Hydrol. Earth Syst. Sci., 30, 1333–1358, 2026
https://doi.org/10.5194/hess-30-1333-2026

Hydrology and Earth System Sciences
Open Access

Climate adaptation-aware flood prediction for coastal cities using Deep Learning

Bilal Hassan, Areg Karapetyan, Aaron Chung Hin Chow, and Samer Madanat

Division of Engineering, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Correspondence: Bilal Hassan (bilal.hassan@nyu.edu)

Received: 23 February 2025 – Discussion started: 27 March 2025
Revised: 18 December 2025 – Accepted: 12 January 2026 – Published: 11 March 2026

Abstract. Climate change and sea-level rise (SLR) pose escalating threats to coastal cities, intensifying the need for efficient and accurate methods to predict potential flood hazards. Traditional physics-based hydrodynamic simulators, although precise, are computationally expensive and impractical for city-scale coastal planning applications. Deep Learning (DL) techniques offer promising alternatives, however, they are often constrained by challenges such as data scarcity and high-dimensional output requirements. Leveraging a recently proposed vision-based, low-resource DL framework, we develop a novel, lightweight Convolutional Neural Network (CNN)-based model designed to predict coastal flooding under variable SLR projections and shoreline adaptation scenarios. Furthermore, we demonstrate the ability of the model to generalize across diverse geographical contexts by utilizing datasets from two distinct regions: Abu Dhabi and San Francisco. Our findings demonstrate that the proposed model significantly outperforms state-of-the-art methods, reducing the mean absolute error (MAE) in predicted flood depth maps on average by nearly 20 %. These results highlight the potential of our approach to serve as a scalable and practical tool for coastal flood management, empowering decision-makers to develop effective mitigation strategies in response to the growing impacts of climate change. Project Page: https://caspiannet.github.io/ (last access: 22 January 2026).

1 Introduction

Of the world's 34 megacities (i.e., those with more than 10 million inhabitants), approximately 70 %[1] are situated on or near the coast. Coastal cities, including megacities, host nearly 10 % of the world's population, are 2.6 times denser populated as compared to inland areas, and are powerhouses of global trade and business activities (the legacy of maritime trade) (Pal et al., 2023). Yet, these cities, especially those in low-elevation coastal zones, are hotspots for climate-induced disasters. Notable examples range from Venice, Italy, and Miami, Florida, to Manila, Philippines (van de Wal et al., 2024; Griggs and Reguero, 2021). In fact, according to a 2021 article by UN-Habitat[2], 90 % of megacities are vulnerable to sea level rise (SLR) and, as analyzed in Hallegatte et al. (2013), the flood risk to coastal cities is expected to rise nine-fold by 2050. The problem is compounded by land subsidence, confronting coastal communities with the challenge of managing multiple, interacting sources of risk (Cao et al., 2021; Ardha et al., 2024; Barnard et al., 2024).

To address the threat of coastal flooding, planners typically consider a portfolio of adaptation strategies, which are broadly classified as measures to protect the shoreline, accommodate rising waters, retreat from vulnerable areas, or avoid new development in hazard zones (Chang et al., 2020; Oppenheimer et al., 2022). This study focuses on *protect* strategies, which involve armoring shorelines with en-

[1]Calculated from the data reported in United Nations' World Cities Report 2024 (https://digitallibrary.un.org/record/4065171?v=pdf, last access: 10 December 2026).

[2]UN-Habitat Press Release (https://unhabitat.org/news/18-nov-2021/busan-un-habitat-and-oceanix-set-to-build-the-worlds-first-sustainable-floating)

Published by Copernicus Publications on behalf of the European Geosciences Union.

gineered structures such as seawalls, levees, and storm barriers (Beagle et al., 2019; Papacharalambous et al., 2013). Rather than evaluating the performance of specific engineering designs, our work addresses the more fundamental strategic question of the optimal spatial configuration of these defenses, that is, determining which shoreline segments are most critical to protect. A recent example of a large-scale protection effort involves New York City, where extensive flood risk modeling led to the fortification of a two-and-a-half mile stretch of Lower Manhattan's shoreline, an intervention largely driven by the aftermath of Hurricane Sandy in 2012 (Lewis, 2023). Construction of these coastal defense structures, however, significantly alters the shoreline geometry, subsequently influencing the local hydrodynamics and potentially creating regional impacts (Hummel et al., 2021; Wang et al., 2018a; Haigh et al., 2020).

Therefore, for effective and responsible flood protection planning, it is essential to account for plausible hydrodynamic changes due to shoreline modifications. To this end, physics-based high-fidelity simulators, such as Delft3D (Deltares, 2025), can be employed to resolve the detailed hydrodynamics. While these tools are accurate, they are computationally expensive, often requiring days to simulate a single shoreline protection scenario (Kyprioti et al., 2021; Rohmer et al., 2023; Karapetyan et al., 2026). This computational burden is not only a barrier for real time, short term forecasting, but it is even more prohibitive for long term strategic planning. Such planning requires exploring a vast combinatorial design space with thousands of possible shoreline adaptation configurations across multiple SLR scenarios, and this level of exploration is computationally infeasible with traditional simulators. Furthermore, in complex urban terrain, flood dynamics are critically dependent on fine-scale features, necessitating high-resolution modeling to accurately capture localized effects on critical infrastructure (Hartnett and Nash, 2017; Wang et al., 2018b; Karapetyan et al., 2026).

In response to these limitations, data-driven methods, including machine learning (ML) and deep learning (DL) techniques, have emerged as promising alternatives for rapid flood prediction (Bentivoglio et al., 2022; Mosavi et al., 2018; Muñoz et al., 2024; Zuhairi et al., 2022; Zhou et al., 2023; Nevo et al., 2022; Zhao et al., 2021). Numerous studies have employed traditional ML methods like random forests and support vector machines (Mosavi et al., 2018; Ali et al., 2022), and more recently, hybrid models that combine ML with hydrodynamic simulations have gained favor (Chen et al., 2024; Du et al., 2024). Compared to ML, DL algorithms have shown enhanced capabilities. While one-dimensional (1D) models like long short-term memory (LSTM) are effective for sequential forecasting, two-dimensional (2D) methods such as Convolutional Neural Networks (CNNs) are particularly well-suited for capturing the complex spatial patterns inherent in flood maps. These surrogate models aim to emulate high-fidelity simulators by learning complex input-output relationships without explicitly modeling the underlying physical processes (Karapetyan et al., 2026).

However, despite these advancements, a significant challenge remains. Many existing flood prediction studies focus on singular triggers or short-term extreme events and, consequently, do not jointly consider the complex, long-term impacts of both SLR and dynamic shoreline adaptation strategies (Jia et al., 2016; Guo et al., 2021). Fulfilling the high-resolution modeling requirement with DL-based models also presents its own complications, chief among them being *data scarcity* and the challenge of handling the *high dimensionality of the output* (predicting an inundation value for every pixel in a large spatial grid) (Kyprioti et al., 2021; Rohmer et al., 2023; Karapetyan et al., 2026). Previously, a 2D DL framework was introduced to address these challenges by recasting flood prediction as a computer vision task (Karapetyan et al., 2026). The core of that approach was to transform discrete shoreline protection scenarios (a list of protected or unprotected segments) into 2D spatial input maps. By treating the problem as an image-to-image translation task, that framework allowed CNNs to inherently learn the geometric relationships between protected areas and the resulting flood patterns. Crucially, that image-based format also enabled the use of random cutouts data augmentation techniques to artificially expand the limited training dataset, a critical advantage in data-scarce domains. While this foundational work demonstrated the viability of the approach, their method was limited to a single location and a particular SLR scenario. Taking a step further, this work introduces a novel DL model designed to generalize across two distinct coastal regions and multiple SLR scenarios. More concretely, the key contributions of this study are as follows:

1. We propose a novel DL model (CASPIAN-v2) designed to accurately predict high-resolution coastal flooding under various SLR scenarios and shoreline adaptation strategies. The architecture is developed as a lightweight CNN for fast and scalable prediction, aiming to significantly reduce computational time compared to traditional high-fidelity hydrodynamic models while maintaining high accuracy.

2. We present two new, comprehensive datasets from vulnerable coastal cities, Abu Dhabi (AD) and San Francisco (SF). These datasets cover different sea-level rise scenarios and shoreline adaptations to facilitate future research in this domain.

3. We conduct a rigorous evaluation of the proposed framework against state-of-the-art (SOTA) ML and DL models to benchmark its performance and test its generalization capabilities across diverse scenarios.

4. We employ explainable artificial intelligence (AI) techniques to validate the outputs of the model, assess the

physical plausibility of its predictions, and offer interpretability to support decision-making in flood risk assessment.

Put together, these contributions can assist urban policymakers in designing more effective and reliable coastal protection programs. Additionally, we open-source the code and datasets in the hope of facilitating further research and attracting greater attention to this problem within the machine learning community.

2 Study Area and Data Description

In this research, we examine two vulnerable metropolitan coastal areas (Abu Dhabi and San Francisco Bay) to predict coastal flooding under various SLR and shoreline protection. Both locations feature low-lying topographies and significant urbanization, making them particularly susceptible not only to direct flooding, but also impacts on transportation links, and specifically whether important arterials such as shoreline highways or freeways will be flooded due to SLR. Our aim is to evaluate the effectiveness and applicability of the DL-based solution for forecasting inundation in these regions.

2.1 Study Area Description

2.1.1 Abu Dhabi

Abu Dhabi, located along the southern coast of the Arabian (Persian) Gulf, faces rising flood risks from climate change-induced SLR, tidal flooding, and storm surges driven by extreme winds such as Shamal (Langodan et al., 2023). Projections estimate that a 0.5 m SLR, expected by 2050–2100 based on IPCC AR6 (IPCC, 2021), could inundate critical ecosystems like mangroves and artificial islands, potentially doubling flood zones when accounting for wind and wave action (Melville-Rea et al., 2021). The shallow bathymetry of the region amplifies these risks, with even minor sea-level increases threatening key infrastructure and densely populated areas, where over 85 % of the population and 90 % of the infrastructure lie just meters above sea level (al Kabban, Marwa, 2019; Melville-Rea et al., 2021).

To assess flood risks, we divided the AD urban coastline into 17 operational landscape units (OLU), based on the Abu Dhabi Urban Structure Framework Plan 2030 (Abu Dhabi Urban Planning Council, 2007), and adopted in previous studies (e.g., Chow and Sun, 2022). In the hydrodynamic model, the protection of a single OLU involves placing an impermeable seawall (that assumes no overtopping) along the coastal boundary of the OLU.

This framework captures unique features of both natural ecosystems and urban zones, enabling detailed flood vulnerability analyses under various shoreline adaptations. Figure 1 illustrates the AD coastline, the OLU divisions, and the inundation points for the 0.5 m SLR scenario.

2.1.2 San Francisco Bay Area

Our second study area is the urban shoreline located along the banks of San Francisco Bay (Fig. 2). Owing to the location of San Francisco Bay as an inland bay, its shoreline communities are relatively sheltered from storm surges by the exterior Californian coastline, with mean significant wave heights within the Bay at about 0.07–0.2 m, in contrast to 2.0–3.0 m at Point Reyes located on the California coast outside the Bay (United States Geological Survey, 2024).

San Francisco Bay faces significant flood risks from SLR and tidal variability, which are exacerbated by climate change (California Energy Commission, 2018; Wang et al., 2018a), which in turn impacts low-lying urban zones, transportation networks, and hydrological systems (such as the Napa River Basin). However, our focus in this study is on tidal flooding within San Francisco Bay in order to highlight the unique tidal behavior within the Bay where the construction of sea walls along certain portions of the shoreline Bay may, in fact, exacerbate the sea level within the Bay to increase by up to 1 m (Holleman and Stacey, 2014).

For San Francisco Bay, the discretization of coastline of the Bay Area into 30 OLUs was based on shoreline morphology, hydrology, and urban infrastructure, originally performed by (Beagle et al., 2019), and used in previous studies (Hummel et al., 2021; Sun et al., 2020). Figure 2 illustrates the 30 OLUS for SF Bay Area, the OLU divisions, and the inundation points for the 0.5 m SLR scenario. In the hydrodynamic model, the protection of a single OLU involves placing an impermeable seawall (that assumes no overtopping) along the coastal boundary of the OLU.

2.2 Data Sources and Hydrodynamic Simulations

The ground truth flood data used for training and evaluating our surrogate model was generated through a series of physics-based hydrodynamic simulations using the Delft3D model. This model integrates key physical processes including SLR and tidal dynamics (see Supplement, Sect. S1). High-resolution bathymetry and digital elevation models (DEM) (with data sources such as TanDEM-X, Landsat-8, and Nautical Charts) were used for both regions to ensure accurate modeling of coastal topography that transitions smoothly between sea and the land. While some authors (De Almeida and Bates, 2013; Neal et al., 2012; Li and Hodges, 2019; Sanders and Schubert, 2019; Nithila Devi and Kuiry, 2024) address subgrid details by using separate subgrid nesting methods, we have retained the same governing equations but used a 30 m model grid in the areas of interest, and Delft3D is capable of automatically modeling wetting and drying of grid cells from one time step to the next.

The accuracy and reliability of these physics-based models were established through rigorous validation against real-world observations. For San Francisco Bay, the Delft3D model was adapted from the CoSMoS model originally de-

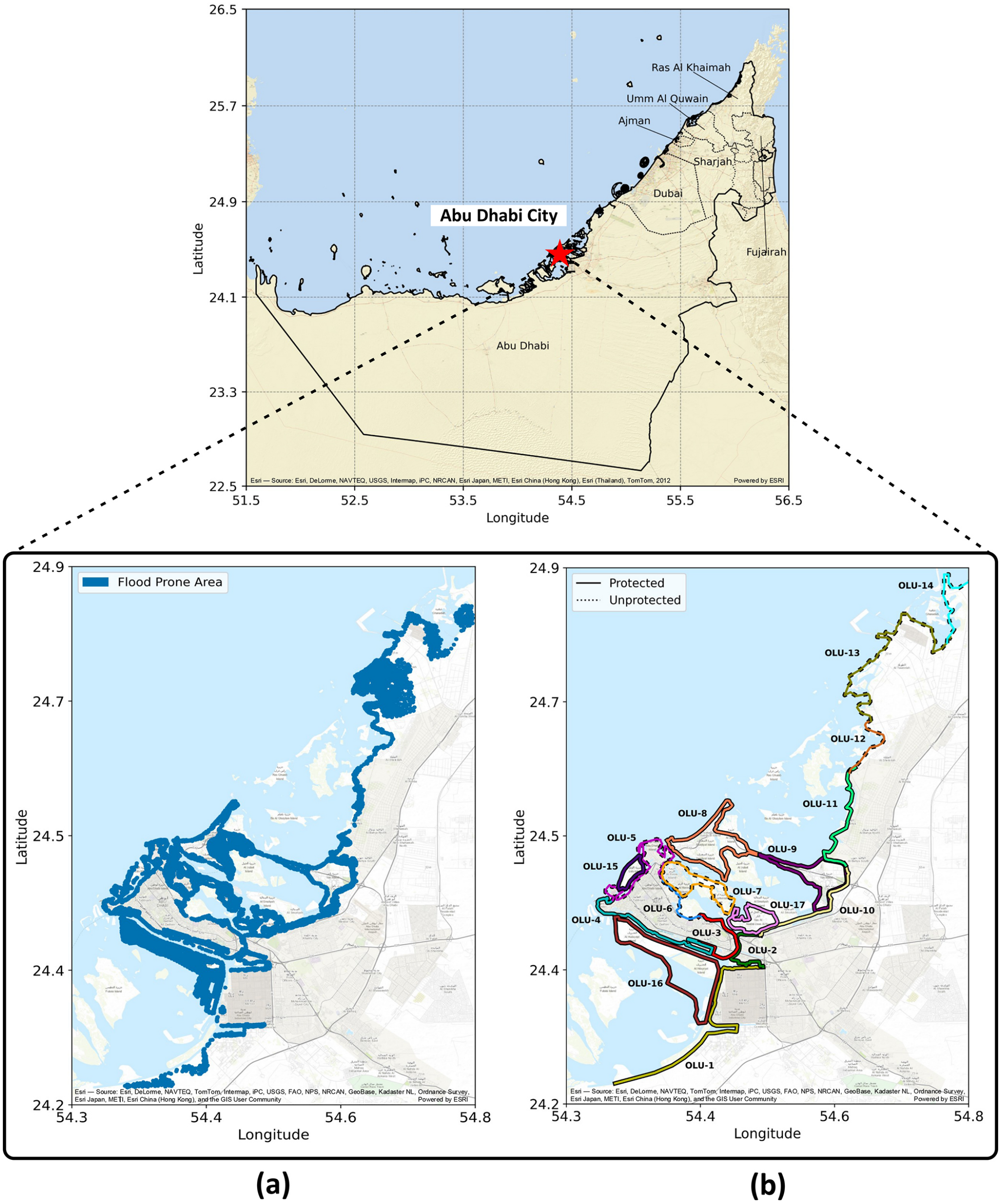

Figure 1. AD study area shown on the map of the United Arab Emirates. **(a)** All areas susceptible to flooding under a 0.5 m SLR scenario without any shoreline protections **(b)** The 17 OLUs defined along the AD shoreline where protections are to be tested for their effectiveness.

veloped by Barnard et al. (2014) and adapted to San Francisco Bay by Wang et al. (2017), and validated in the past using tidal gages at 9 tidal gage locations in and around San Francisco Bay. Pearson correlation coefficients ranged from 0.9862 to 0.9996, while the root mean square (RMS) ratios (the ratio of modeled versus measured RMS amplitudes) ranged from 0.973 to 1.027 (please refer to Wang et al., 2017).

For Abu Dhabi, the Delft3D model was validated using water level data from 196 tidal gage locations throughout the Gulf (as the hydrodynamic model encompassed the entire Gulf in addition to the western portions of the Gulf of Oman). The water levels at these locations were compared with one month's worth of hydrodynamic simulation, and the resulting absolute root mean square error (RMSE) values ranged from 0.0013 to 0.0043 m in the vicinity of Abu Dhabi. More validation details for Abu Dhabi can be found in Chow and Sun (2022). Given this strong validation, the outputs of the hydrodynamic simulations were considered a reliable proxy

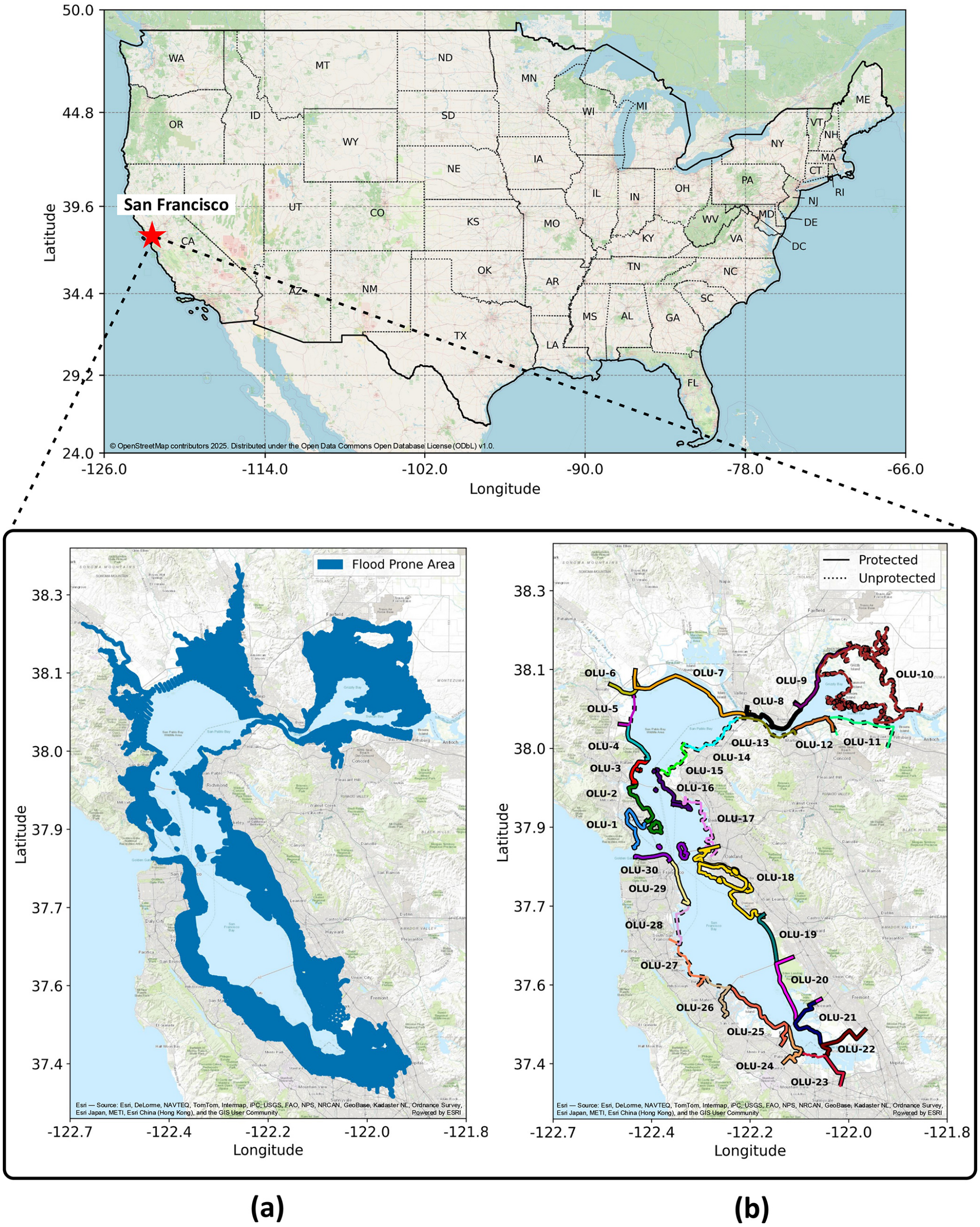

Figure 2. SF Bay study area shown on the map of the United States **(a)** All areas susceptible to flooding under a 0.5 m SLR scenario without any shoreline protections **(b)** The 30 OLUs defined along the SF Bay shoreline where protections are tested for their effectiveness.

for ground truth for the purposes of training and evaluating our deep learning framework.

While the Gulf does not typically experience tropical cyclones, it is known for its northwesterly winds generally occurring with winds at about $20\,\mathrm{m\,s^{-1}}$ with sudden onset and sustained over a period of up to 3–5 d. These are called the Shamal winds (meaning "North" in Arabic) and occur at least 10 times annually, mainly during the winter months (Al Senafi and Anis, 2015; Li et al., 2020). Accordingly, for Abu Dhabi, we applied a nested SWAN model to simulate wind and wave effects, particularly the impact of these Shamal winds, which can significantly intensify tidal flooding risks. Both the SWAN model and Delft3D models were forced using ERA5 meteorological data in the Gulf.

In both geographic locations, our aim was to generate data that correspond to a hypothetical future extreme flooding scenario, where there was little to no flooding observed without SLR. For AD, simulations were based on a 0.5 m SLR scenario, consistent with regional projections for mid-century SLR (as described above) (IPCC, 2021). The 0.5 m SLR scenario was then coupled with storm surges resulting from a sample 3-month long Shamal event. In contrast, flood simulations for the SF Bay Area were conducted under three SLR scenarios: 0.5, 1.0, and 1.5 m, which reflects a possible future scenario for San Francisco Bay in the year (somewhere between 2050-2100 depending on the climate change scenario pathway (between SSP2-4.5 and SSP5-8.5) from IPCC AR6 report (IPCC, 2021). Table 1 provides a comprehensive overview of the datasets generated for this study, which are partitioned into three categories based on their purpose. The *Main Set*, comprising the largest datasets from AD (0.5 m SLR) and SF (1.0 m SLR), was used for the primary training, validation, and testing of the CASPIAN-v2 model. The *Holdout Set* consists of scenarios intentionally curated to be challenging (such as protecting one entire side of the SF Bay while leaving the other exposed) and was used for blind testing of the primarily trained model's performance on complex spatial schemes not seen during training (see Sect. S4). Finally, the *Generalizability Set* includes SF scenarios at different SLR levels (0.5 and 1.5 m) and was used exclusively to evaluate the ability of the model to adapt to new environmental conditions via fine-tuning.

To balance the need to model a larger number of modeled tidal cycles per simulation, with the computational time and storage space used for the simulations, a 3-month simulation period was also applied for San Francisco Bay. Although our San Francisco model includes riverine input from the Sacramento and San Joaquin Rivers, the inflow rates into the Bay were baseline values rather than for extreme fluvial flood events. While we acknowledge that incorporating more hydrodynamic forcing conditions to include pluvial and riverine floods, as well as extreme storm events, can refine the hydrodynamic model to reflect more extreme flooding, our overall scope in this paper is in the use of machine learning to be able to act as a surrogate for a hydrodynamic model running under different SLR scenarios. The detailed protocols for how these datasets were split and used are described in Sect. 4.1.

We ran individual Delft3D scenarios (each with a 3-month simulation time as described above) to collect hourly inland inundation data under different coastal protection scenarios to create a dataset for training and validating our DL model. Our findings highlight the importance of holistic regional flood control measures, especially given the intricate interplay between protected and unprotected zones. Further, the datasets from two regions allowed us to assess the applicability and reliability of the DL model in different vulnerable coastal settings.

2.3 Data Preprocessing

The raw, tabular data generated by the Delft3D simulator, which consists of inundation coordinates and corresponding peak water level (PWL) values, is not directly compatible with our 2D DL model. Therefore, a multi-step preprocessing pipeline was developed to transform this data into a structured grid format suitable for a computer vision task.

The first key step was to map the inundation coordinates onto a standardized 1024×1024 spatial grid. This was achieved by defining the grid boundaries based on the maximum spatial extent of all simulation data and then assigning each inundation point to its nearest grid cell. In cases where multiple inundation points mapped to the same cell due to the high density of the data, a conflict resolution strategy was employed that reassigned the conflicting points to the nearest available empty cell, ensuring a unique one-to-one mapping.

Subsequently, we incorporated the shoreline protection information. For each inundation point, we calculated its proximity to the nearest protected and unprotected OLUs and assigned it a class based on which was closer. This classification, along with the PWL values, was then used to construct the final input and output matrices for training. The shoreline protection scenarios were encoded as binary strings, where "0" indicates unprotected OLUs and "1" denotes protected OLUs. This entire process ensures that the model receives spatially coherent input that encodes not just water levels, but also the crucial context of shoreline defense configurations. A full, detailed breakdown of each step, including the mathematical formulations for grid mapping and OLU classification, is provided in the Supplement (Sect. S2).

3 Method

This section details the proposed deep learning framework for predicting coastal inundation under various SLR depths and shoreline protection scenarios. We first provide a high-level overview of the end-to-end workflow, from data generation to prediction, and then present the specific architecture of the CASPIAN-v2 model and the novel hybrid loss function used for its training.

3.1 Proposed Framework

The proposed framework, illustrated in Fig. 3, provides an end-to-end pipeline for generating, processing, and predicting coastal flood data. The process is organized into several key stages, each represented by a colored path in the diagram.

- *Data Generation and Preprocessing.* The process begins with running physics-based hydrodynamic simulations (e.g., Delft3D) using different shoreline protection scenarios and SLR levels as inputs. This generates raw, tabular flood data containing water levels at specific coordinates. This raw data is then put through a prepro-

Table 1. Dataset details for AD and SF regions, including OLUs, SLR depths, and the number of unique shoreline protection scenarios. The Main Set was used for primary model training and testing. The Holdout Set was used for blind testing on challenging scenarios. The Generalizability Set was used to evaluate model adaptability to new SLR conditions via fine-tuning.

Region	OLUs	SLR	Protection Scenarios
AD	17 OLUs: 1 *(Mussafah)*, 2 *(Bain Al Jesrain)*, 3 *(Grand Mosque District)*, 4 *(AD Island West)*, 5 *(Marina, CBD, Al Mina)*, 6 *(AD Island East)*, 7 *(Al Reem Island)*, 8 *(Saadiyat Island)*, 9 *(Yas Island)*, 10 *(Al Raha Island)*, 11 *(Al Shahama)*, 12 *(Al Rahba)*, 13 *(New Port City)*, 14 *(Ghantoot)*, 15 *(Lulu Island)*, 16 *(Hudayriat Island)*, 17 *(Inner Islands)*	0.5 m	142 (Main Set) 32 (Holdout Set)
SF	30 OLUs: 1 *(Richardson)*, 2 *(Corte Madera)*, 3 *(San Rafael)*, 4 *(Gallinas)*, 5 *(Novato)*, 6 *(Petaluma)*, 7 *(Napa – Sonoma)*, 8 *(Carquinez North)*, 9 *(Suisun Slough)*, 10 *(Montezuma Slough)*, 11 *(Bay Point)*, 12 *(Walnut)*, 13 *(Carquinez South)*, 14 *(Pinole)*, 15 *(Wildcat)*, 16 *(Point Richmond)*, 17 *(East Bay Crescent)*, 18 *(San Leandro)*, 19 *(San Lorenzo)*, 20 *(Alameda Creek)*, 21 *(Mowry)*, 22 *(Santa Clara Valley)*, 23 *(Stevens)*, 24 *(San Francisquito)*, 25 *(Belmont – Redwood)*, 26 *(San Mateo)*, 27 *(Colma – San Bruno)*, 28 *(Yosemite – Visitacion)*, 29 *(Mission – Islais)*, 30 *(Golden Gate)*	1.0 m 0.5 m 1.5 m	285 (Main Set) 46 (Holdout Set) 32 (Generalizability Set) 32 (Generalizability Set)

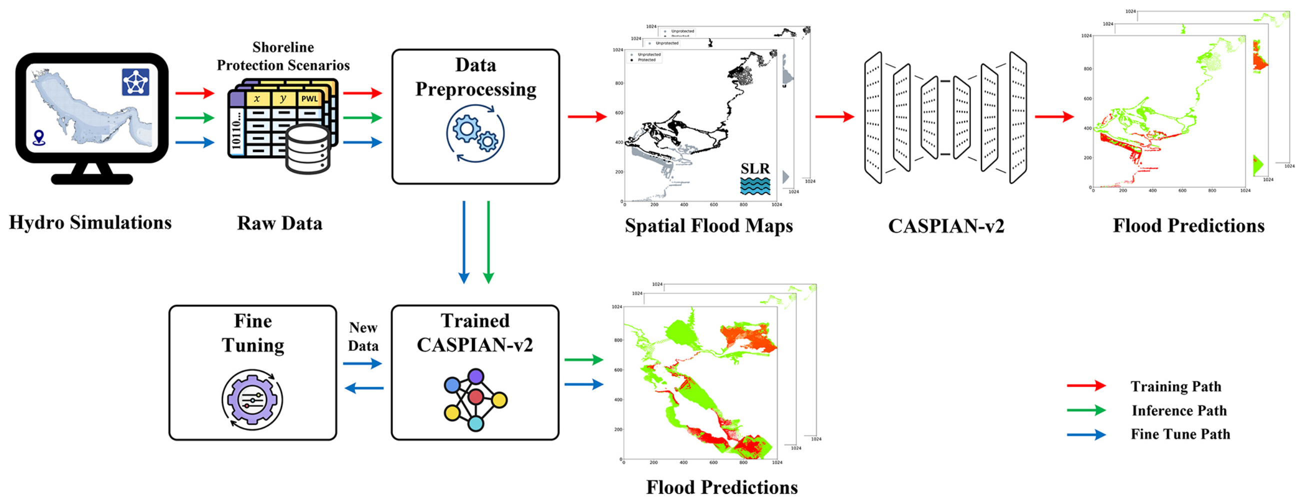

Figure 3. An overview of the proposed framework for coastal flood prediction. It begins with hydrodynamic simulations based on SLR data and coastal protection scenarios to generate raw flood data, which is then processed into spatial flood maps. The CASPIAN-v2 model, trained on these maps, predicts inundation patterns and flood extent. The framework can be fine-tuned with new data for improved adaptability. The different colored paths represent training (red), inference (green), and fine-tuning (blue) stages.

cessing pipeline, where it is transformed into 2D spatial flood maps suitable for a computer vision approach.

- *Training Path (Red).* The preprocessed spatial maps serve as the input-output pairs for training the CASPIAN-v2 model. The model learns the complex, non-linear relationships between the shoreline protection configurations (input) and the resulting flood inundation patterns (output).
- *Inference Path (Green).* Once trained, the model can be used for rapid inference. Given a new, unseen shoreline protection scenario, the model can predict the corresponding high-resolution flood map in a matter of seconds, bypassing the need for computationally expensive hydrodynamic simulations.
- *Fine-Tuning Path (Blue).* To enhance adaptability, the trained CASPIAN-v2 model can be fine-tuned on new data. This is particularly useful for adapting the model to different SLR scenarios or geographical regions for which only limited data might be available, allowing for efficient knowledge transfer without retraining from scratch.

This integrated framework provides a scalable and efficient solution for assessing the impact of diverse coastal adaptation strategies under the threat of climate change.

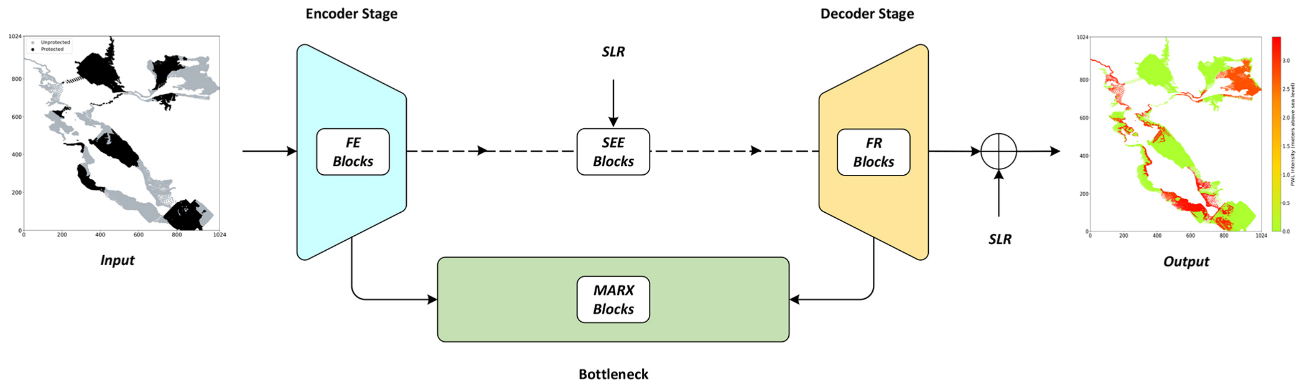

Figure 4. A simplified schematic of the CASPIAN-v2 model architecture, highlighting its key functional components. The model consists of an encoder stage containing feature extraction (FE) blocks that progressively downsample the input map, which feeds into a series of multi-attention ResNeXt (MARX) blocks in the bottleneck. These novel blocks use an attention mechanism to refine the compressed features, focusing on the most critical spatial information. The decoder stage then reconstructs the output using feature reconstruction (FR) blocks. A key innovation is the integration of the scalar SLR value via SLR-enhanced encoding (SEE) blocks, which modulate the features during this reconstruction process. The final output is produced after another integration of the SLR value, allowing the model to generate flood predictions conditioned on different climate scenarios. A detailed, layer-by-layer version of this architecture is provided in the Supplement (Fig. S5).

3.2 CASPIAN-v2 Architecture

The *CASPIAN-v2* model improves and extends the capabilities of the previously developed CNN architecture to predict coastal flooding (Karapetyan et al., 2026). Unlike the previous version, CASPIAN-v2 integrates SLR data and has a more robust yet minimalistic architecture that generalizes across various geographical regions. Figure 4 illustrates the CASPIAN-v2 architecture, which consists of three main stages: encoder, bottleneck, and decoder. The following subsections provide a conceptual overview of the architecture, whereas a detailed exposition of all network layers and operations is presented in the Supplement.

3.2.1 Encoder Stage

The encoder consists of a sequence of convolutional *feature extraction* (FE) blocks that progressively reduce the spatial resolution of the input grid while increasing the depth of the feature maps. This hierarchical feature extraction allows the model to capture multi-scale patterns essential for accurate flood inundation prediction. Each FE block uses depthwise separable convolutions and pooling to condense the input feature maps, followed by pointwise convolutions that expand the feature depth. Moreover, residual skip connections are incorporated to preserve important spatial information and mitigate gradient vanishing, ensuring that critical low-level features are not lost. By the end of the encoder stage, the input grid is transformed into a concise feature representation, encapsulating both localized details, such as inundation patterns in specific regions, as well as a broader spatial context. It should be noted that the scalar SLR input is not passed through the encoder; instead, it is directly incorporated in the decoder stage to globally influence the reconstruction of flood patterns.

3.2.2 Bottleneck Stage

CASPIAN-v2 employs a novel multi-attention ResNeXt (MARX) block at the bottleneck (the deepest part of the network with the smallest spatial dimensions) to refine and enrich the encoded features. The MARX block incorporates ResNeXt blocks (Xie et al., 2017), an aggregated residual structure, alongside the convolutional block attention module (CBAM) (Woo et al., 2018), facilitating the model in concentrating on key features. Specifically, the encoded feature map is first processed by a residual block, then passed through a attention module which sequentially applies channel attention and spatial attention to reweight the feature map, and finally routed through a second residual block. This combination adaptively emphasizes critical features in both the channel and spatial dimensions, thereby enhancing the ability of the model to learn complex flood patterns under various scenarios. The output of the MARX block is a rich high-level representation of the input scenario that serves as input to the decoder.

3.2.3 Decoder Stage

The decoder stage progressively reconstructs the high-resolution flood inundation map through a sequence of *Feature Reconstruction (FR)* blocks. Each block upsamples features using a transpose convolution before fusing them with the corresponding encoder output. This fusion via skip connections is crucial, as it serves to reintroduce fine-scale spatial details that were compressed during encoding. A key enhancement in CASPIAN-v2 is the incorporation of the SLR input into the decoder through a specialized *SLR-Enhanced Encoding (SEE)* block. The SEE mechanism uses the scalar SLR value to modulate decoder features, effectively guiding the upsampling process with global sea-level context. In practice, the SEE block learns a set of weighting coefficients from the encoder's pooled features and the SLR value, which are then applied to the decoder feature maps at each scale. Consequently, regions more susceptible to flooding under a given SLR scenario receive higher weights during reconstruction. After the final upsampling, a convolutional layer produces an initial output grid, which is further refined by adding back the SLR-weighted summed features from the last decoder layer before applying the final activation function. The resulting output is the predicted flood inundation map, where each cell reflects the likelihood or extent of flooding at that location given the input conditions.

3.3 Loss Function

Predicting PWL under different SLR scenarios is challenging due to outliers and the need to balance error sensitivity across multiple regions. To tackle these issues, we introduce a hybrid loss function that combines Huber (Huber, 1992), Log-Cosh (Saleh and Saleh, 2022), and Quantile (Koenker and Bassett, 1978) losses in a weighted setup. The Huber loss L_h aims to robustly minimize small prediction errors while limiting the impact of outliers, and it uses a threshold δ to manage the sensitivity of the error. The L_h for each sample i is computed as expressed in Eq. (1):

$$L_{\mathrm{h},i} = \begin{cases} \frac{1}{2}(y_{\mathrm{p},i} - y_{\mathrm{t},i})^2 & \text{if } |y_{\mathrm{p},i} - y_{\mathrm{t},i}| \leq \delta, \\ \delta \cdot |y_{\mathrm{p},i} - y_{\mathrm{t},i}| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases} \tag{1}$$

where $y_{\mathrm{t},i}$ and $y_{\mathrm{p},i}$ represent the actual and estimated PWL values. We set δ within the range of 0.3 and 0.7, which is dynamically determined to balance sensitivity and robustness. Moreover, we integrate Log-Cosh loss (L_cosh) to smooth gradients in regions with large variations, helping to maintain prediction stability in different areas affected by SLR. The L_cosh is expressed as in Eq. (2):

$$L_{\mathrm{cosh},i} = \log\left(\cosh(y_{\mathrm{p},i} - y_{\mathrm{t},i})\right), \tag{2}$$

In addition, the quantile loss L_q differentiates errors by assigning distinct penalties to underestimation and overestimation, dictated by a quantile parameter $\tau = 0.75$. This loss dynamically adjusts to minimize quantile-specific errors, calculated as in Eq. (3):

$$L_{\mathrm{q},i} = \begin{cases} \tau \cdot (y_{\mathrm{p},i} - y_{\mathrm{t},i}) & \text{if } y_{\mathrm{p},i} \geq y_{\mathrm{t},i}, \\ (1-\tau) \cdot (y_{\mathrm{t},i} - y_{\mathrm{p},i}) & \text{otherwise} \end{cases} \tag{3}$$

To achieve an optimal balance, we linearly combine the three loss components into a comprehensive hybrid loss function L_total, weighted by empirically tuned coefficients $\alpha_\mathrm{h}, \alpha_\mathrm{c}, \alpha_\mathrm{q}$. The final loss is expressed as in Eq. (4):

$$L_\mathrm{custom} = \alpha_\mathrm{h} \cdot L_\mathrm{h} + \alpha_\mathrm{c} \cdot L_\mathrm{cosh} + \alpha_\mathrm{q} \cdot L_\mathrm{q}, \tag{4}$$

where $\alpha_\mathrm{h}, \alpha_\mathrm{c}, \alpha_\mathrm{q} \geq 0$ and $\alpha_\mathrm{h} + \alpha_\mathrm{c} + \alpha_\mathrm{q} = 1$. These weights are empirically determined to optimize predictive performance. By integrating these components, our custom hybrid loss function balances error sensitivity, maintains robustness to outliers, and addresses asymmetric error distributions, enhancing the model's predictive accuracy for PWL under varying SLR scenarios.

4 Experimental Setup

This section outlines the parameters employed to train, validate, and evaluate the proposed DL model. We detail the dataset splits, augmentation strategies, baseline models, and evaluation metrics to validate and compare the performance of the CASPIAN-v2 model.

4.1 Dataset Splits

Our research incorporates datasets from two regions (AD and SF), covering multiple SLR scenarios, as discussed in Sect. 2 and Table 1. The data is divided into sets for primary model training and for subsequent fine-tuning to assess generalization. The composition of these datasets is detailed in Table 2.

To enhance the model's generalization ability and robustness for primary training, we employed a systematic data augmentation strategy on the AD (0.5 m) and SF (1.0 m) training and validation subsets. The augmentation process primarily involves a random remove function, which applies random spatial cutouts and scaling factors to the original samples. Specifically, this technique first identifies the spatial coordinates of the shoreline protection segments and then occludes small, square regions around a random subset of them in the input maps. This process simulates scenarios with imperfect or missing data, forcing the model to learn more robust contextual features rather than memorizing the impact of any single protection segment. We create distinct yet related variants of the original dataset by systematically applying these transformations multiple times (24× for AD and 10× for SF). Compared to the original sparse dataset, this strategy produces a richer dataset for primary training, comprising 2304 training samples and 240 validation samples for AD, along with 2250 training samples and 240 validation samples for SF.

The fine-tuning datasets for SF (0.5 and 1.5 m SLR) consist of 30 protection scenarios where one OLU was protected at a time (more details in Sect. S5). For evaluation, 20 % of the data (6 samples) was reserved, while the remaining 80 % (24 samples) was used for fine-tuning and validation.

Table 2. Dataset details for primary training and fine-tuning.

Type	Region	SLR	Total	Train	Validation	Test
Primary	AD	0.5 m	142	96	10	36
	SF	1.0 m	285	225	24	36
Fine-tuning	SF	0.5 m	30	20	4	6
	SF	1.5 m	30	20	4	6

4.2 Model Optimization and Training Protocol

The CASPIAN-v2 model was implemented in Python 3.10 using TensorFlow 2.10.1 and was trained on a 64 bit Windows operating system. We utilized an Intel Core i9-14900K (3.20 GHz) machine with 64 GB of RAM and an NVIDIA GeForce RTX 4090 GPU. The final CASPIAN-v2 architecture was refined through extensive ablation studies that systematically evaluated the impact of each novel component. Key insights from these studies, which are detailed in Sect. S3, are summarized in Table 3. These experiments confirmed that the optimal design incorporates the custom *Hybrid Loss* function and a bottleneck composed of four *MARX blocks*. This bottleneck design (ResNeXt + CBAM) was empirically shown to be superior to simpler alternatives. Finally, the studies validated that our method of integrating SLR information via the *SEE block* and just before the final output layer was the most effective approach.

4.2.1 Primary Training:

The model was first trained on the combined AD (0.5 m) and SF (1.0 m) datasets using the Adam optimizer and the proposed hybrid loss function. This phase lasted for 200 epochs with a batch size of 2, allowing the model to learn the core relationships between shoreline protection and flood dynamics. The remaining hyperparameters were fine-tuned using Bayesian Optimization and Random Search to ensure optimal performance.

4.2.2 Fine-tuning for Generalization:

To assess adaptability to different SLR conditions, the pre-trained model was then fine-tuned on the new SF datasets (0.5 and 1.5 m SLR). Fine-tuning spanned 100 epochs. To prevent catastrophic forgetting while adapting to the new data, we employed a curriculum-based strategy. This approach involved mixing the new SLR data with holdout data. The training began with batches containing 30 % new data and 70 % old data, with the proportion of new data gradually increasing to 70 % by the end of the fine-tuning process. The final performance on these new SLR levels was evaluated on the reserved test sets (6 samples each), which were not seen during either training or fine-tuning.

4.3 Baseline Models

To ensure a fair and direct comparison, we selected and implemented a suite of SOTA models, as direct benchmarking against many methods in the literature is often not feasible due to a lack of publicly available code or differences in problem formulation. We assessed the performance of CASPIAN-v2 model for coastal flood prediction against several SOTA ML and DL techniques. We considered conventional ML methods, including the Naïve model, which utilizes a dummy regressor to forecast the mean value of the target variable to serve as a basic reference for assessing more advanced models. Additionally, we trained random forest, linear regression, extreme gradient boosting, support vector regression, lasso regression with polynomial features, and kriging with principal component analysis to establish an ML benchmark. The hyperparameters for training these models were optimized through a combination of Bayesian optimization and random search methods, allowing for efficient exploration of the parameter space while preventing overfitting on the validation set.

In addition to traditional ML baselines, we tested several DL models adapted to the flood prediction task. These include a simple feed-forward neural network architecture, specifically a multi-layer perceptron (MLP), and compact convolutional transformers (CCT) (Hassani et al., 2021), which serve as baseline 1D DL models. Furthermore, we evaluated several 2D DL models, including Attention-Unet (Oktay et al., 2018), and Swin-Unet (Cao et al., 2022). To adapt these models for flood prediction, we replaced their segmentation heads with a 1×1 convolution layer followed by activation to output real-valued flood depth predictions. We evaluated two versions of Attention-Unet: one with randomly initialized weights and another (denoted as Atten-Unet*) with an encoder pre-trained on ImageNet (Deng et al., 2009), leveraging transfer learning to improve performance in low-data scenarios. The final DL baseline was CASPIAN, which we previously proposed in (Karapetyan et al., 2026). All DL models were trained using the Adam optimizer and the proposed hybrid loss function (L_{custom}). Additionally, each model was trained for 200 epochs with a batch size of 2, and early stopping based on validation loss. The remaining training hyperparameters for each model were tuned using Bayesian Optimization and Random Search with the Keras Tuner to ensure a fair comparison.

Table 3. Summary of ablation study results identifying the optimal configuration for each key model component. The final configuration of CASPIAN-v2 incorporates all these optimized choices.

Component	Optimal Configuration	Contribution
Bottleneck Architecture	MARX (ResNeXt + CBAM)	Superior spatial feature extraction
Loss Function	Hybrid	Balances error sensitivity and robustness
Number of MARX Blocks	4 Blocks	Optimal balance of complexity and feature learning
Number of SEE Blocks	1 Block*	Improves accuracy while maintaining efficiency
SLR Integration	SEE + Final Output Layer	Most effective placement for leveraging SLR data

* Although 4 SEE blocks yielded the highest accuracy, 1 block was chosen for the final model to balance performance with computational efficiency, as detailed in the supplement.

4.4 Evaluation Metrics

To evaluate the performance of our model in predicting PWL values, we employ a comprehensive suite of metrics. Each metric is chosen to assess a different aspect of predictive accuracy, from point-wise water depth errors to the spatial correctness of the flood extent, ensuring a holistic evaluation relevant to practical flood risk management.

- *Average Relative Total Absolute Error (ARTAE).* In flood modeling, the significance of a prediction error is often relative to the local water depth. An error of 0.2 m is critical in a shallow, 0.5 m flood but less so in a deep, 4 m flood. ARTAE addresses this by measuring error relative to the true value, providing a scale-invariant assessment of the model's accuracy. It quantifies the relative error between the predicted $y_{\mathrm{p},i}$ and true values $y_{\mathrm{t},i}$ using the normalized L_1 difference:

$$\mathrm{ARTAE} \triangleq \frac{1}{N}\sum_{i=1}^{N}\frac{\|y_{\mathrm{t},i}-y_{\mathrm{p},i}\|_1}{\|y_{\mathrm{t},i}\|_1} \tag{5}$$

  where N denotes the total data samples.

- *Average Root Mean Square Error (ARMSE).* For flood risk assessment, large prediction errors can have catastrophic consequences, such as failing to predict the inundation of a key evacuation route or a critical facility like a hospital. ARMSE is highly sensitive to these large deviations because it squares the errors before averaging. It is therefore used to penalize and highlight instances of significant prediction failures. It captures the root mean square error for each sample, as expressed:

$$\mathrm{ARMSE} \triangleq \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{d_y}\sum_{j=1}^{d_y}(y_{\mathrm{t},i,j}-y_{\mathrm{p},i,j})^2} \tag{6}$$

  where d_y indicates the dimensionality of each sample.

- *Average Mean Absolute Error (AMAE).* In contrast to ARMSE, AMAE provides an intuitive measure of the average error magnitude across all spatial points, without being disproportionately skewed by a few extreme outliers. This offers a robust, general assessment of the model's expected performance on a per-pixel basis. The AMAE is calculated as:

$$\mathrm{AMAE} \triangleq \frac{1}{N}\sum_{i=1}^{N}\frac{1}{d_y}\sum_{j=1}^{d_y}|y_{\mathrm{t},i,j}-y_{\mathrm{p},i,j}| \tag{7}$$

- *Coefficient of Determination* (R^2). Beyond average error, it is important to know if the model correctly captures the spatial variability of a flood event. The R^2 metric assesses this by measuring the proportion of variance in the ground truth that is explained by the model. A high R^2 value indicates the model is effective at predicting the location and severity of flood peaks and troughs. It is computed as:

$$R^2 \triangleq \frac{1}{N}\sum_{i=1}^{N}\left(1-\frac{\sum_{j=1}^{d_y}(y_{\mathrm{t},i,j}-y_{\mathrm{p},i,j})^2}{\sum_{j=1}^{d_y}(y_{\mathrm{t},i,j}-\bar{y}_{\mathrm{t},i})^2}\right) \tag{8}$$

  where $\bar{y}_{\mathrm{t}}^{k}$ is the mean of the true values for the kth sample.

- *Threshold Exceedance Metric* ($\delta > \Delta$). This metric is directly tied to operational decision-making. In flood management, specific error thresholds (Δ) often correspond to critical infrastructure limits, such as the floor height of a building or the elevation of a major roadway. This metric quantifies the frequency of "critical failures" (cases where the prediction error exceeds this pre-defined safety margin). It is defined as:

$$\delta > \Delta \triangleq \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\{j : |y_{\mathrm{t},i,j}-y_{\mathrm{p},i,j}| > \Delta\}\right|}{d_y} \tag{9}$$

- *Non-inundated Prediction Accuracy* (Acc[0]). Given the high class imbalance in flood maps (most areas are dry), it is crucial to verify that the model is not prone to false alarms. This metric specifically measures the ability of the model to correctly identify non-inundated (safe) zones. High accuracy is essential for building trust in the model and ensuring the reliability of evacuation and land-use planning. It is computed as:

$$\mathrm{Acc}[0] \triangleq \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\{j : y_{\mathrm{t},i,j}=0\}\right|}{d_y} \tag{10}$$

– *Dice Similarity Coefficient (DSC).* To address spatial fitness, we introduce the DSC, a standard metric for evaluating the spatial overlap between predicted and true flood extents. Unlike the point-wise error metrics above, the DSC assesses the geometric accuracy of the inundation area. To compute the DSC, the continuous model outputs (y_p) and ground truth values (y_t) are first converted into binary inundation masks by applying a threshold (any pixel with a water depth > 0 is considered inundated). From these masks, we calculate the overlap:

$$\mathrm{DSC} \triangleq \frac{2\times|\mathrm{TP}|}{2\times|\mathrm{TP}|+|\mathrm{FP}|+|\mathrm{FN}|} \tag{11}$$

where true positives (TP) represents the area correctly predicted as flooded, false positives (FP) represents the overpredicted (wet where it should be dry) area, and false negatives (FN) represents the underpredicted (dry where it should be wet) area. This metric provides a direct measure of the model's ability to correctly delineate the flood boundaries.

5 Results

In this section, we evaluate the performance of CASPIAN-v2 model through quantitative and qualitative analyses.

5.1 Quantitative Results

5.1.1 Performance Metrics on Test Set

We first report the performance of CASPIAN-v2 on the test set, as shown in Table 4. For AD data, the model achieves an AMAE of 0.0586, ARMSE of 0.4079, and a high average R^2 score of 0.9556, indicating excellent explanatory power. The ARTAE of 4.2793 % and low error percentages ($\delta > 0.5$ %: 1.02 % and $\delta > 0.1$ %: 4.37 %) highlight higher precision in accurately predicting flood inundation levels. Similarly for SF, the model achieves an AMAE of 0.0320, ARMSE of 0.2094, and an average R^2 score of 0.9214. While the ARTAE is higher at 8.8129 %, the model maintains high accuracy metrics with an Acc[0] of 99.76 % compared to 99.04 % in AD.

On the combined dataset, CASPIAN-v2 performs consistently well with an AMAE of 0.0453, ARMSE of 0.3087, and an average R^2 score of 0.9385. The combined ARTAE of 6.5461 % and low error percentages ($\delta > 0.5$: 0.89 % and $\delta > 0.1$: 3.55 %) demonstrate balanced performance across regions. The high Acc[0] of 99.39 % further underscores the reliability of the model in accurately predicting coastal inundation.

5.1.2 Performance Metrics on Holdout Set

In this section, we present CASPIAN-v2 performance on the holdout set. The results are reported in Table 5, where it can be observed that the model achieves an AMAE of 0.0792, an ARMSE of 0.4871, and an average R^2 score of 0.9525 for AD. Furthermore, the small percentages of errors ($\delta >$ 0.5: 1.29 % and $\delta > 0.1$: 5.48 %) underscore its accuracy in predicting flood inundation levels.

Similarly, for SF, CASPIAN-v2 achieves an AMAE of 0.0317, an ARMSE of 0.2259, and an average R^2 score of 0.9694. Compared to AD, the ARTAE of 4.0009 % indicates slightly more predictions that have larger relative errors. However, with Acc[0] of 99.64 %, the model achieves better non-inundated prediction accuracy compared to 99.07 % in AD-Holdout.

Overall, CASPIAN-v2 achieves an AMAE of 0.0512, an ARMSE of 0.3331, and an average R^2 score of 0.9625 on the aggregated holdout dataset. The ARTAE of 3.7167 % and small error percentages ($\delta > 0.5$: 1.04 % and $\delta > 0.1$: 4.17 %) signify consistent performance in both regions. The higher Acc[0] of 99.41 % further confirms its reliability in predicting flood inundation across diverse and challenging shoreline scenarios.

5.1.3 Performance Analysis under Data Imbalance

To quantitatively validate the data imbalance performance, we generated normalized confusion matrices to analyze the model's accuracy specifically on flooded versus non-flooded pixels, as shown in Fig. 5. This analysis confirms the effectiveness of our approach. For the combined dataset, the model correctly identifies non-flooded areas with 99.85 % accuracy and, more importantly, correctly identifies the rare flooded areas with 99.19 % accuracy. The extremely low false negative rate (0.81 % for the combined set) is particularly significant, as it indicates the model rarely fails to predict an existing flood (a critical requirement for any reliable risk assessment tool). Similarly, the low false positive rate (0.15 %) demonstrates that the model does not raise false alarms, further enhancing its practical utility. This quantitative evidence substantiates that our multi-faceted strategy successfully mitigates the effects of data imbalance.

5.1.4 Performance Benchmarking against SOTA Methods

To comprehensively evaluate the performance of CASPIAN-v2, we benchmarked it against a suite of SOTA traditional ML and DL models. The selection and implementation details for these baseline models are described in Sect. 4.3. This section presents a detailed comparison the prediction perfor-

Table 4. Evaluation of CASPIAN-v2 on test set. ↓ indicates that lower values are better, and ↑ indicates that higher values are better.

Dataset	AMAE ↓	ARMSE ↓	ARTAE ↓	Avg. $\delta > 0.5$ ↓	Avg. $\delta > 0.1$ ↓	Avg. R^2 Score ↑	Avg. Acc[0] ↑
AD	0.0586	0.4079	4.2793	1.02 %	4.37 %	0.9556	99.04 %
SF	0.0320	0.2094	8.8129	0.75 %	2.72 %	0.9214	99.76 %
Combined	0.0453	0.3087	6.5461	0.89 %	3.55 %	0.9385	99.39 %

Table 5. Evaluation of CASPIAN-v2 on holdout set.

Dataset	AMAE ↓	ARMSE ↓	ARTAE ↓	Avg. $\delta > 0.5$ ↓	Avg. $\delta > 0.1$ ↓	Avg. R^2 Score ↑	Avg. Acc[0] ↑
AD – Holdout	0.0792	0.4871	3.3081	1.29 %	5.48 %	0.9525	99.07 %
SF – Holdout	0.0317	0.2259	4.0009	0.86 %	3.26 %	0.9694	99.64 %
Combined	0.0512	0.3331	3.7167	1.04 %	4.17 %	0.9625	99.41 %

mance across all models, with the full results presented in Table 6. The analysis is broken down by model class, first comparing against traditional ML methods, and then against other DL architectures.

Comparison with Machine Learning Models. In this section, we compare the performance of CASPIAN-v2 against various traditional ML models for flood prediction, as shown in Table 6. The Naïve model shows high errors with an AMAE of 1.5343, ARMSE of 3.5444, and an average R^2 score of 0.5450. Among traditional approaches, linear regression reduces errors significantly, achieving an AMAE of 0.1272, ARMSE of 0.1946, and an average R^2 score of 0.9464. The lasso with polynomial model further improves performance, giving an AMAE of 0.0937, ARMSE of 0.1202, and the highest average R^2 score of 0.9618 among traditional ML models.

Compared to the best traditional model (lasso with polynomial), CASPIAN-v2 reduces the AMAE by 51.65 % (from 0.0937 to 0.0453). However, CASPIAN-v2 has a higher ARMSE of 0.3087 compared to 0.1202, indicating it minimizes mean errors effectively but may experience larger individual prediction errors. Despite this, CASPIAN-v2 outperforms traditional models across multiple metrics, leveraging DL and multi-dimensional data integration to achieve superior accuracy in flood prediction.

This trend is even more pronounced in the spatial accuracy results. While the lasso model achieved a DSC of 0.6438, CASPIAN-v2 scored 0.8437, representing a 31.05 % improvement. This significant gap underscores the inherent limitations of traditional ML models in capturing the complex geometric shape of flood events, a task for which our deep learning architecture is better suited.

Comparison with Deep Learning Models. Existing 1D and 2D DL models show varied performance, as reported in Table 6. The CCT model achieves an AMAE of 0.9064, an ARMSE of 2.3292, and an average R^2 score of 0.6649, indicating moderate predictive capabilities. Atten-Unet and its variant Atten-Unet* improve performance with AMAE values of 0.1061 and 0.1032 and average R^2 scores of 0.9195 and 0.9210, respectively. Swin-Unet achieves further improvements, reducing the AMAE to 0.0629 and attaining an average R^2 score of 0.9514, reflecting its effectiveness in capturing spatial dependencies.

Compared to the second-best DL model, CASPIAN-v2 reduces the AMAE by 19.96 % (from 0.0566 to 0.0453) and achieves an exceptional average Acc[0] of 99.39 %, surpassing CASPIAN's 98.84 %. These results highlight superior accuracy and robust generalization capabilities of CASPIAN-v2.

In terms of spatial fitness, CASPIAN-v2 (with DSC of 0.8437) also demonstrates a clear advantage over the best-performing DL baseline, CASPIAN (0.8261), representing a 2.13 % improvement in spatial accuracy. Taken together, these results highlight the superior accuracy and robust generalization capabilities of CASPIAN-v2. The integration of advanced components such as the MARX and SEE blocks, combined with an optimized Hybrid loss function, enables the effective modeling of complex flood dynamics.

5.1.5 Computational Efficiency Analysis

A primary motivation for this research is to overcome the significant computational burden of physics-based hydrodynamic simulators. The final three columns of Table 6 provide a comprehensive comparison of the computational efficiency of all evaluated models. To contextualize this comparison, we first summarize the computational cost and hardware requirements of the physics-based simulations used to generate the training and evaluation data.

Hydrodynamic Simulation Cost. The computational cost of generating a peak flood depth map using the coupled hydrodynamic model, which underscores the need for an efficient surrogate, varies significantly between the two study regions. For the coast of Abu Dhabi, the process to generate a map such as the one shown in Fig. 1a takes approximately 71 to 73 h of elapsed runtime, equating to 1500 to 1660 CPU-hours, depending on the specific protection sce-

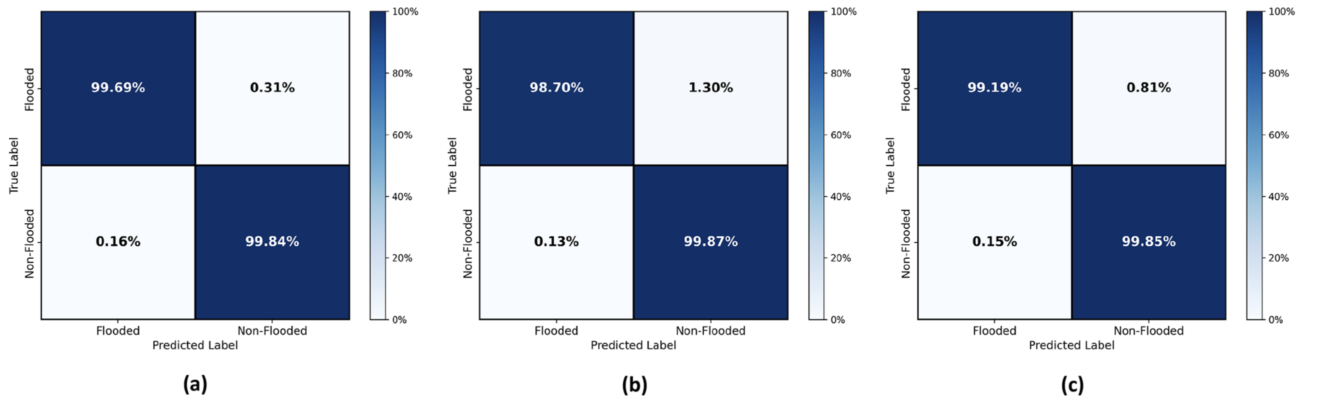

Figure 5. Normalized confusion matrices evaluating the classification performance of the model on flooded versus non-flooded pixels for the **(a)** AD, **(b)** SF, and **(c)** combined test sets. The high values on the main diagonal demonstrate the model's excellent accuracy for both the majority (non-flooded) and, critically, the minority (flooded) classes, confirming its robustness to the severe data imbalance

Table 6. A comprehensive performance comparison between our proposed CASPIAN-v2 and state-of-the-art models, grouped into a baseline physics-based simulator (Delft3D), traditional ML, and DL approaches. Prediction accuracy is evaluated using eight standard metrics, where arrows indicate the desired direction (↑ for higher is better, ↓ for lower is better). Computational efficiency is assessed by three key indicators: the total number of trainable parameters (M = millions), the total training time (TT), and the average inference time (IT) per sample. In the physics-based simulations, PP denotes Post-Processing. The simulation results, which provide the ground truth data, are included for reference. The top-performing result for each metric is highlighted in bold, and the second-best is highlighted in italic.

Type	Model	Prediction Accuracy								Computational Efficiency		
		AMAE ↓	ARMSE ↓	ARTAE ↓	$\delta > 0.5$ ↓	$\delta > 0.1$ ↓	R^2 ↑	Acc[0] ↑	DSC ↑	Param↓	TT ↓	IT ↓
Simulator	AD: Delft3D + SWAN + PP	Served as the ground truth								–	–	71–73 h
	SF: Delft3D + PP									–	–	3.5–6.0 h
ML (1-D)	Naïve	1.5343	3.5444	1746.0693	74.92 %	80.11 %	0.5450	31.01 %	0.3871	–	*62 s*	*0.15 s*
	RF	0.5411	0.7310	264.9505	36.77 %	72.20 %	0.7962	34.19 %	0.4185	–	75 s	0.18 s
	Linear	0.1272	*0.1946*	64.9859	7.87 %	14.03 %	0.9464	59.28 %	0.6279	–	**65 s**	**0.16 s**
	XGBoost	0.2546	0.2446	164.1654	16.27 %	49.88 %	0.9347	44.10 %	0.4711	–	198 s	0.21 s
	SVR	0.2069	0.2423	72.3122	9.24 %	41.17 %	0.9298	45.46	0.4889	–	79 s	0.19 s
	Lasso with Poly	0.0937	**0.1202**	28.1565	4.47 %	15.04 %	**0.9618**	55.78 %	0.6438	–	72 s	0.17 s
	Kriging	0.1098	0.2478	39.9073	5.22 %	11.59 %	0.9414	62.88 %	0.6359	–	76 s	0.18 s
DL (1-D)	MLP	0.6486	2.7247	524.1724	32.82 %	41.94 %	0.6572	36.91 %	0.4356	*0.01 M*	14 h	5.03 s
	CCT	0.9064	2.3292	843.5430	48.08 %	64.63 %	0.6649	34.01 %	0.4228	11.05 M	18 h	0.26 s
DL (2-D)	Atten-Unet	0.1061	0.3714	11.8245	3.14 %	16.70 %	0.9195	95.26 %	0.7390	12.07 M	46 h	0.24 s
	Atten-Unet*	0.1032	0.3627	11.6585	3.31 %	15.62 %	0.9210	94.99 %	0.7469	12.07 M	47 h	0.27 s
	Swin-Unet	0.0629	0.2788	6.7244	1.47 %	12.94 %	*0.9514*	98.10 %	0.8014	8.29 M	26 h	0.24 s
	CASPIAN	*0.0566*	0.3613	**5.8573**	*1.01 %*	*4.79 %*	0.9209	*98.84 %*	*0.8261*	**0.36 M**	22 h	0.22 s
	0.0453	0.3087	*6.5461*	**0.89 %**	**3.55 %**	0.9385	**99.39 %**	**0.8437**	0.38 M	22 h	0.22 s	

* with pre-trained encoder on ImageNet (Deng et al., 2009).

nario. This comprehensive simulation includes Delft3D runs, which require 6 to 7 h on 28 CPU cores (Intel Xeon E5-2680 @ 2.40 GHz; ≈ 168–196 CPU-hours), and SWAN simulations, which take about 10 to 11 h on 128 CPU cores (AMD EPYC 7742 @ 2.25 GHz; ≈ 1280–1408 CPU-hours). Subsequent post-processing and run-up calculations using Matlab scripts add approximately 55 h on a single core.

In contrast, generating a similar map for San Francisco Bay (see Fig. 2a) is computationally less demanding, requiring approximately 3.5 to 6.0 h of elapsed time, or 84.5 to 141 CPU-hours. The Delft3D runs for this region take about 3 to 5 h on 28 CPU cores, and the post-processing of these outputs takes between 0.5 and 1.0 h on a single core. It is important to note that SWAN and run-up calculations were not performed for the San Francisco Bay shoreline, as its relatively sheltered inland location makes these components unnecessary, accounting for the substantial difference in computational cost.

Extrapolating these figures, simulating the full test set of 72 scenarios (36 for each region) using physics-based models would require approximately 2763 h of continuous computation, equivalent to nearly 115 d of uninterrupted runtime on high-performance computing infrastructure.

Computational Efficiency of Data-Driven Models. Against this computational backdrop, the efficiency gains offered by data-driven approaches become particularly evident. As shown in Table 6, the traditional ML models are the fastest to train, typically requiring only a few minutes. However, this speed comes at the cost of significantly lower prediction accuracy. Among the more accurate DL models, CASPIAN-v2 demonstrates a highly favorable balance of performance and efficiency. With only 0.38 million parameters, it is one of the most lightweight 2D models, comparable in size to the original CASPIAN (0.36 M) and substantially smaller than transformer-based models like Swin-Unet (8.29 M) or other U-Net variants (12.07 M). Its training time (22 h) and inference time (0.22 s per scenario) are also highly competitive within this high-performing group.

Most importantly, CASPIAN-v2 can generate predictions for all 72 test scenarios in just under 16 s on a single GPU. This represents a dramatic reduction in computational time compared to physics-based simulations, effectively transforming a months-long simulation effort into a near-instantaneous task. Such a reduction enables large-scale scenario exploration, sensitivity analysis, and long-term coastal adaptation planning that would be computationally infeasible using traditional hydrodynamic models alone. Consequently, CASPIAN-v2 emerges as a practical and scalable surrogate for real-world coastal flood risk assessment.

5.1.6 Numerical Assessment of Generalizability

This section reports the generalization performance of CASPIAN-v2 on unseen data. The model was fine-tuned using new SF data corresponding to 0.5 and 1.5 m SLR depths, encompassing 30 protection scenarios where one OLU was protected at a time (more details in Sect. S5). For evaluation, 20 % of the data (6 samples) was reserved, while the remaining 80 % (24 samples) was used for fine-tuning and validation. Fine-tuning spanned 100 epochs with a progressive gradual recall approach, mixing the new data with the AD and SF holdout data in a 20 : 80 test / train ratio. The training set began with 70 % of the AD and SF holdout set combined with 30 % of the new data, gradually increasing to 70 % by the end of training.

The results in Table 7 demonstrate strong generalization by CASPIAN-v2 across SLR scenarios. For SF 0.5 m data, the model achieved an AMAE of 0.0626, ARMSE of 0.2996, and average R^2 score of 0.9336. An ARTAE of 6.4240 % and low error percentages ($\delta > 0.5$: 1.89 % and $\delta > 0.1$: 7.79 %) highlight its precision. For SF 1.5 m data, the model showed slightly suboptimal performance with an AMAE of 0.1005, ARMSE of 0.4565, and average R^2 score of 0.9196. The ARTAE of 4.3961 % indicates balanced performance, with an average Acc[0] of 98.23 % compared to 97.99 % for 0.5 m data.

When retaining existing knowledge, CASPIAN-v2 achieved an AMAE of 0.0567 and ARMSE of 0.2274 on the AD holdout set for 0.5 m SLR, with an average R^2 score of 0.9901. The ARTAE of 2.5225 % and low error percentages ($\delta > 0.5$: 0.53 % and $\delta > 0.1$: 17.87 %) emphasize its precision. For the SF holdout set at 1.0 m SLR, the model achieved an AMAE of 0.0433, ARMSE of 0.2318, and average R^2 score of 0.9685. The ARTAE of 4.6277 % and error percentages ($\delta > 0.5$: 0.79 % and $\delta > 0.1$: 9.61 %) reflect its ability to balance low absolute and relative errors, with an Acc[0] of 99.34 %.

Overall, the model achieved an AMAE of 0.0652, an ARMSE of 0.3040, and an average R^2 score of 0.9520, revealing robust generalization abilities of the model across various SLR settings. Further, the model achieved an ARTAE of 4.5871 % and low error percentages ($\delta > 0.5$ %: 1.31 % and $\delta > 0.1$ %: 12.07 %), with a high Acc[0] of 98.69 %. These findings highlight the ability of the CASPIAN-v2 model to effectively generalize to new and previously unseen scenarios with minor fine-tuning, making it a reliable tool for real-world inundation prediction.

5.2 Qualitative Results

5.2.1 Visual Performance on Test Set

In this section, we provide a qualitative assessment of the performance of CASPIAN-v2 on the test set. Figure 6 presents two randomly selected scenarios for the AD and SF regions, where it can be observed that the predicted inundation values of the proposed model closely align with the corresponding ground truth values. In single unprotected OLU scenarios (rows 1 and 3), the model accurately captures localized flooding effects, showing sensitivity to minor protection configuration changes. Similarly, CASPIAN-v2 effectively handles the increased complexity of mixed OLU protection statuses (rows 2 and 4). These results highlight the robustness of the model in generalizing across diverse regions and protection patterns. Figure 6c shows the absolute error maps, where it can be observed that the CASPIAN-v2 model produced minimal errors, with deviations occurring mainly in areas with sharp transitions in flood depths. However, these small variations minimally affect the overall prediction accuracy.

To illustrate the local impact of the protection measures on flood dynamics, zoomed-in insets are provided for specific OLUs. For instance, the first inset for AD highlights how inundation patterns are directly controlled by the protection status of the nearest OLU. When OLU-17 is protected, the area behind it remains largely dry, whereas significant flooding occurs inland of the unprotected OLU-14.

5.2.2 Visual Performance on Holdout Set

In this section, we demonstrate the performance of CASPIAN-v2 using a holdout set composed of particularly challenging coastal protection scenarios. Figure 7 showcases the performance of the model on two challenging config-

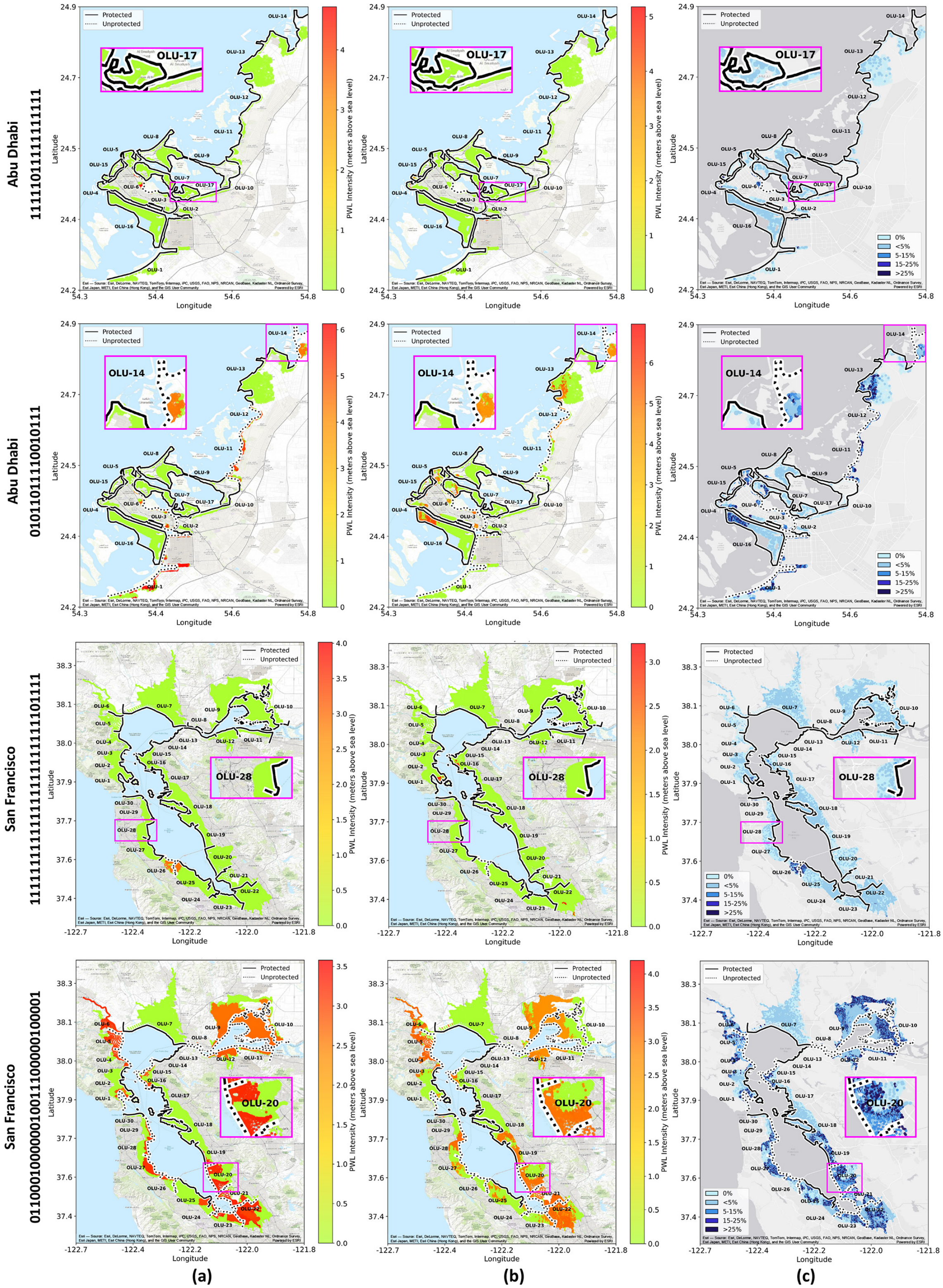

Figure 6. Evaluation of CASPIAN-v2 on the test datasets. **(a)** Ground truth inundation maps for representative AD and SF scenarios. **(b)** Predicted inundation values. **(c)** Absolute error distributions of predicted inundation values. Darker shades of blue indicate higher absolute errors, ranging from near 0 % to greater than 25 %. The magenta insets provide zoomed-in views of specific OLUs to illustrate the effect of protection measures. For instance, the inundation is shown to be minimal inland of the protected OLU-17 in AD, whereas significant flooding occurs near the unprotected OLU-20, a dynamic that the model precisely captures.

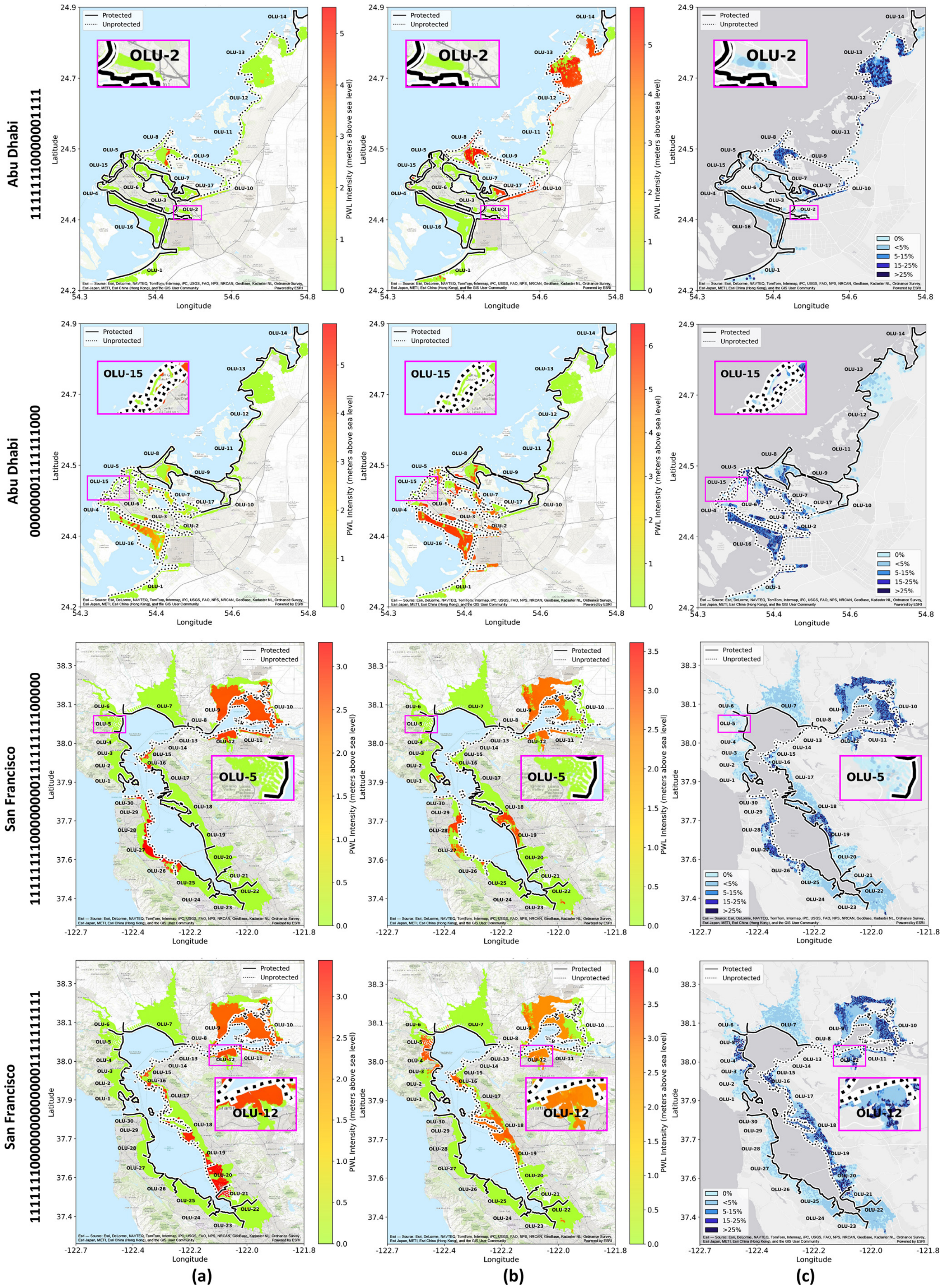

Figure 7. Evaluation of CASPIAN-v2 on the holdout datasets. **(a)** Ground truth inundation maps for representative AD and SF scenarios. **(b)** Predicted inundation values. **(c)** Absolute error distributions of predicted inundation values. Darker shades of blue indicate higher absolute errors, ranging from near 0 % to greater than 25 %. The zoomed-in insets highlight fine-grained hydrodynamic effects. For instance, the successful prevention of inundation by a protected OLU-2 in AD, versus the widespread inland flooding resulting from an unprotected OLU-12 in SF.

Table 7. CASPIAN-v2 generalizability evaluation using different SLR data.

Dataset *(SLR)*	AMAE ↓	ARMSE ↓	ARTAE ↓	Avg. $\delta > 0.5$ ↓	Avg. $\delta > 0.1$ ↓	Avg. R^2 Score ↑	Avg. Acc[0] ↑
SF – Generalizability *(0.5 m)*	0.0626	0.2996	6.4240	1.89 %	7.79 %	0.9336	97.99 %
SF- Generalizability *(1.5 m)*	0.1005	0.4565	4.3961	1.97 %	14.51 %	0.9196	98.23 %
AD – Holdout *(0.5 m)*	0.0567	0.2274	2.5225	0.53 %	17.87 %	0.9901	99.18 %
SF – Holdout *(1.0 m)*	0.0433	0.2318	4.6277	0.79 %	9.61 %	0.9685	99.34 %
Overall	0.0652	0.3040	4.5871	1.31 %	12.07 %	0.9520	98.69 %

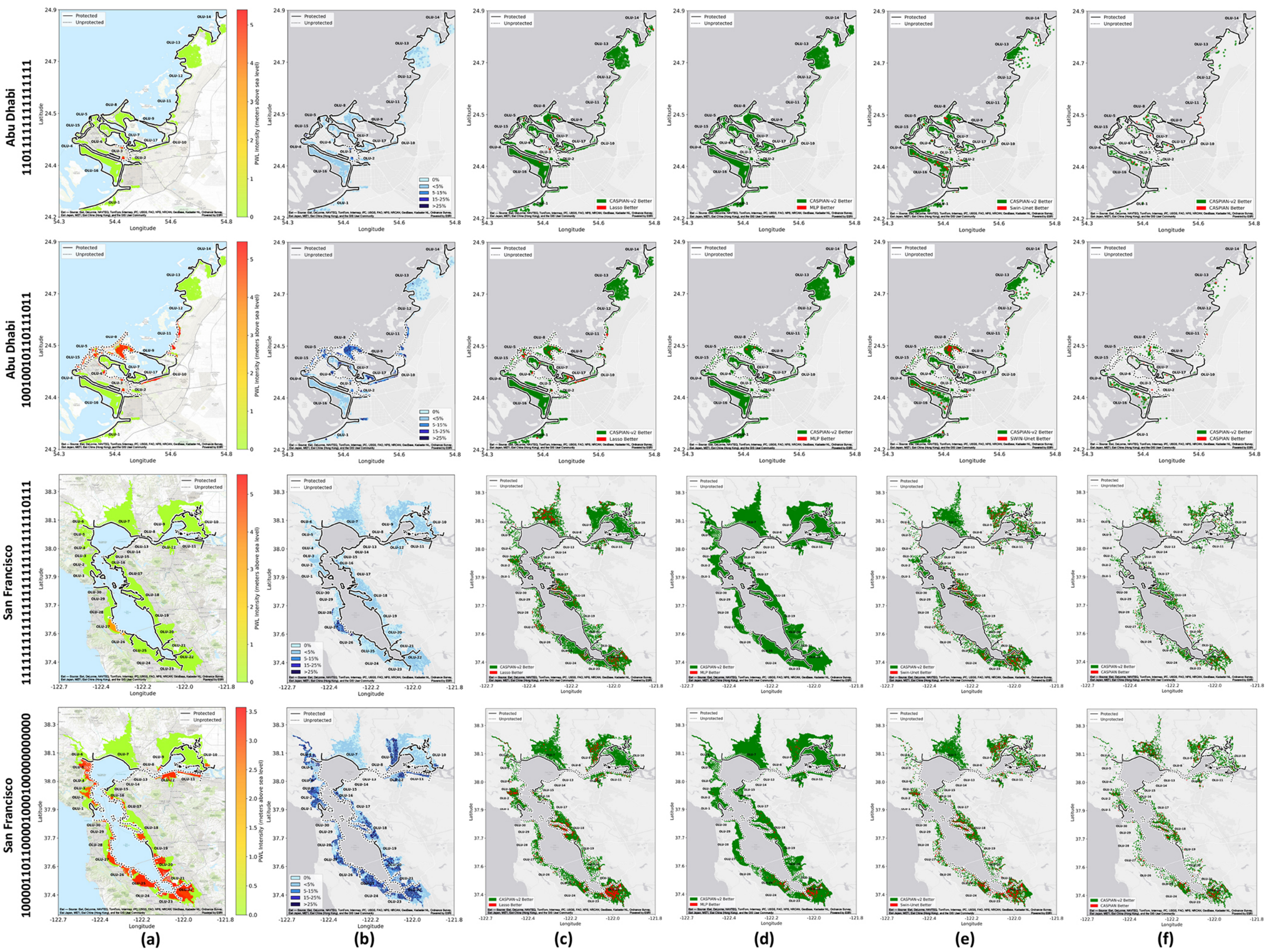

Figure 8. Qualitative comparison of CASPIAN-v2 with SOTA approaches in predicting coastal flood inundation **(a)** Ground truth inundation maps for representative AD and SF scenarios. **(b)** Absolute error map for our proposed CASPIAN-v2 model, with darker blue indicating higher error. **(c–f)** Error difference maps comparing CASPIAN-v2 to key baselines. In these maps, green indicate regions where CASPIAN-v2 is more accurate than the baseline, red areas show where the baseline performed better, and transparent regions denote similar performance. The visualization clearly shows that CASPIAN-v2 provides a substantial improvement over the **(c)** Lasso, **(d)** MLP, **(e)** Swin-Unet, and **(f)** original CASPIAN models.

urations from the holdout set, which was specifically designed to test generalization across complex protection scenarios. These scenarios feature intricate mixes of protected and unprotected OLUs, creating sharp inundation boundaries where flooded and non-flooded regions meet. CASPIAN-v2 demonstrates high fidelity in these cases, accurately capturing these abrupt changes in local flood behavior. For instance, it correctly captures the inundation dynamics when one side of the SF bay is protected and the other is not (last row of Fig. 7).

The strong performance of the model here is particularly noteworthy given that it was trained on only a small subset of the thousands of possible protection combinations (2^n, where n is the number of OLUs). This success on unseen, complex configurations indicates that CASPIAN-v2 is not merely memorizing training data but is learning the underlying spatial logic of how flood defenses influence inundation patterns. This affirms its robustness and reliability for real-world application.

5.2.3 Visual Comparison with SOTA Methods

We qualitatively evaluated the performance of the proposed CASPIAN-v2 by visually comparing its prediction errors with those of key SOTA baselines. Figure 8 presents this analysis for representative scenarios in both Abu Dhabi and San Francisco. Figure 8b shows the absolute error map for our proposed CASPIAN-v2 model, demonstrating that errors are generally low and confined to complex hydraulic transition zones. The key insights, however, come from the error difference maps (Fig. 8c–f), which directly compare the spatial accuracy of CASPIAN-v2 to each baseline. In these maps, green areas highlight regions where CASPIAN-v2 is more accurate, while red indicates where the baseline had a lower error, and transparent areas denote regions where both models performed similarly.

Compared to the Lasso with polynomial features Fig. 8c and MLP Fig. 8d baselines, CASPIAN-v2 offers a dramatic improvement, with vast green areas indicating its superior ability to capture the fundamental flood patterns that these simpler models miss. The comparison with the more advanced Swin-Unet Fig. 8e and the original CASPIAN Fig. 8f models is also convincing. While these models are more competitive, the difference maps still show a clear and consistent advantage for CASPIAN-v2, which successfully reduces errors in many of the most deeply inundated and complex areas.

Moreover, Fig. 9 visualizes the flood extents predicted by CASPIAN-v2 against the best-performing ML and DL baseline model. The map breaks down the predictions into correctly matched areas (green), over-predicted areas (orange), and under-predicted areas (purple). The visualization reveals that while the baseline models produce a more fragmented prediction with significant patches of both over- and under-prediction, the output of the proposed CASPIAN-v2 model aligns much more closely with the ground truth. Its predicted flood extent is more coherent and captures the true inundation boundaries with far fewer spatial errors. These qualitative comparisons align with the quantitative results in Table 6, highlighting the ability of the proposed model to achieve higher accuracy and visually superior predictions.

5.2.4 Visual Assessment of Generalizability

We next evaluate the generalizability of CASPIAN-v2 under different environmental conditions by fine-tuning the model on two additional SLR data of 0.5 and 1.5 m. Figure 10 shows the prediction results, illustrating that while the fine-tuned model exhibits some localized discrepancies (Fig. 10c), these deviations remain modest given the minimal training data and limited fine-tuning epochs. In the 0.5 m SLR scenario, the model yields relatively lower absolute errors in predicting flood extents. By contrast, the 1.5 m scenario exhibits slightly higher errors, likely due to the increased variability in PWL values. Nonetheless, the predictions generally align well with the ground truth inundation patterns.

Overall, these findings underscore adaptability of the proposed model to evolving coastal conditions, suggesting that with sufficient training data and appropriately tuned hyperparameters, the model can maintain robust performance across a broad range of SLR scenarios.

6 Discussion and conclusion

This research presents a novel DL model to predict coastal inundation across two geographical locations (AD and SF). The effectiveness of the proposed CASPIAN-v2 model is validated through extensive experiments, where it outperforms the existing SOTA methods, as shown in Table 6. Although traditional ML approaches are relatively fast to train, these methods lack the ability to capture complex spatial patterns in the data, thus producing less accurate results. Similarly, we found that 1D DL approaches do not scale effectively to large, spatially focused grids. Furthermore, jointly training these methods on AD and SF datasets was less successful and yielded poor results due to inconsistent input features, particularly the different number of OLUs across regions and the need to address a broader array of shoreline adaptation scenarios. In comparison, the proposed 2D DL model can learn complex input patterns, enabling it to produce superior prediction results. Additionally, our data augmentation strategy, which involved creating new training samples by applying random spatial cutouts and scaling factors (as mentioned in Sect. 4.1), exposes the model to a wider variety of conditions. This enhances its resilience to noise, missing data, and varying shoreline configurations.

Moreover, CASPIAN-v2 demonstrates strong generalizability across different levels of SLR, which underlines its utility for future resilience planning. As shown in our experiments, the model can be fine-tuned to generalize to both lower and higher SLR scenarios (0.5 and 1.5 m). In principle, and given the availability of suitable training data, CASPIAN-v2 could also be extended to SLR values of 0 m, enabling applications to short-term coastal flooding prediction. However, the two study regions considered in this work do not experience storm-surge-driven flooding from tropical cyclones. As a result, hydrodynamic simulations at 0 m SLR produce negligible or no coastal inundation, leading to trivial all-zero predictions regardless of shoreline protection configurations. Consequently, explicitly demonstrating generalization to 0 m SLR would not yield additional insights in the present setting and is therefore not included.

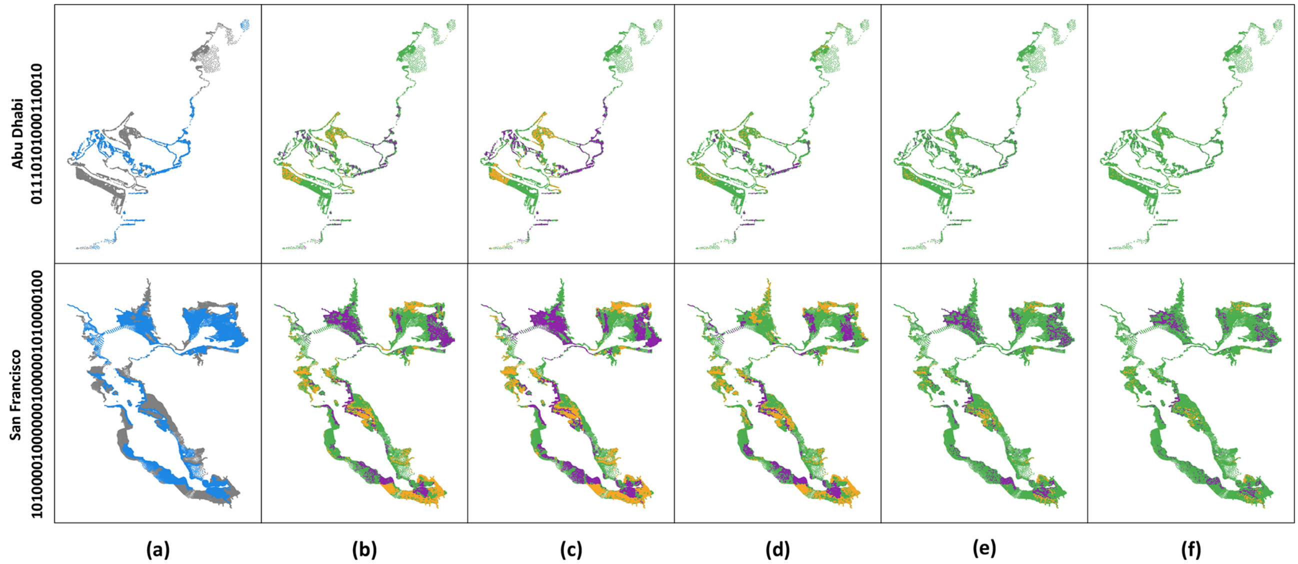

Figure 9. Visual comparison of spatial prediction performance for CASPIAN-v2 and baseline models on representative AD and SF test scenarios. **(a)** Ground-truth inundation maps, where blue denotes inundated regions and gray denotes non-inundated regions. **(b–f)** Model predictions for Lasso, MLP, Swin-UNet, the original CASPIAN model, and CASPIAN-v2, respectively. Green represents correctly predicted inundated areas (true positives), orange represents over-predicted regions (false positives), and purple represents under-predicted regions (false negatives). Across both cities, CASPIAN-v2 produces the most accurate and spatially coherent inundation patterns, with larger regions of correct agreement and fewer misclassified areas.

A critical aspect influencing model performance is the underlying data distribution. As is common in flood modeling, our dataset is highly imbalanced, with a vast majority of non-inundated (zero value) points compared to the relatively rare inundated points (see Fig. S6). To address this significant challenge, our framework employs a multi-faceted strategy. First, our Hybrid Loss function is inherently designed to handle this skew. The Quantile loss component allows us to place more weight on correctly predicting the less frequent, but more critical, positive flood values, while the Huber loss prevents the numerous small errors in non-inundated areas from dominating the training process. Second, the attention mechanism within the MARX block is crucial. This theoretical benefit is substantiated by empirical evidence from our Grad-CAM analysis (Fig. 11), which shows that the model focuses highly around the vulnerable, unprotected shoreline segments where inundation originates. This focus on salient regions prevents the model's learning from being diluted by the vast areas of non-inundated points. Finally, our choice of evaluation metrics, particularly the DSC and non-inundated accuracy, provides a more balanced assessment of performance. This combination of a tailored loss function, an attentive architecture with demonstrated focus, and robust evaluation allows CASPIAN-v2 to maintain high predictive fidelity despite the challenging data distribution.

Beyond its technical role in the model, this demonstrated interpretability provides critical insights for stakeholders. The clear spatial alignment between the focus of the model and known vulnerabilities (Fig. 11c) serves to empirically validate its decision-making process. This level of transparency is instrumental for planners and policymakers, as it clarifies why specific areas are identified as high-risk, thereby fostering trust in DL-based solutions and aiding in the design of targeted resilience strategies.

An essential element for any model designed for risk assessment, aside from interpretability, is the measurement of its uncertainty. To address this, we implemented a deep ensemble method to estimate the predictive uncertainty of CASPIAN-v2 (Lakshminarayanan et al., 2017). We trained five independent models and used the pixel-wise standard deviation of their predictions as a direct proxy for model uncertainty (see Sect. S8 for full quantitative results). The resulting maps, shown in Fig. 12, reveal a crucial characteristic of our model, which is a strong spatial correlation between predictive uncertainty and prediction error. The bright, high-uncertainty regions in panel Fig. 12c closely align with the areas of higher absolute error shown in panel Fig. 12b, while the dark, low-uncertainty regions correspond to areas of high accuracy. This indicates that the model demonstrates a valuable form of self-awareness and it effectively learns to identify regions where its own predictions are less reliable. This is invaluable for coastal planners, as it allows them to trust the high-certainty predictions for general assessments while flagging the high-uncertainty zones as areas that require a higher margin of safety or further, more detailed hydrodynamic study. This ability to not only make accurate predictions but also to reliably signal its own confidence is in-

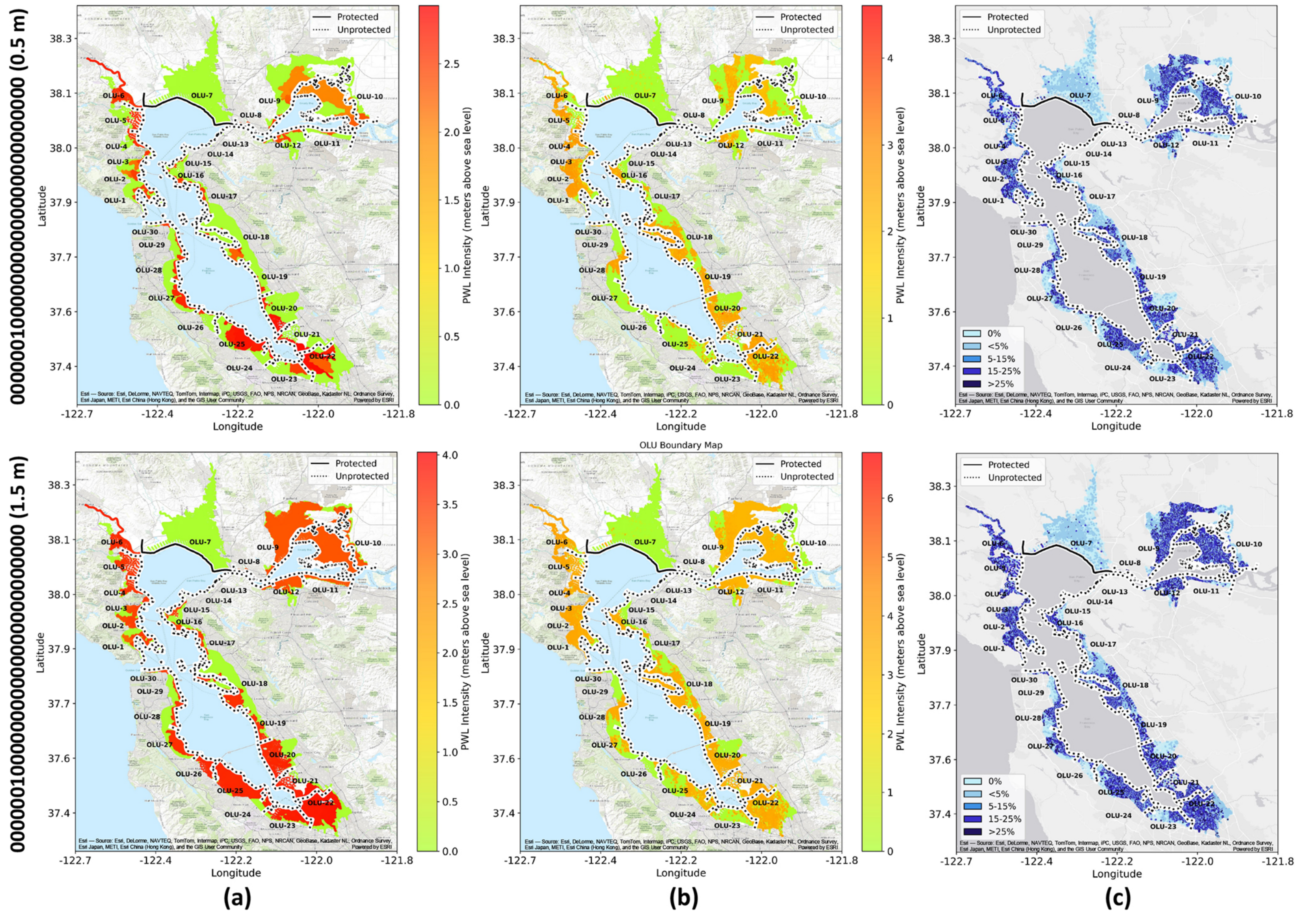

Figure 10. Generalizability evaluation of CASPIAN-v2 fine-tuned for 0.5 and 1.5 m SLR scenarios. **(a)** Ground truth inundation maps. **(b)** Predicted inundation values. **(c)** Absolute error distributions of predicted inundation values. Darker shades of blue indicate higher absolute errors, ranging from near 0 % to greater than 25 %.

strumental for fostering trust and supporting real-world, risk-informed decision-making.

We provide practical, actionable guidance by empirically deriving an uncertainty threshold directly linked to our critical error metric, $\delta > 0.5$. For every scenario in our test set, we first identified all pixel locations where the absolute prediction error exceeded 0.5 m. We then extracted the corresponding normalized uncertainty values from our ensemble's standard deviation map for just those specific critical failure pixels. By averaging these uncertainty values across all identified pixels and all test scenarios, we determined a representative uncertainty level that corresponds to significant model error. This analysis revealed that a normalized uncertainty value of approximately 0.75 or greater is a strong indicator of a potential critical failure ($\delta > 0.5$). Therefore, we recommend a practical guideline for coastal planners that any region where the model's predictive uncertainty exceeds 0.75 should be flagged as a high-priority zone, necessitating either the use of more conservative safety margins or further investigation with detailed hydrodynamic simulations.

A key motivation for this study is the application of DL surrogates to long-term coastal resilience planning. While the high efficiency of such models is often associated with short-term, real-time forecasting (e.g., storm surges), their primary value in the context of strategic SLR planning lies in navigating the vast combinatorial design space of possible adaptation measures. Coastal planners must evaluate not just one future, but thousands of potential combinations of shoreline protection configurations across multiple plausible SLR scenarios. For instance, considering the 30 OLUs in San Francisco, there are over a billion possible protection combinations. Even if a planner wanted to test a mere 10 000 of these scenarios at a single SLR level, doing so with a hydrodynamic model (at 5 h per scenario) would require over 5 years of continuous computation. This is the prohibitive barrier that currently limits comprehensive, data-driven planning.

Our framework addresses this directly by enabling a two-tiered decision-making workflow. CASPIAN-v2 serves as a rapid scenario-assessment tool, allowing planners to explore this vast design space in a matter of hours, not decades, to

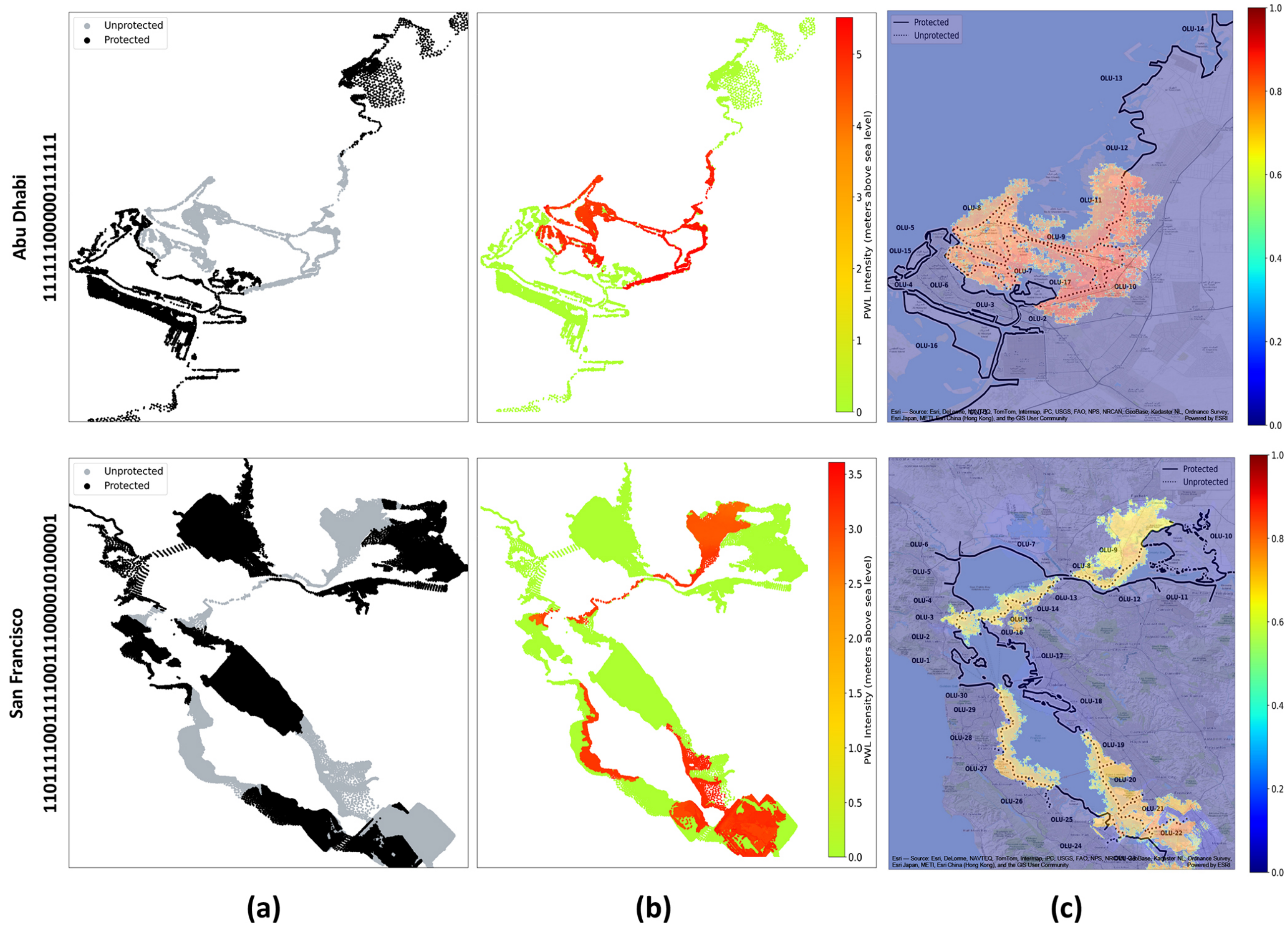

Figure 11. CASPIAN-v2 inundation prediction for AD (top) and SF (bottom): **(a)** Input images representing protected and unprotected OLUs, **(b)** Predicted inundation with PWL intensity, **(c)** Grad-CAM visualizations highlighting model attention, where warmer colors indicate regions the model focused on most during prediction, aligning with unprotected and vulnerable areas.

identify a small shortlist of the most effective and efficient protection strategies. These promising candidates can then be subjected to rigorous validation using the precise, but slow, physics-based models. While the accuracy of our model is indeed bounded by the hydrodynamic model it emulates, its high fidelity (as demonstrated in our results) confirms its value as a reliable proxy for this broad-scale exploration. This synergy, using the surrogate for rapid exploration and the physics-based model for targeted validation, is what establishes the significance of our work as a practical and essential tool for future coastal resilience planning.

Although CASPIAN-v2 represents a significant advancement in surrogate modeling, it is essential to highlight certain limitations and avenues for future research. The prediction accuracy of the model is fundamentally contingent on the quality of the underlying hydrodynamic simulations, where any inaccuracies in land surface conditions or atmospheric data can introduce biases. Furthermore, while we demonstrate generalizability across two distinct regions, applying the model to a new coastal environment with unique bathymetry and hydrodynamic characteristics would still require a dedicated dataset of simulations for that specific location. Overcoming this data dependency is a key challenge for the broad-scale deployment of such surrogate models.

Future research could address these limitations in several ways. Incorporating more diverse geographical contexts and additional input channels, such as detailed elevation and hydro-connectivity data, could enhance predictive reliability. To address the data requirements for new regions, exploring advanced transfer learning techniques, such as few-shot or zero-shot learning, could be a promising direction. These methods could allow the model to be fine-tuned for a new location with a drastically reduced number of new simulations, significantly lowering the barrier to entry for practical application. Moreover, domain adaptation and incremental learning could accelerate implementation, while model compression and distributed training could further enhance scalability and operational utility.

In conclusion, the CASPIAN-v2 model offers a robust, adaptable, and comprehensible approach to predicting

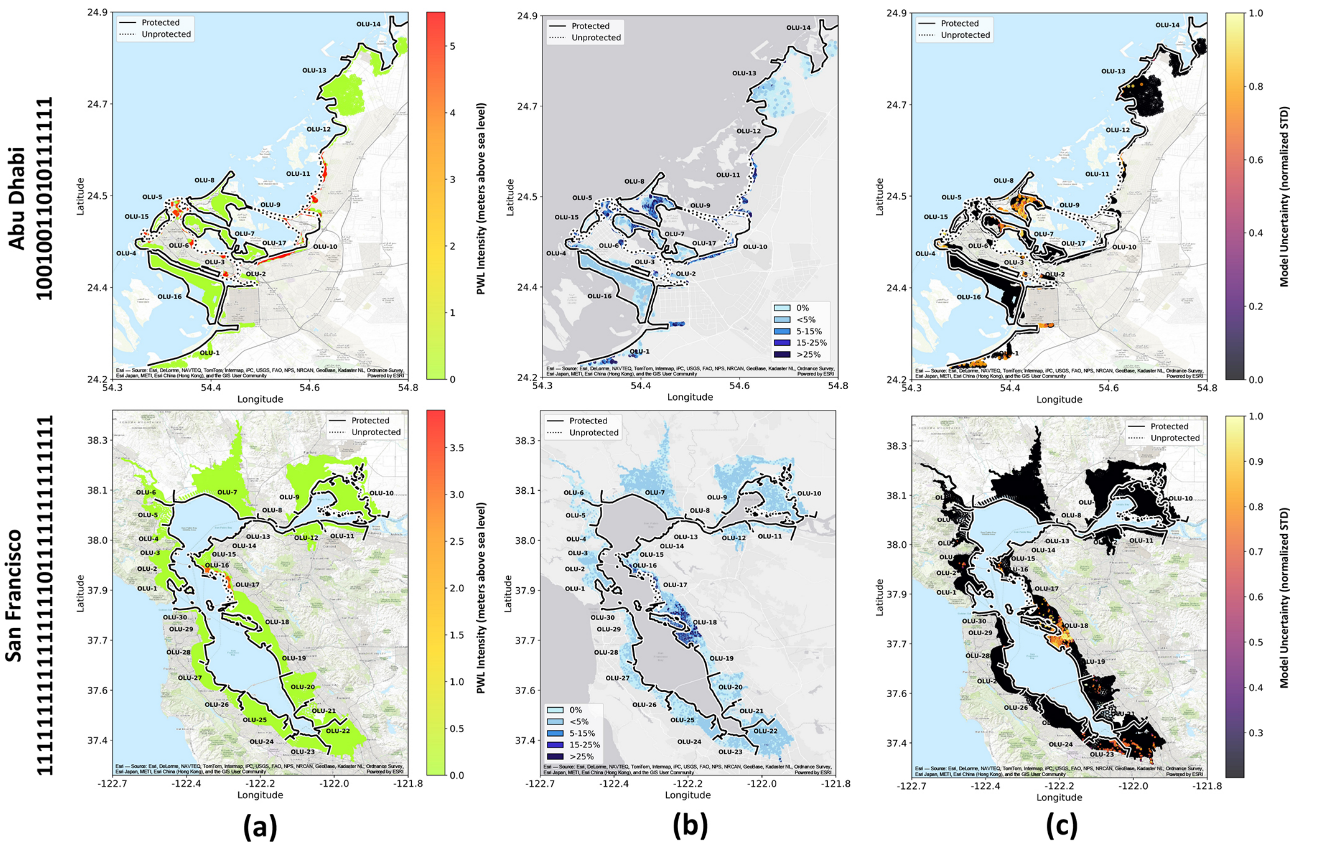

Figure 12. Predictive uncertainty maps derived from the deep ensemble for representative AD and SF scenarios. **(a)** Ground truth inundation. **(b)** Absolute error of the ensemble mean prediction. **(c)** Pixel-wise predictive uncertainty, calculated as the normalized standard deviation of the five models output. lighter colors indicate higher uncertainty.

coastal floods. The proposed model incorporates computer vision and a DL-inspired framework to address the complexities of diverse geographical regions, protection scenarios, and climate variability. The CASPIAN-v2 model effectively identifies critical inundation areas, handles uneven data distribution, and provides a clear rationale for its predictions. These strengths position CASPIAN-v2 as an essential tool for coastal resilience planning, helping decision makers, engineers, and legislators address current and future flood risks in the context of rapidly rising sea levels and changing coastal conditions.

Code and data availability. The code and data supporting the findings of this study are available on the project page: https://caspiannet.github.io (last access: 8 March 2026) and at https://doi.org/10.7910/DVN/RPHXGV (Hassan, 2026).

Supplement. The supplement related to this article is available online at https://doi.org/10.5194/hess-30-1333-2026-supplement.

Author contributions. BH, AK, ACHC, and SM contributed to the conceptualization and validation of research. BH and AK performed the formal analysis, investigation, and visualization; BH also developed the software. ACHC managed data curation, contributed to the methodology, and provided essential resources. SM secured funding, oversaw project administration, and offered supervision. All authors participated in writing the original draft and in reviewing and editing the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Financial support. This research has been supported by the New York University Abu Dhabi (NYUAD).

Review statement. This paper was edited by Lelys Bravo de Guenni and reviewed by three anonymous referees.

References

Abu Dhabi Urban Planning Council: Abu Dhabi 2030: Urban Structure Framework Plan, https://www.ecouncil.ae/PublicationsEn/plan-abu-dhabi-full-version-EN.pdf (last access: 6 October 2024), 2007.

Ali, M. H. M., Asmai, S. A., Abidin, Z. Z., Abas, Z. A., and Emran, N. A.: Flood prediction using deep learning models, Int. J. Adv. Comput. Sci. Appl., 13, 972–981, https://doi.org/10.14569/IJACSA.2022.01309112, 2022.

al Kabban, Marwa: Sea Level Rise Vulnerability Assessment for Abu Dhabi, United Arab Emirates, Student Paper, http://lup.lub.lu.se/student-papers/record/8998495 (last access: 19 December 2024), 2019.

Al Senafi, F. and Anis, A.: Shamals and climate variability in the Northern Arabian/Persian Gulf from 1973 to 2012, Int. J. Climatol., 35, 4509–4528, https://doi.org/10.1002/joc.4302, 2015.

Ardha, M., Chulafak, G. A., Anggraini, N., Syetiawan, A., and Khomarudin, M. R.: Flood inundation prediction model related to land subsidence with Lidar in North Coastal Jakarta, in: Eighth Geoinformation Science Symposium 2023: Geoinformation Science for Sustainable Planet, vol. 12977, 392–401, SPIE, https://doi.org/10.1117/12.3009685, 2024.

Barnard, P. L., van Ormondt, M., Erikson, L. H., Eshleman, J., Hapke, C., Ruggiero, P., Adams, P. N., and Foxgrover, A. C.: Development of the Coastal Storm Modeling System (CoSMoS) for predicting the impact of storms on high-energy, active-margin coasts, Natural Hazards, 74, 1095–1125, https://doi.org/10.1007/s11069-014-1236-y, 2014.

Barnard, P. L., Befus, K. M., Danielson, J. J., Engelstad, A. C., Erikson, L. H., Foxgrover, A. C., Hayden, M. K., Hoover, D. J., Leijnse, T. W. B., Massey, C., McCall, R., Nadal-Caraballo, N. C., Nederhoff, K., O'Neill, A. C., Parker, K. A., Shirzaei, M., Ohenhen, L. O., Swarzenski, P. W., Thomas, J. A., van Ormondt, M., Vitousek, S., Vos, K., Wood, N. J., Jones, J. M., and Jones, J. L.: Projections of multiple climate-related coastal hazards for the US Southeast Atlantic, Nat. Clim. Change, 15, pages 101–109, https://doi.org/10.1038/s41558-024-02180-2, 2024.

Beagle, J., Lowe, J., McKnight, K., Safran, S., Tam, L., and Szambelan, S. J.: San Francisco Bay shoreline adaptation atlas: Working with nature to plan for sea level rise using operational landscape units, SFEI publication# 915, SFEI, https://trid.trb.org/View/1605924 (last access: 25 November 2024), 2019.

Bentivoglio, R., Isufi, E., Jonkman, S. N., and Taormina, R.: Deep learning methods for flood mapping: a review of existing applications and future research directions, Hydrol. Earth Syst. Sci., 26, 4345–4378, https://doi.org/10.5194/hess-26-4345-2022, 2022.

California Energy Commission: San Francisco Bay Area Report, California's Fourth Climate Change Assessment, https://www.energy.ca.gov/sites/default/files/2019-11/Reg_Report-SUM-CCCA4-2018-005_SanFranciscoBayArea_ADA.pdf (last access: 4 October 2024), 2018.

Cao, A., Esteban, M., Valenzuela, V. P. B., Onuki, M., Takagi, H., Thao, N. D., and Tsuchiya, N.: Future of Asian Deltaic Megacities under sea level rise and land subsidence: current adaptation pathways for Tokyo, Jakarta, Manila, and Ho Chi Minh City, Current Opinion in Environmental Sustainability, 50, 87–97, https://doi.org/10.1016/j.cosust.2021.02.010, 2021.

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., and Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation, in: European conference on computer vision, 205–218, Springer, https://doi.org/10.1007/978-3-031-25066-8_9, 2022.

Chang, S. E., Yip, J. Z., Conger, T., Oulahen, G., Gray, E., and Marteleira, M.: Explaining communities' adaptation strategies for coastal flood risk: Vulnerability and institutional factors, J. Flood Risk Manage., 13, e12646, https://doi.org/10.1111/jfr3.12646, 2020.

Chen, G., Hou, J., Liu, Y., Xue, S., Wu, H., Wang, T., Lv, J., Jing, J., and Yang, S.: Urban inundation rapid prediction method based on multi-machine learning algorithm and rain pattern analysis, J. Hydrol., 633, 131059, https://doi.org/10.1016/j.jhydrol.2024.131059, 2024.

Chow, A. C. and Sun, J.: Combining Sea level rise inundation impacts, tidal flooding and extreme wind events along the Abu Dhabi coastline, Hydrology, 9, 143, https://doi.org/10.3390/hydrology9080143, 2022.

De Almeida, G. A. and Bates, P.: Applicability of the local inertial approximation of the shallow water equations to flood modeling, Water Resour. Res., 49, 4833–4844, https://doi.org/10.1002/wrcr.20366, 2013.

Deltares: Delft3d, https://oss.deltares.nl/web/delft3d, last access: 12 January 2025.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, 248–255, Ieee, https://doi.org/10.1109/CVPR.2009.5206848, 2009.

Du, B., Wang, M., Zhang, J., Chen, Y., and Wang, T.: Urban flood prediction based on PCSWMM and stacking integrated learning model, Natural Hazards, 121, 1971–1995, https://doi.org/10.1007/s11069-024-06893-7, 2024.

Griggs, G. and Reguero, B. G.: Coastal adaptation to climate change and sea-level rise, Water, 13, 2151, https://doi.org/10.3390/w13162151, 2021.

Guo, Z., Leitao, J. P., Simões, N. E., and Moosavi, V.: Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks, J. Flood Risk Manage., 14, e12684, https://doi.org/10.1111/jfr3.12684, 2021.

Haigh, I. D., Pickering, M. D., Green, J. A. M., Arbic, B. K., Arns, A., Dangendorf, S., Hill, D. F., Horsburgh, K., Howard, T., Idier, D., Jay, D. A., Jänicke, L., Lee, S. B., Müller, M., Schindelegger, M., Talke, S. A., Wilmes, S.-B., and Woodworth, P. L.: The tides they are a-Changin': A comprehensive review of past and future nonastronomical changes in tides, their driving mechanisms, and future implications, Rev. Geophys., 58, e2018RG000636, https://doi.org/10.1029/2018RG000636, 2020.

Hallegatte, S., Green, C., Nicholls, R. J., and Corfee-Morlot, J.: Future flood losses in major coastal cities, Nat. Clim. Change, 3, 802–806, https://doi.org/10.1038/nclimate1979, 2013.

Hartnett, M. and Nash, S.: High-resolution flood modeling of urban areas using MSN_Flood, Water Science and Engineering, 10, 175–183, https://doi.org/10.1016/j.wse.2017.10.003, 2017.

Hassan, B.: San Francisco Bay Area Coastal Flood Prediction Dataset for Deep Learning, Harvard Dataverse, V1 [data set], https://doi.org/10.7910/DVN/RPHXGV, 2026.

Hassani, A., Walton, S., Shah, N., Abuduweili, A., Li, J., and Shi, H.: Escaping the big data paradigm with compact transformers, arXiv [preprint], arXiv:2104.05704, https://doi.org/10.48550/arXiv.2104.05704, 2021.

Holleman, R. C. and Stacey, M. T.: Coupling of sea level rise, tidal amplification, and inundation, J. Phys. Oceanogr., 44, 1439–1455, https://doi.org/10.1175/JPO-D-13-0214.1, 2014.

Huber, P. J.: Robust estimation of a location parameter, in: Breakthroughs in statistics: Methodology and distribution, 492–518, Springer, https://doi.org/10.1007/978-1-4612-4380-9_35, 1992.

Hummel, M. A., Griffin, R., Arkema, K., and Guerry, A. D.: Economic evaluation of sea-level rise adaptation strongly influenced by hydrodynamic feedbacks, P. Natl. Acad. Sci. USA, 118, e2025961118, https://doi.org/10.1073/pnas.2025961118, 2021.

IPCC: Climate change 2021: the physical science basis, Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change, 2, 2391, https://doi.org/10.1017/9781009157896, 2021.

Jia, G., Taflanidis, A. A., Nadal-Caraballo, N. C., Melby, J. A., Kennedy, A. B., and Smith, J. M.: Surrogate modeling for peak or time-dependent storm surge prediction over an extended coastal region using an existing database of synthetic storms, Natural Hazards, 81, 909–938, https://doi.org/10.1007/s11069-015-2111-1, 2016.

Karapetyan, A., Chow, A. C., and Madanat, S.: Deep vision-based framework for coastal flood prediction under sea level rise and shoreline protection, Sci. Rep., 16, 3663, https://doi.org/10.1038/s41598-025-33803-z, 2026.

Koenker, R. and Bassett Jr., G.: Regression quantiles, Econometrica, 46, 33–50, https://doi.org/10.2307/1913643, 1978.

Kyprioti, A. P., Taflanidis, A. A., Nadal-Caraballo, N. C., and Campbell, M.: Storm hazard analysis over extended geospatial grids utilizing surrogate models, Coastal Eng., 168, 103855, https://doi.org/10.1016/j.coastaleng.2021.103855, 2021.

Lakshminarayanan, B., Pritzel, A., and Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles, Advances In Neural Information Processing Systems, 3, 6402–6413, 2017.

Langodan, S., Cavaleri, L., Benetazzo, A., Bertotti, L., Dasari, H. P., and Hoteit, I.: The peculiar wind and wave climatology of the Arabian Gulf, Ocean Eng., 290, 116158, https://doi.org/10.1016/j.oceaneng.2023.116158, 2023.

Lewis, A.: After a Decade of Planning, New York City Is Raising Its Shoreline, Yale School of the Environment, https://e360.yale.edu/features/new-york-city-climate-plan-sea-level-rise (last access: 2 December 2024), 2023.

Li, D., Anis, A., and Al Senafi, F.: Physical response of the Northern Arabian Gulf to winter Shamals, J. Marine Syst., 203, 103280, https://doi.org/10.1016/j.jmarsys.2019.103280, 2020.

Li, Z. and Hodges, B. R.: Modeling subgrid-scale topographic effects on shallow marsh hydrodynamics and salinity transport, Adv. Water Resour., 129, 1–15, https://doi.org/10.1016/j.advwatres.2019.05.004, 2019.

Melville-Rea, H., Eayrs, C., Anwahi, N., Burt, J. A., Holland, D., Samara, F., Paparella, F., Al Murshidi, A. H., Al-Shehhi, M. R., and Holland, D. M.: A roadmap for policy-relevant sea-level rise research in the United Arab Emirates, Front. Marine Sci., 8, 670089, https://doi.org/10.3389/fmars.2021.670089, 2021.

Mosavi, A., Ozturk, P., and Chau, K.-W.: Flood prediction using machine learning models: Literature review, Water, 10, 1536, https://doi.org/10.3390/w10111536, 2018.

Muñoz, D. F., Moftakhari, H., and Moradkhani, H.: Quantifying cascading uncertainty in compound flood modeling with linked process-based and machine learning models, Hydrol. Earth Syst. Sci., 28, 2531–2553, https://doi.org/10.5194/hess-28-2531-2024, 2024.

Neal, J., Schumann, G., and Bates, P.: A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas, Water Resour. Res., 48, https://doi.org/10.1029/2012WR012514, 2012.

Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D., Kratzert, F., Elidan, G., Dror, G., Begelman, G., Nearing, G., Shalev, G., Noga, H., Shavitt, I., Yuklea, L., Royz, M., Giladi, N., Peled Levi, N., Reich, O., Gilon, O., Maor, R., Timnat, S., Shechter, T., Anisimov, V., Gigi, Y., Levin, Y., Moshe, Z., Ben-Haim, Z., Hassidim, A., and Matias, Y.: Flood forecasting with machine learning models in an operational framework, Hydrol. Earth Syst. Sci., 26, 4013–4032, https://doi.org/10.5194/hess-26-4013-2022, 2022.

Nithila Devi, N. and Kuiry, S. N.: A novel local-inertial formulation representing subgrid scale topographic effects for urban flood simulation, Water Resour. Res., 60, e2023WR035334, https://doi.org/10.1029/2023WR035334, 2024.

Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D.: Attention u-net: Learning where to look for the pancreas, arXiv [preprint], arXiv:1804.03999, https://doi.org/10.48550/arXiv.1804.03999, 2018.

Oppenheimer, M., Glavovic, B. C., Hinkel, J., van de Wal, R., Magnan, A. K., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., DeConto, R. M., Ghosh, T., Hay, J., Isla, F., Marzeion, B., Meyssignac, B., and Sebesvari, Z.: Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities, 321–446, Cambridge University Press, https://doi.org/10.1017/9781009157964.006, 2022.

Pal, I., Kumar, A., and Mukhopadhyay, A.: Risks to Coastal Critical Infrastructure from Climate Change, Annu. Rev. Environ. Resour., 48, 681–712, https://doi.org/10.1146/annurev-environ-112320-101903, 2023.

Papacharalambous, M., Davis, M., Marshall, W., Weems, P., and Rothenberg, R.: Greater New Orleans Urban Water Plan: Implementation, Waggonner & Ball Architects: New Orleans, LA, USA, https://wbae.com/wp-content/uploads/2021/11/GNO-Urban-Water-Plan_Implementation_03Oct2013.pdf (last access: 8 March 2026), 2013.

Rohmer, J., Sire, C., Lecacheux, S., Idier, D., and Pedreros, R.: Improved metamodels for predicting high-dimensional outputs by accounting for the dependence structure of the latent variables: application to marine flooding, Stochastic Environmental Research and Risk Assessment, 37, 2919–2941, https://doi.org/10.1007/s00477-023-02426-z, 2023.

Saleh, R. A. and Saleh, A.: Statistical properties of the log-cosh loss function used in machine learning, arXiv [preprint], arXiv:2208.04564, https://doi.org/10.48550/arXiv.2208.04564, 2022.

Sanders, B. F. and Schubert, J. E.: PRIMo: Parallel raster inundation model, Adv. Water Resour., 126, 79–95, https://doi.org/10.1016/j.advwatres.2019.02.007, 2019.

Sun, J., Chow, A. C., and Madanat, S. M.: Multimodal transportation system protection against sea level rise, Transportation Research Part D: Transport and Environment, 88, 102568, https://doi.org/10.1016/j.trd.2020.102568, 2020.

United States Geological Survey: Modeled surface waves from winds in South San Francisco Bay, https://www.usgs.gov/data/modeled-surface-waves-winds-south-san-francisco-bay (last access: 18 September 2024), https://doi.org/10.5066/P9QH0GU5, 2024.

van de Wal, R., Melet, A., Bellafiore, D., Camus, P., Ferrarin, C., Oude Essink, G., Haigh, I. D., Lionello, P., Luijendijk, A., Toimil, A., Staneva, J., and Vousdoukas, M.: Sea Level Rise in Europe: Impacts and consequences, in: Sea Level Rise in Europe: 1st Assessment Report of the Knowledge Hub on Sea Level Rise (SLRE1), edited by: van den Hurk, B., Pinardi, N., Kiefer, T., Larkin, K., Manderscheid, P., and Richter, K., Copernicus Publications, State Planet, 3-slre1, 5, https://doi.org/10.5194/sp-3-slre1-5-2024, 2024.

Wang, R.-Q., Herdman, L. M., Erikson, L., Barnard, P., Hummel, M., and Stacey, M. T.: Interactions of estuarine shoreline infrastructure with multiscale sea level variability, J. Geophys. Res.-Oceans, 122, 9962–9979, https://doi.org/10.1002/2017JC012730, 2017.

Wang, R.-Q., Stacey, M. T., Herdman, L. M. M., Barnard, P. L., and Erikson, L.: The influence of sea level rise on the regional interdependence of coastal infrastructure, Earth's Future, 6, 677–688, https://doi.org/10.1002/2017EF000742, 2018a.

Wang, Y., Chen, A. S., Fu, G., Djordjević, S., Zhang, C., and Savić, D. A.: An integrated framework for high-resolution urban flood modelling considering multiple information sources and urban features, Environ. Model. Softw., 107, 85–95, https://doi.org/10.1016/j.envsoft.2018.06.010, 2018b.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S.: Cbam: Convolutional block attention module, in: Proceedings of the European conference on computer vision (ECCV), 3–19, https://doi.org/10.1007/978-3-030-01234-2_1, 2018.

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K.: Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 1492–1500, https://doi.org/10.1109/CVPR.2017.634, 2017.

Zhao, G., Bates, P., Neal, J., and Pang, B.: Design flood estimation for global river networks based on machine learning models, Hydrol. Earth Syst. Sci., 25, 5981–5999, https://doi.org/10.5194/hess-25-5981-2021, 2021.

Zhou, Q., Teng, S., Situ, Z., Liao, X., Feng, J., Chen, G., Zhang, J., and Lu, Z.: A deep-learning-technique-based data-driven model for accurate and rapid flood predictions in temporal and spatial dimensions, Hydrol. Earth Syst. Sci., 27, 1791–1808, https://doi.org/10.5194/hess-27-1791-2023, 2023.

Zuhairi, A. H., Yakub, F., Zaki, S. A., and Ali, M. S. M.: Review of flood prediction hybrid machine learning models using datasets, in: IOP Conference Series: Earth and Environmental Science, IOP Publishing, 1091, 012040, https://doi.org/10.1088/1755-1315/1091/1/012040, 2022.