Development of a Composite Drought Index using deep learning: A Unified Framework for Multi-Dimensional Drought Characterization

Mostafa Khosh Chehreh[1*], Carlo De Michele[1] | [1]Politecnico di Milano, Dept. of Civil and Environment Engineering, Milano, Italy

Drought is a complex, multidimensional phenomenon, and each drought index captures different aspect of its occurrence and impact. A prolonged drought can reduce water availability, leading to substantial declines in both crop yields and livestock productivity. For example, for example, recent severe droughts in Southern Europe have resulted in yield losses of up to **30%** for certain crops such as maize and wheat .

What did we do ?

1- Developing a new composite drought index based on a cutting-edge **deep learning** algorithm using existing drought indices.

2- **Impact-oriented** analysis of the developed model instead of the model-by-model evaluations.

Why is it important ?

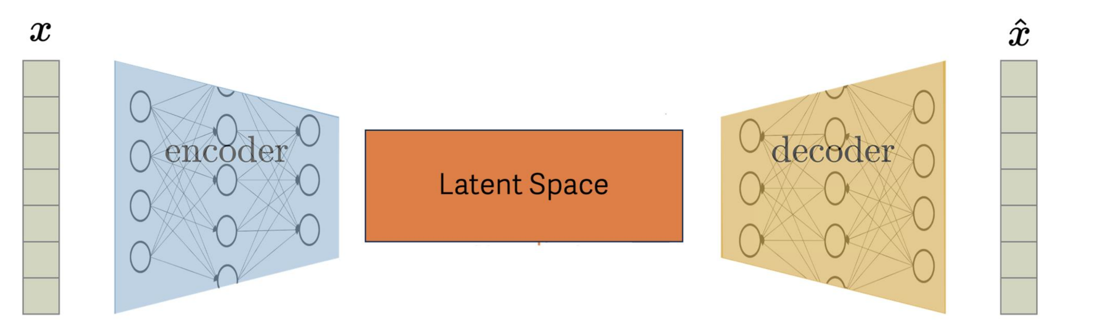
It makes a new stride toward shifting the paradigm from the estimate "what the drought event **will be**" to "what the drought event **will do**", and this approach enables decision-makers to plan and implement targeted, actionable measures.

Method

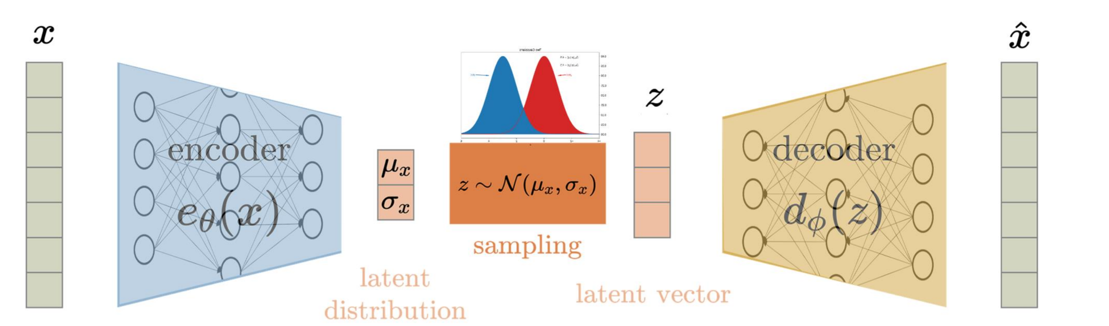
In this study, we developed three distinct autoencoder architectures to derive a novel drought index: a fully connected neural network autoencoder (NN-AE), a convolutional neural network autoencoder (CNN-AE), and variational autoencoder (VAE).

The autoencoder architecture comprised of three sectors: the encoder layers, the decoder layers and a latent space component(S. Chen & Guo, 2023).[1] The first unit, the encoder maps an input data into the latent space, where typically enforces a compressed representation of the original input, after that the decoder network aims to reconstruct the original data from the latent representation

Using three widely adopted monthly drought indicators—Standardized Precipitation Index (SPI), Standardized Precipitation Evapotranspiration Index (SPEI), and Soil Moisture Deficit Index (SMDI)—as input features, we trained four separate autoencoder models under distinct architectural assumptions to learn robust, low-dimensional latent representations of drought severity over the period 1989–2024.



$$\text{reconstruction loss} = \|x-\hat{x}\|_2 = \|x-d_\phi(z)\|_2 = \|x-d_\phi(\mu_x+\sigma_x\epsilon)\|_2$$



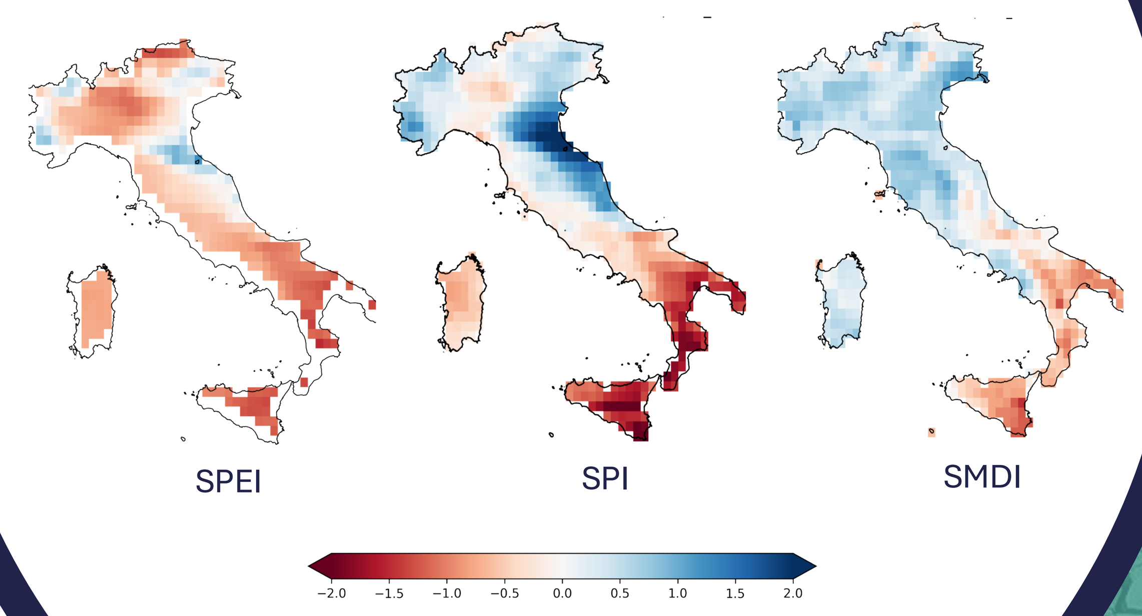
$$\text{reconstruction loss} = \|x-\hat{x}\|_2 = \|x-d_\phi(z)\|_2 = \|x-d_\phi(\mu_x+\sigma_x\epsilon)\|_2$$

$$\mu_x, \sigma_x = e_\theta(x), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$similarity\ loss = KL\ Divergence = D_{KL}(\mathcal{N}(\mu_x,\sigma_x) \parallel \mathcal{N}(\mathbf{0},\mathbf{I}))$$

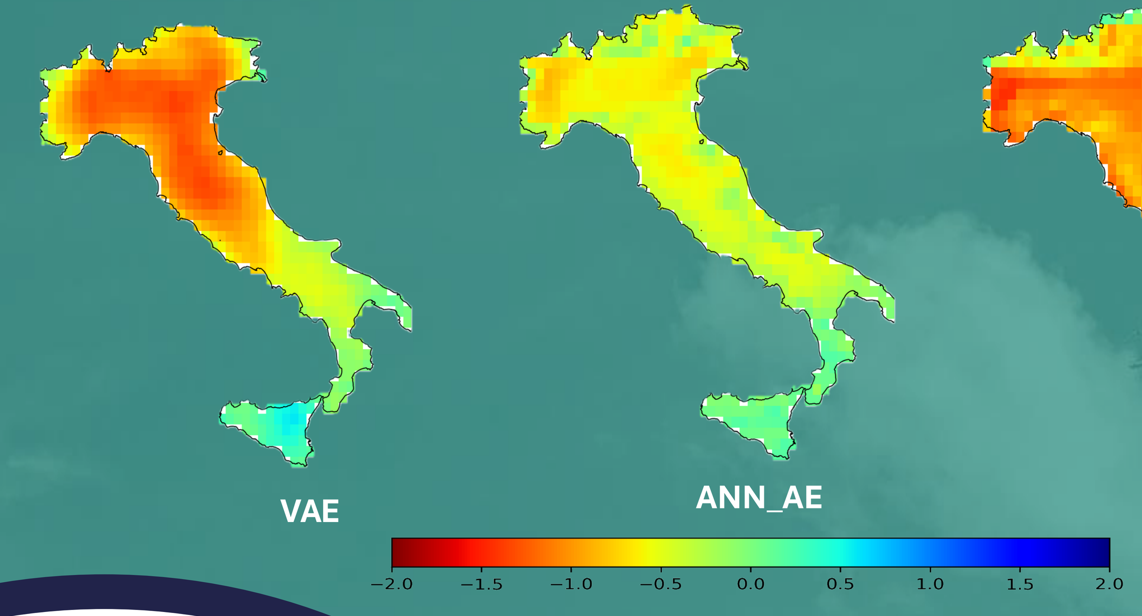
$$loss = reconstruction\ loss + similarity\ loss$$

For performance evaluation of DL model, we selected Italy—a Southern European country characterized by a Mediterranean climate and recently affected by severe droughts.



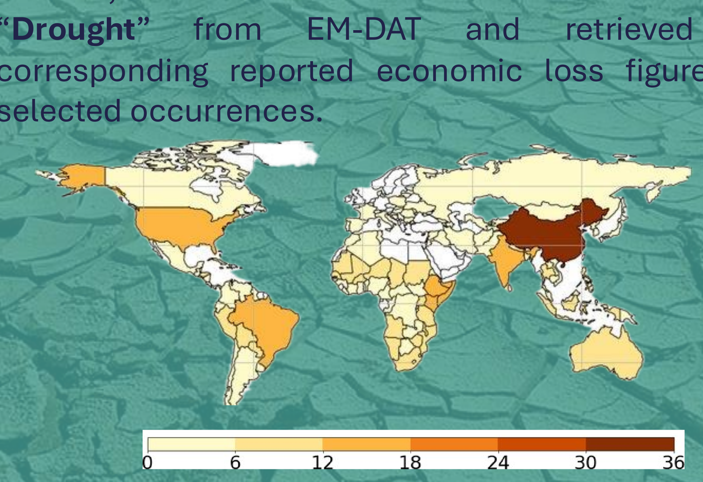
We computed the SPI, SPEI and SMDI from the ERA5 reanalysis dataset for the period 1983–2024

DL-based drought index over Italy for 2003 drought event



These results illustrate how architectural priors—global versus local connectivity and symmetric versus asymmetric encoding—affect each model's ability to reconstruct and emphasize critical spatial features of severe drought. The VAE excels at capturing extreme anomalies and preserving smooth, general patterns, the ANN-AE at conveying a broad overview, and the CNN-AE at detecting medium-scale spatial coherence. However, visual inspection alone cannot tell us which model performs best in practice.how can we find it ?

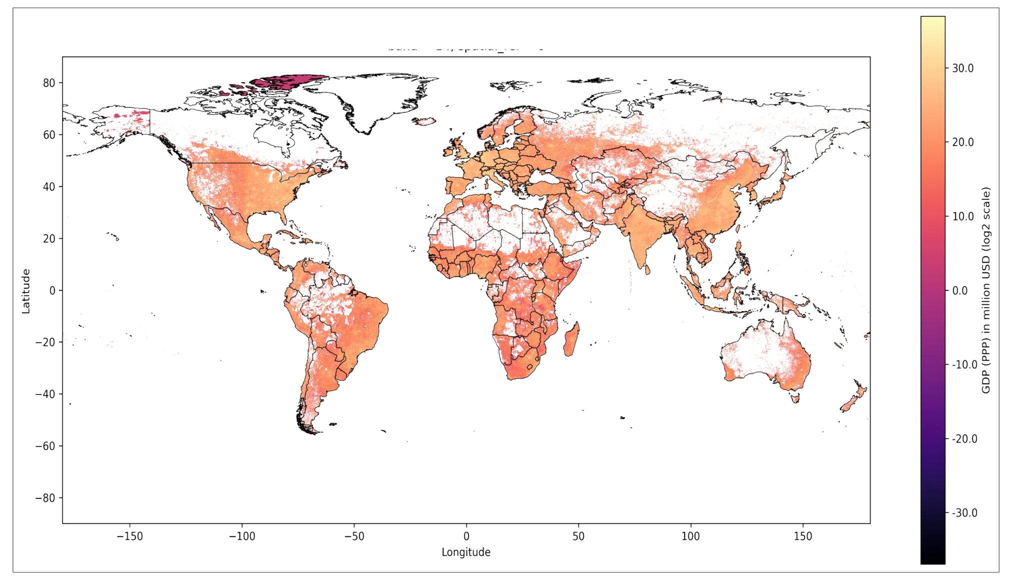
EM-DAT[2] (Delforge et al., 2023) is an open-access, country-level database of global disaster events. For local-scale performance evaluation of our new drought index The number of drought events based on EM-DAT, we extracted all events classified as "**Drought**" from EM-DAT and retrieved the corresponding reported economic loss figures for selected occurrences.



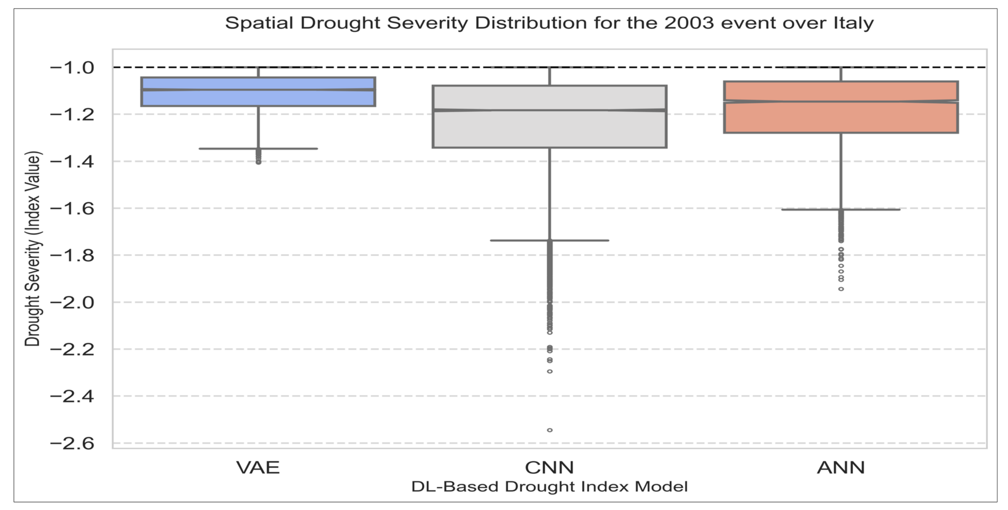
The number of drought events based on EM-DAT

For estimating potential economic losses, we used a high-resolution gridded GDP per capita (PPP)[3] dataset downscaled to 43 501 Admin-2 units worldwide (1990–2022), based on official subnational statistics from 89 countries (2 708 units). We first isolated agricultural GDP by intersecting the GDP grid with a land-use map, producing an Admin-2 map of agriculture-specific GDP-PPP. Then, for each autoencoder model, we applied a drought-severity threshold of –1 to identify drought-affected pixels. By overlaying these thresholded drought maps with the agricultural GDP grid, we quantified the exposed economic value and computed the potential losses attributable to drought events over the study period.

Global Spatial Distribution of GDP per Capita (PPP), 2003



+



2480.810	3757.144	755.64

Potential economic loss -M€

2146.776

Reported economic loss for 2003 drought event -M€

The VAE is the clear choice: its €2 480.81 M projection is within 16 % of the €2 146.8 M reported loss, whereas the CNN-AE overestimates by 75 %, and the ANN-AE underestimates by 65 %. This tells us that an encoder–decoder imbalance—more capacity devoted to the encoder—better preserves extreme drought signals without exaggerating or smoothing them away. In practical terms, VAE strikes the optimal balance between sensitivity to severe events and spatial specificity, making it the most reliable model for translating latent drought representations into realistic economic-loss estimates.

This presentation participates in OSPP

Outstanding Student & PhD candidate Presentation contest

POLITECNICO MILANO 1863

Eco Hydro System Group

Reference

1-Chen, S., & Guo, W. (2023). Auto-Encoders in Deep Learning—A Review with New Perspectives. *Mathematics*, 11(8), 1777. https://doi.org/10.3390/math11081777

2-Delforge, A., & Guha-Sapir, D. (2023). EM-DAT: The International Disaster Database. Centre for Research on the Epidemiology of Disasters (CRED), Université catholique de Louvain, Brussels, Belgium. Available online: https://www.emdat.be

3-Kummu, M., Kosonen, M., & Sayyar, S. M. (2025). Downscaled gridded global dataset for gross domestic product (GDP) per capita PPP over 1990–2022. Scientific Data, 12, Article 178. https://doi.org/10.1038/s41597- 025-04487-x

OSPP

Let's Start!

Email :mostafa.khosh@polimi.it