Received: 30 April 2025 | Revised: 25 December 2025 | Accepted: 23 March 2026 | Published online: 7 April 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–17
DOI: 10.47852/bonviewAIA62026053

BON VIEW PUBLISHING

Toward Resilient Communities: Integrating Predictive Flood Models with Natural Language Processing for Actionable Insights

Divas Karimanzira[1,*]

[1] *Department of Surface Water, Fraunhofer Institute for Optronics, System Technologies and Image Exploitation, Germany*

Abstract: Communities worldwide are increasingly concerned about flooding, making accurate forecasting crucial. This paper introduces two innovative models to improve the mapping of flood inundation areas and depths using large language models (LLMs) and advanced computational techniques. The first model analyzes historical gauge data to establish distinct inundation thresholds for each pixel, significantly enhancing forecast accuracy. The second model employs digital elevation models (DEMs) alongside projected water levels to determine water depth in real time. We tested these models in a flood-prone region, comparing results with traditional physical models. The threshold-based model achieved an impressive average F1-score of 0.87, outperforming the physical model's score of 0.75. Additionally, our DEM-based model maintained a mean absolute error of only 0.15 m for water depth predictions, while the physical model's error was 0.30 m. These findings demonstrate that our models can predict floods more accurately and efficiently. The integration of LLMs enhances computational effectiveness, enabling rapid processing of large datasets and facilitating real-time flood forecasting. LLMs simplify complex numerical data into actionable insights, generating tailored reports and alerts for city planners and emergency responders. Unlike traditional models that require extensive time and resources for calibration, our approach allows for quick adjustments to varying hydrological conditions and real-time updates. Overall, these innovative models represent a significant advancement in flood mapping, providing a more accurate, scalable, and economical alternative while enhancing resilience in flood-prone areas.

Keywords: flood inundation, manifold model, digital elevation model (DEM), real-time forecasting, large language models

1. Introduction

One of the worst natural disasters that we encounter worldwide is flooding. It causes serious harm to ecosystems, infrastructure, and the livelihoods of those impacted [1, 2]. Accurate flood prediction and mapping are now essential since climate change is making weather patterns more unpredictable and urban areas more vulnerable [3, 4]. Deterministic models that evaluate flood risks using historical data and hydrological parameters are the foundation of many conventional flood management techniques [5]. Nevertheless, the complex dynamics of flooding over large geographic areas are frequently difficult for these models to depict. They may therefore make it difficult to make decisions during significant floods [6, 7]. In this regard, flood risk management requires the use of inundation modeling. It aids in our comprehension of how floods can cause water from rivers and other bodies to overflow into nearby areas [7, 8]. These models create maps of the extent of flooding from forecasted river stages, which are essentially predictions of water levels at particular locations along a river. These maps provide us with a tangible image of the areas that are in danger and clearly depict the anticipated flooded areas surrounding gauging stations [9].

Forecasting the distribution of floodwaters based on expected changes in river levels is the primary objective of inundation modeling. Inundation models can produce comprehensive maps that assist stakeholders in comprehending possible flood impacts by utilizing hydrological data, topographical information from digital elevation models (DEMs), and real-time gauge readings [10, 11]. Emergency management teams, urban planners, and communities at risk can all benefit greatly from these maps, which give them the knowledge they need to anticipate flooding and respond to it efficiently. These models can predict the depth of the water at different locations within the impacted areas in addition to displaying the extent of flooding. For assessing the possible effects of flooding on ecosystems, infrastructure, and public safety, this depth data is essential [12, 13]. Authorities can make critical decisions regarding evacuation plans, resource allocation, and infrastructure protection by knowing how deep the water might get.

In large rivers, changes in water levels are influenced by substantial volumes of water, which typically lead to slower fluctuations over time. As a result, inundation models often determine flood extents based mainly on the predicted river stage at a specific moment, without considering historical river stages from earlier times. This method operates under the assumption that river flow

***Corresponding author:** Divas Karimanzira, Department of Surface Water, Fraunhofer Institute for Optronics, System Technologies and Image Exploitation, Germany. Email: divas.karimanzira@iosb-ast.fraunhofer.de

is in a quasi-steady state, meaning that changes in water levels can be forecasted based on current conditions, with little regard for past fluctuations [7, 8]. This assumption simplifies the computational process, allowing for quicker updates to flood predictions as new data becomes available. This is especially useful in real-time flood forecasting, where making timely decisions is crucial to reduce the impacts of flooding [8]. However, despite advancements in flood modeling, traditional physical models often have difficulty capturing the complexities of flood dynamics across large areas. These models tend to be limited by their scale, as creating and maintaining physical models for extensive regions can be costly and time-consuming [9]. Furthermore, they may not respond well to changing environmental conditions or extreme weather events that differ significantly from past occurrences [7].

This paper explores how we can combine advanced predictive models, particularly those that concentrate on inundation and depth predictions, with a large language model (LLM) that employs natural language processing (NLP) techniques in order to get around the drawbacks of conventional flood prediction models [14]. This novel strategy seeks to improve the interpretability and accessibility of flood management outputs, which can significantly improve decision-making and communication among community members, urban planners, and emergency responders [8]. As described by Nevo et al. [15], we present two computational models for inundation and water depth prediction. By using historical gauge data to determine specific inundation thresholds for each pixel, the first model enables us to make more precise predictions about the potential extent of floodwaters. This approach allows for a more thorough comprehension of local flood dynamics and takes into consideration the variability that can occur over short distances as a result of various topographical features and land uses. Our model can produce high-accuracy flood extent predictions by concentrating on historical data, balancing recall and precision.

The second model uses real-time gauge data and DEMs to determine the water depth. This ability is crucial for delivering precise and timely water depth estimates, which are necessary for estimating the potential effects of flooding on communities and infrastructure. Our computational models are made to be efficient, in contrast to conventional physical models, which frequently take a lot of time and money to set up. They are particularly well-suited for real-time flood forecasting because they allow for the quick processing of large datasets and can quickly adjust to changing conditions. Furthermore, we improve the predictive power of both models by incorporating a wide variety of data sources, such as historical records and data from remote sensing. Because it enables us to update the models with new data and modify them to reflect changing hydrological conditions, this flexibility is especially crucial in the context of climate change. Being able to quickly adjust to new information is a big plus in emergency management scenarios where quick decisions can have a big impact. Additionally, we can translate complicated numerical results into actionable insights that are communicated in plain, natural language by combining predictive models with NLP technologies [16]. Our integrated system can generate insightful reports, alerts, and customized recommendations for multiple stakeholders by leveraging historical flood data, real-time gauge readings, and pertinent contextual information [17–19]. In order to promote a more knowledgeable and adaptable flood management process, this paper also presents a framework for an interactive decision support system that allows users to interact with the integrated model in natural language [20]. In the face of rising flood risks, our research ultimately seeks to improve flood resilience tactics and facilitate efficient communication [21].

2. Materials and Methods

2.1. Study area: Baden-Württemberg, Germany

The federal state of Baden-Württemberg, which is situated in southwest Germany, is renowned for its varied landscapes, which include the Black Forest, the Swabian Jura, and a system of important rivers like the Neckar, Rhine, and Danube. The region, which is home to about 11 million people and spans an area of about 35,751 square kilometers, is distinguished by a mix of rural and urban landscapes, with significant areas of industrialization, forests, and agricultural land (Figure 1). Baden-Württemberg is

Figure 1
Study area with areas of interest (AOIs) defined by the gauges

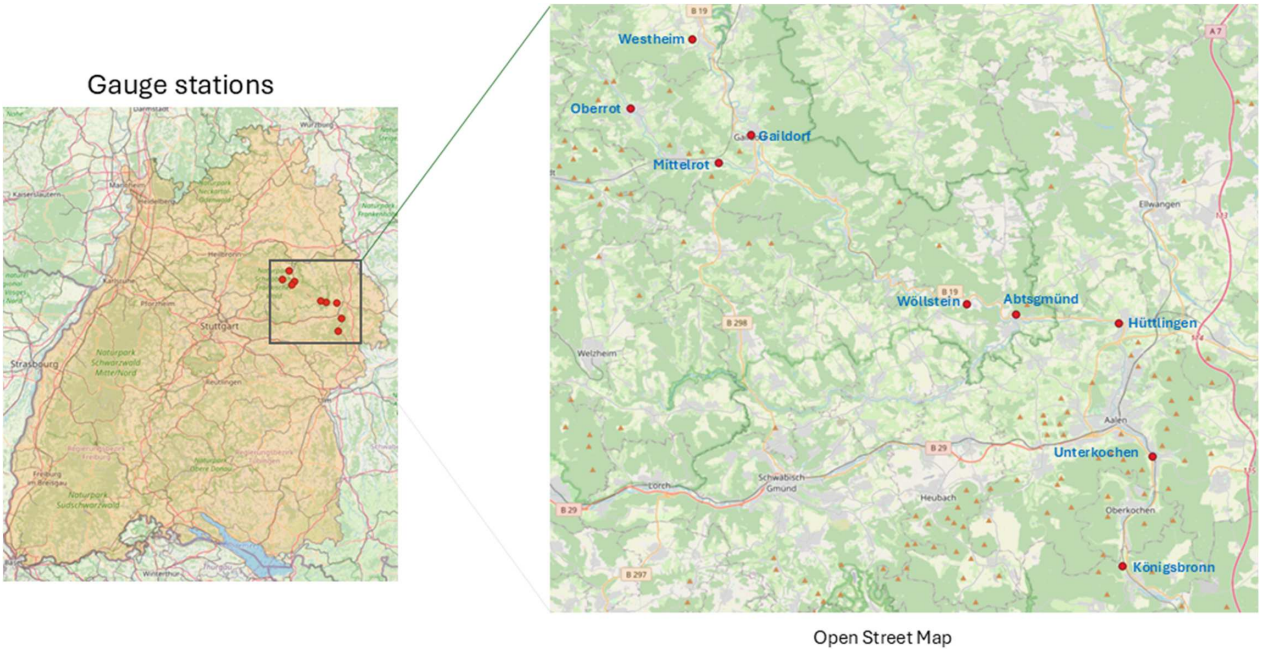
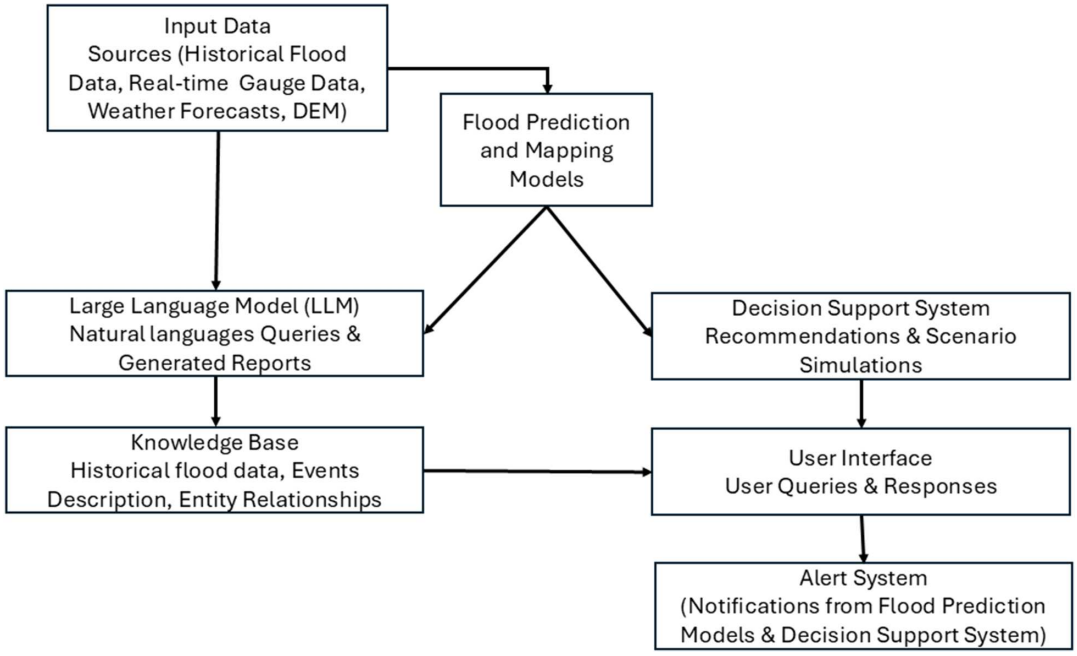


Figure 2
Integrated flood prediction and management system from data collection to flooding warning and communication



especially vulnerable to flooding due to its diverse topography and climate, particularly during periods of intense precipitation or swift snowmelt [1].

With significant rivers traversing both urban and rural regions, Baden-Württemberg boasts a vast hydrological network. While the Rhine forms the state's western border and is a vital waterway for trade and transportation, the Neckar River, a significant tributary of the Rhine, winds through cities like Heidelberg and Stuttgart. The second-longest river in Europe is the Danube, which rises in the Black Forest area. The intricate hydrological dynamics of the region are influenced by the interaction of these rivers and the surrounding terrain, underscoring the necessity of efficient flood modeling to control the risks of flooding. Several data sources were used to create the threshold and manifold inundation models (MIMs) especially for Baden-Württemberg. The Landesamt für Geoinformation und Landentwicklung (LGL) Baden-Württemberg provided a high-resolution DEM. With a resolution of 1 m, this DEM offers comprehensive topographic data that is necessary for precise flood modeling[1].

The German Federal Waterways Engineering and Research Institute (BAW) and regional water authorities provided historical hydrological data in addition to the DEM. This dataset, which spans several years and provides information on past flood events (BAW), consists of water stage records from gauging stations along the Neckar, Rhine, and Danube rivers. In addition, data on previous flood occurrences, including peak water levels and inundation levels, was obtained from the Baden-Württemberg State Ministry of the Environment, Climate Protection, and the Energy Sector. This data is essential for the calibration and validation of the model (Baden-Württemberg State Ministry of the Environment). The CORINE Land Cover (CLC) database, which offers thorough details on land use types throughout Europe, was also used to collect data on land use and land cover. A better understanding of how various surfaces impact flood dynamics is made possible by this data, which classifies land use into pertinent types such as urban, agricultural, and forested areas [20]. Additionally, the Deutscher Wetterdienst (DWD), the German National Meteorological Service, provided climate data, such as temperature and precipitation records[2]. Understanding how hydrology reacts to weather events that could cause flooding requires knowledge of this information.

2.2. Methods

Traditional hydraulic models, such as HEC-RAS and MIKE FLOOD, have long been the standard for flood prediction. However, they often require extensive data and computational resources, making them less accessible for real-time applications. In response to these challenges, the MIM introduces a novel approach that integrates machine learning techniques to enhance predictive capabilities and operational efficiency. A comprehensive and well-organized approach to flood prediction and management is presented by the methodology shown in the block diagram (Figure 2), which combines a number of components to facilitate data-informed decision-making. The first step is to collect input data from various sources, such as weather forecasts, gauge station data in real time, and historical flood records. Following data collection, it is subjected to advanced flood prediction models, which evaluate the data and generate important results, including the anticipated depth of water and the extent of flooding. While depth predictions help predict how high the water will rise in those areas, inundation extent shows which geographic areas are most likely to be impacted by floodwaters. Understanding the possible effects of floods is crucial for enabling stakeholders to adequately prepare and react.

We employ an LLM to simplify these intricate predictions. The model outputs are transformed into actionable insights and presented in natural language by the LLM. It creates reports that

[1] https://opengeodata.lgl-bw.de/#/(sidenav:product/3

[2] https://www.dwd.de/DE/leistungen/cdc_portal/cdc_portal.html

give a summary of the forecasts and context for different stakeholders, making the important information easier for everyone to understand. The LLM is enhanced by a knowledge base that contains historical flood information, past event descriptions, and connections between different geographical features and flood behavior. This makes it possible for the LLM to provide more precise insights that are suited to particular flood management scenarios. A user interface serves as the primary means of communication between stakeholders and the system. Through this interface, users can ask the LLM questions and get personalized answers. Users can make well-informed decisions by using this interface to access real-time alerts and recommendations based on the most recent predictions. Additionally, it connects to a decision support system that combines data from the flood prediction models and the LLM.

2.2.1. Data generation

The process of generating, preparing, and preprocessing data for the threshold flood risk model (TFRM) and the manifold water depth model (MWDM) involves several critical steps to ensure accurate and effective flood modeling. Below is a general description of these steps, which can be adapted to various regions.

1) Data compilation

The initial phase involves compiling a comprehensive dataset that encompasses various elements necessary for efficient flood modeling. This includes:

a. Digital elevation model (DEM): A high-resolution DEM is obtained, typically from a relevant geoinformation authority, providing a detailed representation of the landscape. A spatial resolution of 1 m is ideal for accurately simulating flood dynamics.
b. Hydrological data: Historical data is gathered from gauging stations situated along key rivers. This data includes water stage records over multiple years, collected at various intervals (e.g., hourly or daily).
c. Flood event records: Information regarding past flood events, including peak water stages and inundation extents, is collected from environmental agencies. This data is essential for calibrating and validating the inundation models.
d. Land use and land cover data: The CLC database provides detailed records of land use types (e.g., urban, agricultural, forested areas). Understanding how different land surfaces affect flood dynamics is vital for effective modeling.
e. Climate data: Meteorological data, including temperature and precipitation records, is sourced from national weather services. This information is crucial for evaluating the influence of weather events on flood risks and understanding the effects of rainfall duration and intensity on river stages.

2) Data integration and processing

Once the necessary datasets are collected, the next step involves integrating and processing the information to prepare it for modeling:

a. Geographic information system (GIS) integration: All datasets—including the DEM, hydrological data, land use information, and climate data—are combined within a GIS framework. This integration facilitates spatial analysis and allows for the seamless interaction of diverse data types, enhancing the predictive capabilities of the models.
b. DEM preparation: The DEM undergoes various preprocessing procedures, such as:

Reprojection: Aligning the DEM with other datasets by converting it into an appropriate coordinate system.

Smoothing: Reducing noise in the DEM to enhance its usability.

Sink filling: Addressing any depressions in the DEM that could collect water, ensuring a more accurate representation of hydrological features.

Threshold setting for the TFRM: Historical gauge data is analyzed to establish inundation thresholds necessary for the TFRM. These thresholds are critical for predicting flood risks based on real-time conditions.

c. Depth calculation for the MWDM: Real-time gauge readings are incorporated into the MWDM to estimate flood depths. The DEM is utilized to determine the elevation of each pixel, enabling the calculation of water heights.

3) Data generation for computational models

To generate the additional data required for the computational models, various modeling approaches are employed:

a. Hydrodynamic models: These models simulate the movement of water over different surfaces using principles of fluid dynamics. They provide detailed information on water surface elevations and flow velocities during various climatic scenarios (e.g., heavy rainfall or snowmelt). The outputs of these models, including floodplain dynamics and inundation extents, enhance the historical gauge data used in the TFRM, leading to more precise inundation thresholds.
b. Numerical models: Utilizing mathematical equations, these models solve the shallow water equations (SWE) to simulate water flow and inundation processes. They offer spatially distributed predictions of potential flood occurrences and water depths over time. Data generated from numerical simulations can be integrated with real-time gauge readings and high-resolution DEMs to dynamically update inundation depths based on current hydrological conditions.
c. Physical models: In controlled environments, physical models allow researchers to observe flood dynamics directly. By simulating different flood scenarios, these models generate empirical data on inundation patterns and water flow characteristics. This data is crucial for validating and calibrating the threshold and manifold models, ensuring they accurately represent real-world conditions.

2.2.2. Threshold-based flood risk model (TFRM)

By setting pixel-specific thresholds based on historical gauge data, the threshold model [15], which is illustrated in Figure 3, is a computational method intended to forecast flood inundation extents. The underlying assumption of this model is that, depending on the water stage measured at adjacent gauging stations, each pixel in a given area can be categorized as either "wet" (inundated) or "dry." The main goal of the model is to maximize these classifications to improve the accuracy of flood predictions. The inputs and outputs of the model are listed in Table 1.

The threshold model consists of several key components that work together to produce inundation predictions.

Threshold calculation

The model starts by examining historical data from gauging stations, such as measurements of water stages and the

Figure 3
Flood inundation mapping threshold model

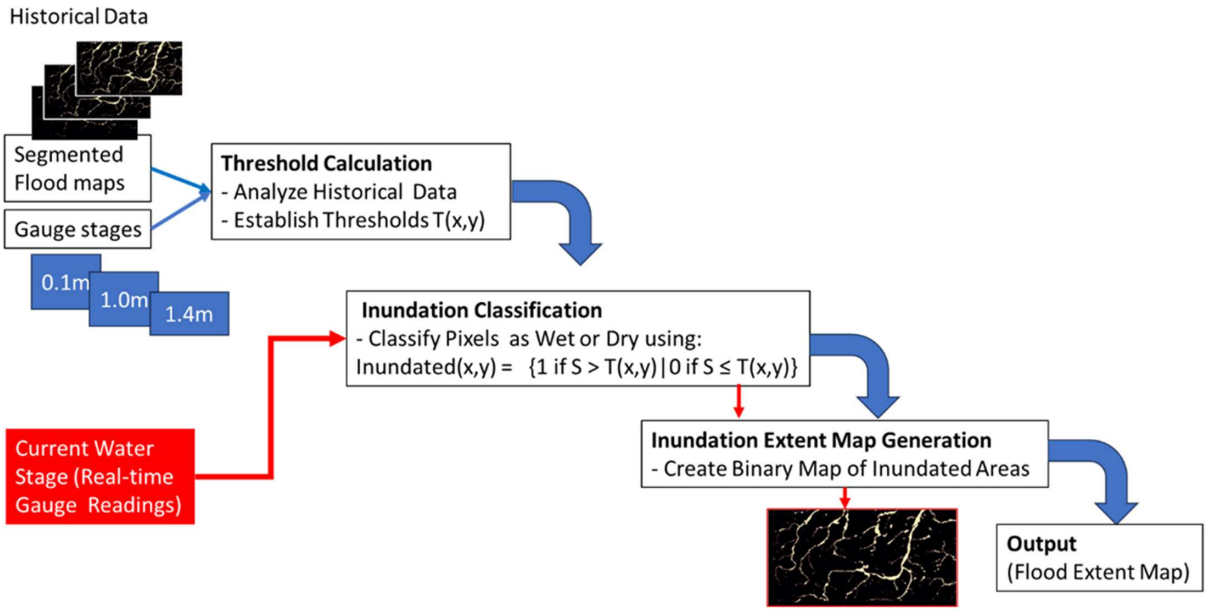


Table 1
Inputs and outputs of the model

Component	Type	Description
Historical Gauge Data	Input	Gauging stations that track water levels over time provide time series data. This data is critical for understanding past flood events and establishing thresholds.
Digital Elevation Model (DEM)	Input	A digital representation of the terrain in the area of interest, providing elevation data necessary for calculating water heights above the ground.
Land Use and Land Cover Data	Input	Information on how the land is used (e.g., urban, agricultural, forested). This data helps assess how different surfaces may influence flood dynamics.
Current Water Stage	Input	Real-time water level measurements from gauging stations that inform the model about the current conditions affecting inundation predictions.
Inundation Extent Map	Output	A binary map indicating which pixels are classified as inundated (wet) or dry based on established thresholds.
Threshold Values	Output	Pixel-specific inundation thresholds that define the water stage at which a pixel transit from dry to wet. These values are crucial for accurate inundation classification.

corresponding levels of inundation seen during previous floods. The DEM determines a particular inundation threshold for every pixel, which is represented as $(T(x, y))$, where $((x, y))$ is the pixel's coordinates. By analyzing the correlation between each pixel's inundation status and the historical water stages, the threshold is established.

Classification algorithm

The classification of each pixel is based on the relationship between the water stage S at the gauge and the established threshold $T(x, y)$. The inundation status $I(x, y)$ can be mathematically expressed as follows:

$$I(x,y) = \begin{cases} S > T(x,y) & 0 \\ S \leq T(x,y) & 1 \end{cases} \tag{1}$$

In this case, $(I(x, y))$ is a binary indicator where 1 denotes that the pixel at (x, y) is wet, and 0 denotes that it is dry. To determine how to balance the model's precision and recall using the beta parameter, the threshold $T(x, y)$ is optimized using historical data to maximize the F_β-score. The definition of F_β-score is:

$$F_1 = (1 + \beta^2) \cdot \frac{Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \tag{2}$$

where precision measures the proportion of true positive predictions among all positive predictions and recall measures the proportion of true positive predictions among all actual positive cases are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

True positives, or correctly predicted wet pixels, are denoted by TP in these equations, false positives, or incorrectly predicted wet pixels, by FP, and false negatives, or missed wet pixels, by FN.

In order to ensure that the model successfully differentiates between wet and dry areas, the optimization process entails modifying the thresholds for each pixel until the F_β-score reaches its maximum value. Once the thresholds are established, the model can be applied to real-time gauge data. When a new water stage $S_{current}$ is recorded, the model classifies each pixel based on the previously determined thresholds:

$$I(x,y) = \begin{cases} S_{current} > T(x,y) & 0 \\ S_{current} \leq T(x,y) & 1 \end{cases} \tag{5}$$

Authorities can identify areas at risk of flooding thanks to this classification, which offers vital information for flood risk assessment and management. Because of its optimization process and dependence on historical data, the threshold model is able to adjust to the particularities of the flood-prone areas it is used in, improving its predictive power.

An iterative process is employed to select the threshold where, in each iteration, the algorithm identifies the threshold that maximizes the ratio of true wet events to false wet events. True wet events are instances where the gauge stage is above the threshold, and the pixel is wet, while false wet events are instances where the gauge stage is above the threshold, but the pixel is dry. The process continues until the true–false ratio drops below a defined minimal ratio parameter, indicating a cost-effective threshold has been found. Furthermore, the model ensures that the selected pixel-specific thresholds are Pareto optimal, meaning that no other set of thresholds can simultaneously improve both precision and recall for a given minimal ratio parameter. By testing various values of the minimal ratio parameter, the model identifies the thresholds that maximize the desired F_β-score.

It is important to note that the model does have certain limitations. The quality and accessibility of historical stage data have a significant impact on its efficacy. The model might have trouble correctly identifying thresholds in regions with sparse data. These thresholds may not adjust well to changing environmental conditions or extreme weather events that differ greatly from past experiences because they are based on historical data. Furthermore, it may be difficult to predict wet pixels accurately if there are significantly fewer wet events than dry ones, which could distort the model's performance metrics.

2.2.3. Manifold water depth model (MWDM)

The MIM [15] presented in the context of flood forecasting provides a machine learning alternative to traditional hydraulic models. It is designed to predict flood inundation depths and extents using data in Table 2, which include DEMs and real-time water stage data from gauging stations as shown in Figure 4.

The flood extent calculation, flood extent to water height algorithm, and water depth estimation from gauge stage are the three separate components that make up the model structure. The model's structure enables it to offer thorough insights that are essential for managing flood risk.

1) Flood extent calculation

Calculating the extent of inundation using the water stage measured at a gauge is the main goal of the first section of the MIM. The model compares the water height $h(x,y)$ to the DEM elevation $z(x,y)$ to determine if each pixel in the DEM is wet or flooded. The following is a mathematical expression for the relationship:

$$h(x,y) = S - z(x,y) \tag{6}$$

where $h(x,y)$ is the water height above the terrain at coordinates (x,y), S is the water stage recorded at the nearest gauging station, and $z(x,y)$ is the elevation of the terrain at coordinates (x,y).

The inundation classification for each pixel is determined using the following criterion:

$$I(x,y) = \begin{cases} h(x,y) > 0 & 0 \\ h(x,y) \leq 0 & 1 \end{cases} \tag{7}$$

This classification indicates that a pixel is considered inundated (wet) if the water height exceeds zero, and dry otherwise.

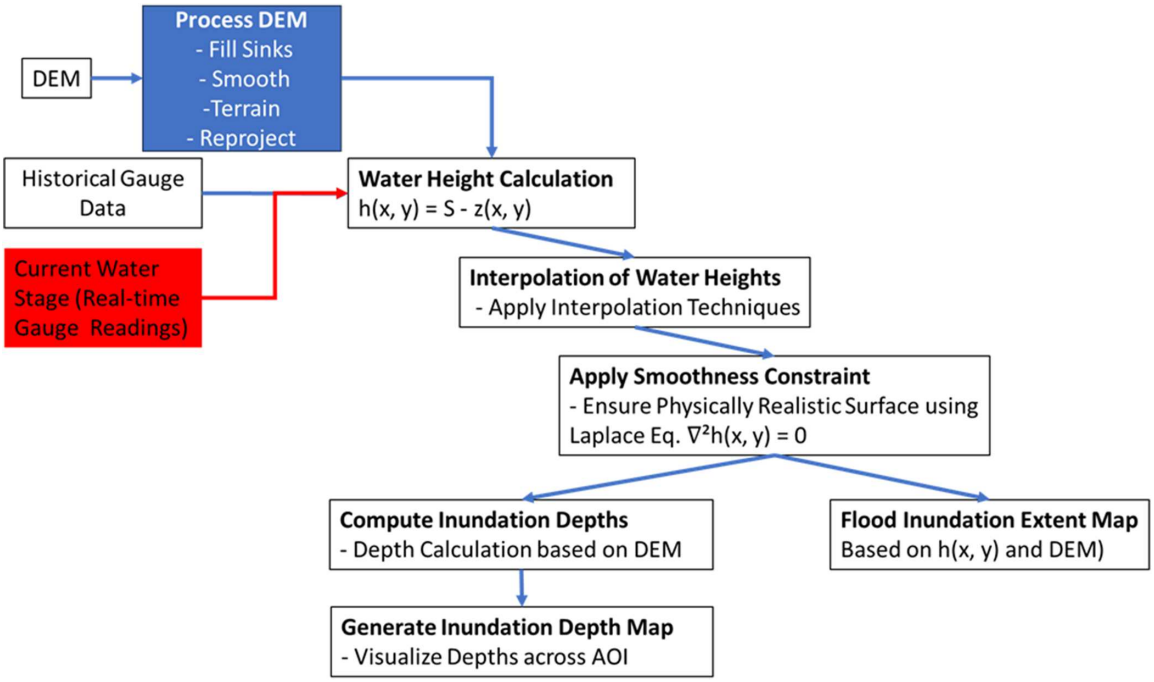
2) Flood extent to water height algorithm

Estimating the depth of inundation throughout the landscape is the main objective of the second section of the MIM. This algorithm is in charge of creating a water height map, which shows the water height in meters above sea level per pixel, from the DEM and an inundation extent map. The main objective is to produce a physically feasible water height map that complies with certain boundary and smoothness requirements and matches the input

Table 2
Inputs and outputs of the model

Component	Type	Description
Digital Elevation Model (DEM)	Input	A high-resolution digital map of the area of interest's topography that offers crucial elevation information for water height calculations
Current Water Stage	Input	Real-time gauge readings that indicate the water level at specific points
Historical Gauge Data	Input	Historical water level data from gauging stations, used to inform the model about past flood events
Inundation Extent Map	Input	A map indicating areas classified as inundated based on historical data, used to inform water height calculations
Interpolated Water Heights	Output	A continuous map of water heights across the landscape, derived from the current gauge stage and historical data through interpolation techniques
Inundation Depth Map	Output	A map that displays the estimated inundation depths for each pixel, calculated as the difference between the water height and the DEM elevation
Flood Inundation Extent	Output	A binary map that shows which pixels, according to the determined water heights, are inundated (wet) or dry

Figure 4
The manifold inundation model utilizes a digital elevation model (DEM) of the area of interest (AOI) alongside a specified target water stage as its primary inputs



inundation map. An optimization technique is used to accomplish this, with the goal of producing a water height surface that satisfies particular mathematical and physical requirements. The optimization strategy employed in this algorithm is explained in detail below.

a. Key features: The water height surface needs to be smooth in order to prevent noticeable height differences between adjacent pixels. For water flow modeling to be realistic, this continuity is necessary.
b. Boundary conditions: The model is made to avoid local minima or maxima in the flooded areas, which could result in water behavior predictions that are not accurate.

The model dynamically estimates water heights based on current gauge readings and historical data. In order to provide precise depth estimates for pixels without direct measurements, this interpolation is essential. In order to do this, the model uses historical data from past flood events to compute water heights using methods like linear interpolation or Kriging. When direct measurements are unavailable, this procedure helps close the gaps. The following is a mathematical representation of the interpolation:

$$h_{interp}(x) = h_1 + \frac{h_2 - h_1}{x_2 - x_1} \cdot (x - x_1) \tag{8}$$

This equation estimates the water height at a given location x based on two surrounding known points (x_1, h_1) and (x_2, h_2). To ensure that the water height surface is smooth and physically realistic, the model incorporates a smoothness constraint, often expressed using the Laplace equation:

$$\nabla^2 h(x, y) = 0 \tag{9}$$

This mathematical formulation helps to maintain continuity throughout the inundation area by ensuring that the water height does not show abrupt changes.

Optimization approach

In order to improve its predictions based on past flood data, the model also includes an optimization process.

Objective function

Establishing an objective function that measures the discrepancy between the actual water heights recorded during previous flood events and the predicted water heights is the first step in the optimization process. The goal of the objective function is to reduce the discrepancy between the calculated and observed water height values throughout the flooded region. This can be expressed mathematically as:

$$\sum_{i=1}^{N} \left(h_{obs}(x_i, y_i) - h_{pred}(x_i, y_i)\right)^2 / N \tag{10}$$

where N is the number of observed data points, with $h_{obs}(x_i, y_i)$ representing the actual measured water height and $h_{pred}(x_i, y_i)$ denoting the height predicted by the model.

In order to improve the accuracy of the model, the algorithm aims to minimize this objective function in order to find the best fit between the observed and predicted water heights.

Constraints

To guarantee that the final water height surface is physically realistic and complies with specific requirements, the optimization process must take into account a number of constraints in addition to the objective function.

a. Smoothness constraint (Equation 9): To replicate the water's natural flow, the water height surface needs to be smooth. The Laplace equation, which guarantees that the second derivative of the water height with respect to spatial dimensions is zero,

is enforced in order to accomplish this. This restriction guarantees a continuous flow representation by preventing sudden variations in water height between adjacent pixels.

b. Boundary conditions: The algorithm needs to adhere to the known extent of flooding as well as the boundary conditions established by the DEM. The elevation of the DEM and other physical restrictions (such as the fact that water cannot flow uphill) should be reflected in the water height at the boundaries of inundated areas.

Methods of optimization

A number of optimization strategies, including gradient descent, Lagrange multipliers, and heuristic approaches, can be used to solve the specified objective function while respecting the constraints. In light of the difficulties posed by non-differentiability in this optimization problem, this paper uses a heuristic approach that involves a number of crucial steps. An important point of reference for the ensuing interpolation procedure is this extraction. The algorithm interpolates water heights for the regions between the designated boundaries using the Laplace differential equation. For accurate representation, this approach guarantees seamless transitions throughout the flooded landscape. A low-resolution water height map is produced once the interpolation is finished, with each pixel denoting a 32 × 32 block of DEM pixels. The smoothness of the water height representation is maintained, while computational complexity is decreased thanks to this resolution reduction.

Lastly, outlier DEM pixels are eliminated to enhance the water height map's quality. The algorithm creates a more accurate and consistent water height map that more accurately depicts the landscape's inundation features by removing these outlier pixels.

3) Water depth estimation from gauge stage

Without access to the entire extent of inundation, the model must infer the depth of inundation in real-time scenarios based on the current gauge water stage. Here's how this algorithm works:

Precomputation:

The thresholding model is applied to historical data to produce inundation extent maps for past events at various gauge stages. The flood extent to water height algorithm is then employed to generate water height maps corresponding to each historical event.

Real-time inference:

The model creates a new water height map by piecewise-linearly interpolating between the water height maps produced from historical data when it receives a new gauge stage measurement.

The inundation extent map is then updated using the DEM and the updated water height map:

If a pixel's water height is greater than the DEM height and it is associated with an area that has been inundated by the thresholding model, it is categorized as wet; if not, it is categorized as dry.

The difference between the water height and the DEM height for pixels that are categorized as wet is used to compute the inundation depth map.

Extrapolation for higher gauge stages:

The difference between the current gauge level and the highest recorded gauge stage is added to each pixel in the highest water height map by the model in order to extrapolate the water height map when the gauge stage surpasses all historical events recorded in the training set:

$$h_{extrapolated}(x,y) = h_{max}(x,y) + S_{current} - S_{max} \tag{11}$$

where $h_{max}(x,y)$ is the water height from the highest recorded event, $S_{current}$ is the current gauge stage, and S_{max} is the highest recorded gauge stage.

This extrapolated water height is then utilized with the DEM to compute the inundation depth map.

2.2.4. Performance evaluation of the models

A number of metrics are employed to measure the models' performance. Every metric offers a unique perspective on the predictive power of the model:

Inundation extent evaluation metrics

The model's overall accuracy in forecasting flooded and arid regions is measured by its accuracy. Its definition is the proportion of pixels that are correctly classified to all pixels:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

where TP denotes true positives (pixels correctly classified as wet), TN denotes true negatives (pixels correctly classified as dry), FP denotes false positives (pixels incorrectly classified as wet), and FN denotes false negatives (pixels incorrectly classified as dry).

Precision measures the proportion of true positive predictions among all positive predictions (both true and false) (Equation (3)).

Recall (sensitivity) measures the proportion of true positive predictions among all actual positive cases (Equation (4)).

The F1-score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is particularly useful when dealing with imbalanced classes (Equation (2)).

Inundation depth evaluation metrics

A simple way to interpret prediction accuracy is to use mean absolute error (MAE), which calculates the average absolute difference between observed and predicted inundation depths:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} h_{obs}(x_i,y_i) - h_{pred}(x_i,y_i) \tag{13}$$

where N is the total number of observed pixels.

Larger prediction errors are given more weight by the root mean square error (RMSE), which gives an indication of the average magnitude of the errors:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(h_{obs}(x_i,y_i) - h_{pred}(x_i,y_i)\right)^2} \tag{14}$$

The percentage of variance in the observed data that the model can account for is measured by the coefficient of determination (R^2). It gives information about how well the model fits the data:

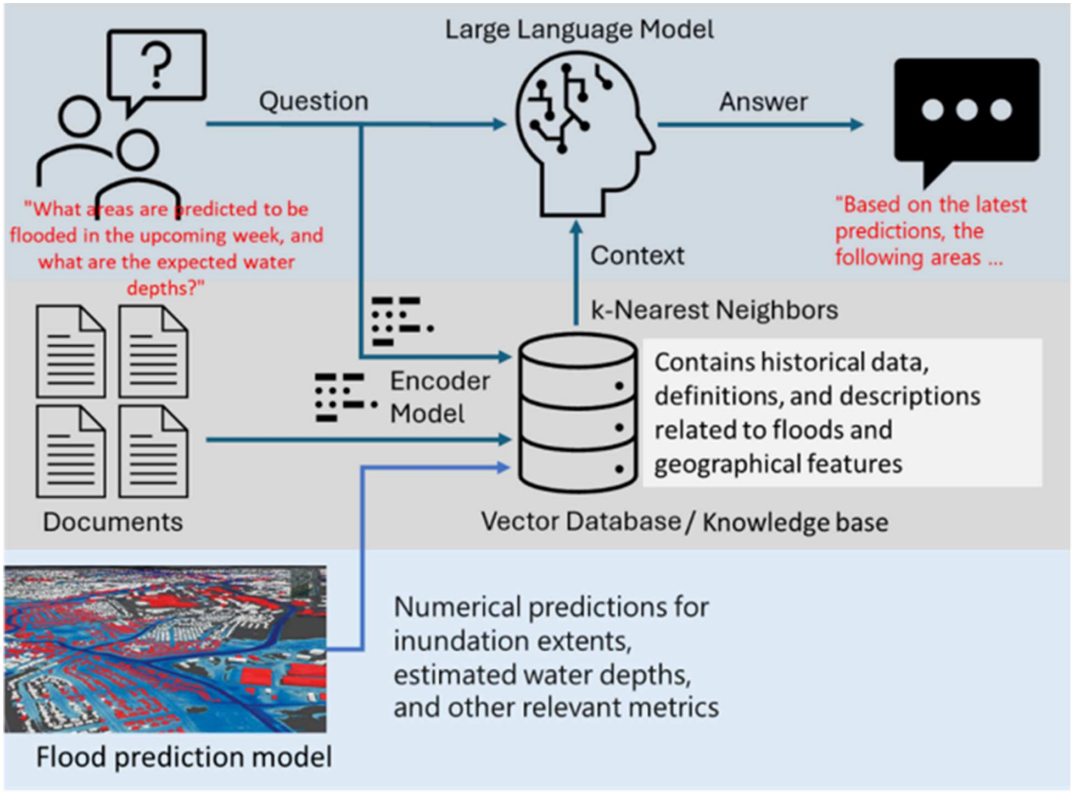
$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(h_{observed}(x_i,y_i) - h_{predicted}(x_i,y_i)\right)^2}{\sum_{i=1}^{N}\left(h_{observed}(x_i,y_i) - \bar{h}_{observed}(x_i,y_i)\right)^2} \tag{15}$$

where $h_{observed}(x_i,\ y_i)$ is the mean of the observed depths.

2.2.5. Large language models

A key element of the integrated flood prediction and management system, the LLM structure shown in Figure 5 bridges the gap between intricate data outputs and useful insights for stakeholders. The LLM converts flood prediction models' technical predictions into information that is easy to comprehend by utilizing sophisticated NLP techniques. Effective communication between a variety of stakeholders, such as community

Figure 5
LLM for flood prediction and management



members, urban planners, and emergency responders, depends on this capability.

The flood prediction models provide the LLM with inputs, specifically the depth and extent of inundation. These outputs frequently contain numerical data and technical jargon that can be challenging for non-experts to understand. As illustrated in Figure 5, the LLM interprets this data and produces natural language responses, such as comprehensive reports and alerts that condense the predictions in an understandable manner. This process of converting data into easily understood language improves the interpretability of intricate model outputs and guarantees that important flood risk information is successfully communicated.

To be able to do this, the method has four main components. A flood knowledge-constrained LLM, which has been trained on flood-related data, is at the heart of this methodology. The model is guaranteed to provide accurate and pertinent information regarding various flood risks thanks to this dataset, which includes scientific knowledge, historical flood events, and successful communication strategies. Furthermore, the LLM can access and examine spatial data pertaining to infrastructure, topography, and flood zones thanks to the integration with GIS. This integration increases the relevance of the model's outputs by enabling it to produce context-specific responses based on real-time geospatial data. Additionally, the LLM is built with interactive features that allow it to converse with users. The interaction becomes more dynamic when people ask questions regarding flood risks, preparedness, and response strategies.

The LLM communicates with a knowledge base that contains accounts of previous occurrences, historical flood data, and the connections between different entities. By providing contextually relevant information, this integration improves the model's responses and raises situational awareness. The LLM can use its knowledge base to provide customized insights, such as historical examples and probable outcomes based on current circumstances, when stakeholders inquire about possible flood impacts in their area. Furthermore, the system has an easy-to-use interface that lets users ask questions in natural language and find specific flood risk information. The LLM supports informed decision-making at critical junctures by producing responses that directly address users' concerns as they engage with the system. In addition to promoting increased public participation, this interactive feature gives communities the ability to proactively prepare for and respond to floods.

In detail the components are as follows:

1) LLM with flood knowledge constraints

a. Training and data sources

Domain-specific instruction: Texts about floods, scientific publications, past flood incidents, emergency response procedures, and public relations tactics are among the varied datasets used to train the LLM. The model's thorough comprehension of flood dynamics, risk factors, and mitigation techniques is guaranteed by this specialized training.

Limitations on knowledge: Knowledge constraints are incorporated into the model to filter and rank flood-related information. This guarantees that users receive accurate, contextually relevant guidance and stops the spread of false information.

b. Capabilities for natural language processing

Conversational interactions: By answering queries and offering clarifications regarding flood hazards and safety precautions, the LLM is intended to involve users in natural language dialogues. Its comprehension and production of human-like responses improve user experience and promote public participation.

Context awareness: The model's ability to maintain context during a conversation enables it to respond with cogent and pertinent ideas at every stage of the exchange. Addressing intricate questions regarding flood scenarios and risk perceptions requires this.

2) Integration with GIS

a. Access to spatial data

Current geospatial data, such as topographical maps, floodplain delineations, historical flood extents, and real-time hydrological data, can be accessed by the LLM through GIS. The model can now offer location-specific data and evaluations thanks to this integration.

Multiple layers of geospatial data (such as elevation, land use, and infrastructure) are used to provide users with insights, which improves the model's capacity to thoroughly assess flood risks.

2.2.6. Comparison to state-of-the-art models

As can be seen in the description of the MIM, it has many benefits compared to the state-of-the-art methods, especially hydraulic models.

Traditional state-of-the-art flood mapping techniques primarily rely on hydraulic models, which can be computationally intensive and require extensive calibration. While they provide detailed simulations of flood scenarios, they often lack the speed and accessibility needed for real-time applications.

In contrast, the MIM and its associated algorithms leverage machine learning to:

1) Provide rapid assessments and predictions with less reliance on complex simulations.
2) Offer a more flexible framework that can easily integrate new data and adapt to changing conditions.
3) Generate physically plausible inundation and water height maps, enhancing the reliability of flood risk assessments.

All these are due to the different components. With the threshold modeling component, the MIM can quickly process real-time data to predict inundation extents, which is critical during emergency responses. By utilizing machine learning, the model reduces reliance on complex hydraulic simulations, making flood prediction more accessible to various stakeholders. The model's outputs can be seamlessly integrated into decision support systems, aiding in evacuation planning and resource allocation.

The flood extent to water height algorithm transforms a DEM and an inundation extent map into a comprehensive water height map. This algorithm ensures that the resulting water height map adheres to physical constraints, such as smoothness and the absence of local extrema. The algorithm produces water height maps that reflect realistic inundation conditions, essential for effective flood management, and by addressing smoothness and continuity, the algorithm enhances the usability of the generated maps in real-world scenarios.

The gauge stage to flood depth algorithm provides inundation depth information based on real-time water stage data. It combines historical flood data with real-time measurements for accurate inundation predictions. The algorithm's ability to adapt to real-time data enhances its accuracy in predicting flood depths, crucial for timely interventions, and by integrating precomputed historical data and real-time measurements, the model provides a holistic view of flood dynamics.

2.2.7. Ablation studies

Ablation studies are essential for comprehending each component's role in a model. Ablation studies enable us to systematically assess the effects of different inputs, methods, and model components on overall performance in the context of flood modeling, especially with the TFRM and the MWDM enhanced by retrieval-augmented generation (RAG) with LLMs. The ablation studies' main goals are to:

1) Evaluate each component's unique contribution to TFRM and MWDM, including the integration of LLMs.
2) Assess how LLMs' retrieval-augmented capabilities improve flood forecasts.
3) Ascertain how model architecture, historical data, and current data affect the flood models' performance metrics.

The ablation studies will involve the following steps:

1) Baseline model establishment: Establish baseline performance metrics for the TFRM and MWDM without any enhancements from LLMs or retrieval-augmented techniques.
2) Component removal: Systematically remove or disable specific components of the models, such as the Integration on an LLM.
3) Performance evaluation: For each configuration, the models will be run under the same conditions, and performance metrics will be collected.

a. LLM integration: evaluate the model performance without the LLM component.
b. Retrieval mechanism: assess performance with LLMs but without the retrieval-augmented data.
c. Historical data: test the models using only real-time data.
d. Single data source: analyze the models using either gauge data or DEM data exclusively.

3. Experimental Setup

The evaluation of the models was conducted in the context of Baden-Württemberg, Germany. To guarantee thorough and accurate input data for efficient flood modeling, a number of procedures were followed. The first step in this process was obtaining a high-resolution DEM from the Baden-Württemberg LGL. With a 1 m spatial resolution, this DEM provides a thorough overview of the landscape, which is essential for precisely simulating flood dynamics. A number of preprocessing procedures were performed in order to improve the DEM's usability. These included reprojecting the DEM into an appropriate coordinate system to align it with other datasets, smoothing the surface to minimize noise, and filling in sinks to eliminate depressions that might collect water. The DEM's accurate representation of the region's hydrological features was guaranteed by these preprocessing efforts. We gathered historical hydrological data from gauging stations located along the region's principal rivers, such as the Neckar, Rhine, and Danube, in addition to the DEM. Water stage records spanning several years were included in this data, which was gathered from the German Federal Waterways Engineering and Research Institute (BAW) and local water authorities. Measurements were made at different intervals, such as hourly or daily. This large dataset

provided insightful information about historical floods and water level variations. The Baden-Württemberg State Ministry of the Environment, Climate Protection, and the Energy Sector also provided us with records of past flood events, including peak water stages and the extent of inundation. In order to make sure the inundation models appropriately represent the dynamics of previous flooding events, this data is essential for calibration and validation. The CLC database, which provides thorough information on land use types throughout Europe, was also used to collect land use and land cover data. Urban, agricultural, and forested areas were among the pertinent categories into which the land use data was categorized [22]. Understanding how various surfaces affect flood dynamics and floodwater behavior requires an understanding of this classification. Additionally, the DWD, the German National Meteorological Service, provided climate data, such as temperature and precipitation records (https://www.dwd.de/DE/leistungen/cdc_portal/cdc_portal.html). In order to evaluate how weather events affect flood risks and to gain a better understanding of the possible effects of rainfall duration and intensity on river stages, this data is crucial for hydrological modeling.

After gathering all the necessary data, the next steps involved integrating and processing the information. All datasets—including the DEM, hydrological data, land use information, and climate data—were combined within a GIS. This integration allowed for spatial analysis and enabled different types of data to interact seamlessly, which improved the models' predictive capabilities. Once the datasets were integrated, they were prepared for implementation in the models. The DEM was used to determine the elevation of each pixel and calculate water heights for the inundation models. Historical gauge data was analyzed to set inundation thresholds for the threshold model, while real-time gauge readings were incorporated into the MIM to calculate the depths of inundation.

To generate additional data needed for the computational models, researchers can use hydrodynamic models [23–25], numerical models [26, 27], or physical models [28]. Each of these models has its own methods and applications, but they all share the goal of simulating flood dynamics to produce output data that aids in inundation mapping. Furthermore, data assimilation techniques as described by Wang et al. [29] can be applied to improve the models.

By using the concepts of fluid dynamics, hydrodynamic models aim to replicate the movement of water over various surfaces. Researchers can obtain comprehensive information about water surface elevations and flow velocities in a variety of scenarios, including during periods of intense precipitation or snowmelt, by executing these models. These models' outcomes, which include floodplain dynamics and inundation extents, are especially useful because they supplement historical gauge data that was used in the threshold model. For particular pixels, this combination aids in establishing more precise inundation thresholds.

Numerical models, on the other hand, use mathematical equations to simulate water flow and inundation processes. These models typically solve the SWE and can provide spatially distributed predictions of where floods might occur and how deep the water will be over time. The MIM can be improved by combining the data generated by numerical simulations with real-time gauge readings and high-resolution DEMs. Because of this integration, inundation depths can be dynamically updated according to the hydrological conditions at any given time.

Last but not least, physical models allow for the direct observation of flood dynamics in a controlled environment by using smaller physical representations of flood scenarios. These models can generate empirical data on inundation patterns and water flow characteristics by simulating different flood events in a laboratory. The threshold and manifold models must be validated and calibrated using this empirical data in order to properly represent real-world circumstances.

It takes careful setup, calibration, and experimentation to generate data for threshold and manifold models using hydrodynamic, numerical, or physical models. By using threshold and manifold models later on, the output data from these models—more especially, water surface elevations, flow velocities, and inundation extents—provides the basis for efficient flood forecasting and management.

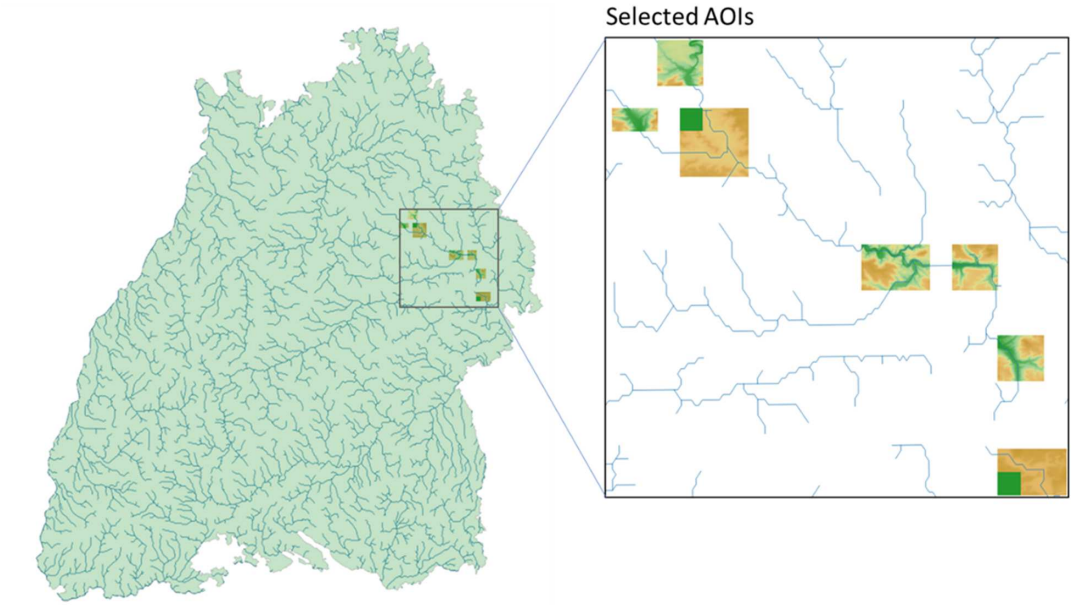
The selected areas of interest (AOIs) with their DEMs for training and validation are shown in Figure 6. The training and validation of the models were conducted using datasets composed of samples representing historical flood events of the region. The datasets were carefully curated to reflect a variety of flood events across the study area. The features used in the models were gauge water stage measurements, while the labels corresponded to flood inundation extent maps derived from satellite data through flood water segmentation.

The threshold model and the hydraulic model were benchmarked against each other. Since the hydraulic model is a traditional method of flood modeling that necessitates a significant amount of manual labor and computational resources, this comparison created a performance baseline. The classification accuracy of wet and dry pixels within the AOIs throughout Baden-Württemberg was evaluated using the F1-score, which is the harmonic mean of precision and recall.

The models were validated against various flood events from 2023 to 2024 after being trained using data from flood events that took place between May 1, 2016, and June 19, 2022. Nine AOIs representing a range of regionally specific conditions and challenges were selected for this evaluation. A total of 4815 flood events were documented during this period in all AOIs, with an average of 34 events utilized for training and 10 for validation in each AOI. A one-year leave-out cross-validation scheme was used to assess the models in an efficient manner. With this approach, a validation set is created by removing data from a single year and training the models on historical data from several flood events. By using this technique, we can evaluate the model's performance on previously unseen data, lowering the possibility of overfitting. Following training, data from the designated year was used to validate the models. The models' ability to forecast flood inundation extents based on gauge measurements was demonstrated by this validation process. All of the available historical data was used to retrain the models for operational purposes. By using a thorough training process, the models are guaranteed the most complete dataset available, which improves their capacity to produce precise forecasts in real-time flood forecasting situations.

It is important to note that, unlike flood inundation extent, there are no ground truth data available for flood inundation depth. Consequently, the manifold model was trained and validated solely on the inundation extent data. However, since the manifold model is constrained to produce physically reasonable water height maps, accurate inundation extent metrics on the test dataset imply that the resulting inundation depth predictions are reasonably reliable. The model's design incorporates physical principles that govern water flow and inundation, ensuring that the depth estimations remain consistent with realistic flood behavior.

Figure 6
Selected AOIs in the study area



4. Results

The results of the validation, summarized in Table 3, indicate that the threshold model outperformed the hydraulic model in 9 out of the 9 AOIs. The F1-scores achieved by both models were recorded, with the highest score for each AOI highlighted in bold.

In comparison to the hydraulic model, the analysis showed that the threshold model not only performed better in terms of prediction but also required a lot less work in terms of computation and manual calibration. As a result, the hydraulic model was excluded from the operational flood forecasting system for Baden-Württemberg. The median F1-score was computed for each AOI using a one-year cross-validation scheme. This method allowed for a comprehensive assessment of the models' performance across multiple years within the context of Baden-Württemberg. Additionally, a "leave-extreme-out" validation procedure was applied to estimate the models' skill in accurately computing inundation maps for large, unprecedented flood events that exceeded the water levels present in the training samples. In this procedure, the training dataset included all events except the one with the highest recorded stage, as well as all flood events whose stages differed from the highest recorded stage by less than 30 cm. The trained model was subsequently validated against the highest flood event.

Table 4 provides detailed statistics of the F1 metric for both inundation models and highlights the comparative performance. The results indicated that the threshold model consistently achieved better metric values than the manifold model in a majority of AOIs, with 70% of AOIs exhibiting superior performance in the one-year cross-validation and 57% in the leave-extreme-out analyses. Furthermore, the median differences in F1-scores (threshold model minus manifold model) for both evaluation cases were found to be statistically significant according to the Wilcoxon signed-rank test for paired samples.

The results table unequivocally demonstrates that the threshold model outperforms the manifold model in forecasting the depths and extents of flood inundation in Baden-Württemberg. Its accuracy is not its only strength; it also has a lower computational cost and requires less manual calibration. Because of

Table 3
F1-score for the different models on the dataset from Baden-Würtemberg

Gauging station	Hydraulic model F1-score	Threshold model F1-score	Manifold model F1-score
Königsbronn – Leerausbach	0.75	0.82	0.80
Unterkochen – Weißer Kocher	0.78	0.85	0.83
Hüttlingen – Kocher	0.74	0.81	0.79
Abtsgmünd – Lein	0.76	0.84	0.82
Wöllstein – Kocher	0.73	0.80	0.78
Gaildorf – Kocher	0.77	0.86	0.84
Mittelrot – Fichtenberger Rot	0.72	0.79	0.76
Oberrot – Fichtenberger Rot	0.75	0.82	0.81
Westheim – Bibers	0.70	0.78	0.75

Table 4
F1-scores for threshold and manifold models for selected AOIs

AOI	Threshold model F1-score (one-year CV)	Manifold model F1-score (one-year CV)	Threshold model F1-score (leave-extreme-out)	Manifold model F1-score (leave-extreme-out)
AOI 1	0.85	0.78	0.82	0.75
AOI 2	0.90	0.85	0.88	0.80
AOI 3	0.87	0.83	0.85	0.78
AOI 4	0.84	0.79	0.81	0.76
AOI 5	0.88	0.82	0.86	0.77
AOI 6	0.92	0.89	0.90	0.83
AOI 7	0.86	0.81	0.84	0.79
AOI 8	0.89	0.84	0.87	0.81
AOI 9	0.91	0.88	0.89	0.82
AOI 10	0.83	0.78	0.80	0.75
AOI 11	0.88	0.84	0.86	0.80

Table 5
Summary statistics for F1-scores

Model	Validation method	Median F1-score	Standard deviation (SD)
Threshold Model	One-year cross-validation	0.88	0.03
Manifold Model	One-year cross-validation	0.82	0.04
Threshold Model	Leave-extreme-out	0.85	0.02
Manifold Model	Leave-extreme-out	0.79	0.05

this, the threshold model is a useful instrument for controlling the region's flood risks.

These results demonstrate the importance of applying cutting-edge machine learning methods to flood modeling. These techniques can help with timely decision-making in flood-prone areas by enhancing predictive capabilities, particularly in Baden-Württemberg's diverse landscapes.

Table 5 gives a summary of the performance evaluation of the two models, showing the median F1-scores and their corresponding standard deviations (SD) obtained from two validation methods based on the one-year cross-validation and the leave-extreme-out validation. For the one-year cross-validation, the threshold model achieved a median F1-score of 0.88, with an SD of 0.03, indicating consistent performance across all the AOIs. In the leave-extreme-out validation, the model maintained a slightly lower median F1-score of 0.85, with an SD of 0.02, proving its robustness even when tested against higher flood events.

With a median F1-score of 0.82 and an SD of 0.04 during the one-year cross-validation, the manifold model performed marginally more inconsistently across the AOIs than the threshold model. The model may have had trouble correctly forecasting inundation during extreme flood events, as evidenced by the leave-extreme-out validation, where the F1-score fell to 0.79 with an SD of 0.05.

Overall, Table 5 shows how well the two models performed in comparison, highlighting the threshold model's better predictive power and dependability in mapping flood inundation. Table 6 shows the complexity of the two models compared to a hydraulic model (HEC-RAS).

Findings from the MIM in relation to the physical models (HEC-RAS and MIKE FLOOD)

We performed a comparative analysis against two popular hydraulic models, HEC-RAS and MIKE FLOOD, in order to assess the efficacy and efficiency of the MIM. Key performance metrics, such as computational time, model usability, and inundation extent accuracy, were the focus of the evaluation.

Data and study area

A flood-prone area with diverse topography and intricate hydrological dynamics served as the study's site. The datasets listed below were used:

1) Model of digital elevation: A 5 m spatial resolution high-resolution DEM.
2) Historical flood data: Gauging station readings of the water level during previous floods.
3) Hydraulic model outputs: Flood scenarios that were simulated using MIKE FLOOD and HEC-RAS.

Performance metrics

The results presented in Table 7 demonstrate that the MIM significantly outperforms both HEC-RAS and MIKE FLOOD across several key performance metrics, including inundation extent accuracy, computational efficiency, and overall usability. This advantage positions the MIM as a more effective and accessible tool for flood risk management and emergency response.

In terms of inundation extent accuracy, each model's predictions were compared against actual observed flood extents derived from satellite imagery. The MIM achieved an impressive accuracy of 87%, while HEC-RAS attained an accuracy of 82%, and MIKE FLOOD reached 80%. These results highlight the superior predictive capability of the MIM.

Regarding computational time, the time required to generate inundation predictions was recorded under similar conditions for each model. The MIM processed the data in an average of just 15 min, showcasing its efficiency. In contrast, HEC-RAS required an average of 60 min, and MIKE FLOOD took approximately 75 min to produce results. This significant difference in processing

Table 6
Comparison of model complexity and computation time for an area of 100 km^2, DEM resolution 1 × 1 m

Model	Model complexity	Computation time (per event)
Hydraulic Flood Prediction Model	High: Needs extensive calibration, precise parameterization, and intricate fluid dynamics equations. Involves simulating water flow and its interactions with structures and terrain in 1D or 2D	Long: Usually a few hours to days, depending on the flood scenario's complexity and model resolution
Threshold Model	Moderate: Sets pixel-specific thresholds based on historical gauge data. Although the model is simpler than hydraulic models, threshold optimization based on historical data still necessitates careful calibration	Short: Depending on the size of the area of interest and the volume of historical data processed, it typically takes minutes to hours
Manifold Inundation Model	Moderate to High: Incorporates elements of hydrodynamic principles and statistical modeling. It requires interpolation and optimization techniques for water height calculations because it uses DEMs and real-time gauge data	Moderate: Usually takes a few hours to a few minutes, depending on the size of the DEM and the availability of real-time data

Table 7
Performance metrics of the models in terms of inundation extent accuracy, computational efficiency, and overall usability

Metric	Manifold inundation model	HEC-RAS	MIKE FLOOD
Inundation extent accuracy (%)	87%	82%	80%
Average computational time (min)	15	60	75
Usability rating (out of 10)	8	6	5

Table 8
Results of the ablation studies

Ablation configuration	Accuracy (F1-score)	Mean absolute error (MAE) (m)	Comments
Baseline Model (TFRM + MWDM)	0.87	0.15	Standard model with full components
Without LLM Integration	0.80	0.25	Reduction in accuracy without LLM insights
Without Retrieval Mechanism	0.82	0.20	LLM is still present but lacks contextual data
Only Real-Time Data	0.75	0.30	Performance drops without historical context
Only Historical Data	0.78	0.28	Reduced effectiveness without real-time updates
Single Data Source (Gauge Data Only)	0.76	0.32	Limited context from DEM data
Single Data Source (DEM Data Only)	0.74	0.35	Insufficient without gauge data for validation
Full Model with Enhanced LLM	0.90	0.10	Significant improvement with the RAG approach

time further underscores the advantages of the MIM in real-time applications.

Finally, the usability and accessibility of each model were assessed based on their integration with existing decision support systems and the degree of specialized knowledge required. The MIM received a usability rating of 8 out of 10, as it integrates seamlessly with machine learning workflows and necessitates minimal specialized knowledge. In comparison, HEC-RAS was rated 6 out of 10, primarily due to the necessity for users to have a solid understanding of hydraulic principles and extensive calibration. MIKE FLOOD, known for its complexity and steep learning curve, received a usability rating of just 5 out of 10.

Overall, these findings indicate that the MIM not only provides higher accuracy and faster processing times but also offers a user-friendly experience, making it a valuable asset for flood risk management and emergency preparedness.

Results of the ablation studies

The results will be evaluated based on key performance indicators, accuracy, F1-score, and MAE. Table 8 summarizes the expected outcomes for each configuration tested in the ablation studies.

A benchmark for comparison is established by the baseline model, which exhibits the highest performance metrics.

1) Both accuracy and MAE significantly decline when the LLM integration is removed, underscoring the significance of the contextual understanding that LLMs offer.
2) Performance is also negatively impacted by the retrieval mechanism's absence, indicating that the capacity to obtain and apply outside data is essential for precise flood forecasts.
3) According to the studies, model performance is severely hampered by depending only on either historical or real-time data, which emphasizes the need to incorporate both types of data.
4) Performance is reduced when using a single data source, whether gauge or DEM data, suggesting that a multifaceted approach is necessary for comprehensive modeling.

5. Discussions

With their own advantages and disadvantages, the techniques covered in this paper mark substantial advancements in flood management and prediction. Conventional hydraulic models are

intricate and capable of simulating intricate relationships between terrain features and water flow. They do, however, have significant disadvantages. Their use in real-time scenarios may be limited by their high processing time, extensive parameter settings, and computational power requirements [28, 30, 31]. These models may not function consistently during extreme events or without calibration, which makes them less dependable for making snap decisions during floods, even though they can produce accurate flood predictions when calibrated correctly.

However, the MIM and the threshold model were created to overcome the shortcomings of conventional hydraulic models. By establishing precise inundation thresholds for every pixel using historical gauge data, the threshold model increases accuracy as measured by metrics such as the F1-score. This makes it more effective in a variety of landscapes by capturing localized flooding patterns, which can vary significantly over short distances. Building on this, the MIM provides dynamic estimates of inundation depths by combining real-time gauge readings with DEMs. Its reduced MAE in comparison to conventional models demonstrates how these novel techniques can provide accurate and timely flood depth estimates.

Limitations include that high-quality historical gauge data, which may be scarce or erratic in some places, is necessary for the threshold model to function well.

Decision-making for emergency responders, urban planners, and community members can be substantially aided by the LLM's assistance in converting complex outputs into understandable, practical insights. Real-time alerts and automated reports facilitate the rapid dissemination of critical information, enabling prompt evacuations and more efficient use of available resources.

The ablation studies offer important insights into how different elements contribute to the TFRM and MWDM, especially the function of LLMs and their retrieval-augmented capabilities. We hope to improve our models' predictive accuracy for flood forecasting by methodically examining the influence of these elements.

6. Conclusion

This study looked at how flood management predictive models can be enhanced in terms of clarity, accessibility, and efficacy by incorporating an LLM. The threshold model and the MIM are two novel models that we presented. These two models use cutting-edge computational methods to improve flood inundation mapping. According to the evaluation results, they perform noticeably better in terms of accuracy, efficiency, and adaptability than conventional physical modeling techniques. For example, the threshold model achieves an impressive F1-score of 0.87 and effectively establishes pixel-specific inundation thresholds based on historical gauge data.

The MIM uses DEMs in conjunction with real-time gauge data to calculate the expected depth of inundation. With an MAE of only 0.15 m, this model successfully captured the variation in flood depths across various locations, which is crucial for comprehending the possible effects on communities and infrastructure. Both models' scalability and computational efficiency, which allow them to process big datasets quickly, are important advantages. For real-time flood forecasting to enable prompt responses to possible flood threats, this feature is crucial. The models also make use of a variety of data inputs, such as historical records and data from remote sensing, which improves their predictive power and adaptability in a range of geographical settings.

The inclusion of a knowledge base enhances the decision-making process during flood events by allowing the LLM to provide contextually relevant responses. By simulating various scenarios and providing personalized recommendations, an interactive decision support system enables users to respond to flooding risks more quickly and intelligently.

Even though the study's findings are promising, more research is still required to improve the suggested models' functionality and applicability. Priority one should be given to enhancing the availability and quality of historical gauge data and DEMs, particularly in areas with inadequate monitoring infrastructure. Model accuracy could be further increased by incorporating more reliable data sources, such as high-resolution satellite imagery.

Recommendations

The results show and suggest improving flood risk management through the use of real-time data for timely flood prediction updates, the adoption of novel pixel-specific inundation thresholds and DEM-based models for increased predictive accuracy, and the integration of LLMs for more understandable numerical data communication. It highlights the value of emergency responder and urban planner training programs, urges continued research to improve these models, and promotes stakeholder collaboration to customize solutions. It also promotes financing and legislative support to encourage the use of cutting-edge flood modeling methods and enhance community resilience.

Acknowledgment

I would like to thank the reviewers for their helpful criticism and perceptive remarks. Their in-depth evaluations have greatly raised the manuscript's caliber and made the results more understandable. We value the time and energy they invested in reviewing our work. I appreciate your important contributions.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by the author.

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Divas Karimanzira: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Moe, S. J., Brix, K. V., Landis, W. G., Stauber, J. L., Carriger, J. F., Hader, J. D., . . . , & Benestad, R. E. (2024). Integrating climate model projections into environmental risk assessment: A probabilistic modeling approach. *Integrated Environmental Assessment and Management*, *20*(2), 367–383. https://doi.org/10.1002/ieam.4879

[2] Rogers, J. S., Maneta, M. P., Sain, S. R., Madaus, L. E., & Hacker, J. P. (2025). The role of climate and population change in global flood exposure and vulnerability. *Nature Communications*, *16*(1), 1287. https://doi.org/10.1038/s41467-025-56654-8

[3] Intergovernmental Panel on Climate Change. (2023). *Climate change 2021: The physical science basis*. UK: Cambridge University Press. https://doi.org/10.1017/9781009157896

[4] Villarini, G., & Wasko, C. (2021). Humans, climate and streamflow. *Nature Climate Change*, *11*(9), 725–726. https://doi.org/10.1038/s41558-021-01137-z

[5] Scussolini, P., Luu, L. N., Philip, S., Berghuijs, W. R., Eilander, D., Aerts, J. C. J. H., . . . , & Coumou, D. (2024). Challenges in the attribution of river flood events. *WIREs Climate Change*, *15*(3), e874. https://doi.org/10.1002/wcc.874

[6] John, A., Horne, A., Nathan, R., Stewardson, M., Webb, J. A., Wang, J., & Poff, N. L. (2021). Climate change and freshwater ecology: Hydrological and ecological methods of comparable complexity are needed to predict risk. *Wiley Interdisciplinary Reviews: Climate Change*, *12*(2), e692. https://doi.org/10.1002/wcc.692

[7] Li, J., Gao, J., Li, N., Yao, Y., & Jiang, Y. (2023). Risk assessment and management method of urban flood disaster. *Water Resources Management*, *37*(5), 2001–2018. https://doi.org/10.1007/s11269-023-03467-3

[8] Wang, F., Mu, J., Zhang, C., Wang, W., Bi, W., Lin, W., & Zhang, D. (2025). Deep learning model for real-time flood forecasting in fast-flowing watershed. *Journal of Flood Risk Management*, *18*(1), e70036. https://doi.org/10.1111/jfr3.70036

[9] Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., . . . , & Lawson, K. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. *Nature Reviews Earth & Environment*, *4*(8), 552–567. https://doi.org/10.1038/s43017-023-00450-9

[10] Todini, E. (2004). Role and treatment of uncertainty in real-time flood forecasting. *Hydrological Processes*, *18*(14), 2743–2746. https://doi.org/10.1002/hyp.5687

[11] Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., . . . , & Matias, Y. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature*, *627*(8004), 559–563. https://doi.org/10.1038/s41586-024-07145-1

[12] Tabasi, N., Fereshtehpour, M., & Roghani, B. (2025). A review of flood risk assessment frameworks and the development of hierarchical structures for risk components. *Discover Water*, *5*(1), 10. https://doi.org/10.1007/s43832-025-00193-2

[13] Rentschler, J., Salhab, M., & Jafino, B. A. (2022). Flood exposure and poverty in 188 countries. *Nature Communications*, *13*(1), 3527. https://doi.org/10.1038/s41467-022-30727-4

[14] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[15] Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., . . . , & Matias, Y. (2022). Flood forecasting with machine learning models in an operational framework. *Hydrology and Earth System Sciences*, *26*(15), 4013–4032. https://doi.org/10.5194/hess-26-4013-2022

[16] Gao, T., Fisch, A., & Chen, D. (2021). Making pre-trained language models better few-shot learners. In *Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 3816–3830. https://doi.org/10.18653/v1/2021.acl-long.295

[17] Awah, L. S., Belle, J. A., Nyam, Y. S., & Orimoloye, I. R. (2024). A systematic analysis of systems approach and flood risk management research: Trends, gaps, and opportunities. *International Journal of Disaster Risk Science*, *15*(1), 45–57. https://doi.org/10.1007/s13753-024-00544-y

[18] Gordon, C. A., Foulon, E., & Rousseau, A. N. (2023). Deriving synthetic rating curves from a digital elevation model to delineate the inundated areas of small watersheds. *Journal of Hydrology: Regional Studies*, *50*, 101580. https://doi.org/10.1016/j.ejrh.2023.101580

[19] Hamouda, M. A., Awadallah, A. G., & Abdel-Maguid, R. H. (2024). Extension of the Geomorphic Flood Index classifier to predict flood inundation maps for uncalibrated rainfall depths in arid regions. *Nature Hazards*, *120*(5), 4633–4655. https://doi.org/10.1007/s11069-023-06393-0

[20] Lawless, C., Schoeffer, J., Le, L., Rowan, K., Sen, S., St. Hill, C., . . . , & Sarrafzadeh, B. (2024). "I want it that way": Enabling interactive decision support using large language models and constraint programming. *ACM Transactions on Interactive Intelligent Systems*, *14*(3), 22. https://doi.org/10.1145/3685053

[21] Zevenbergen, C., Gersonius, B., & Radhakrishan, M. (2020). Flood resilience. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *378*(2168), 20190212. https://doi.org/10.1098/rsta.2019.0212

[22] European Environment Agency. (2019). CORINE land cover 2018 (vector). *Europe*. https://doi.org/10.2909/71C95A07-E296-44FC-B22B-415F42ACFDF0 2020 6-yearly—Version 2020_20u1 [Data set]

[23] Rangari, V. A., Umamahesh, N. V., & Bhatt, C. M. (2019). Assessment of inundation risk in urban floods using HEC RAS 2D. *Modeling Earth Systems and Environment*, *5*(4), 1839–1851. https://doi.org/10.1007/s40808-019-00641-8

[24] Fraehr, N., Wang, Q. J., Wu, W., & Nathan, R. (2022). Upskilling low-fidelity hydrodynamic models of flood inundation through Spatial analysis and Gaussian Process learning. *Water Resources Research*, *58*(8), e2022WR032248. https://doi.org/10.1029/2022WR032248

[25] Fraehr, N., Wang, Q. J., Wu, W., & Nathan, R. (2023). Supercharging hydrodynamic inundation models for instant flood insight. *Nature Water*, *1*(10), 835–843. https://doi.org/10.1038/s44221-023-00132-2

[26] Motta, M., de Castro Neto, M., & Sarmento, P. (2021). A mixed approach for urban flood prediction using Machine Learning and GIS. *International Journal of Disaster Risk Reduction*, *56*, 102154. https://doi.org/10.1016/j.ijdrr.2021.102154

[27] Dhote, P. R., Joshi, Y., Rajib, A., Thakur, P. K., Nikam, B. R., & Aggarwal, S. P. (2023). Evaluating topography-based approaches for fast floodplain mapping in data-scarce complex-terrain regions: Findings from a Himalayan basin.

Journal of Hydrology, *620*, 129309. https://doi.org/10.1016/j.jhydrol.2023.129309

[28] Bates, P. D., Horritt, M. S., & Fewtrell, T. J. (2010). A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *Journal of Hydrology*, *387*(1-2), 33–45. https://doi.org/10.1016/j.jhydrol.2010.03.027

[29] Wang, K., Bertoli, G., Cheng, S., Schröter, K., Caporali, E., Piggott, M. D., . . . , & Arcucci, R. (2025). Ai-empowered latent four-dimensional variational data assimilation for river discharge forecasting. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *18*, 24676–24689. https://doi.org/10.1109/JSTARS.2025.3611136

[30] Murphy, S., Wang, M., Cheng, C.-S., Passalacqua, P., & Leite, F. (2025). Land-use analysis using infrastructure representations and high-resolution flood inundation mapping techniques. *International Journal of Disaster Risk Reduction*, *123*, 105518. https://doi.org/10.1016/j.ijdrr.2025.105518

[31] Shi, P., Lyu, K., Li, Z., Yang, T., Xu, C. -Y., Hao, X., & Xiao, J. (2025). A novel topography-based approach for real-time flood inundation mapping. *Water Resources Research*, *61*(2), e2024WR037851. https://doi.org/10.1029/2024WR037851

How to Cite: Karimanzira, D. (2026). Toward Resilient Communities: Integrating Predictive Flood Models with Natural Language Processing for Actionable Insights. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62026053