scientific data

Article in Press

Check for updates

OPEN

DATA DESCRIPTOR

A high-resolution global flood susceptibility dataset derived from multi-source earth observation and geospatial data

Mirza Waleed[1,2], Muhammad Sajjad[3]✉, Sami G. Al-Ghamdi[2] & Meng Gao[1]

Floods are among the most damaging disasters worldwide, necessitating high-resolution modeling of where terrain and environmental conditions predispose societies to flooding. While geospatial data are invaluable for flood susceptibility mapping, comprehensive and globally consistent datasets remain scarce. We present the Global Flood Susceptibility Map (GFSM v1), a globally harmonized 30 m flood susceptibility dataset produced by incorporating multi-source flood conditioning factors accounting for topographic, hydrological, meteorological, and anthropogenic characteristics within a machine learning framework. A gradient-boosted tree model was trained on ~17,000 curated grid tiles with ~30.45 million training samples, using a global flood inundation dataset as training labels and nine conditioning factors (elevation, slope, aspect, wetness index, vegetation index, height above nearest drainage, water/roads proximity, and rainfall frequency). The model was deployed based on geopolitical and climate strata (units n = 192) to generate global pixel-level susceptibility. Rigorous cross-validation indicates strong model predictive accuracy (median area-under-curve ~0.95), and the output captures localized flood-prone regions worldwide. GFSM v1 offers a global open-source, high-resolution resource to support informed flood risk management, particularly in data-sparse regions.

Background & Summary

Of all disasters, flooding poses some of the worst threats to communities, infrastructure, and economies worldwide. Identifying areas inherently susceptible to flooding – irrespective of specific flood event frequency or magnitude – is crucial for disaster risk reduction, informed decision-making, and climate change adaptation[1]. Flood susceptibility mapping (FSM) provides such information by highlighting zones where physiographic and environmental conditions favor flood occurrence[2]. Traditionally, FSM studies have been conducted at local or regional scales, combining flood conditioning factors (FCFs) like topography, land cover, and drainage characteristics to infer flood-prone areas using statistical or machine learning (ML) models[2–7]. However, no high-resolution FSM product exists at a global scale, representing a significant gap for comparative flood risk assessment across regions—particularly in data scarce countries. We address this gap by producing the Global Flood Susceptibility Map (GFSM) v1[8], a global high-resolution 30 m flood susceptibility dataset derived through a harmonized, unified, and data driven methodology.

Various global flood hazard models and maps have been produced in recent years (e.g., the World Resources Institute Aqueduct Flood Hazard Maps[9] and JRC's flood hazard layers[10]). Such hazard models primarily rely on hydrodynamic simulations and historical flow data. While invaluable, they often operate at ~1–10 km or coarser resolution. This situation results in under- or overestimation as such approaches do not inherently account for local terrain effects[11]. In contrast, FSM focuses on inherent landscape propensity to flooding, usually by analyzing static factors (e.g., elevation, slope, drainage proximity) that contribute to inundation when extreme rainfall or overflow occurs. FSM thus complements hazard models: it does not predict when a flood will occur, but

[1]Department of Geography, Hong Kong Baptist University, Hong Kong, Hong Kong SAR, China. [2]Environmental Science and Engineering Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia. [3]Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China. ✉e-mail: muhammadsajjad@cuhk.edu.hk

rather where conditions are favorable for flooding, serving as a baseline indicator of flood risk potential. Modern FSM approaches typically achieve this by using ML to model the complex, non-linear relationships between known flood occurrences and a diverse set of earth observation based FCFs. This data-driven methodology, which our work adopts at a global scale, learns from these patterns to generalize flood likelihood in high detail, even in data-poor regions.

Recent global flood products also illustrate the distinction between susceptibility screening and physically based hazard simulation. Liu *et al.* (2023) developed a global flood susceptibility mapping framework by integrating deep learning with hydrological information[12], whereas Wing *et al.* (2024) presented a 30 m global flood inundation/hazard modeling framework for multiple flood perils and climate scenarios[13]. GFSM v1 differs in scope and intended use. It does not simulate event-specific inundation depth, return-period hazard, or flood damages. Instead, it provides a globally harmonized 30 m susceptibility screening layer derived from high-resolution conditioning factors and Aqueduct baseline flood-hazard labels. Thus, GFSM v1 is complementary to global hydrodynamic flood models and is intended primarily for screening, exposure overlay, and prioritizing areas where more detailed local or scenario-specific flood analyses may be warranted.

Previous studies have demonstrated the efficacy of ML for FSM at sub-national scales. For example, a recent work provides a comparative evaluation of numerous ML algorithms for high-resolution FSM and reported that tree-based ensemble methods (particularly gradient-boosted trees like XGBoost) achieved the highest predictive performance[14]. Another study on high-resolution FSM for Pakistan combined topographic indices, earth observation data, and socio-environmental indices to identify national flood susceptibility hotspots[15]. They found that approximately 29% of Pakistan's area fell under critical flood susceptibility, with ~95 million people (~47% of the population) living in high-susceptibility zones. However, implementing such frameworks on a global scale is challenging due to constraints, such as computational power, lack of high-resolution conditioning factors datasets, issues related to data harmonization, and absence of globally applicable FSM schemas. We here resolve such scientific and implementation bottlenecks via leveraging state-of-the-art computational technology integrated with ML and multi-source geospatial data to produce GFSM v1.

GFSM v1 was generated through a coupled workflow combining advanced geospatial data processing, high-performance computing, and ML. We acquired nine global FCF layers and preprocessed them to produce high-resolution (30 m) (details provided later in subsection *Input Datasets and Preprocessing*) input layers that represent the principal physical and environmental determinants of flooding. These include topographic factors—derived from a high-accuracy digital elevation model—such as slope and aspect. Hydrological factors include proximity to rivers and other water bodies, indicating potential exposure to overflow, Topographic Wetness Index (TWI), and Height Above Nearest Drainage (HAND), which describe terrain control on water accumulation. Meteorological conditions are represented by long-term rainfall frequency, serving as a proxy for climate-driven flood likelihood. Finally, anthropogenic influences are captured through distance to roads and the Normalized Difference Vegetation Index (NDVI), the latter provides information regarding vegetation/surface-cover conditions influencing infiltration and runoff. All final input layers were prepared globally at 30 m resolution and rigorously co-registered. We used global flood inundation maps (Aqueduct Flood Hazard Maps) with a moderate return period (5-year flood) under present climate conditions[9]. By merging coastal and riverine inundation depth maps from these datasets, we obtained binary flood labels (flooded vs not flooded) at ~0.5–1 km scale, which then were further resampled and aligned with the 30 m FCF grids.

To manage the enormous data volume and geographic extent, we implemented a tiling and regional modeling scheme. The planet's land surface was partitioned into ~17,176 grid tiles based on the Sentinel-2 global tiling system[16] (Military Grid Reference System, 100 km UTM zones). After filtering out tiles lacking necessary data (e.g., ocean tiles or areas without satellite coverage), we retained 17,069 tiles covering all inhabited landmasses. For modeling purposes, adjacent tiles were aggregated into "units" corresponding to countries or sub-national regions (to ensure sufficient training samples per model and avoid edge discontinuities). Specifically, each country was treated as a unit, except a few very large countries (e.g., USA, China, Russia, India, Brazil, Australia) which were subdivided by major climate zones using the Köppen climate classification (e.g., "USA_Temperate", "USA_Arid")[17]. This resulted in 192 modeling units worldwide, each representing a reasonably homogeneous region in terms of geography and climate (see Fig. 1 in methods section). We trained a separate ML model for each unit, allowing the susceptibility relationships to adjust to regional characteristics while still being automated and standardized globally.

The outcome of this framework – GFSM v1 – is a flood susceptibility dataset at 30 m resolution stored as individual tiles globally. To facilitate usability, we provide the susceptibility output in a five-class discrete format (Very Low, Low, Moderate, High, Very High), which is a common scheme in susceptibility mapping[3,7,15]. The classification was derived by slicing the continuous susceptibility score output of the ML model into equal intervals (0–0.2 for Very Low, 0.2–0.4 Low, etc.), thus maintaining consistency of class breakpoints across the globe. This enables a straightforward comparison of susceptibility levels between different areas. GFSM v1 is rigorously validated through cross-validation and performance evaluation against the training labels (see Technical Validation). Overall, the ML models achieved high accuracy in distinguishing flooded vs non-flooded locations within each region (median AUC ~0.95 and F1-score ~0.89 across all 192 units). The consistency of performance across diverse environments – from arid climates to tropical monsoons – underscores the robustness of the chosen FCFs and modeling approach. While some regional variations exist (e.g., slightly lower accuracy in highly engineered landscapes), the GFSM dataset is broadly reliable and highlights meaningful localized spatial patterns of flood susceptibility on a global scale.

GFSM v1 is released as an open dataset[8] to support various applications. By overlaying the susceptibility map with population, infrastructure, or land-use data, stakeholders can identify communities at risk and prioritize flood mitigation investments. The dataset can also serve as a baseline for global flood risk models or be incorporated into compound hazard analyses (e.g., flooding combined with socioeconomic vulnerability). We anticipate that GFSM will be especially valuable for data-sparse regions where detailed flood studies are lacking. In the following sections, we describe in detail the data sources, sampling and model training, and validation of GFSM v1.

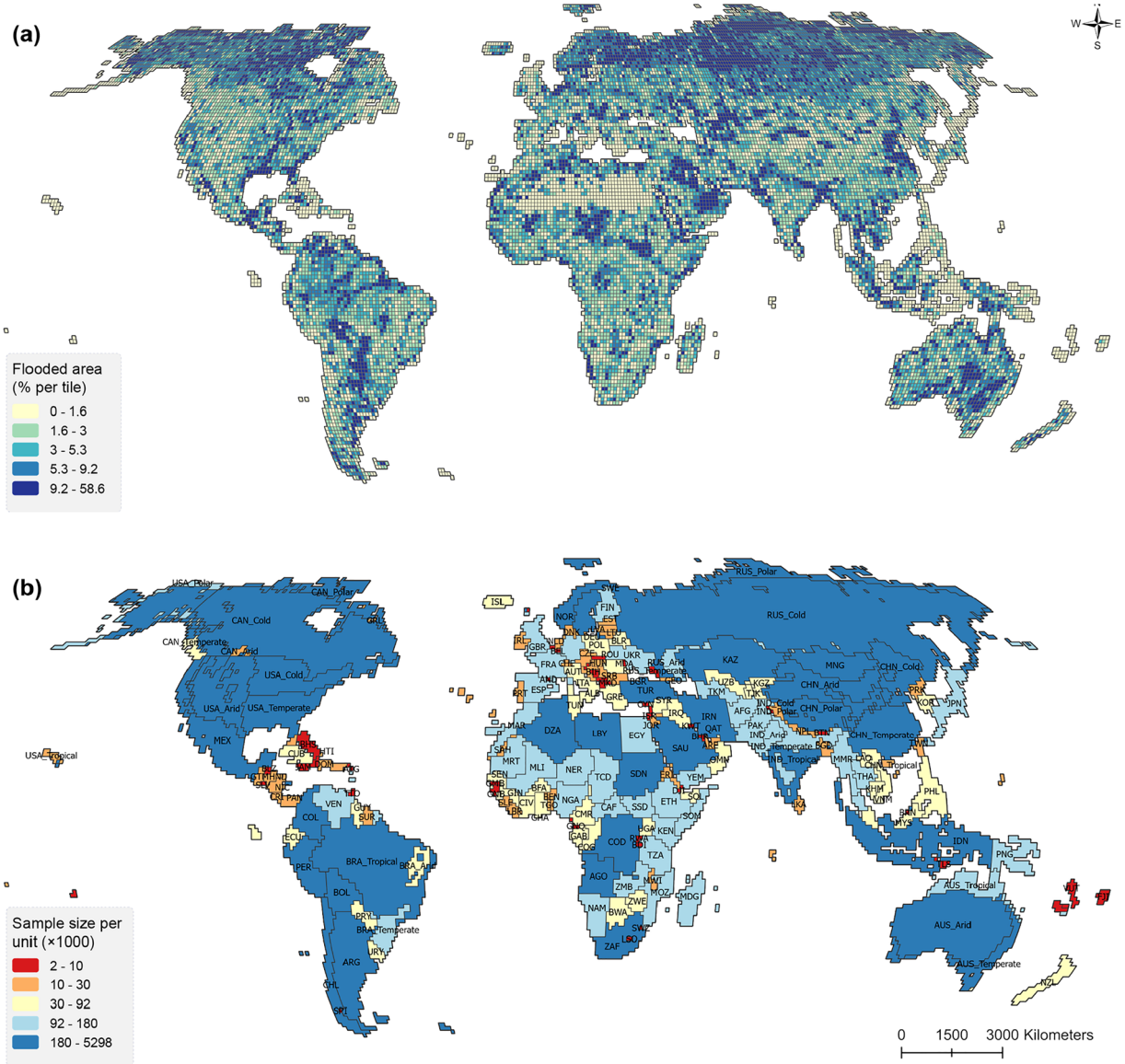

Fig. 1 (**a**) Percentage of each tile covered by the Aqueduct flood-hazard layer, ranging from 0 to 58.6%. (**b**) Global modeling units and the number of training samples aggregated within each unit, illustrating the spatial distribution of sample sizes across the 192 units.

Methods

Input Datasets and Preprocessing. *Flood hazard training data.* Rather than relying on sparse and event-driven flood observations, we deliberately adopted the modeled hazard layers from the World Resources Institute Aqueduct Floods dataset[9]. These products offer globally consistent coverage, harmonized return periods, and physically based representation of both riverine and coastal flooding, making them more suitable as training labels for a uniform, global-scale susceptibility model. Specifically, we used the Aqueduct Flood Hazard Maps Version 2, which provides global inundation extents and depths for both riverine and coastal flooding under baseline (historical) conditions. We selected the relatively frequent flood scenario of a 5-year return period, representing a high-probability, lower-magnitude flood event that still causes significant flooding in low-lying areas. Two separate hazard layers – coastal flood inundation (including land subsidence effects) and riverine flood inundation (baseline scenario using the WATCH climate forcing) – were obtained from the Aqueduct dataset using the Google Earth Engine (GEE) platform. We merged these by taking the maximum inundation depth at each location (depth > 0), thereby capturing any flooding from either source (this approximates compound flood susceptibility in estuaries and deltas). The merged flood hazard data has a native resolution of ~1 km (dependent on latitude) and provides a binary indication of flooded vs non-flooded area for the given event severity. For instance, any pixel with a reported inundation depth > 0 was labeled as "flood" (positive class), and all other areas as "non-flood" (negative class). This binary flood mask was then resampled to 30 m and aligned to each tile's grid (using nearest-neighbor assignment to preserve the flood extent shapes). The result is a training label raster per tile indicating where flooding would occur in a 5-year event under the present climate. While this Aqueduct

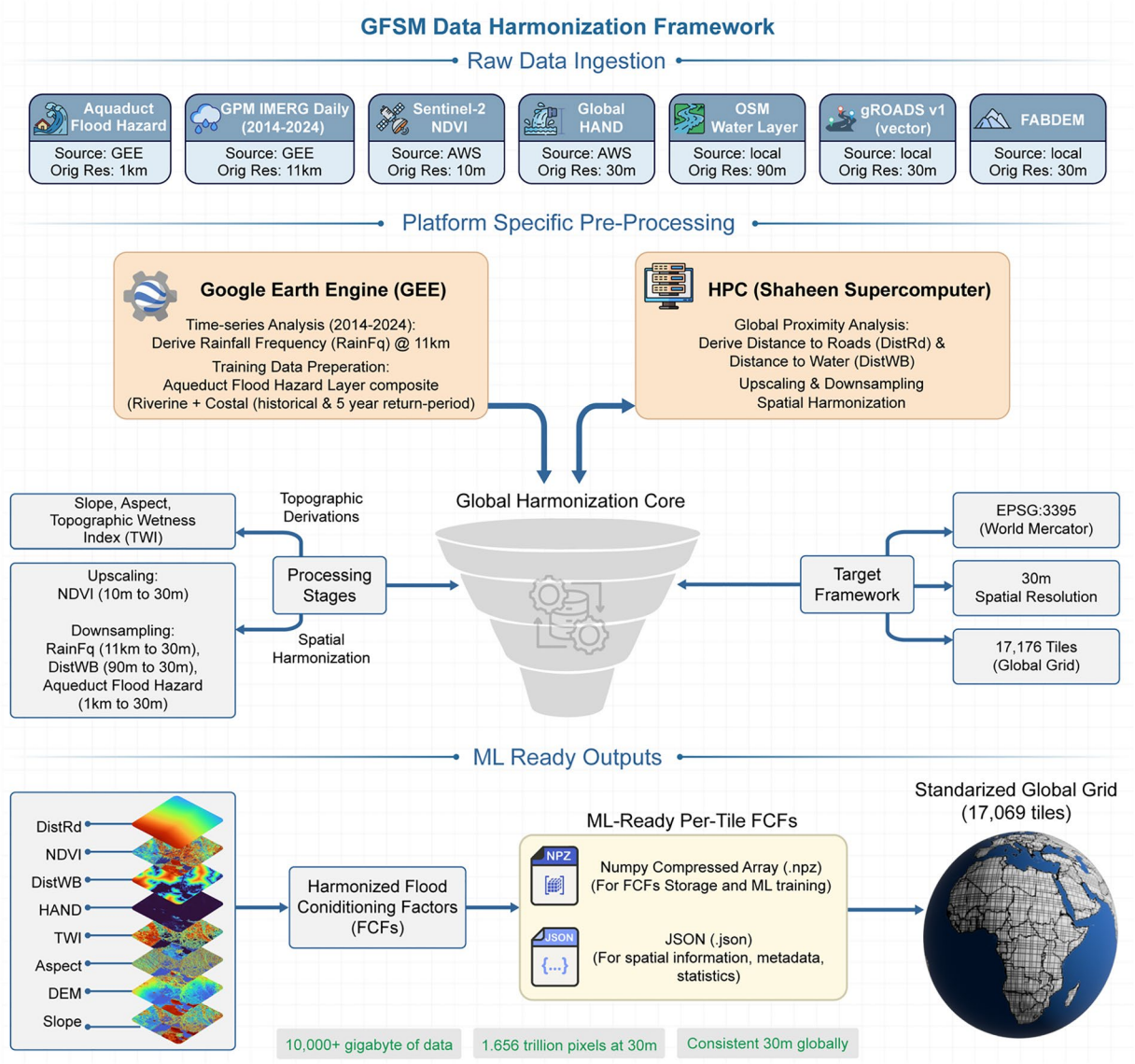

Fig. 2 Schematic showing workflow for preprocessing and harmonizing global input datasets into machine-learning-ready flood conditioning factors (FCFs). Blue arrows indicate data flow from raw data ingestion to platform-specific preprocessing, spatial harmonization, and final outputs. Orange boxes distinguish processing conducted in Google Earth Engine (GEE) and on the Shaheen high-performance computing (HPC) system. The central funnel represents the global harmonization core, where datasets with different projections, resolutions, and formats are standardized to a common target framework: World Mercator projection (EPSG:3395), 30 m spatial resolution, and a global tile grid. The stacked raster layers represent harmonized FCFs, while the NPZ and JSON icons indicate compressed per-tile feature arrays and metadata/statistics for machine learning.

hazard-based label has limitations (being model-derived and scenario-specific), it provides a consistent global baseline for training the susceptibility model, which can further be enhanced for future scenarios and extreme cases (i.e., a 100 or 500-year flood). Figure 1 shows the total tiles with the Aqueduct flood hazard area (Fig. 1a), and units with sample size (Fig. 1b). Details of units are provided later in the Sampling Strategy and Model Training section.

Flood conditioning factors. We compiled nine globally consistent FCFs based on their documented influence on flood occurrence, and open-source availability[6,14,15,18] (see Table S2 for more details regarding sources of these FCFs). All FCFs were processed and harmonized to a spatial resolution of 30 meters and a common projection (World Mercator, EPSG:3395) to ensure spatial alignment. The processing involved standardizing the FCFs to a uniform scale, which included converting float decimal values to a suitable range of up to 65,000, and changing the data type to 16-bit unsigned integer (uint16). In general, these FCFs are grouped into four categories representing the principal physical drivers of flooding (Fig. 2).

- **Anthropogenic and Surface Factors:** These factors represent human modification and surface-cover conditions that may influence local hydrology. We included distance to the nearest road (DistRd), calculated from the Global Roads Open Access Dataset (gROADS v1)[19], as an indirect indicator of human access, settlement patterns, and possible drainage disruption. Vegetation/surface-cover influence was represented by the Normalized Difference Vegetation Index (NDVI), derived from the ESA WorldCover Sentinel-1 and Sentinel-2 10 m Annual Composites dataset hosted on AWS[20] (see Table S2 for details).
- **Hydrological Factors:** This category captures the landscape's interaction with water, including proximity to flood sources and propensity for water accumulation. We derived the Euclidean distance to the nearest surface water body (DistWB) using the OpenStreetMap global water layer[21]. We also incorporated two critical terrain-hydrology indices: the Height Above Nearest Drainage (HAND)[22], which normalizes elevation relative to the local drainage network to delineate floodplains, and the Topographic Wetness Index (TWI), derived from the digital elevation model (DEM) which estimates soil saturation propensity.
- **Topographic Factors:** These variables describe the terrain's control on water flow. We used the FABDEM dataset[23], a global 30 m DEM with vegetation and building artifacts removed, as the primary data source. From the DEM, we derived slope and aspect to characterize local steepness and orientation.
- **Meteorological Factors:** To account for the climatic drivers of flooding, we included a rainfall frequency (RainFq) layer. This was derived from the Global Precipitation Measurement (GPM IMERG) dataset[24], representing the average annual number of days with precipitation exceeding 10 mm between 2014 and 2024.

Sampling Strategy and Model Training. Preparing an unbiased and representative training sample was critical given the global scale and the imbalanced nature of flood occurrence in different regions (floods cover only a fraction of land). We adopted a stratified random sampling approach with balancing to ensure that the model sees both flooded and non-flooded examples in equal measure. For each tile, we first applied a mask to exclude tiles that had a negligible flood presence: any tile with less than 0.6% of its area marked as flooded in the hazard layer was dropped from training sampling. The 0.6% flood-area threshold was applied only to training-sample extraction and not to the final prediction domain. Tiles below this threshold generally contained too few positive-label pixels to support stable tile-level balanced sampling, particularly when maintaining spatial separation among samples. Excluding these label-sparse tiles from sample extraction reduced the influence of near-zero positive-label areas on model training. However, tiles below the threshold were still predicted in the final GFSM product when they belonged to a trained modeling unit and had complete flood-conditioning-factor coverage. Thus, the threshold controlled training-sample reliability rather than removing dry or arid regions from the released map.

Within each tile, we drew 2,000 point samples stratified by class: 1,000 points randomly from flooded pixels and 1,000 from non-flooded pixels. This 1:1 ratio produces a balanced local training set, which is important because the true prevalence of flood pixels is very low (often $< 1\%$) and a model trained on raw proportions would be biased toward predicting "no flood" everywhere. By design, each tile thus contributed an equal number of flood and non-flood samples, and roughly equal total samples, regardless of its size or flood percentage. The sampling was random within each class mask, ensuring a good spread of locations (to avoid clustering, we enforced a minimum spacing distance of 500 m between sample points when possible). For each sampled point, the values of all FCF layers were extracted from the tile's npz data. These per-point feature vectors, along with the binary class label, formed the raw training data.

Moving forward, we aggregated samples at the unit level for model training. Rather than train separate models on each 100 km tile (which could cause edge artifacts and inconsistent class definitions between tiles), we combined all samples from tiles belonging to the same regional unit (country or country-climate zone). This yielded one training dataset per unit, typically containing tens of thousands of samples. Because we had pre-balanced samples within each tile, the combined unit sample was also roughly class-balanced. Importantly, this unit-based approach suggests that the model learns patterns from a broader spatial context, reducing overfitting individual tile idiosyncrasies and ensuring that susceptibility gradients are continuous across tile boundaries. As mentioned earlier in Background and Summary section, each country was treated as a unit, except a few very large countries (e.g., USA, China, Russia, India, Brazil, and Australia), which were subdivided by major climate zones using the Köppen climate classification (e.g., "USA_Temperate", "USA_Arid")[17]. This resulted in 192 modeling units worldwide, each representing a reasonably homogeneous region in terms of geography and climate (see Fig. 1b in methods section).

For each modeling unit, we trained a gradient boosting ML model to classify flood susceptibility. We chose the Extreme Gradient Boosting algorithm (XGBoost) due to its high accuracy, efficiency on large datasets, and ability to handle nonlinear relationships. XGBoost has been shown to perform on par with or better than other algorithms (e.g., random forests, neural networks, etc.) in flood susceptibility applications with the ability to scale up while being computationally efficient[14,15,25–27]. Existing regional experiments[14,15] and recent studies[25,28–31] show that tree-based ensembles like XGBoost consistently rank among the top in predictive skill for flood mapping, which justifies their use for this global implementation. We enabled GPU acceleration for XGBoost (using the gpu_hist parameter) to expedite training on large datasets. Training was conducted on an NVIDIA CUDA-compatible GPU – in our case, we utilized an NVIDIA GPU on a local workstation and also had access to GPU and CPU nodes on the Shaheen HPC cluster at the King Abdullah University of Science and Technology for parallel processing. The detailed methodology framework workflow is illustrated in Fig. 3.

Hyperparameter optimization. We used a single globally optimized XGBoost parameter set for all modeling units to maintain comparability, reproducibility, and computational tractability across the global workflow. Unit-specific hyperparameter tuning could improve performance in some regions, but it could also increase

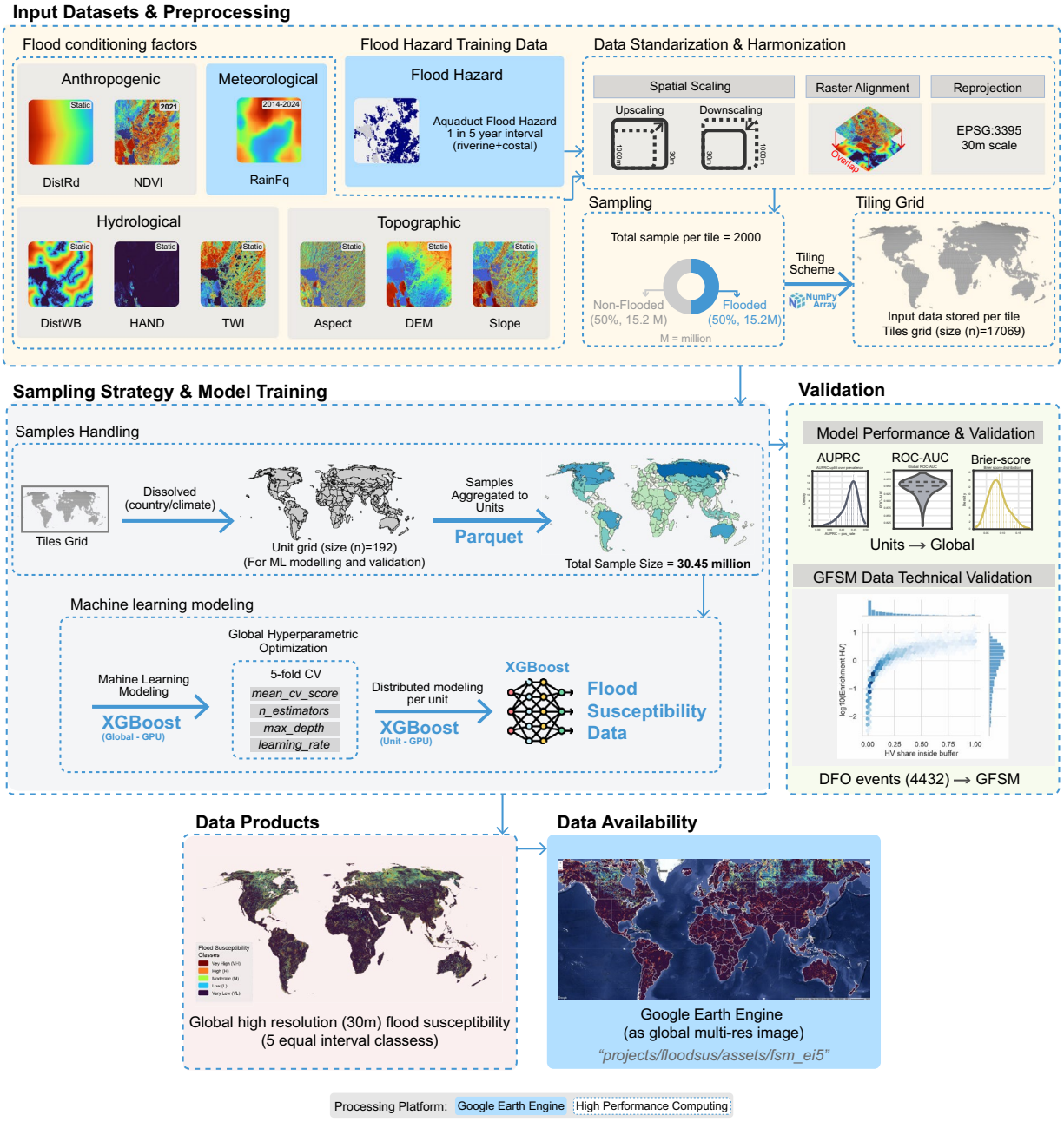

Fig. 3 Methodology framework to scale high-resolution flood susceptibility mapping globally.

overfitting risk in smaller or label-sparse units and would reduce consistency between units. A stratified random subset of approximately 200,000 samples was drawn from across different modeling units worldwide to represent diverse conditions. Using five-fold cross-validation, we explored variation in maximum tree depth, learning rate, subsampling of rows and columns, and L1/L2 regularization terms. The optimization revealed that a balance between model complexity and learning speed was crucial; moderate learning rates (~0.1) paired with a sufficient number of estimators (e.g., 1,000–1,500) consistently yielded the highest performance. Based on these evaluations, we selected a near-optimal yet computationally scalable configuration to apply uniformly across all 192 modeling units, ensuring comparability and reproducibility. The final parameters used were: *learning rate = 0.11, maximum depth = 13, subsample = 0.85, colsample_bytree = 0.90, reg_alpha = 3.6, and reg_lambda = 0.46.*

Model training and validation. For each modeling unit, the aggregated sample data were split into training and held-out test subsets using a deterministic stratified partition. This internal holdout was designed to evaluate model-label discrimination under the tile-balanced sampling design. Because samples from the same or neighboring tiles can share similar environmental conditions, the internal test should not be interpreted as a

fully spatially independent validation. We therefore report ROC-AUC, AUPRC, F1-score, and related metrics as internal discrimination/classification metrics against withheld Aqueduct-derived labels under the sampled-label distribution. Independent Dartmouth Flood Observatory event correspondence is reported separately as a complementary, non-training-label plausibility check. The training set was used to fit the XGBoost model, and the hold-out test set was kept entirely unseen during training to later evaluate performance. We did not use a separate validation set for hyperparameter tuning per unit (since we had globally optimized the parameters as described); however, during training, we did monitor performance on a subset of data for early stopping if needed. In practice, we trained a fixed number of boosting rounds (up to ~1200 trees) with early stopping not heavily utilized, to ensure full model convergence. Each model training over the Shaheen supercomputer, typically completed in few minutes on GPU (for ~100k samples per unit) under our chosen parameters, which was efficient enough to scale to ~192 units in a reasonable time.

Each unit-level model produced a continuous flood-class susceptibility score (ranging 0 to 1) for a given feature vector. We used gradient-boosted decision trees (XGBoost) with a binary logistic objective and stratified, class-balanced sampling to mitigate prevalence bias in the training data. For each unit, samples were split into training and held-out test subsets using a stratified partition; models were trained on the former and evaluated on the latter without post-hoc threshold tuning. We summarized discrimination with the area under the receiver operating characteristic curve (ROC-AUC) and the area under the precision–recall curve (AUPRC; informative under class imbalance). To report label-based skill, we applied a single operating threshold to the predicted probabilities (kept fixed across units unless otherwise noted) and computed precision, recall, F1-score, overall accuracy, Intersection-over-Union (IoU), and the confusion matrix. To assess probabilistic quality, we computed Brier scores and expected calibration error (ECE) from reliability curves (standard binned approach). Overfitting was screened by comparing training vs test loss and by checking that unit-level performance remained within the cross-unit interquartile ranges. All per-unit artifacts (metrics tables, ROC/PR curves, threshold sweeps, calibration curves, and feature importance) were exported to support reproducibility.

Beyond held-out model validation tests, we also designed external validation tracks to check spatial and temporal plausibility of the susceptibility maps. For external validation, we performed an event-based correspondence analysis using the Dartmouth Flood Observatory Global Active Archive of Large Flood Events (1985–2016). Event centroids (n = 4,432) were buffered at 5, 10, 20, and 50 km to accommodate geolocation uncertainty and footprint–centroid mismatch. For each buffer we recorded "coverage" as a binary indicator of whether the buffer intersected at least a minimal contiguous area of GFSM H + VH ($\geq$300 m^2), then (a) computed annual coverage time series (1985–2016) with Wilson 95% binomial confidence intervals, (b) summarized coverage as a function of buffer radius, and (c) quantified local concentration by relating the within-buffer H + VH share to its enrichment over the global H + VH prevalence (log-ratio). This track was designed to test whether historically reported floods tend to occur within or near the high-susceptibility corridors rather than to reproduce exact event footprints.

Together, these two layers of evaluation—(i) held-out unit tests (discrimination, classification, and calibration), (ii) independent event-based correspondence with uncertainty treatment—provide complementary checks on model skill, spatial plausibility, and robustness without circularity. No external validations were used to tune models or thresholds; all were conducted on fixed outputs generated from the unit-level training described above. More details are provided in the Technical Validation section.

Finally, with a trained model for each unit, we proceeded to generate the flood susceptibility prediction for each 30 m pixel at a global scale. For a given unit, we took each tile belonging to that unit and loaded its FCF arrays. We then fed the feature vectors of all pixels in the tile (in batches) into the XGBoost model to predict flood susceptibility. This step was computationally intensive (applying millions of predictions per tile), but we parallelized it on the HPC and took advantage of XGBoost's efficient prediction on GPU. The output for each tile is a raster of continuous values between 0 and 1, indicating the model-derived susceptibility score learned against the 5-year Aqueduct labels. It is noted that these values represent the relative likelihood of an area flooding compared to other areas within the study region (global in our case). These rasters were then post-processed into the final product format as described below.

Output Map Generation—Producing GFSM v1. We produced two forms of the output flood susceptibility map from the model: (1) a continuous susceptibility score map (stored in scaled integer format for efficiency), and (2) a discrete five-class map. The continuous map retains the full 0–1 value from the model, providing nuanced information about relative susceptibility. To store this at scale (~billions of pixels) efficiently, we linearly scaled the 0–1 susceptibility score to an integer 1–255 range and saved as 8-bit GeoTIFF tiles. The value 0 was reserved for "NoData" (areas with no prediction, e.g., outside land, water bodies, or where input data was missing), and 1–255 corresponds to lowest to the maximum potential for flooding. Users can convert these back to original 0–1 susceptibility scores by (*value*-1)/254. This high-resolution susceptibility score dataset is very large (~200 + GB for the global mosaic) and thus is mainly intended for expert use or on-demand analysis (it can be provided via specialized request or accessed on cloud platforms).

This primary released product is the five-class equal-interval map, referred to as *GFSM v1 (EI-5)*. We classified the values of the map into five categories by equal intervals of width 0.2. The classes are defined as: 0–0.2 = Very Low (VL, Class 1), 0.2–0.4 = Low (L, Class 2), 0.4–0.6 = Moderate (M, Class 3), 0.6–0.8 = High (H, Class 4), and 0.8–1.0 = Very High (VH, Class 5) susceptibility. These thresholds were chosen for simplicity and consistency; equal intervals allow the classes to be interpreted in a uniform way globally (each class represents a 20% range of model-estimated flood susceptibility score). While one could select other classification schemes (e.g., quantiles, standard deviation, or natural breaks), the equal-interval approach ensures that *Very High* in one region means the same level of modeled flooding potential as *Very High* elsewhere, which is advantageous for localized comparisons on a global scale. The resulting categorical raster (see Fig. 4 in the Data Records

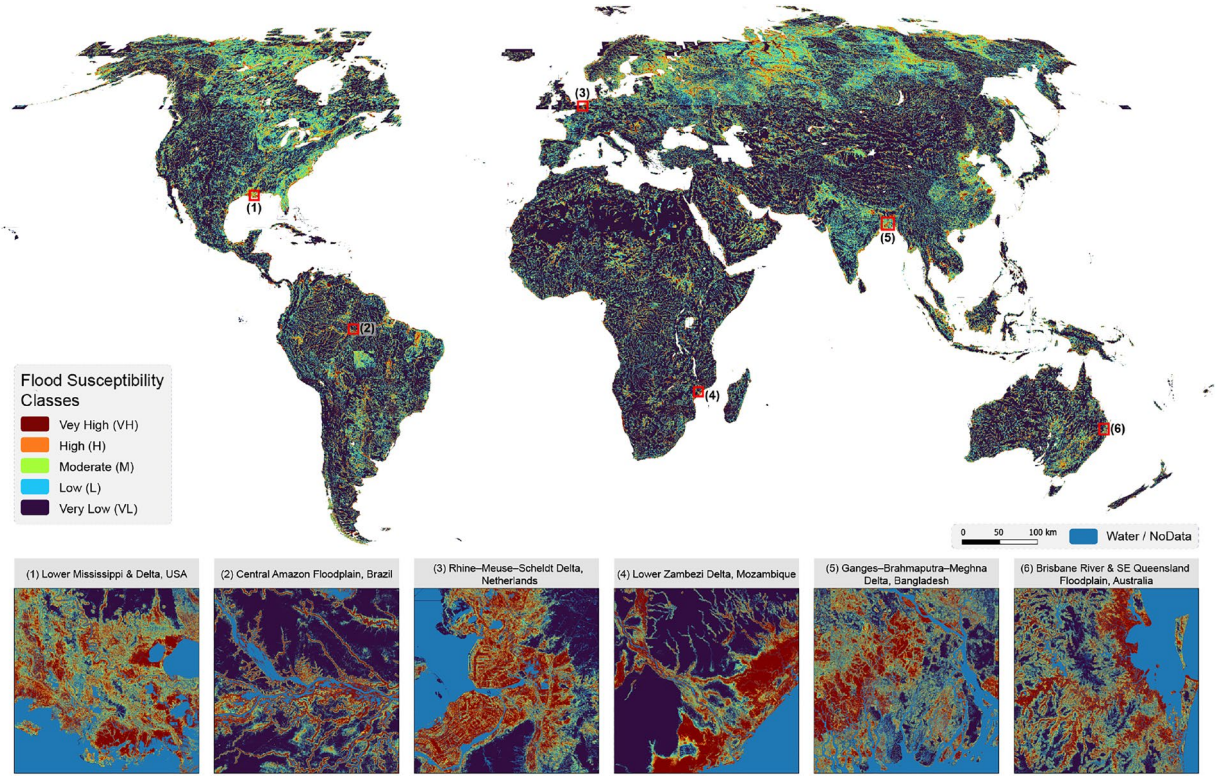

Fig. 4 Global flood susceptibility map (GFSM) data product shown globally, and for six key regions. Note the global visualization is based on GFSM equal interval 5 classes data resampled to 250 m (for visualization), whereas the six inset maps show high-resolution (3llll0m) flood susceptibility data.

section) provides an intuitive map where, for example, Class 5 (Very High) highlights pixels with the maximum potential to be flooded (value >0.8), whereas Class 1 (Very Low) denotes areas with <20% predicted potential to be flooded. We assigned descriptive labels and a standard color scheme (blue for Very Low through red for Very High) to the classes for visualization purposes.

Both the continuous and class maps were produced for every tile. The tile-based approach allows easier handling and download of subsets (each tile is ~100 km square) for regional to local utilization. For distribution, we primarily provide the five-class map as it is ~1/4 the data volume of the continuous map and sufficient for most use cases. The continuous map (in 8-bit scaled form) is retained as an internal product for further analysis and can be shared for detailed analysis (e.g., if a user wants to apply a custom threshold other than our class breaks). By default, we expect users to use the five-class GFSM v1 dataset for identifying high-susceptibility zones. This data format decision was made to maximize accessibility – the categorical map is easier to load on web platforms (like GEE) and compare with other discrete maps (e.g., land cover, administrative boundaries).

Data Records

The Global Flood Susceptibility Map (GFSM v1) is openly available via multiple platforms to facilitate both scientific analysis and applied use. The core dataset is the equal-interval five-class (EI-5) flood susceptibility product at 30 m resolution, which classifies each pixel into five categories: Very Low (1), Low (2), Moderate (3), High (4), and Very High (5). Pixels outside modeled land areas, open water bodies, or locations lacking sufficient input coverage are assigned to a NoData value of 0 and are displayed as transparent. Figure 4 illustrates the global GFSM EI-5 dataset, including both the full global visualization and high-resolution examples for six key floodplains. The GFSM v1 dataset provides near-global coverage from ~80°N to ~60°S, with gaps limited to small islands where one or more input datasets were unavailable. The EI-5 collection at 30 m resolution occupies approximately 50 GB in tiled GeoTIFF form, with smaller sizes for coarser mosaics.

The GFSM v1 flood susceptibility dataset is freely and publicly accessible through multiple platforms to maximize usability and long-term preservation.

Google Earth Engine (primary access). An interactive GEE-based web application has been developed for GFSM, allowing users to explore the dataset visually without writing code. Users can access the application here (https://waleedgis.users.earthengine.app/view/gfsm) or from the project GitHub page.

The dataset can also be accessed directly in Earth Engine using:

$$\text{var } fsm = ee.\ ImageCollection('projects/floodsus/assets/fsm\_ei5')$$

More details on how to use GFSM dataset with GEE are provided at https://github.com/waleedgeo/GFSM.

Statistic	Value	Interpretation/note
Spatial tiles	17,069	Number of processed global tiles used for near-global 30 m GFSM generation.
Modeling units	192	Country or country–climate units used for regionalized model training and prediction.
Training samples	~30.45 million	Stratified, tile-balanced flood/non-flood samples used for model development.
Output resolution	30 m	Final GFSM output grid resolution; training labels were derived from coarser hazard layers.
Output format	Five-class EI-5 GeoTIFF tiles	Classes represent Very Low, Low, Moderate, High, and Very High susceptibility.
Training label source	Aqueduct Flood Hazard Maps v2, 5-year baseline	Model-derived riverine/coastal flood labels, not direct flood observations.
Median ROC-AUC	~0.95	Internal held-out model-label discrimination across modeling units.
Median AUPRC	~0.94	Precision–recall performance under the sampled-label distribution.
DFO correspondence	79% within 5 km; 92% within 10 km	Main independent event-based plausibility result using DFO flood-event locations.

Table 1. Summary statistics and key characteristics of the GFSM v1 dataset.

Zenodo (archival access). The five-class GFSM v1 EI-5 product is archived and available for bulk download via Zenodo under the https://doi.org/10.5281/zenodo.18137661.

The Zenodo repository[8] is organized into the following components:

- **Raster Data:** Per-tile GeoTIFF files aligned to the global tiling scheme, projected in World Mercator (EPSG:3395) and grouped as 20 by 20 degree zip files.
- **Metadata:** Comprehensive documentation for each tile, detailing the Coordinate Reference System (CRS), band schema, nodata conventions, and variable statistics.
- **Spatial Index:** A Global Index GeoPackage (.gpkg) file containing spatial footprints and metadata to facilitate tile identification and bulk download.

To make the core dataset characteristics and validation outcomes directly accessible to readers, Table 1 summarizes the main spatial, modeling, training, output, and validation statistics for GFSM v1. More detailed unit-level performance metrics, calibration diagnostics, feature-importance outputs, and validation tables are provided through the project repository, but the key dataset-level values are reported here for completeness.

Technical Validation

We performed extensive technical validation to confirm the quality, reliability, and scientific rigor of the GFSM v1 dataset. The validation focused on two parts. First, evaluating the internal performance of the ML models used to generate the susceptibility map, ensuring they accurately learned the relationships between the FCFs and flood occurrence from the training data. All metrics were calculated on a held-out test set (20% of samples) for each of the 192 modeling units, providing a robust assessment of the models' generalization capabilities. The second part consists of independent cross-comparison with existing data (whether the output maps behave as expected, compared with existing flood data products).

Model Performance Evaluation. The ML models achieved consistently stronger predictive performance across all modeling units, demonstrating the robustness of the chosen methodology. Figure 5 provides a global snapshot of key performance metrics aggregated from the 192 individual unit-level models. The models showed excellent discrimination between flooded and non-flooded labels. The AUPRC was 0.93 (median 0.94), and the ROC-AUC was 0.94 (median 0.95) when averaged over 192 units. These values indicate excellent discrimination between flood susceptible and low or non-susceptible samples – for reference, an AUC of 0.5 would be no-skill (random guessing), and even 0.8+ is considered good in most classification problems. The precision-recall metric is particularly meaningful here because of the class imbalance (few flooded pixels in reality); a median AUPRC ~0.94 far exceeds the base positive rate (~0.5 in our balanced sample, or much lower in real landscapes), confirming the models are skillfully identifying the susceptible minority. The F1-score (harmonic mean of precision and recall) had a median of ~0.89 across units, reflecting a well-balanced tradeoff – the models achieved ~91% recall (sensitivity) and ~87% precision (positive predictive value) on average (median recall = 0.92, precision = 0.87). In practical terms, this means that a large majority of derived referenced labels were correctly flagged as susceptible (few false negatives), and most of the predicted susceptible areas were indeed in flood zones per the reference data (few false positives). The overall accuracy (percentage of correctly classified pixels) was about 88% on average; however, we focus more on F1, AUC, and error rates given our balanced training approach. Detailed tables on accuracy metrics, model performance, and validation on various scales (Global, Continents, and Units), are available at the GFSM GitHub repository (see Data Records section for links).

Both violin plots (Fig. 5, Global AUPRC and Global ROC-AUC) show a tight concentration near the high end (with interquartile ranges around 0.90–0.97 for AUPRC, and 0.93–0.98 for AUC), and only few of the units fall below 0.9 in either metric. Notably, even the 5th percentile of AUC across units is above 0.85, indicating that every regional model performed at least reasonably well, and most performed exceptionally well. This uniform accuracy gives confidence that no unit was left behind by the model (the approach generalizes globally). Next, F1-score based iso-curve plot provides a complementary view via a precision-recall plot aggregated over units. The cloud of points (each representing a model's precision and recall at the chosen operating threshold) lies near the upper-right corner, with many models achieving both high recall and high precision. For instance, a typical model might have recall ~0.92 and precision ~0.88 (as noted above). Additionally, we also show the

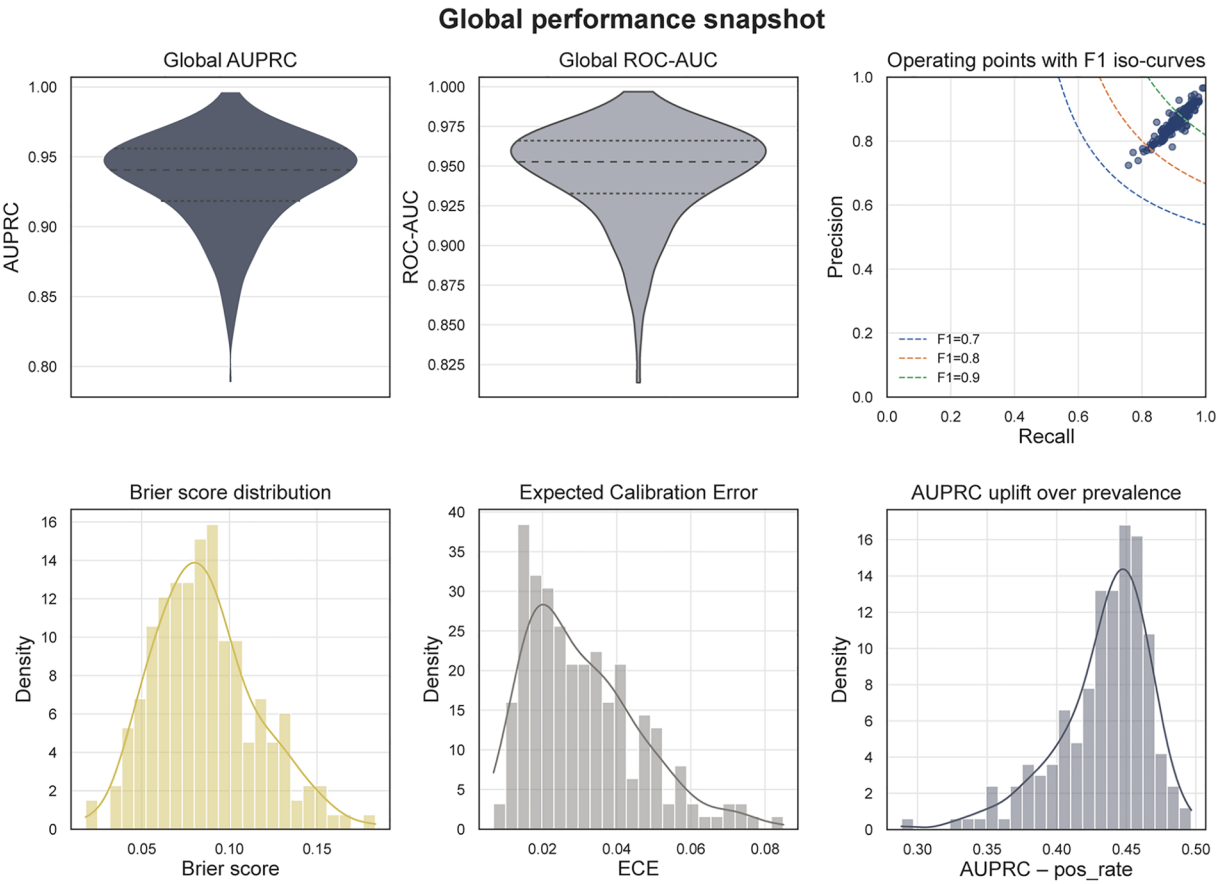

Fig. 5 Technical validation of GFSM v1 model performance. Global performance snapshot showing distributions of area under the precision–recall curve (AUPRC) and receiver operating characteristic curve (ROC-AUC) across 192 units (violin plots), precision–recall operating points with F1 iso-curves, and histograms of Brier score, expected calibration error (ECE), and AUPRC uplift over baseline prevalence. Metrics confirm high discrimination, well-calibrated model scores under the sampled-label distribution, and strong uplift relative to chance.

iso-F1 contours for 0.7, 0.8, 0.9; almost all operating points are above the 0.8 line, again underscoring the high F1-scores. This implies that since we balanced the data (flood vs non-flood sampling ratio) and did not adjust the decision rule post-training, is yielding a good balance – models are neither over-predicting flood (which would hurt precision) nor under-predicting (which would hurt recall) to an extreme degree.

To examine calibration and skill in more detail, we calculated additional metrics. The Brier score, which is the mean squared error of the probability predictions, had a median of ~0.083 across units (mean ~0.086). This is low on a 0–1 scale, indicating well-calibrated and confident predictions. For context, a perfectly confident and correct prediction would have Brier 0, and a non-informative prediction (always 0.5 probability) would have Brier 0.25 for a balanced dataset. Our scores closer to 0.08 show that predictions are both confident and usually correct. We also computed calibration error measures: the ECE, which was around 0.03 (3%) on average (median ~2.8%). We additionally assessed the skill above chance by comparing AUPRC to the prevalence (positive rate) in each unit's test data. Since we enforced a 50% prevalence in test sets (balanced), this particular comparison is trivial (all models' AUPRC ~0.93 vs baseline 0.5 yields an "uplift" of ~0.43). In a real prevalence scenario (say 1% flood rate), an AUPRC of 0.93 would be astonishingly high; even though our sampling is artificial, it demonstrates that the model can pick out rare positives effectively.

Correspondence with Historical Flood Events. To assess the real-world plausibility of the GFSM v1 dataset, we compared the mapped susceptibility patterns against an independent, long-term record of observed flood events. We used the Dartmouth Flood Observatory (DFO) Global Active Archive of Large Flood Events, which documents 4,432 major floods between 1985 and 2016[32]. We tested whether the geographic locations of these historical events spatially coincided with areas classified as "High" or "Very High" susceptibility in GFSM v1.

To account for positional uncertainty in the historical records, we created buffers of varying radii (5 km to 50 km) around each DFO event point and calculated the percentage of events whose buffers intersected high-susceptibility zones. Coverage was summarized overall, and annually (1985–2016). Furthermore, binomial 95% confidence intervals were computed using the Wilson score method, which provides more reliable coverage than the normal approximation, especially when probabilities are near 0 or 1.

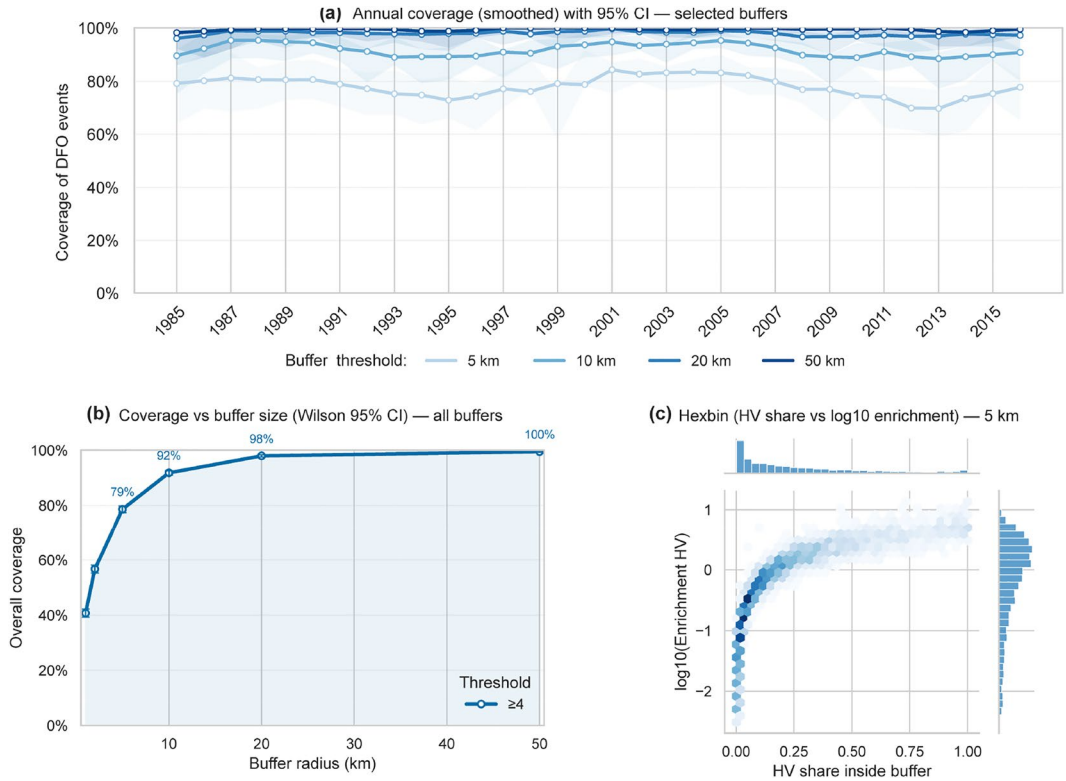

Fig. 6 Validation of GFSM v1 against global flood event points. Comparison with 4,432 DFO flood events (1985–2016) using buffered event points. (**a**) Annual coverage (95% CI), showing consistent capture of 75–85% of events at 5 km and near-complete coverage at larger buffers. (**b**) Overall coverage vs. buffer size, rising steeply from ~79% at 5 km to >90% at 10 km and saturating near 100% by 20 km. (**c**) Hexbin plot of 5 km buffers, showing systematic enrichment of H + VH pixels relative to global prevalence. The three views confirm GFSM's temporal robustness, spatial accuracy, and local concentration of susceptibility around independent flood events.

Figure 6a shows annual coverage (smoothed) with Wilson 95% CIs for selected buffer sizes. Even at 5 km, coverage remained high (approximately 75–85% across most years), indicating that the majority of documented floods fell within short distances of GFSM H + VH pixels. With larger buffers (10–50 km), annual coverage increased further and generally approached 95–100%, with minimal long-term drift, suggesting temporal robustness of the susceptibility patterns across three decades of independent events.

Figure 6b aggregates across years and summarizes coverage vs. buffer size (Wilson 95% CIs). Coverage rises steeply from ~79% at 5 km to ~92% at 10 km, and then plateaus near 98–100% for 20–50 km. The DFO buffer comparison was used as an independent event-centroid correspondence test rather than an exact event-footprint validation. The main interpretation is based on the 5 km and 10 km buffers, which provide stricter tests of whether reported historical flood-event locations occur near GFSM High and Very High susceptibility corridors. Larger 20–50 km buffers are reported only as sensitivity bounds for event-location uncertainty and footprint–centroid mismatch.

To assess intensity of overlap, Fig. 6c relates the share of each 5 km buffer mapped as H + VH to the log10 enrichment of H + VH relative to its global baseline prevalence (values > 0 denote enrichment). Most events cluster at low-to-moderate H + VH shares (0–0.3) but still exhibit positive enrichment, implying that GFSM concentrates high susceptibility disproportionately around observed floods rather than intersecting them by chance. A smaller set of events shows high shares (>0.6), consistent with floods centered in large, contiguous H + VH corridors (e.g., wide alluvial floodplains). The three complementary views (i.e., temporal coverage, dose response with buffer size, and local enrichment) indicate that GFSM v1 reliably captures the "*where*" of flood proneness at global scale: most historical DFO events lie within or near H + VH zones, and buffers are systematically enriched in high-susceptibility pixels. Because DFO points represent event centroids spanning diverse reporting sources and years, some residual mismatches are expected; the Wilson-based confidence bounds, and buffer sensitivity analyses explicitly account for these uncertainties.

Uncertainty and limitations. No dataset is without limitations, and we acknowledge several factors in interpreting GFSM v1. First, the model is only as good as the training data – the Aqueduct flood maps used for training are themselves outputs of large-scale models (with their own uncertainties). If the Aqueduct model underestimates flooding in certain areas (e.g., due to missing local rain-flood mechanisms or levee failures), then GFSM may also under-identify those. Conversely, any bias or false flood signals in Aqueduct

could lead to slight overprediction in GFSM. We mitigated this by focusing on a frequent flood (5-year) scenario, which is relatively well constrained and likely to reveal all the obviously flood-prone terrain. Second, our model does not explicitly include *temporal* factors – it's a static susceptibility map for baseline conditions. Changes in land use or climate beyond the training period might shift actual flood risk. For example, rapid urbanization (more roads, less permeable land) in some regions could increase flood susceptibility beyond our static NDVI/Road data captured (which are for ~2021). Similarly, our RainFq feature is based on recent climate (2014–2024); in a future warming scenario, some regions may experience heavier rainfall, which would effectively increase susceptibility (our map would need updating to reflect that if using future climate data). Third, the equal-interval classification, while globally consistent, means that the "High" class covers the same susceptibility range everywhere; but the real-world implication of a given susceptibility score may differ by context. In some places, even a 30% flood susceptibility (Moderate class, 0.3 susceptibility score) might be considered a serious risk if a large population is present, whereas in other places 30% might be seen as moderate. Users should be aware of these context nuances. We opted not to use unequal class breaks because that would involve subjective choices or region-specific thresholds, which complicate global comparisons.

The Aqueduct layers used for model training are hydrodynamic model outputs rather than direct observational flood inventories. We selected them because a globally consistent, harmonized, observed flood/non-flood inventory with comparable riverine and coastal coverage and return-period context is not currently available for global training. The resulting labels therefore represent Aqueduct-modeled inundation for the selected baseline 5-year scenario, not observed flood event footprints. Consequently, GFSM v1 may inherit uncertainties, assumptions, or spatial biases from Aqueduct, including omissions related to local flood defenses, levee failure, pluvial drainage processes, or poorly represented local topography. In addition, the native Aqueduct label resolution is coarser than the 30 m GFSM output grid. The 30 m product should therefore be interpreted as a high-resolution susceptibility estimate conditioned by 30 m terrain and environmental predictors and constrained by coarser modeled flood labels, not as direct 30 m observational evidence of inundation. Future versions could incorporate additional return periods or multiple hazard products to produce scenario-specific susceptibility layers.

We also note that GFSM v1 does not explicitly include soil texture, permeability, detailed land-use classes, impervious-surface fraction, drainage density, or engineered drainage infrastructure. These variables can be important controls on runoff generation and flood response, especially in urban and agricultural regions. They were not included in this first global version because the workflow prioritized high-resolution variables that could be harmonized consistently across all modeling units and redistributed in an open global product. Future GFSM versions should evaluate these additional predictors where globally consistent and openly reusable datasets are available.

Finally, we emphasize that GFSM v1 is not a flood inundation map for any specific event, but rather a susceptibility baseline. It is best applied as a screening tool to identify hotspots that warrant more detailed analysis. We have confidence in the technical validity of GFSM given the high validation scores and logical behavior, but we encourage users to consider these limitations and where possible combine GFSM with local data (such as drainage infrastructure, historical flood records) for comprehensive risk assessments.

Scope, Implications and Usage Notes. GFSM v1 provides a globally consistent baseline of flood susceptibility at 30 m resolution and is best interpreted as a relative indicator of where the landscape is naturally prone to flooding. The five classes (Very Low to Very High) reflect modeled susceptibility score based on terrain, hydrology, climate, and anthropogenic factors, and should be used for screening, exposure assessment, and comparative analyses rather than as event forecasts. High and Very High zones typically correspond to floodplains, deltas, and coastal lowlands, and are useful for prioritizing areas for more detailed hazard studies or planning measures. Users should note that, at this stage, the dataset does not incorporate flood defenses, fine-scale urban drainage, or future climate change; susceptibility therefore reflects natural propensity under present-day baseline conditions. The data are well suited for overlay with population, infrastructure, or land-use layers to quantify exposure, but results should be validated against local knowledge and complemented with engineered flood models where available.

Figure 7 provides a perspective of this data integration workflow and its key decision-making applications. For precise spatial analysis, we recommend using an equal-area projection (e.g., EPSG:6933) to ensure correct area estimates. Additionally, GFSM v1 does not include a formal per-pixel uncertainty layer. The reported validation metrics and DFO correspondence provide unit/tile level quality checks, but they do not quantify pixel-level uncertainty. Producing robust global pixel-level uncertainty would require ensemble labels, alternative return periods, bootstrapped/ensemble models, or scenario-specific predictions, which are planned as future development. Additionally, future versions of GFSM will aim to incorporate additional globally consistent predictors, including soil, land-use/impervious-surface, and drainage characteristics, together with explicit uncertainty analysis and pixel-level uncertainty estimates where feasible.

Data availability

The Global Flood Susceptibility Map (GFSM v1) is openly available via multiple platforms to facilitate both scientific analysis and applied use.

The core dataset is the equal-interval five-class (EI-5) flood susceptibility product at 30 m resolution, which classifies each pixel into five categories: Very Low (1), Low (2), Moderate (3), High (4), and Very High (5). Pixels outside modeled land areas, open water bodies, or locations lacking sufficient input coverage are assigned to a NoData value of 0 and are displayed as transparent. Figure 5 illustrates the global GFSM EI-5 dataset, including

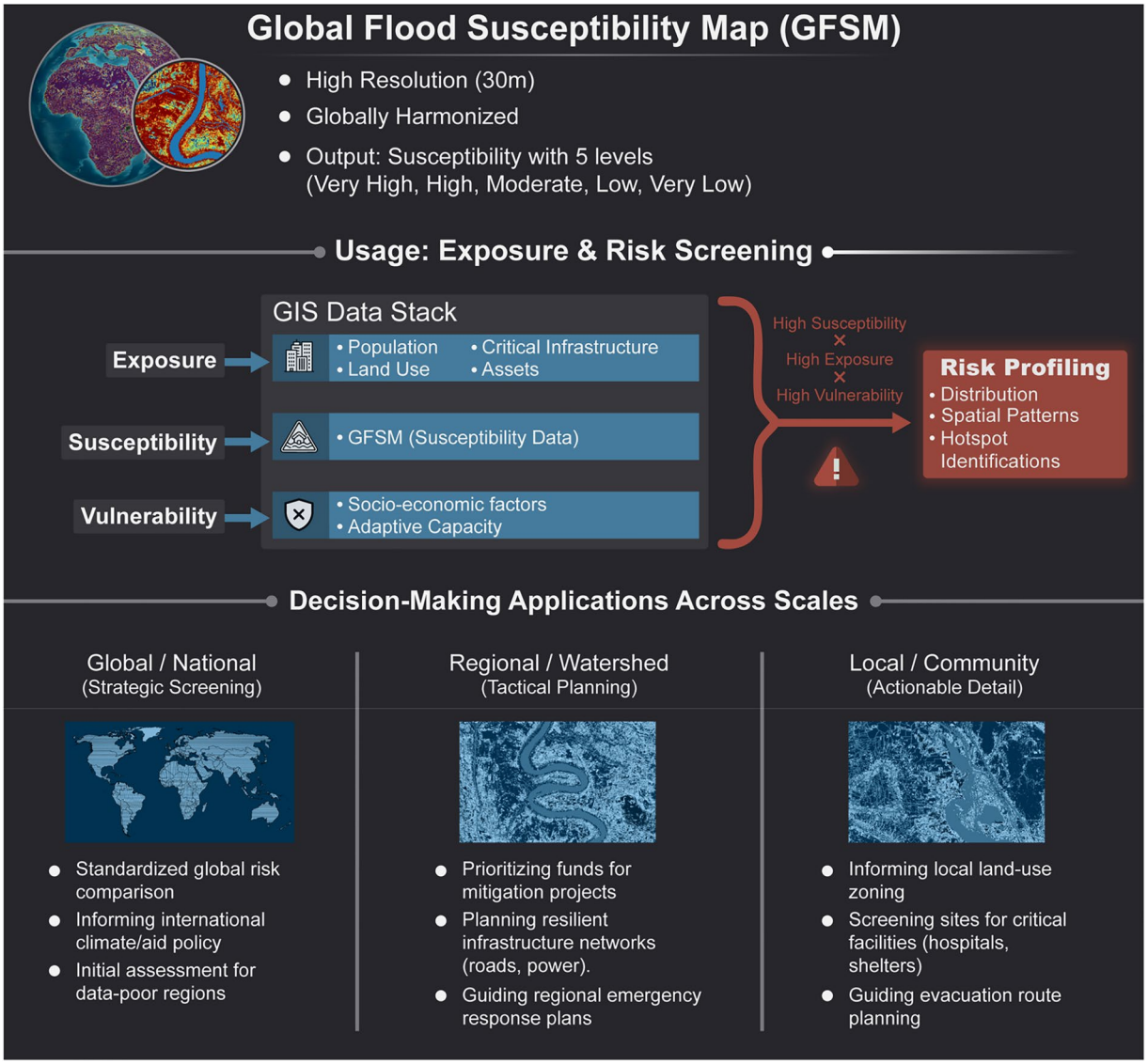

Fig. 7 Conceptual framework for the implications of the Global Flood Susceptibility Map (GFSM) dataset. The infographic outlines the dataset's key features (e.g., 30 m resolution, 5-level classification), the risk screening workflow for identifying "Hotspots" by integrating the susceptibility layer (GFSM) with exposure (e.g., population, infrastructure) and vulnerability data, and its decision-making applications across global (strategic screening), regional (tactical planning), and local (actionable detail) scales.

both the full global visualization and high-resolution examples for six key floodplains. The GFSM v1 dataset provides near-global coverage from ~80°N to ~60°S, with gaps limited to small islands where one or more input datasets were unavailable. The EI-5 collection at 30 m resolution occupies approximately 50 GB in tiled GeoTIFF form, with smaller sizes for coarser mosaics.

The GFSM v1 flood susceptibility dataset is freely and publicly accessible through multiple platforms to maximize usability and long-term preservation.

Google Earth Engine (primary access):

An interactive GEE-based web application has been developed for GFSM, allowing users to explore the dataset visually without writing code. Users can access the application here (https://waleedgis.users.earthengine.app/view/gfsm) or from the project GitHub page.

The dataset can also be accessed directly in Earth Engine using:

var fsm = ee.ImageCollection('projects/floodsus/assets/fsm_ei5');

More details on how to use GFSM dataset with GEE are provided at https://github.com/waleedgeo/GFSM.

Zenodo (archival access):

The five-class GFSM v1 EI-5 product is archived and available for bulk download via Zenodo under the https://doi.org/10.5281/zenodo.18137661.

The Zenodo repository8 is organized into the following components:

Raster Data: Per-tile GeoTIFF files aligned to the global tiling scheme, projected in World Mercator (EPSG:3395) and grouped as 20 by 20 degree zip files.

Metadata: Comprehensive documentation for each tile, detailing the Coordinate Reference System (CRS), band schema, nodata conventions, and variable statistics.

Spatial Index: A Global Index GeoPackage (.gpkg) file containing spatial footprints and metadata to facilitate tile identification and bulk download.

Code availability

Related processing scripts, trained models per unit, and example analysis notebooks associated with GFSM v1 are openly available at: https://github.com/waleedgeo/GFSM.

Received: 16 December 2025; Accepted: 11 September 2026;
Published: xx xx xxxx

References

1. Amiri, A., Soltani, K., Ebtehaj, I. & Bonakdari, H. A novel machine learning tool for current and future flood susceptibility mapping by integrating remote sensing and geographic information systems. *J. Hydrol.* **632**, 130936 (2024).
2. Dodangeh, E. *et al*. Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Sci. Total Environ.* **705**, 135983 (2020).
3. Fang, Z., Wang, Y., Peng, L. & Hong, H. Predicting flood susceptibility using LSTM neural networks. *J. Hydrol.* **594**, 125734 (2021).
4. Gudiyangada Nachappa, T. *et al*. Flood susceptibility mapping with machine learning, multi-criteria decision analysis and ensemble using dempster shafer theory. *J. Hydrol.* **590**, 125275 (2020).
5. Misra, A., White, K., Nsutezo, S. F., Straka, W. & Lavista, J. Mapping global floods with 10 years of satellite radar data. *Nat Commun* **16**, 5762 (2025).
6. Luu, C. *et al*. GIS-based ensemble computational models for flood susceptibility prediction in the quang binh province. *vietnam. J. Hydrol.* **599**, 126500 (2021).
7. Pham, B. T. *et al*. Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? *J. Hydrol.* **592**, 125615 (2021).
8. Waleed, M. Global Flood Susceptibility Map (GFSM v1): A high resolution (30m) flood susceptibility dataset derived from multi-source Earth observation and geospatial data. *Zenodo* https://doi.org/10.5281/ZENODO.18137661 (2026).
9. WRI. Aqueduct flood hazard maps version 2. (2020).
10. Baugh, C. *et al*. Global river flood hazard maps. *European Commission, Joint Research Centre (JRC)* (2024).
11. Yan, L., Rong, H., Yang, W., Lin, J. & Zheng, C. A novel integrated urban flood risk assessment approach based on one-two dimensional coupled hydrodynamic model and improved projection pursuit method. *J. Environ. Manage.* **366**, 121910 (2024).
12. Liu, J., Liu, K. & Wang, M. A Residual Neural Network Integrated with a Hydrological Model for Global Flood Susceptibility Mapping Based on Remote Sensing Datasets. *Remote Sens.* **15**, 2447 (2023).
13. Wing, O. E. J. *et al*. A 30 m global flood inundation model for any climate scenario. *Water Resour. Res.* **60**, e2023WR036460 (2024).
14. Waleed, M. & Sajjad, M. Advancing flood susceptibility prediction: a comparative assessment and scalability analysis of machine learning algorithms via artificial intelligence in high-risk regions of Pakistan. *J. Flood Risk Manag.* **18**, e13047 (2025).
15. Waleed, M. & Sajjad, M. High-resolution flood susceptibility mapping and exposure assessment in Pakistan: an integrated artificial intelligence, machine learning and geospatial framework. *Int. J. Disaster Risk Reduct.* **121**, 105442 (2025).
16. Tiling System – Harmonized Landsat Sentinel-2 https://hls.gsfc.nasa.gov/products-description/tiling-system/.
17. Beck, H. E. *et al*. Present and future köppen-geiger climate classification maps at 1-km resolution. *Sci. Data* **5**, 180214 (2018).
18. Seleem, O., Ayzel, G., de Souza, A. C. T., Bronstert, A. & Heistermann, M. Towards urban flood susceptibility mapping using data-driven models in Berlin, germany. *Geomat. Nat. Hazards Risk* **13**, 1640–1662 (2022).
19. International Earth Science Information Network-CIESIN-Columbia University, C. & NASA gROADS. Global Roads Open Access Data Set, Version 1 (gROADSv1). *NASA Socioeconomic Data and Applications Center* (SEDAC, Palisades, NY, 2013).
20. Zanaga, D. *et al*. ESA WorldCover 10 m 2021 v200. *Zenodo* https://doi.org/10.5281/ZENODO.7254221 (2022).
21. Yamazaki, D. *et al*. A high-accuracy map of global terrain elevations. *Geophys. Res. Lett.* **44**, 5844–5853 (2017).
22. Donchyts, G. *et al*. Global 30m Height Above the Nearest Drainage. in. https://doi.org/10.13140/rg.2.1.3956.8880 (Vienna, Austria, 2016).
23. Hawker, L. *et al*. A 30 m global map of elevation with forests and buildings removed. *Environ. Res. Lett.* **17**, 24016 (2022).
24. Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J. & Tan, J. GPM IMERG final precipitation L3 half hourly 0.1 degree × 0.1 degree V06. in *Earth Sciences Data and Information Services Center (GES DISC)*. (Greenbelt, MD, Goddard, 2019).
25. Dawson, G., Butt, J., Jones, A. & Fraccaro, P. Flood susceptibility mapping at the country scale using machine learning approaches. *Appl. AI Lett.* **4**, e88 (2023).
26. Seydi, S. T. *et al*. Comparison of machine learning algorithms for flood susceptibility mapping. *Remote Sens.* **15**, 192 (2023).
27. Saravanan, S. & Abijith, D. Flood susceptibility mapping of northeast coastal districts of tamil nadu india using multi-source geospatial data and machine learning techniques. *Geocarto Int.* **37**, 1–30 (2022).
28. Nguyen, H. D. Flood susceptibility assessment using hybrid machine learning and remote sensing in quang tri province, vietnam. *Trans. GIS* **26**, 2776–2801 (2022).
29. Deroliya, P. *et al*. A novel flood risk mapping approach with machine learning considering geomorphic and socio-economic vulnerability dimensions. *Sci. Total Environ.* **851**, 158002 (2022).
30. Shah, S. A. & Ai, S. Flood susceptibility mapping contributes to disaster risk reduction: a case study in sindh, pakistan. *Int. J. Disaster Risk Reduct.* **108**, 104503 (2024).
31. Panahi, M. *et al*. A country wide evaluation of Sweden's spatial flood modeling with optimized convolutional neural network algorithms. *Earths Future* **11**, e2023EF003749 (2023).
32. Kettner, A. J., Brakenridge, G. R., Schumann, G. J. & Shen, X. DFO—flood observatory. in *Earth Observation for Flood Applications* 147–164 https://doi.org/10.1016/b978-0-12-819412-6.00007-9 (Elsevier, 2021).

Acknowledgements

We are grateful to KAUST for providing access to the Shaheen supercomputing facilities, which were instrumental in processing the global datasets. The computational simulations utilized the resources of the Supercomputing Core Laboratory at KAUST. We also thank the Earth Engine team for their GEE platform that facilitated several analyses in this research. We acknowledge the developers and data providers of all open datasets used: the World Resources Institute for Aqueduct Flood maps, the ESA WorldCover team for NDVI composites, the developers of FABDEM (Laurence Hawker and colleagues), NASA SEDAC for gROADS, OpenStreetMap contributors and

Dai Yamazaki's team for the water layer, and NASA for the GPM precipitation data, and for their commitment to open data which helped this study. We are also sincerely grateful to our colleagues for their valuable feedback, with special thanks to Dr. Safi Ullah from the Urban Lab at KAUST for insightful feedback that significantly enhanced the GFSM dataset.

Author contributions

Waleed M. and Sajjad M. conceived the study. Waleed M. led the development of the GFSM v1 dataset, which included data preprocessing tasks, such as the collection and harmonization of input datasets. He further implemented the machine learning modeling framework, performed hyperparameter tuning, and conducted the technical validation analyses. Additionally, Waleed M. and Sajjad M. prepared the first draft and created the visualizations for this study. SAJJAD M. supervised the overall work. Sajjad M., Sami G., and Meng G. contributed to revising the initial draft and reviewing the manuscript. All authors approved the final version.

Funding

While no specific funds were available to conduct this research, the authors acknowledge the financial and technical support that made this work possible. Waleed M. is supported by a postgraduate studentship from the HKBU Research Grant Committee (PhD studentship, 2022–2026). Sajjad M. is supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Grant Number: RFS2223-2H02). The first author is also thankful to the Graduate School of the Hong Kong Baptist University for providing financial support to visit King Abdullah University of Science and Technology (KAUST) in Thuwal, Saudi Arabia to conduct this research.

Competing interests

The authors declare no competing interests. The dataset and analyses were produced in an academic setting without any commercial or financial conflicts of interest.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-026-08338-1.

Correspondence and requests for materials should be addressed to M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.