would contribute to an enhanced overall effectiveness of risk management processes. The proposed modeling framework demonstrates significant potential for replication across diverse geographical regions. As new datasets emerge, this approach holds promise for widespread applicability, making it a versatile and adaptable tool for disaster risk assessment across various contexts.

[1] Department of Civil, Environmental and Architectural Engineering, University of L’Aquila