A deep learning approach to evaluate individual predictors for extreme precipitation in Greece

Vasileios Vatellis[1,2,*], Stelios Karozis[3], Iraklis A. Klampanos[4], Antonis Troumpoukis[1] and Antonis Gkanios[1]

[1]Institute of Informatics and Telecommunications, National Centre for Scientific Research "Demokritos", Greece
[2]School of Rural, Surveying and Geoinformatics Engineering, National Technical University of Athens, Greece
[3]Institute of Nuclear & Radiological Sciences & Technology, Energy & Safety, National Centre for Scientific Research "Demokritos", Greece
[4]University of Glasgow, UK

Abstract

Deep learning has become an increasingly powerful tool in climate science, enabling advances in tasks ranging from the identification of atmospheric circulation patterns to weather forecasting and extreme-event classification. Yet the inherent complexity of atmospheric processes—particularly those driving rare, high-impact precipitation extremes—continues to challenge the ability of neural models to generalise robustly across different regimes. In this study, we evaluate seven single-level predictors—spanning standard ERA5 fields (total precipitation, total cloud cover, 10 m u- and v-wind components) and physics-enriched diagnostics (convective precipitation, K-index, vertically integrated moisture divergence)—to forecast extreme precipitation (> 95th percentile) over Greece. Using inputs from a broad European domain (34°–72° N, 25° W–65° E), we train a representative deep-learning architecture on different subsets of these variables to isolate the single predictor that maximises classification skill. To corroborate our findings, we then apply an XGBoost classifier and analyse its split-gain importances. We find that vertically integrated moisture divergence consistently yields the highest skill in the deep-learning framework across all three forecast lead times (2, 4, and 6 days), whereas the XGBoost model most frequently splits on total precipitation. Through this dual-model approach, we pinpoint the ERA5 fields and diagnostic indices that carry the strongest signal for local precipitation extremes.

Keywords

Extreme precipitation, ERA5, Single-level predictors, Deep learning, XGBoost

1. Introduction

Over the past decades, the frequency and severity of climate-related hazards—such as floods, flash floods, and intense rainstorms—have risen sharply, underscoring the multifaceted impacts of global warming [1]. These extremes not only threaten lives and livelihoods but also strain the capacity of authorities to plan for and mitigate their consequences. In particular, the spatial heterogeneity of extreme rainfall—driven by local topography, land-cover, and microphysical processes—means that global and even regional climate signals can translate into highly localized impacts.

Municipal, regional, and national decision-makers therefore urgently need forecasting tools that can resolve local hazard impacts and enable proactive adaptation. In recent years, machine-learning (ML) and deep-learning (DL) approaches have shown great promise for predicting precipitation and related extremes (e.g., heavy rainfall, hail), in some cases achieving performance comparable to traditional numerical weather prediction (NWP) in both speed and skill [2, 3, 4]. State-of-the-art models—such as GraphCast [5]—leverage hundreds of reanalysis fields to jointly forecast multiple atmospheric variables, while specialized ML architectures (e.g., U-Nets [6], convolutional neural networks, ConvLSTMs [7]) have been trained to predict precipitation alone, drawing on a diverse assortment of inputs. These models demonstrate that data-centric forecasting can capture complex nonlinear relationships and rapidly assimilate new observations.

Workshop on AI-driven Data Engineering and Reusability for Earth and Space Sciences (DARES'25), co-located with the 28th European Conference on Artificial Intelligence (ECAI 2025), Bologna, Italy, October 25, 2025
vasileios.vatellis@iit.demokritos.gr (V. Vatellis)
0000-0001-8565-8343 (V. Vatellis)

Yet, an open question remains as to which predictor variables contribute most to extreme-rain classification. While many studies either assemble broad pools of reanalysis fields for general forecasting [4, 5, 8] or rely on a small set of popular inputs—such as geopotential height and wind components [9]—few have systematically evaluated each variable's predictive power, alone or in combination [10]. Consequently, a research gap remains in identifying the minimal, physically interpretable feature sets that yield the highest classification skill for rare, high-impact precipitation events.

From physics prespective, extreme precipitation events in Greece often arises from synoptic-scale disturbances (e.g., Mediterranean cyclones, atmospheric rivers) that traverse a broader European domain before impacting local systems [11]. By drawing predictor fields from a broad European domain (34°–72° N, 25° W–65° E), we aim to capture these upstream teleconnections and moisture pathways, under the assumption that large-scale atmospheric states exert dominant control on local extremes.

In this work, we fill the gap in variable-selection methodology by conducting an evaluation of ERA5 reanalysis variables as single predictors of extreme precipitation (> 95th percentile) over Greece. We assemble a core set of single-level predictor fields—encompassing thermodynamic, dynamic, and moisture-flux drivers—and including an anomaly indexes known to flag extremes. By training a representative deep-learning architecture on varying predictor inputs, we seek to identify the single variable that maximises classification accuracy for extreme precipitation events. To validate and interpret these results, we complement our neural model with an XGBoost classifier, whose split-gain feature importances provide an independent measure of each variable's influence. This dual-model framework ensures that our conclusions about predictor relevance are robust across both deep-learning and tree-based paradigms.

2. Related Work

Predicting precipitation has become a core task within the new generation of "foundation" weather models—large-scale (MLWP) systems trained on decades of ERA5 reanalysis—that aim to represent the full state of the Earth's atmosphere. Notable examples include GraphCast [5], GenCast [4], and Pangu-Weather [8], which together leverage architectures such as graph neural networks, conditional diffusion models, 3D Earth-specific transformers, and autoregressive forecasting loops to deliver global medium-range predictions with unprecedented speed and accuracy. Besides these global "foundation" systems, specialized deep-learning models have been developed expressly for short-term precipitation nowcasting. One prominent example is SmaAt-UNet [6], which adapts the classic U-Net segmentation architecture by integrating spatial attention modules and depthwise-separable convolutions to efficiently process sequences of radar or satellite-derived precipitation fields. Evaluated on real-world datasets—precipitation maps over the Netherlands and binary cloud-coverage images of France—SmaAt-UNet achieves comparable nowcasting accuracy to much larger networks while using only one quarter of their trainable parameters, demonstrating that lightweight, attention-augmented CNNs can deliver high-fidelity short-range rainfall forecasts.

An ensemble of machine-learning methods has also been applied to identify the dominant drivers of extreme-precipitation intensity and frequency on a regional scale [12], Random Forest, XGBoost, and feedforward Artificial Neural Networks were trained on meteorological and land-surface variables to predict monthly extremes across six U.S. regions, while separate emulators were built to estimate return periods of these events. Using Shapley Additive Explanations to interpret model outputs, the authors found that latent heat flux, near-surface relative humidity, soil moisture, and large-scale subsidence consistently ranked among the top predictors for both extreme intensity and frequency. Their results highlight the compound—and often non-linear—interactions of moisture, energy fluxes, and atmospheric stability in governing precipitation extremes, underscoring the value of systematic feature-importance analyses when selecting inputs for downstream classification or forecasting models.

Another complementary approach is provided by a recent flash-flood study in central western Europe, which links extreme precipitation events—defined as radar-derived hourly totals exceeding 40 mm h^{-1} from the RADOLAN dataset—to a small set of ERA5 proxy variables via linear modeling

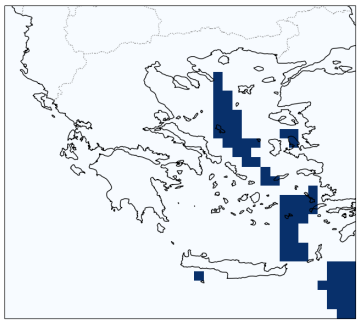

Figure 1: Binary map of extreme-precipitation events over Greece (34°–42° N, 19°–28° E) on a 33×37 grid. Dark blue cells indicate locations where the ERA5 total-precipitation value exceeded the 95th climatological percentile for the period of 1980–2023, while white cells fell below this threshold. This mask defines the "extreme" precipitation events.

over 1981–2020 [13]. In that work, high lower-tropospheric specific humidity ($q \geq 0.004 kg^{-1}$), sufficient instability (CAPE $\geq 327 Jkg^{-1}$), and low vertical wind shear between the surface and 500 hPa ($WS_{10}m$ 500hPa $\leq 6ms^{-1}$) emerged as the key atmospheric conditions favoring flash-flood-producing rainfall. Although they documented rising trends in moisture content and instability, no coherent trend was found in convective organization or event frequency—underscoring the intricate, non-linear pathways from large-scale atmospheric state to local precipitation extremes and the need to incorporate additional factors (e.g., intra-annual rainfall patterns, catchment characteristics) when selecting predictors for extreme-rainfall classification.

3. Data

We base our experiments on the ERA5 reanalysis [14], but rather than ingest its full catalog of hundreds of variables, we focus on a hand-picked suite of seven single-level predictors (Table 1) that dominate the precipitation-forecasting and extreme-event classification in the literature. These include the 10 m wind components (u10 and v10) to capture large-scale moisture advection; total precipitation (tp), convective precipitation (cp), and vertically integrated moisture divergence (vimd) to represent both accumulated rainfall and the fluxes that supply it; total cloud cover (acc) as a proxy for large-scale saturation; and the K-index (kx), a composite stability measure widely used to flag the posibility of a thunderstorm development. By blending these meteorological variables, our model will be tested with large-scale dynamics and localized effects that give rise to extreme precipitation.

Our predictor fields cover a broad European domain (34°–72° N, 25° W–65° E) on a 144×261 grid, sampled at two consecutive times frames—t_0 and $t_2 = t_1 + 6$ hours—and used to forecast a third time $t = t_2 + n$, where n is 2, 4, or 6 days depending on the forecast windo. We draw these samples at four synoptic times (00, 06, 12, 18 UTC) over the period of 1980 to 2023.

From the same ERA5 dataset we derive our target variable: a binary mask of extreme precipitation over the Greek domain (34°–42° N, 19°–28° E) on a 33×37 grid (Fig. 1). For each grid cell, we compare its total precipitation value against the the 95th-percentile threshold—computed from the 1980–2023 ERA5 record over the Greek domain (Table 2)—to determine if it exceeds this extreme cutoff. At forecast time t, any cell whose ERA5 tp exceeds its threshold 3 is labeled "1" (extreme), while all others are labeled "0."

$$tp(t_3) > P_{95}(tp).$$

To train and evaluate our models, we split this dataset chronologically into three subsets: training (1980–2010), validation (2011–2020), and testing (2021–2023). All results presented in the next section are computed on the held-out testing data, ensuring that our performance metrics reflect the models' ability to generalize to unseen, recent extreme-precipitation events.

Single Level
Total precipitation (tp), 10 metre u wind component, 10 metre v wind component, K index, Vertically integrated moisture divergence (vimd) [15], Total cloud cover (tcc) [15], Convective precipitation (cp)

Table 1
Seven single-level ERA5 variables handpicked as candidate predictors for classifying extreme precipitation (> 95th percentile) events over Greece. These fields span dynamic (10 m u- and v-wind), moisture (total precipitation, vertically integrated moisture divergence), convective (convective precipitation, K-index), and cloud-cover (total cloud cover) drivers used in our ML experiments.

Variables	Min Value	Max Value	Threshold
Total precipitation (tp)	0.0	0.023895263671875	0.0015935897827148438

Table 2
Climatological statistics for total precipitation (tp) over the Greek domain (34°–42° N, 19°–28° E) based on ERA5 hourly data (1980–2023). The 95th-percentile threshold (0.00159 m) was used to generate the binary extreme-precipitation mask in our experiments, alongside the minimum and maximum tp values observed.

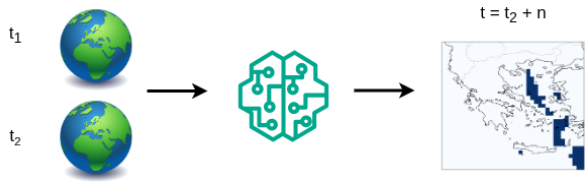

Figure 2: Schematic of our forecasting framework. Two consecutive six-hour snapshots of ERA5 fields at times t_1 and t_2 are fed into the neural network, which learns spatio-temporal patterns in the European domain. The model then produces a binary mask over the Greek subdomain at lead time $t = t_1 + n$ (where n =2,4,6 days),

4. Methodology

Figure 2 illustrates our modeling framework. The network ingests two consecutive 6-hour snapshots of atmospheric fields—at times t_1 and t_2—each defined over a 144×261 grid spanning Europe. From these inputs, it produces at a later time $t = t_2 + n$ (with $n = 2, 4$ or 6 days) a binary mask of extreme-precipitation events over the 33×37 Greek subdomain.

To assess the predictive power of individual ERA5 variables, we keep the network architecture and hyperparameters fixed for all experiments Figure 3. This way, any differences in forecast skill can be traced directly to the input variable set rather than changes in model complexity or training procedure. Practically, we train the same neural network repeatedly on input configurations ranging from each single predictor on its own up to the complete suite of seven fields, and evaluate performance using held-out validation metrics at lead time t.

By systematically comparing these single-variable and multi-variable runs, we reveal which atmospheric fields—notably thermodynamic, dynamic, or moisture-flux diagnostics—carry the strongest signal for classifying > 95th-percentile precipitation over Greece, without confounding effects from architectural or tuning differences.

To corroborate these findings with a complementary methodology, we also train an XGBoost classifier on the same input–output pairs. Because XGBoost expects tabular feature vectors rather than full spatial maps, we first apply an adaptive average pooling layer to downsample each 144 × 261 European predictor field to the 33 × 37 Greek grid, and then flatten these pooled maps into one-dimensional feature vectors. Unlike the deep-learning experiments—where we tested variables individually and in combinations—in the XGBoost workflow we present the full set of seven predictors simultaneously.

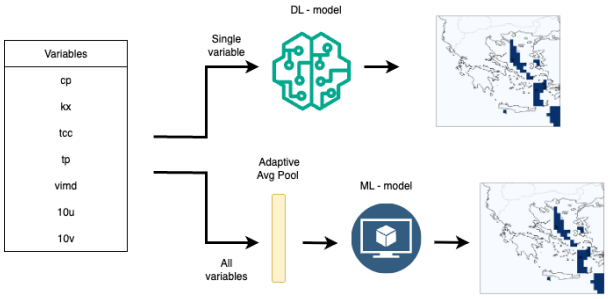

Figure 3: Dual-model pipeline for extreme-precipitation classification in Greece. The left panel lists the seven hand-picked ERA5 predictors. In the top row, each variable (alone) is passed through the deep-learning model—receiving full 144×261 European maps at two times—and outputs a 33×37 extreme-precipitation mask. In the bottom row, the same predictors are first downsampled via adaptive average pooling to the Greek grid, flattened into feature vectors, and then fed into an XGBoost classifier.

Furthermore, recognizing that tree-based ensembles require substantially less data to converge, we train XGBoost on a smaller time window (only the 2021–2023 samples) rather than the entire 1980–2023 record. This ensures that XGBoost sees exactly the same data representation as our neural network's final output layer, while also capitalizing on the efficiency of gradient-boosted trees. The resulting split-gain feature importances then provide an independent, model-agnostic ranking of which ERA5 fields carry the strongest signal for anticipating > 95th-percentile precipitation events over Greece.

5. Network

Our model builds on the SmaAt-UNet [6] backbone—a compact U-shaped encoder–decoder that combines depthwise-separable convolutions with convolutional block attention modules (CBAM) to efficiently extract multiscale spatial features. To capture short-term temporal dependencies, we insert a Convolution layer compained with an LSTM (ConvLSTM) at the bottleneck: the network ingests two consecutive time-step feature maps from the encoder, processes them through the ConvLSTM to learn spatiotemporal dynamics, and then feeds the recurrent output into the decoder path. Finally, after the last 1×1 convolution produces a high-resolution logit map, we apply 2D adaptive average pooling to exactly match the 33×37 grid of the Greek domain. This hybrid design lets us leverage both UNet's spatial hierarchies and ConvLSTM's temporal memory Figure 4.

In highly imbalanced binary-classification tasks (e.g., detecting rare "positive" pixels against an abundant "negative" background), standard binary cross-entropy tends to be dominated by easy negatives. To address this issue, we introduced the focal loss [16], which is an extension of standard binary cross-entropy.

$$\mathcal{L}_{FL}\left(p_t\right) = -\alpha\left(1-p_t\right)^{\gamma}\log p_t \tag{1}$$

6. Results

We evaluate our classification models at three forecast lead times—2, 4, and 6 days—ahead, with results tabulated in (Tables 3, 4, and 5), respectively. In each case, we report the held-out test loss, precision, recall, and F_1 score for the 2021–2023 period. As expected for a highly imbalanced task (few extreme-rain events vs. many non-extremes), recall remains low across all lead times, even as precision stays comparatively higher. Moreover, all four metrics steadily degrade as the forecast horizon lengthens—loss rises, while precision, recall, and F_1 decline—highlighting the increasing difficulty of detecting rare events further into the future as it can be show Figure 6, where the deficalty of the model to predict iextremes increases with time.

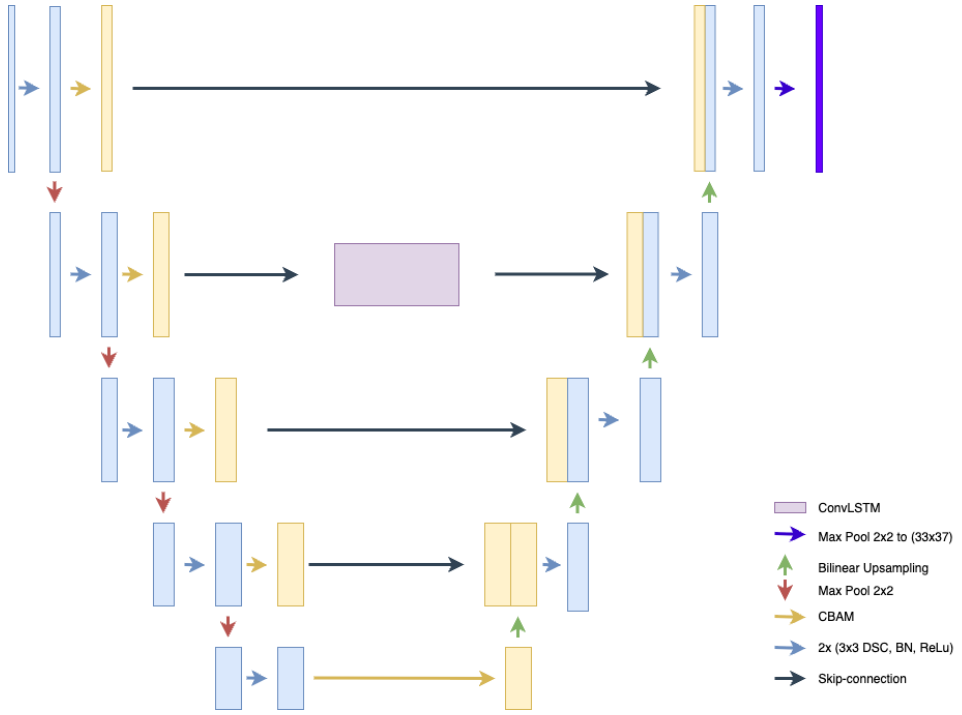

Figure 4: Architecture of our ConvLSTM-augmented SmaAt-UNet backbone [6]. Two consecutive multi-channel feature maps (at times t_1 and t_2) are encoded via depthwise-separable convolutions and CBAM attention blocks (gold bars), then merged through a ConvLSTM layer (purple) at the network bottleneck. Skip connections (black arrows) carry spatial features across the U-shaped encoder–decoder, which uses bilinear upsampling (green) and max pooling (red) to adjust spatial resolution. A final 1×1 convolution produces logits that are adaptively average-pooled (dark blue arrow) to the 33×37 Greek grid.

Despite this overall decay, certain predictors consistently stand out. In our deep-learning architecture, vertically integrated moisture divergence (vimd) invariably ranks among the top three variables, alongside total precipitation (tp) and the K-index (kx). Figure 5 illustrates how vimd's spatial patterns at t_2 closely resemble the eventual extreme-rain mask at t, underscoring its physical relevance. Although the precise ordering of tp, kx, nd vimd can shift slightly depending on whether we optimise for precision, recall, or F_1, all three deliver balanced performance across metrics, confirming that their prominence is not an artifact of a particular lead time, metric choice, or data split.

When we turn to the XGBoost experiments (Table 6), the stability of variable importance is even more pronounced. Although we do not report XGBoost's loss or precision—since our primary aim is to leverage its split-gain importances rather than its predictive scores—total precipitation and the K-index nonetheless occupy the first and second slots across all three horizons, followed consistently by total cloud cover (acc). In the middle ranks, vimd and convective precipitation (cp) trade places for the fourth position, while the 10 m u-wind component (u10) remains least important. This concordance between the tree-based and neural approaches—each with very different inductive biases—reinforces our conclusion that tp, kx, and vimd carry the strongest, most reliable signal for anticipating >95th-percentile precipitation in Greece.

Across all forecast horizons, the XGBoost feature-importance results mirror the deep-learning findings: vertically integrated moisture divergence (vimd) and total precipitation (tp) emerge as the top two predictors for classifying >95th-percentile precipitation events over Greece. This concordance between the tree-based and neural models underscores the robustness of these variables' predictive power. Although each methodology—ConvLSTM-U-Net and XGBoost—learns from different inductive biases, they both identify vimd and tp as carrying the strongest signal for upcoming extremes. The full split-gain importances for each lead time are presented below.

Variables	loss	precision	recall	f1
Convective precipitation (cp)	0.2003	0.2292	0.0219	0.0400
K index (kx)	0.2042	0.2228	0.0765	0.1139
Total cloud cover (tcc)	0.2044	0.2195	0.1477	0.1766
Total precipitation (tp)	0.2355	0.2027	0.0851	0.1199
Vertically integrated moisture divergence (vimd)	0.1999	0.2213	0.2402	0.2304
10 metre u wind component (10u)	0.2043	0.2224	0.2251	0.2237
10 metre v wind component (10v)	0.1974	0.2442	0.2021	0.2212

Table 3
Performance of the 2-day extreme-precipitation classification models when trained separately on each single predictor variable.

Variables	loss	precision	recall	f1
Convective precipitation (cp)	0.2188	0.1831	0.0001	0.0001
K index (kx)	0.2286	0.2120	0.0056	0.0109
Total cloud cover (tcc)	0.2304	0.1853	0.0373	0.0621
Total precipitation (tp)	0.2382	0.1003	0.0909	0.0954
Vertically integrated moisture divergence (vimd)	0.2211	0.2159	0.0426	0.0711
10 metre u wind component (10u)	0.2338	0.1696	0.0722	0.1012
10 metre v wind component (10v)	0.2158	0.2132	0.0432	0.0718

Table 4
Performance of the 4-day extreme-precipitation classification models when trained separately on each single predictor variable.

Variables	loss	precision	recall	f1
convective precipitation (cp)	0.2476	0.1080	0.1027	0.1052
K index (kx)	0.2286	0.2120	0.0056	0.0109
Total cloud cover (tcc)	0.2191	0.2126	0.0019	0.0038
Total precipitation (tp)	0.2443	0.0943	0.0002	0.0003
Vertically integrated moisture divergence (vimd)	0.2251	0.1990	0.0261	0.0462
10 metre u wind component (10u)	0.2243	0.0800	0.0008	0.0015
10 metre v wind component (10v)	0.2211	0.2115	0.0217	0.0393

Table 5
Performance of the 6-day extreme-precipitation classification models when trained separately on each single predictor variable.

forecast/Variable	cp	kx	tcc	tp	vimd	u10	v10
2d	44.6	132	62.3	235.7	49.2	38.1	49.2
4d	58.5	124.8	78.1	191.9	32.9	42.3	58.7
6d	49.8	117.6	64.4	186.3	47.9	38.6	72.6

Table 6
Mean XGBoost split-gain importances for each predictor variable and forecast horizon (2, 4, and 6 days). Each cell gives the average gain contributed by that feature toward reducing the model's training error, with larger values indicating more informative splits in the decision trees.

7. Conclusion

While our objective was not to build a state-of-the-art precipitation forecasting system, these experiments nonetheless revealed significant hurdles. Predicting a rare, binary extreme-event label—compounded as the precipitation which is also a cumulative variable and a heavily imbalanced

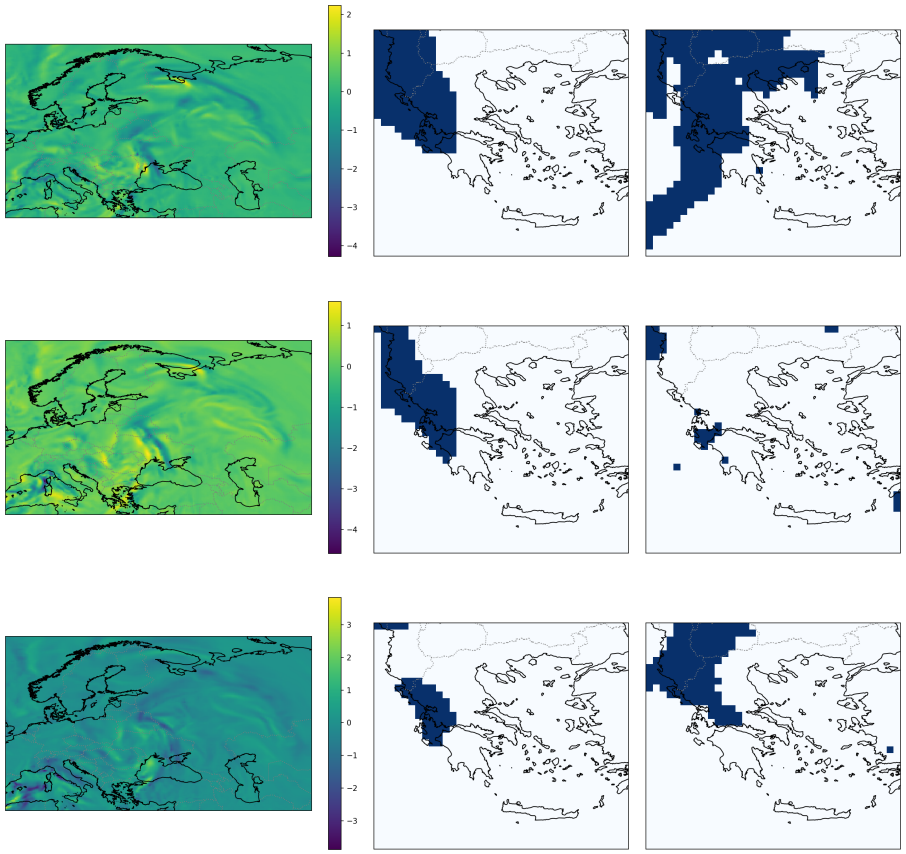

Figure 5: Left to right: input, predicted, and ground truth images for vimd. Start to bottom: 2 days, 4 days, 6 days forecast.

dataset—proved challenging even for our fixed deep-learning architecture. As anticipated, skill declined at longer lead times, with classification performance dropping off beyond the first forecast horizon (2 days).

Notably, total precipitation (tp) emerged as the single most important predictor across all three lead times when considering the XGBoost. The K-index also delivered consistently balanced performance, underscoring its robustness for extreme-rain classification. When we examine the XGBoost results, tp and K-index occupy the top two importance slots for every forecast horizon, reflecting their dominant role in the tree-based splits. In contrast, the ConvLSTM-U-Net's variable ranking is somewhat more fluid: its top three predictors vary by lead time but when considering the whole picture kx, vimd, and tp where the most important predictors. This agreement on the leading variables—despite the different inductive biases of the two methods—reinforces our confidence that these fields carry the strongest signal for predicting >95th-percentile precipitation events over Greece.

In addition to these methodological insights, our results also align with physical expectations. Total precipitation (tp) intuitively carries direct information about extreme-precipitation and therefore should strongly influence event classification. In the XGBoost experiments—where all seven predictors were presented simultaneously—tp indeed emerges as the most important split feature, suggesting that its signal is amplified by the presence of complementary variables. By contrast, in the deep-learning tests using tp alone, it did not rank as the top predictor, indicating that tp is less informative in the absence of other fields such as the 10 m wind components (u10, v10). From a physical perspective, the finding that tp combined with wind components improves predictive skill is consistent with the idea that extreme precipitation in Greece is driven by large-scale weather systems propagating across Europe, where the wind fields govern the trajectories and moisture transport of those events.

Together, these findings highlight both the difficulty of rare-event forecasting in a pure deep-learning

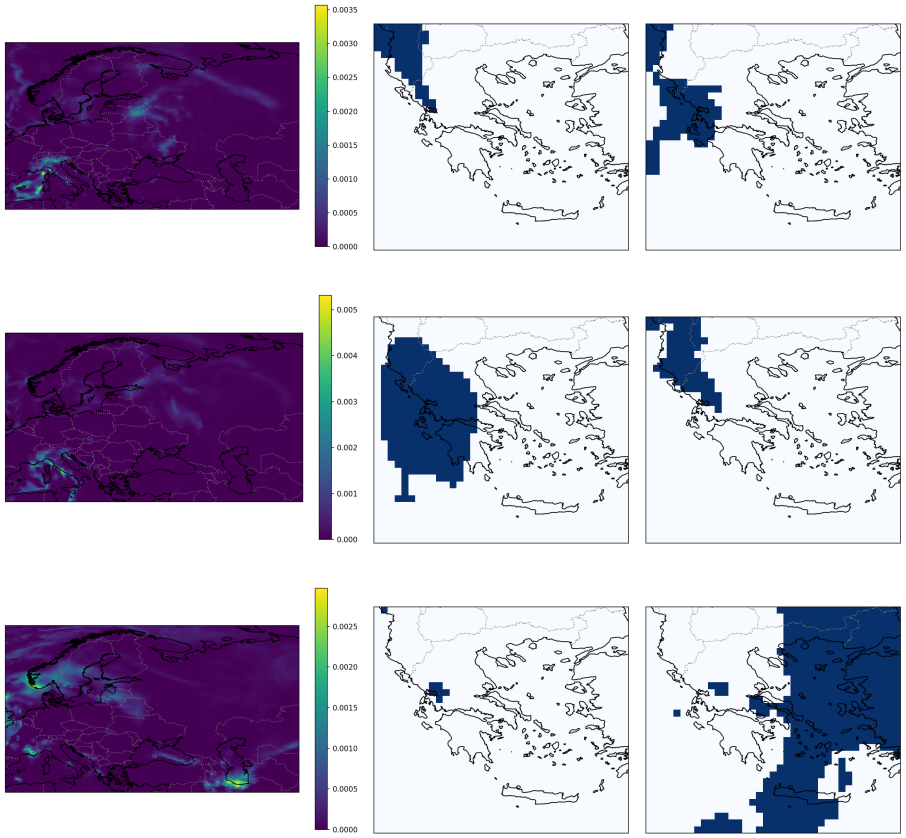

Figure 6: Left to right: input, predicted, and ground truth images for tp. Start to bottom: 2 days, 4 days, 6 days forecast.

framework and the critical importance of leveraging physically meaningful indices like the K-index, vimd, which are not that common in deep learning forecastings. By identifying and focusing on the variables that carry the strongest signal—rather than relying on ever-larger input pools—future work can build more interpretable, efficient, and ultimately more reliable predictors of extreme precipitation.

8. Future work

Building on the insights gained in this study, we plan to pursue three main directions. First, we will extend our predictor set by incorporating key pressure-level variables—such as geopotential height, humidity, and wind fields at multiple atmospheric levels—to evaluate their added value for extreme-precipitation classification. Second, we will systematically explore a wider array of variable combinations and interaction effects, using our fixed-architecture framework to pinpoint the most informative subsets. Finally, we aim to compare our data-driven extreme-event classifier with outputs from numerical weather prediction models, testing its skills.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT to: Grammar and spelling check. After using the service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

Acknowledgements

This work has received funding from the European Union's Digital Europe Programme (DIGITAL) under grant agreement No 101146490.

References

[1] F. Khan, Y.-A. Liou, G. Spöck, X. Wang, S. Ali, Assessing the impacts of temperature extremes on agriculture yield and projecting future extremes using machine learning and deep learning approaches with cmip6 data, International Journal of Applied Earth Observation and Geoinformation 132 (2024) 104071. URL: https://www.sciencedirect.com/science/article/pii/S1569843224004254. doi:https://doi.org/10.1016/j.jag.2024.104071.

[2] Z. Ben Bouallègue, M. C. A. Clare, L. Magnusson, E. Gascón, M. Maier-Gerber, M. Janoušek, M. Rodwell, F. Pinault, J. S. Dramsch, S. T. K. Lang, B. Raoult, F. Rabier, M. Chevallier, I. Sandu, P. Dueben, M. Chantry, F. Pappenberger, The rise of data-driven weather forecasting: A first statistical assessment of machine learning–based weather forecasts in an operational-like context, Bulletin of the American Meteorological Society 105 (2024) E864–E883. URL: http://dx.doi.org/10.1175/BAMS-D-23-0162.1. doi:10.1175/bams-d-23-0162.1.

[3] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, et al., Learning skillful medium-range global weather forecasting, Science 382 (2023) 1416–1421.

[4] I. Price, A. Sanchez-Gonzalez, F. Alet, T. R. Andersson, A. El-Kadi, D. Masters, T. Ewalds, J. Stott, S. Mohamed, P. Battaglia, R. Lam, M. Willson, Gencast: Diffusion-based ensemble forecasting for medium-range weather, arXiv preprint arXiv:2312.15796 (2023).

[5] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, et al., Learning skillful medium-range global weather forecasting, Science 382 (2023) 1416–1421.

[6] K. Trebing, T. Stanczyk, S. Mehrkanoon, Smaat-unet: Precipitation nowcasting using a small attention-unet architecture, 2021. URL: https://arxiv.org/abs/2007.04417. arXiv:2007.04417.

[7] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. kin Wong, W. chun Woo, Convolutional lstm network: A machine learning approach for precipitation nowcasting, 2015. URL: https://arxiv.org/abs/1506.04214. arXiv:1506.04214.

[8] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, Q. Tian, Pangu-weather: A 3d high-resolution model for fast and accurate global weather forecast, 2022. URL: https://arxiv.org/abs/2211.02556. arXiv:2211.02556.

[9] D. S. Civitarese, D. Szwarcman, B. Zadrozny, C. Watson, Extreme precipitation seasonal forecast using a transformer neural network, 2021. URL: https://arxiv.org/abs/2107.06846. arXiv:2107.06846.

[10] X. Lin, J. Fan, Z. J. Hou, J. Wang, Machine learning of key variables impacting extreme precipitation in various regions of the contiguous united states, Journal of Advances in Modeling Earth Systems 15 (2023) e2022MS003334.

[11] N. Mastrantonas, L. Magnusson, F. Pappenberger, J. Matschullat, What do large-scale patterns teach us about extreme precipitation over the mediterranean at medium-and extended-range forecasts?, Quarterly Journal of the Royal Meteorological Society 148 (2022) 875–890.

[12] X. Lin, J. Fan, Z. J. Hou, J. Wang, Machine learning of key variables impacting extreme precipitation in various regions of the contiguous united states, Journal of Advances in Modeling Earth Systems 15 (2023) e2022MS003334.

[13] J. Meyer, M. Neuper, L. Mathias, E. Zehe, L. Pfister, Atmospheric conditions favouring extreme precipitation and flash floods in temperate regions of europe, Hydrology and Earth System Sciences 26 (2022) 6163–6183.

[14] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey,

R. Radu, I. Rozum, et al., Era5 hourly data on single levels from 1940 to present, copernicus climate change service (c3s) climate data store (cds)[data set], 2023.

[15] E.-M. Walz, P. Knippertz, A. H. Fink, G. Köhler, T. Gneiting, Physics-based vs data-driven 24-hour probabilistic forecasts of precipitation for northern tropical africa, Monthly Weather Review 152 (2024) 2011–2031.

[16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, 2018. URL: https://arxiv.org/abs/1708.02002. arXiv:1708.02002.