The Robustness Gap: Real-World Challenges in AI-Based Flood Severity Assessment

Received dd.mm.yyyy, Revised dd.mm.yyyy, Accepted dd.mm.yyyy

ABSTRACT: Flood events are becoming more frequent and severe with rapid urbanization and climate change. With rising risks, accurate real-time assessments are critical for disaster management and operational decision-making, such as emergency response and resource allocation. Implementing artificial intelligence (AI), particularly deep learning (DL), has advanced the automation of estimating the damage from floods using satellite and sensor data. However, despite strong results in controlled environments, these models often suffer significant performance degradation under real-world conditions. This article assesses the state of the art in using AI to measure the severity of floods, paying specific attention to the robustness of the models and their general applicability to various environmental conditions. More than fifty primary studies published from 2019 through 2024 were analysed within the systematic literature review framework. The review proposes a diagnostic framework centered on three interconnected domains of difficulty: the quality and diversity of data, the model structure and interpretability, and the authenticity of the environment, which includes turbidity, meteorological interference, and intricate topographical challenges. Results show that predictive reliability decreases significantly under sensor noise and unseen environmental conditions, although benchmark performance is usually high. There is a lack of analytic tools to measure the degree of these shortcomings. This review focuses on the growing "robustness gap" in artificial intelligence literature on flooding. It presents a vision for developing precise, adaptable, and functionally resilient AI capable of reliable deployment in operational flood disaster management.

1. INTRODUCTION

The global hydrological cycle is changing and becoming less predictable, more natural and similar to its original chaotic state. Catastrophic flooding is a global concern due to the meteorological phenomena of climate change, including the increase in extreme rainfall, coupled with rapid, often poorly planned urban development [1] [2] [3]. Floods now make up nearly 40% of all weather-related disasters, and the number of major flood events has more than doubled since 2000 compared to the previous 20 years [4]. Scientists also warn that climate change is causing heavier rainfall and stronger storms, which means floods are becoming more likely and damaging worldwide. Therefore, finding innovative methods to monitor and control floods in a changing climate is crucial because the number of vulnerable areas, including several river basins, has grown significantly [5]. Consequently, assessing the magnitude of the floods includes what areas of land the water pours onto, at what rate, the depth of the flooding, and the potential impact of its destruction, which becomes invaluable as the bedrock of capable disaster management [6]. This information is critical for the life and wealth of many, and even the management of billions in relief funds rests on its responsiveness and correctness.

Historical dependence on physics-driven hydrodynamic models was justified owing to their reliability as accurate representations of the physical world [7]. The failure of traditional hydrological models to keep pace with the rapidity and precision required for flood impact assessments has cast AI as a potential game-changer. In addition, the conventional approach to remote sensing, including satellite data analysis and sparse ground interpolation, has always been described as inefficient and discontinuous. Flood measuring practices have recently been transformed with the help of AI.Deep learning (DL) models can quickly process huge amounts of satellite images or river sensors data. For instance, CNNs can effectively locate relevant features in the images, thereby facilitating the production of accurate flood maps [8]. Using LSTM models enables one to understand the temporal variations of floods [9]. Vision Transformers are being adopted to fuse complex, multi-modal datasets like optical and Synthetic Aperture Radar (SAR) readings [10] [11] [12] [13]. The management of floods in the real world has already benefited from Copernicus Emergency Management Service (CEMS) tools and NASA's FloodSENS, which integrates AI with intensive satellite and sensor data analysis. Producing flood inundation and flood risk maps faster than existing techniques, this technology greatly aids authorities in formulating evacuation and response strategies [14] [15].

However, the ground realities are more complex than that. Although these AI models can perform well during tests, they can weaken very quickly when they have to handle situations from the real world. Frequently, they overlook significant details or wrongly handle complex situations.

1.1. The Robustness Gap: Ideal Performance vs. Real-World Failure

The robustness gap exists between the performance of the AI under optimal conditions, which in this case would be the use of clean and well-curated datasets, and the performance of the AI in real-world applications. The ideal scenario would be one in which a strong AI flood model could produce cross-domain flood severity estimates regardless of whether the input data originates from radar data or optical satellite images, irrespective of whether the target domain is a city in the Ganges Delta or Europe. The reality is that such models cannot handle the nuances and unpredictability of real floods.

There are multiple tangible factors that may affect the practical applications of the models under consideration during real-life catastrophes. For instance, optical sensors are much less effective at locating the boundary of a flood when turbidity is high due to suspended sediments. Furthermore, heavy rain accompanied by strong winds disturbs the satellite radar, resulting in data loss and leading to missed flood zones or mistakenly alarming a flood zone. Cities introduce their own problems, as the roads, buildings, and various surface types confuse models that have predominantly been trained on rural environments. More recent research indicates that problems such as diminished visibility during storms, rain on a camera lens, and other issues significantly impact the accuracy of AI-based flood mapping.

1.2. Purpose and Objectives

The present study is not merely a summarization of AI methodologies used to measure flood severity. In fact, it attempts to analyze the techniques' shortcomings concerning the real-world variability of the environment. The survey covers more than 50 studies published between 2019 and 2024, examining the AI models' behavior changes with varying data and adverse environmental conditions. The review is constructed around a conceptual model that organizes the pitfalls into three interrelated aspects: data and model quality, model diversity, and environmental realism. This approach makes it possible to evaluate not just the precision of the models, but also resilience—how they respond to variability and noise in the real world. This review aims to analyze the current techniques applied in measuring flood severity using AI, focusing on the critical gaps of versatility and strength about real world data.

The contribution of this review is threefold: (i) to quantify the performance drop of CNN, LSTM, and transformer architectures under the raw environmental stresses of high-turbidity water, persistent meteorological noise (such as cloud/rain cover), and complex terrain features; (ii) to design a scalable framework that helps researchers and practitioners pinpoint and attribute the root causes of robustness failures in flood severity models; and (iii) to design and recommend practical addition and modification approaches (e.g., domain adaptation, physics-informed training) that quantifyably enhance model resilience and generalizability across varying data ecosystems. This Fig. 1 signifies the major key words in this paper.

This paper has a well-structured approach to assist in improving the development of AI in flood analysis. In Section 2 of the document, the fundamentals are clarified, flood severity is calculated, and the primary concepts of deep learning are addressed. In section 3, the author analyzes various research papers and aligns several artificial intelligence techniques to the flood problem they are designed to address. Section 4 utilizes a diagnostic framework to assess the relative effectiveness of the techniques and outlines where AI continues to have challenges in performing the relevant real-world tasks. In the last section, the author presents the conclusion and outlines the additional work needed to provide practical, reliable, and fully automated artificial intelligence for flood severity analysis.

Figure 1. Visualizing the lexicon of AI-driven flood severity measurement

2. BACKGROUND THEORY

2.1. Hydrological Basis

Flood severity is considered as a complex concept comprising several physical parameters that collectively demonstrate the magnitude and potential impact of the flood event. The leading indicators cover water depth, area (inundation extent), discharge volume, and event duration, reflecting different aspects of flood occurrence and hazard.

- Water depth or stage is the most visible indicator directly impacting the buildings' structural integrity and the people's safety. It is measured as the vertical distance from a fixed reference point to the water's surface. Stream gauges are installed on the banks of rivers and streams to measure and monitor the water level stage. These devices transmit data in real-time, which is a valuable tool for emergency services and the rest of the community to monitor floods and send a flood warning when this stage goes beyond certain limits.

- Inundation extent is typically derived from satellite images captured by Synthetic Aperture Radar (SAR) or optical sensors and forms the basis for estimating the area covered by floodwaters. Perfect estimation of such an area determines the population and infrastructure at risk, and thus influences the speed and precision of disaster relief efforts. Presently, spatial analysis uses pre- and post-event images with techniques such as thresholding, water detection algorithms (NDWI and MNDWI), ML, etc., to improve flood delineation in various land cover types and circumstances. The integration of optical and SAR data has improved inundation mapping in cases of cloud cover or complex terrain where a single sensor fails. These maps serve as crucial inputs for the loss assessment and emergency response, and help the decision makers by providing timely situational awareness.

- Peak flow is measured at a stream gauge as discharge volume (in m^3/s) and shows how intense the flood is at a specific spot in the river. However, the overall flood volume by summing all the water passing the gauge during the entire event gives a fuller picture of the water that needs to be transported, absorbed, or stored by the river and floodplain. Duration tells us how long the floodwaters remain above a critical level, which is crucial for estimating crop damage and the disruption and permanent harm to businesses, soil, and infrastructure. The flood hydrograph is constructed from the relationship between peak flow and duration. For example, a flood with a sharp peak and short duration can cause rapid erosion, while a flood with a lower peak but longer duration can cause soil deep wetness and slope failures. To estimate flood intensity and determine the level of risk to the populace and property, hydrologists gauge the duration of a hydrograph above established damage thresholds.

The relationship among key flood severity parameters can be expressed as Eqn. 1:

$$V_f \approx \sum (A_i \times D_i) \tag{1}$$

where V_f is the flood volume, A_i the inundation area, and D_i the average water depth of segment i. This formulation underscores that overall flood impact is governed by the interaction between spatial extent and hydraulic magnitude, where the inundation area defines the geographical spread and average depth quantifies the intensity of submergence.

Differentiation between damage assessment and flood severity measurement from flood forecasting is significant because they are different stages in the disaster management process. Measurement is primarily about recognizing and quantifying an event that is either ongoing or has happened (a diagnostic process). On the other hand, forecasting is a predictive exercise relying on numerical weather prediction and hydrodynamic modeling [16] [17] and damage assessment, which translates the measured hydrological metrics into socio-economic terms like estimated losses [18]. Therefore, measurement is the vital empirical link connecting the flood event to subsequent predictive modeling and post-event assessment.

Determining the severity of a flood from the hydrological perspective involves combining water-related parameters with the context of earth sciences [19]. Factors like soil saturation, catchment or watershed topography (from high-resolution DEMs), and previous moisture strongly influence flood propagation. The spatial and temporal limitations of traditional data sources (e.g., river gauges) are often the main reasons for the lack of precision [20]. Thus, AI-based surrogate models are necessary. Realistic measurement of these models requires AI fusion with a practical hydrological approach [21], where the models obey the basic hydrological constraints of the outlined system, understanding that flood processes are dynamic physical processes, not pixel-correlated behaviors.

2.2. AI and Remote Sensing Fundamentals

The effectiveness of any flood severity measuring AI model highly depends on the remote sensing data that it uses. Optical imagery and SAR are currently dominant but complementary forms of satellite data. Each is defined by distinct physical principles that govern its relative strengths and weaknesses in the context of a disaster.

2.2.1. Optical Imagery: A multispectral view, obscured by clouds

The European Space Agency's Sentinel 2 and NASA USGS's Landsat 8/9 operate within the optical spectrum as they analyze reflected sunlight in visible, near-infrared (NIR), and shortwave infrared (SWIR) segments. The spectral resolution is the main advantage, facilitating the computation of spectral indices, such as the Normalized Difference Water Index (NDWI). This index captures the water and dry land interface by analyzing the light emitted and absorbed in specific spectral bands. Multispectral satellite data is also valuable in assessing vegetation and soil moisture changes after floods. The problem regarding cloud cover during floods becomes a serious issue because optical sensors cannot penetrate clouds. In these situations, satellites may not be able to provide useful images for days at a time. This limitation has created far-reaching implications for any attempts to maintain reasonable emergency protocols during a flood and has been a significant bottleneck for flood monitoring [22] [23].

2.2.2. SAR: The all-weather Sentinel

The European Space Agency's Sentinel 1 is unique in flood monitoring because its active radar captures images by sending out radar microwaves and measuring their return from the Earth's surface. This means that images can be captured 24/7, under any weather conditions. This capability is particularly advantageous during storms when failure in optical satellites is likely. In these images, flood zones with smooth water surfaces that drastically attenuate most of the radar energy incident upon the sensor appear black because they are easily distinguishable from land. Like any other satellite sensor, the radar sensor also has its challenges. The most concerning challenge is speckle, a grainy noise that conceals other details in an image and increases the difficulty of automatic processes. The radar return may be signal-cluttered in urban areas due to the radar reflections from the structures, the water below, and the land in the foreground. In addition, the reflection from wind-agitated water surfaces can be mistaken for land, which results in mapping errors. For these reasons, converting raw SAR images to flood extent maps is a complex problem requiring sophisticated AI models trained to recognize various backscatter patterns to image flood instead of signal noise artifacts [24] [25].

2.2.3. AI processing and tasks

AI models are designed to turn the pixel values into measurable, actionable severity metrics and estimates as soon as data is obtained through the experimental configurations. Such tasks involve multiple levels of complexity.

Preprocessing: Once a spectral index has been determined (e.g. NDWI and MNDWI), optical images undergo atmospheric correction, and additional steps of cloud and shadow masking and other optical techniques are used to enhance the signals of water. The processing of the SAR of water bodies includes radiometric calibration, geometric correction of the terrain, multi-look speckle filtering, and, in some instances, polarimetric decomposition. Wind, vegetation, and water turbidity affect SAR backscatter. Therefore, careful calibration and context-sensitive filtering are required [26]. High-quality auxiliary data (e.g. Digital Elevation Models (DEMs), landcover maps, and in-situ gauges) are frequently combined with remote sensing data to enhance the spatial and physics-based coherence of the AI outputs [27].

Tasks: In the domain of flood remote sensing, there are three predominant tasks for the application of AI technology:

- **Semantic segmentation:** This step is to classify an image at a pixel level into "water" and "non-water" categories. U-Net architectures have become the state-of-the-art, learning to identify water's spectral (optical) or backscatter (SAR) patterns despite the noise and other contextual variations [28]. The result generated is a binary inundation map.

- **Regression for Water Depth Estimation:** The continuous approximation of flood depth and other severity measures is a problem of regression, not classification. This is usually accomplished by fusing the AI-generated inundation map and a pre-flood DEM. The model conducts a gross geometric operation by estimating the altitude of the water and the bottom terrain. However, this can be improved with hydraulic models or trained end-to-end with CNNs [29]. Models are adjusted to predict per-pixel or per-patch water depth/stage (deep learning patch models) by adding segmentation regression heads instead of segmentation heads and applying MAE or RMSE-based loss functions. Moreover, the model regression framework is more complex than binary mapping because the model has to predict physical magnitudes from spectral or backscatter proxies.

- **Spatio-temporal modeling:** Once the waters have settled, the focus shifts to impact. Concrete structure identification and classification from high-resolution imagery is carried out via object detection models (Faster R-CNN and YOLO). Moreover, change detection algorithms can analyze the differences between the pre- and post-event images and recognize the most important changes autonomously, distinguishing between collapsed, submerged, or intact buildings. This reflects the social and economic effects necessary for recovery planning [30].

2.3. Deep learning architectures in flood mapping

Advanced AI algorithms, such as artificial neural networks, enhance their potential to analyse satellite images for flood mapping. CNNs deploy multiple filters to the image to extract features and detect the orientation of edges and textures, and then shift to complex shapes and patterns on deeper levels. This cascading style helps CNNs recognise patterns in the different landforms under which water may exist and then capture their ionic or radar signatures, thus enabling water detection and flood mapping [31].

CNNs are used for performing semantic segmentation and classifying each pixel of an image. The widely used U-Net architecture for flood mapping consists of an encoder, a decoder and a skip connection. An encoder that compresses context and a decoder that reconstructs the context, separated by skip connections that capture and retain important features within the image. This architecture is compelling for delineating complex flood margins, including river channels. DeepLabV3+ facilitates segmentation by capturing context at multiple scales, while HRNet achieves impressive results with high-resolution images by maintaining the data through all layers [32].

Supervised learning is used with labelled satellite image datasets to train the models. The training process works with different prediction outputs based on the loss function. Although Cross-Entropy loss is standard, Dice Loss is used for flood mapping as it mitigates the disparity between the limited water pixels (foreground) and the much larger non-water (background) pixels [33]. Typical evaluation includes Intersection over Union (IoU), which measures the predicted and actual water pixels, and the F1-Score, which measures the overall performance based on precision and recall. In flood mapping, these approaches to evaluation ensure complete rigour and reliability for model performance to be compared.

The self-attention based Vision Transformers (ViTs) advances flood mapping, particularly concerning global context image analysis when compared to classical methods like CNNs [10] [34]. VLMs take this further by associating satellite data with supplementary text, like weather observations or topographical descriptions. This multi-model integration lets these models respond to text commands, like 'locate urban areas with flooding', bringing a new level of interactivity and context to flood assessments. Such models integrate

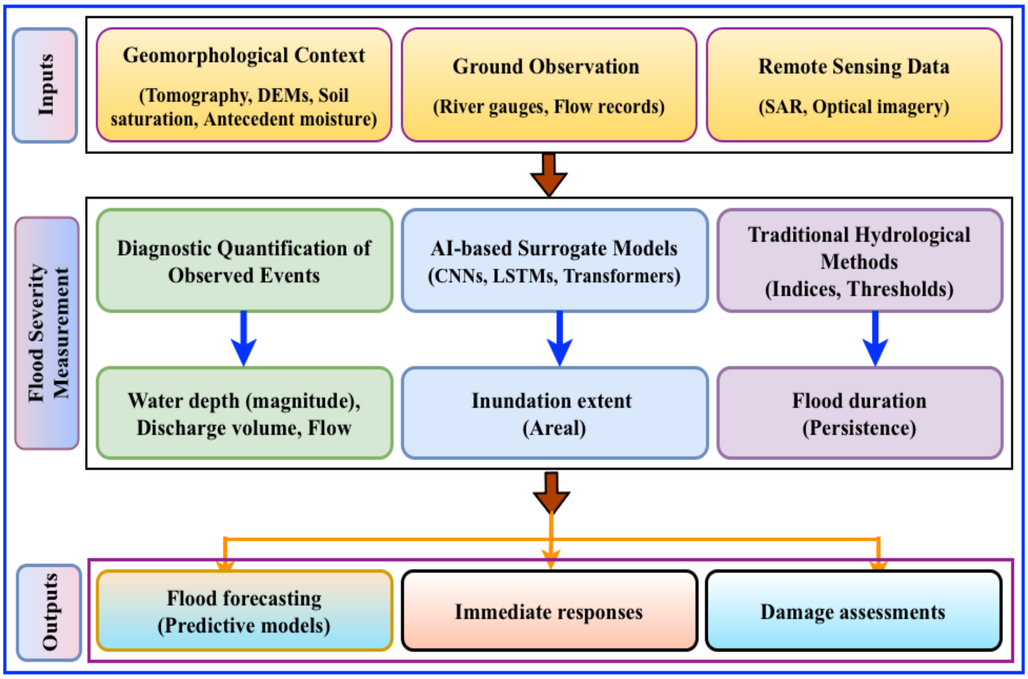

Figure 2. Integrated AI–hydrology framework for flood severity estimation

vision and language in flood monitoring, pioneering new methods to target focused, hyper-resolution analysis, while including visual and semantic context for more flexible interpretation [35].

Fig. 2 illustrates this integrated AI–hydrology framework for flood severity estimation, linking physical observations with data-driven modeling to support forecasting, response, and damage assessment.

3. LITERATURE REVIEW

This literature review reframes progress in the AI domain for flooding applications in inundation mapping, water surface estimation, and damage assessment within the context of the observed robustness gap and its reasons. These works collectively highlight AI's strengths and weaknesses in providing dependable, scalable, and physically consistent flood intelligence systems.

3.1. AI for Inundation Mapping (Binary segmentation)

Binary segmentation of flood areas in satellite, aerial, and ground images using CNN architectures, such as U-Net, DeepLab, SegNet, can achieve very high accuracy levels (most of the time above 85–95%). This makes them superior to the conventional thresholding and rule-based methods.

AI-assisted flood monitoring utilizes IoT and CNNs to detect and segment floods in real-time satellite images. The DeepLabv3 model [36] can provide an excellent level of segmentation (up to 87%) by determining the class of each pixel in order to create flood maps. A significant problem with such methods is the reliance on satellite images and the requirement for precise labeling. Additionally, models struggle with varying geographical areas and weather conditions. The next step is to employ more adaptable methods to address these issues by integrating satellite images with real-time data from IoT sensors and weather forecasts, and by utilizing transfer learning to enhance the model's ability to generalize to new or evolving flood situations.

CNN-based attention models and U-nets with MobileNet backbones have achieved high scores in both accuracy and segmentation metrics, including Mean IoU [37]. The major drawback is that the training heavily relies on augmented non-flood images, which may not necessarily represent the complex visual features of a real flood situation. Incorporating diverse, real-world flood datasets sourced from multiple environments and conditions remains a significant challenge that models must address to remain generalizable and robust. Exploring further advanced architectures, such as transformer-based models or multimodal data fusion, to achieve higher segmentation accuracy in challenging contexts is also worth considering.

In [38], the power of a U-Net CNN combined with multi-source geospatial data (Sentinel-1 SAR, DEM, and rainfall) to predict urban floods in Near Real-Time (NRT) was demonstrated, achieving an impressive testing accuracy of 93.73%. The temporal gaps due to the 6-day revisit interval of Sentinel-1 and the simplification of flood causes by using daily rainfall data only are the primary limitations. To address the limitations, NRT flood studies should focus on data fusion that involves multiple satellites and ground-based IoT sensors, as well as the use of hydrodynamic simulation or Physics-Informed Neural Networks (PINNs) with high-resolution temporal rainfall data to close the observation frequency gap and accurately capture the watershed's dynamic response.

This paper [39] presents an alternative PINN model to replace the much slower hydrodynamic simulators, such as LISFLOOD-FP and Delft3D, used in flood inundation and coastal modelling. The authors address one of the primary limitations of data-driven surrogates in terms of efficiency and physical realism. The new model utilizes a custom loss function and incorporates mass conservation into the neural network, eliminating the need for automatic differentiation. This methodology eliminates the need to compute continuous relational derivatives, thereby reducing the overall cost, enhancing generalization, and achieving high computational efficiency. In doing so, the method improves the accuracy of the standard CNN surrogates by over 25%. However, it has drawbacks, including sensitivity to the initial weight configuration, increased complexity of the model under development, and the absence of inherent uncertainty quantification. Future research could investigate the integration of Bayesian methodologies with PINNs to provide robust uncertainty estimates and create more streamlined routes, thereby reducing the need for specialization in physics-informed regularisation and ultimately decreasing the complexity of model development.

ML has been used in recent studies as standalone data-driven models or as surrogates for physics-based hydrodynamic models to improve computational efficiency and prediction accuracy in urban flood prediction [40]. Nevertheless, the current methods have drawbacks; for instance, they rely on single-event training, which makes them less generalizable, do not consider physical processes such as rain-on-grid and green-grey infrastructure, utilize limited hydrodynamic features, and do not adequately quantify predictive uncertainty. Developing physics-informed ML surrogate models trained on various storm events, utilizing high-resolution geospatial and infrastructural data, and implementing ensemble ML methods along with uncertainty quantification techniques, such as quantile regression forests, are some suggested ways to overcome these limitations. These upgrades can make the models more transferable, robust, and capable of being deployed in real-time for street-scale flood forecasting.

The research in [32] addressed the limitations of feature localization in typical CNNs and the high computational cost of Vision Transformers (ViTs) by developing the Residual Wave Vision U-Net (WVResU-Net) for precise flood mapping using dual-polarization Sentinel-1 SAR data. This model combines sophisticated Vision MLPs with a ResU-Net backbone, treating image patches as wave operations to capture complex features in a highly effective manner, resulting in a significant performance improvement over existing state-of-the-art architectures. This model's main limitation is the excessive computation and training time associated with the Vision MLPs in the decoder. Future advancements should focus on developing techniques to significantly reduce the computational expense of the WVResU-Net architecture while maintaining its high segmentation accuracy.

DeepFlood is a multi-modal dataset containing aerial and SAR images with high detail and involved obstruction annotations, such as flooded vegetation [41]. While the dataset enables the training of robust models and facilitates model construction for multi-modal data fusion, a significant drawback remains the substantial difference in resolution between high-resolution optical imagery (1.5cm) and Sentinel-1 SAR (10m) data.Development of innovative multi-resolution data fusion strategies and deep learning architectures that focus on the effective utilization of disparate datasets to solve real-world challenges associated with advanced flood mapping.

A UNet with an EfficientNet-B7 backbone showed its adaptability across different areas using a NASA benchmark dataset [42] . The major limitation remains the models' varying performance across different polarisation band combinations and geographical settings. Currently, no single architecture outperforms the rest. Future work can combine multi-source data, such as topographical and land cover information, as well as broaden training data to challenging environments (i.e., urban and mountainous) to help improve model robustness and generalisation.

Flood mapping with Sentinel-1 SAR data is accomplished through the use of probabilistic deep learning models such as Bayesian Convolutional Neural Networks (BCNNs) and Monte Carlo Dropout Networks

(MCDNs) [43]. These models not only generate flood maps, but also provide estimates of the model's confidence in its predictions. Between BCNN and MCDN, BCNN offers better segmentation and more accurate quantification of uncertainty. While MCDN tends to be overconfident, misclassifying and overestimating erroneous doubtful areas up to 253 sq. km, BCNN ensures that emergency response plans remain justifiable while concurrently avoiding and minimising overconfidence errors. The models, as well as the fulfilment of segmentation precision and uncertainty estimation confidence, also present challenges related to prior definitions and the issue of imbalance.

3.2. Data integration and model optimization

DL has been focusing on the problem of distinguishing floodwaters from permanent water bodies using post-flood imagery. Recent advances in DL take advantage of multi-source data fusion approaches utilizing the large Sen1Flood11 dataset, including Sentinel-1 SAR and Sentinel-2 optical imagery. A single CNN model (BASNet [44]), enhanced by focal loss to address class imbalance issues in the data, achieved an impressive overall accuracy of 92.8% and demonstrated robustness under varied conditions. Future studies will focus on utilizing high spatial resolution images and designing more advanced network modules that can better balance the information efficiency bottleneck of transfer learning from pre-trained RGB networks, thereby minimizing the mIoU drop during layer freezing.

The OmbriaNet model [45] demonstrates the effectiveness of utilising bitemporal and multisensor satellite images (Sentinel-1 SAR and Sentinel-2 optical) within a supervised deep learning framework. As a result, the model achieves state-of-the-art performance in distinguishing between flooded and permanent water bodies. However, the model's generalisation remains limited, as it was trained on a dataset with a narrow spatial and temporal focus, covering only 23 flood events worldwide. Future research can address the limitations of the dataset by incorporating more geographically and topologically diverse flood events, augmenting feature representation with additional Sentinel spectral bands, and exploring transfer learning from very high-resolution UAV images to improve local-scale mapping, thereby enhancing robustness and accuracy.

A trained model segmenting very high-resolution optical images of water bodies using U-Net and DeepLabV3+ has been reported in [46]. It showcases a U-Net with a MobileNet-V3 backbone, providing the optimal tradeoff between speed and accuracy for flooded area mapping. Deep learning models tend to underperform beyond the domain of the training dataset. The models lack generalizability due to cross-sensor environmental differences, such as in floodwaters with high sediment loads or complicated urban areas, as well as their reliance on regionally confined datasets. To improve the robustness and transferability of the models, future research should focus on curating multi-sensor and geographically diverse training datasets, incorporating domain adaptation, and utilizing additional context data, such as digital elevation models, near-infrared bands for segmentation, and addressing complex situations.

In flood mapping using SAR imagery, Modified DeepLabv3+ with a MobileNetv2 encoder have demonstrated high accuracy and efficiency in segmenting water bodies under all-weather conditions [47]. However, a significant limitation of these supervised approaches is their reliance on extensive, manually curated datasets, which are expensive and potentially lack variety in ground features, thereby limiting model generalisation. Future work should focus on automating or semi-automating these data labelling processes to expand the breadth of training datasets and incorporate other data sources, such as digital elevation models, into the model to improve segmentation under challenging terrains. The additional use of new, more affordable satellite SAR systems scheduled for deployment in the near future will also enhance the temporal coverage of the model, which is crucial for prompt and effective responses in high-need emergencies.

In SAR-based flood mapping, CNNs(AlbuNet-34, U-Net, and DeepLabV3+) classify water bodies with greater accuracy than traditional rule-based techniques, especially when trained on dual-polarized Sentinel-1 data and optimized using a blending of cross-entropy and Lovász loss functions [48]. These models still struggle to accurately classify water bodies in complex geographies, such as dry and mountainous regions. Furthermore, the models' accuracy is sensitive to the data augmentation techniques used. Subsequent research should consider incorporating land cover and slope data to enhance model generalization, as well as increasing the variability and complexity of the geographical training data to improve model robustness.

One example of progress in flood mapping is the development of the STURM-Flood dataset [49], which combines Sentinel-1 and Sentinel-2 imagery with ground truth data and specialist annotations. This has resulted in the training of U-Net CNN that achieve superb segmentation accuracy. Models using Sentinel-2 data outperform those using Sentinel-1 data (over 92% accuracy). The ability of these models, however, is limited due to their overly simplistic network architecture, the suboptimal fixed classification thresholds deemed suboptimal by these models, and the persistent class imbalance in the dataset, which is caused by the low proportion of water pixels. Progress in these areas needs to be complemented with dataset expansions that include more flood events and sensor types. Model generalization needs to be improved through the implementation of multi-class segmentation, transfer learning, and model ensembling to achieve generalizability across varying geographical and climatic conditions.

3.3. Post-flood damage assessment

The research [50] demonstrates the effectiveness of DCNN models, such as Deeplabv3+ and LinkNet, in segmenting urban flood extent from surveillance camera images. Consequently, these models have achieved high validation scores. Nevertheless, the research has been cautious about these limitations in applying the models to real-life situations, including the loss of performance due to specular reflections on still water or wet roads, low-light conditions, increased obscuration by storms, rain on camera lenses, and other environmental factors. To address such issues and enhance the model's generalization, the authors recommend camera-specific model training, which involves supplementing a base training set with a small number of localized flood images collected from the targeted camera. Likewise, the installation of smart rain wipers with HD cameras is a feasible way to remove the obstruction to visibility and prevent damage to the lens.

The integration of LSTM-SegNet and multi-head self-attention with EfficientNet-based segmentation models [51] demonstrates potential in augmenting flood dynamics prediction and land-use economic loss estimation. However, these methods are constrained by their dependence on the quality and regional specificity of the input data, particularly in the case of manually labeled land use classifications and the transferability of depth-damage curves adapted from other regions. The generalizability and accuracy of these models can be enhanced by developing region-specific loss models for socioeconomically diverse datasets and incrementally learning structured, dynamic, and enhanced spatial and temporal risk assessment data.

YOLOv3 has been used successfully for rapid post-flood evaluations by detecting flooded road sections in aerial photographs [52]. These models perform exceptionally well in combination with elevation data, particularly in high-risk areas. However, their use in practice may be hampered by problems such as cloud cover and the variability of an image's brightness. In addition, static elevation benchmarks are not the most suitable for flat terrains, nor for other disasters such as earthquakes. To enhance the robustness of these models, adequate image preprocessing should be performed to minimise the effects of atmospheric conditions. Adaptive elevation thresholds should be applied for use in specific regions. The scope of training could be broadened to encompass other water hazard disasters, such as tsunamis, to expand the utility of the models.

The ”FloodImageClassifier” is a CNN architecture that incorporates YOLOv3, Mask R-CNN, and EfficientDet for automated flood image analysis [53]. The model was trained over 9,000 annotated images collected from social media, transportation cameras, and USGS river cameras to detect floods and estimate water levels through object detection and segmentation. With the help of Canny Edge Detection and aspect ratio analysis, ”FloodImageClassifier” can calculate inundation areas and assess flood depth and severity. The tool enables real-time monitoring of river and road flooding conditions through integration with existing camera networks, providing critical early intelligence to emergency responders.

Using YOLO-based models to detect human body parts (chest, thigh, shoulder, head) in social media images for urban flood depth estimation, resulting in high detection accuracy (mAP: 0.967) and quite good depth estimation (MAE: 10.22 cm) [54]. Limitations of the model include the use of average body dimensions without gender differentiation, detection confidence from occluded body parts, and image quality issues. For gender consideration, integrating YOLOv6/7 and refining image selection criteria for detection will enhance facial recognition biometry. Among the proposed solutions are using facial recognition to determine gender for more accurate scaling, switching to a more advanced detection model (YOLOv6/7), and applying stricter criteria for selecting images. This method serves as a proof-of-concept for using social media images as an alternative data source for emergency response; however, it still requires some refinements to improve measurement accuracy and operational stability.

After the devastating flood in Kashmir in 2014, authors [55] developed a post-disaster assessment structural damage index (DI) for residential constructions, which emerged as a possible solution for quickly and rationally assessing structural damage visually. The damage index (DI) for a structure is calculated using a systematic, multi-step approach that quantifies damage based on crack characteristics and their structural significance. The process begins with the evaluation of individual cracks using the formula (Eqn. 2):

$$DI_i = IF_{primary} \times IF_{secondary} \times IF_{tertiary} \times DF \quad (2)$$

Here, DI_i represents the damage contribution of a single crack, determined by multiplying three importance factors (IFs) with a damage factor (DF). The primary importance factor ($IF_{primary}$) accounts for the floor level where the crack is located (e.g., ground floor = 1.0, first floor = 0.8). The secondary importance factor ($IF_{secondary}$) relates to the structural element type (e.g., walls = 1.0, beams = 0.8, plinth = 0.5), and the tertiary importance factor ($IF_{tertiary}$) considers the crack's specific location on the element (e.g., wall corners = 1.8, lintel level = 1.5). The damage factor (DF) is assigned based on the measured crack width (e.g., 1 mm crack = DF 2, 2 mm crack = DF 4). The total damage index for the structure is then obtained by summing the DI_i values of all identified cracks, providing a comprehensive numerical assessment of the overall structural damage. The proposed DI faces several limitations. Its additive nature cannot distinguish between widespread and critically concentrated damage, leading to oversimplified analysis and anomalous trends in the data. Moreover, the methodology is highly subjective and relies on engineering judgment, is restricted in scope to a small area, and therefore lacks broad applicability.

Authors [56] proposed the Disaster Vegetation Damage Index (DVDI) as a remote sensing-based technique for quickly assessing flood-related crop damage. The method begins with the calculation of the Normalized Difference Vegetation Index (NDVI) using Eqn. 3.

$$NDVI = \frac{\lambda_{nir} - \lambda_{red}}{\lambda_{nir} + \lambda_{red}} \quad (3)$$

where λ_{nir} and λ_{red} represent the near-infrared and red reflectance values, respectively. This is followed by the computation of the median Vegetation Condition Index (mVCI), which normalizes the current NDVI against its historical median and maximum values for the same day of the year:

$$mVCI = \frac{NDVI_{present(x,y)} - NDVI_{median(x,y)}}{NDVI_{maximum(x,y)} - NDVI_{median(x,y)}} \quad (4)$$

The DVDI is then derived as the difference between the post-event mVCI (calculated over a 7-day window) and the pre-event mVCI (calculated over a 14-day window):

$$DVDI = mVCI_{after} - mVCI_{before} \quad (5)$$

Damage to crops is classified into six categories based on negative values of the DVDI, with the lowest category of "No Damage" assigned to $DVDI \geq 0$ and the highest category of "Very Severe Damage" assigned to $DVDI < -0.4$. The final damage assessment is obtained by overlaying DVDI values with flood inundation maps and crop-type layers to ensure crop-specific and flood-attributable loss estimates.

A separate study [57] attempted to address the need for post-disaster impact assessment on specific crops by developing the Crop Flood Damage Assessment Index (CFAI). To compute the CFAI, the Analytic Hierarchy Process (AHP) is employed to assign weights to each parameter, which include the NDVI change ($\Delta NDVI$), cumulative precipitation (P_{cum}), growth period (T_{grow}), relative elevation (E_{rel}), and slope gradient (S_{grad}). The CFAI was formulated using Eqn. 6.

$$CFAI = (\Delta NDVI) \times W_{NDVI} + P_{cum} \times W_P + T_{grow} \times W_T + E_{rel} \times W_E + S_{grad} \times W_S \quad (6)$$

While CFAI effectively classifies crop impact (e.g., sub-slight to severe), a primary limitation is the underestimation of damage compared to the DVDI. Further constraints include subjectivity in parameter grading and difficulty in optical data extraction due to cloud cover. To improve reliability, the methodology requires refinement to reduce subjectivity and adapt the parameter grading for diverse geographic settings and crop types.

A Google Earth Engine (GEE)-based framework was developed in [58] for rapid monetary loss estimation by integrating Synthetic Aperture Radar (SAR)-derived flood extent, FwDET-GEE flood depth, and a Random Forest-generated land cover map. Sector-specific depth–damage functions were employed for damage computation, with housing damage modeled using a sigmoid relationship as shown in Eqn. 7.

$$DR_{\text{housing}} = \frac{0.24}{1 + e^{-1.50(FD-1.45)}} - 0.024 \quad (7)$$

Crop damage for rice and corn were modeled using the following equations (Eqn. 8 and Eqn. 9):

$$DR_{\text{rice}} = \frac{0.505}{1 + e^{-7.513(FD-0.854)}} - 0.001 \quad (8)$$

$$DR_{\text{corn}} = \frac{1.44}{1 + e^{-2.11 \cdot FD}} - 0.720 \quad (9)$$

Public infrastructure losses were estimated using a linear regression model between damage cost and floodwater volume. However, the method faced limitations, including the lack of crop-stage-specific damage curves, coarse DEM resolution leading to flood depth inaccuracies, and temporal misalignment between satellite acquisitions and peak inundation. The authors recommend employing higher-resolution DEMs, refining depth–damage relationships using local crop calendars, and incorporating multi-temporal satellite observations to improve accuracy and support Build Back Better recovery planning.

A District Flood Severity Index (DFSI) was developed in [59] to address India's critical need for a standardized flood assessment tool at the administrative district level. The index synthesizes multiple impact dimensions through a multiplicative logarithmic formula (Eqn. 10):

$$DFSI = \log_{10}\left[(1 + N_{\text{events}}) \times (1 + \bar{D}) \times (1 + N_{\text{fatalities}})^2 \times (1 + N_{\text{injured}}) \times (1 + A_{\text{flooded}}) \times P_{\text{total}}\right] \quad (10)$$

where N_{events} denotes the number of flood events, $\bar{D}$ the mean flood duration, $N_{\text{fatalities}}$ the human fatalities (squared to emphasize loss of life), N_{injured} the number of injured persons, A_{flooded} the percentage of district area flooded, and P_{total} the total population. While this approach successfully identifies high-risk districts such as Patna and Murshidabad, the authors acknowledge significant limitations: the index relies on historical data that may underestimate short-duration floods, excludes critical economic damage metrics due to data unavailability, and uses static population and land use data that cannot capture dynamic vulnerability. To enhance the robustness of the DFSI, the authors recommend integrating economic loss data, establishing standardized data collection protocols across agencies, incorporating near-real-time remote sensing for dynamic flood mapping, and adopting adaptive management strategies such as the "Room for the River" concept to address future climate and land use changes.

Multi-modal middle fusion architecture that leverages self-attention and cross-modal mechanisms to enhance the detection of disaster events by fusing textual and visual content from social media [60] [61] [62]. The proposed method [62] achieves an accuracy of more than 91% in informativeness and disaster type classification. Nevertheless, the research encounters obstacles in the form of overfitting by transformer architectures, class imbalance in damage severity data, and the inability to resolve contradictory information across modalities. The limitations can be solved by utilising data augmentation methods, such as SMOTE and GANs, employing explainable AI for model interpretability, and investigating sophisticated fusion methods with hierarchical attention mechanisms. Moreover, incorporating temporal and spatial data, as well as semi-supervised learning, can help future research generalise better and utilise unlabeled social media content more effectively.

The research [63] contributed to the advancement of flood severity measurement from a purely technocratic approach to a hybrid socio-physical approach that captures the 'human element' often overlooked in AI models. By combining satellite imagery and a large-scale field survey of 2,382 households, the study demonstrates that the community's adaptive capacity, particularly its financial resilience, significantly moderates the physical exposure (e.g., topography and drainage) of the community. Through the use of PCA and Cluster Analysis, the study establishes that real flood severity is the combination of the magnitude of the flood hazard and the social vulnerability, offering a statistically validated "ground-truth" methodology necessary to close the robustness gap in automated disaster management systems.

The Flood Resilience Index (FRI) combines hydrodynamic modelling with socioeconomic factors to quantify the changes in urban flood resilience at the ward level over time [64]. The analysis revealed remarkable inequalities and identified areas near the Tondano River with the slowest recovery, which are the most alarming. The use of secondary demographic data primarily limits the study, due to the insufficient detail of the slum areas and the reliance on climate change projections. Collecting primary data through community surveys, incorporating additional features of slum areas (e.g., inadequate drainage, poor housing), and developing climate-resilient, nature-based solutions in conjunction with appropriately scaled governance for climate change will collectively become an effective solution.

A more comprehensive representation of flood severity is made possible through the introduction of dual flood severity trigger indices for economic loss aggregates and maximum submerged housing counts. The proposed methodology [65] enhances the realism and equity in flood bond pricing, facilitating improved flood risk assessment.

4. COMPARISON AND DISCUSSION

The literature demonstrates the advanced development of artificial intelligence (AI) and deep learning (DL) applications in flood mapping, data integration, and flood impact assessment during the post-flood phase. There has been notable advancement in the methodologies of binary segmentation, data fusion, and optimisation, as well as post-disaster analysis. The findings, however, suggest that the development in all three classes remains limited due to data availability, temporal resolution, and model flexibility.

For instance, U-Net-like models attain high prediction accuracy. However, they may suffer from certain limitations, such as considering numerous varieties based on cutting-edge models, which are often viewed as having enormous inertia. Similarly, a range of damage assessment methodologies vary, from small-scale vegetation-based indices like DVDI and VSI, to various indices for flood disasters, which require robustness. The table also exhibits a compromise between the complexity of the model and its usability, demonstrating that, for instance, HDF and Spi3DNet present higher classification performance but are characterized by increased complexity and computational requirements. The analysis reveals the three key contributory challenges to the robustness gap.

- **Data limitations:** Poor temporal resolution, lack of heterogeneity and quality of available diverse datasets also have a negative impact on model generalization [66]. Model performance is often limited to the training domains due to the use of geography-bound training datasets and the proprietary step of manual training.
- **Model architecture constraints:** The existing frameworks continue to fail to address the issues of physical consistency, computational efficiency, and uncertainty quantification simultaneously. The issue of balancing the model's complexity with the practicality of its operations still needs to be explored.
- **Dynamic context awareness:** The majority of systems fail to integrate real-time situational awareness, socio-economic context, and the effects of climate change, resulting in analyses that do not adequately address the changing conditions of floods.

The pathway to robust AI systems, illustrated in Fig. 3, requires addressing these interconnected challenges through:

- **Integrated Data Ecosystems:** Establishing a multi-source data collection framework that incorporates satellite images, IoT sensor data, crowd-sourced data, and socio-economic data, to be adopted as protocols. This unified protocol addresses the fundamental data shortfall identified in every field, particularly in damage assessment, where there is a lack of data on economic and agricultural losses.
- **Physics-consistent AI frameworks:**Incorporating physical constraints via hybrid modeling or PINNs is a step beyond purely data-driven approaches. This ensures that predictions, while retaining the pattern recognition capabilities of deep learning, are still physically plausible.
- **Dynamic adaptation mechanism:** Systems designed to implement dynamic adaptation mechanisms are capable of continuous learning and self-updating in response to new and changing environments. Instead of being "static," these systems implement new techniques like dynamic learning paired with real-time feedback so that they change their forecasts and decisions as the behavior of the world shifts, for example, during global warming or city development and change. These models are valuable in sustaining accurate and reliable model behaviour over an extended period when new information is added continuously. These models are used for very complex situations of wayward behaviour, as system dynamics applies to complex and unpredictable evolving situations.

In addition, the development of advanced systems should be the result of operational collaboration among AI scientists, specialised domain practitioners, and policymakers at the community level. This extends beyond

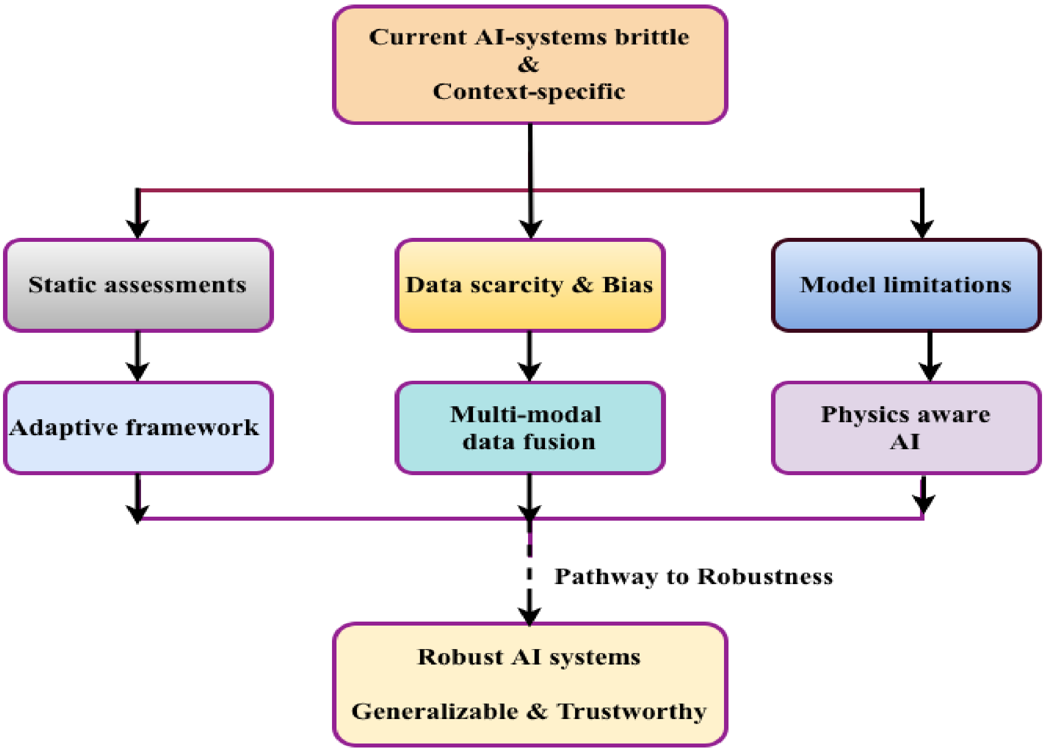

Figure 3. Pathway from brittle to robust AI systems for flood severity measurement

technical issues to consider social and contextual realities, leading to more effective and equitable strategies for managing floods.

A recent study [67] aimed to address the gap in post-disaster assessment caused by the limited availability of very high-resolution optical and SAR imagery. Coupled with the lack of any rapid automation response in disaster response caused by traditional optical data, the novel dataset enables the refinement of AI models. The lack of benchmarks and training data for disaster scenarios has significantly hindered progress in the domain. This work aims to utilise BRIGHT to benchmark foundational models and advanced AI tasks in domain adaptation and change detection, to build robust models for damage assessment in disaster scenarios.

The advanced flood damage modeling AI no longer restricts itself to purely hydrological data. These include hazard data (precipitation and river stage), topographical data (DEM and HAND), exposure data (the built infrastructure), and impact data (claims and field surveys). This integrated approach offers a comprehensive and high-resolution understanding of flood risk.

Despite achieving moderate performance (F1 score of 0.598) on four damage classes, DamageScope's DL framework [68] for automated damage assessment of buildings using publicly available datasets ([67], [69], [70]) within a post-disaster context spans a wide range of practical usefulness by employing geolocation and integration visualisation capabilities. Its utility is, however, restricted by poor accuracy in classifying damage levels, reliance on single-resolution satellite images, and the inability to distinguish between minor and moderate levels of damage. The authors propose the use of damage-aware loss functions and multi-task learning to classify target images more sensitively, as well as the dendritic integration of multi-source remote sensing data (such as UAV and aerial) to refine the spatial resolution of the image and capture more structural elements. Furthermore, geolocation accuracy can be enhanced to spatially correct the data used for visualisation, thereby transforming the system from a static assessment mechanism to a dynamic support resource for decision-making during emergency management, predictive of on-site actions.

The FloodDamageCast framework [71], developed for 500-meter resolution flood damage nowcasting, follows this integrated strategy. It combines diverse datasets, including NFIP claims, FEMA assistance records, and built-environment and topographic features. A key improvement in FloodDamageCast is the use of a Conditional Tabular GAN (CTGAN) to generate synthetic samples, which helps address the severe class imbalance often found in flood damage data. However, some challenges remain: the model relies on dense local sensor networks, synthetic data may differ from real data distributions, and the categorization results can vary depending on the clustering method employed. Future research should focus on utilizing remote sensing data to reduce sensor dependency and enhancing GAN training to improve generalizability and robustness.

5. CONCLUSION

The effectiveness of AI techniques for flood monitoring, mapping, and damage assessment, and their capabilities, limitations, and future directions, have all been analyzed extensively. The literature indicates that specialized deep learning techniques, particularly CNNs such as U-Net, SegNet, and DeepLabv3+, have significantly improved flood segmentation compared to traditional methods based on rules and thresholds, and that these methods are more accurate than traditional techniques. The combination of various geospatial data, including Sentinel-1 SAR and Sentinel-2 optical data, DEM, rainfall, and hydrologic soil data, enhances the accuracy and robustness of urban and regional flood mapping models. Additionally, the probabilistic and hybrid models have refined the quantification of uncertainty and the real-time applicability of the models, which represents a significant advancement for operational flood intelligence systems. Additionally, the review highlights ongoing issues related to temporal imaging gaps, class imbalance, data variety limitations, and the computational complexity of calculations. Most models are still regionally focused and do not perform well under different hydro-climatic conditions. Furthermore, the next area of development in this field is the integration of satellite, UAV, and IoT data streams using Physics-Informed Neural Networks (PINNs), which embed hydrodynamic concepts into the neural network learning process. Furthermore, the integration of multi-temporal datasets along with probabilistic approaches will improve the accuracy and reliability of flood models, which is crucial for disaster management decision-making. The dimensionality of physically consistent, scalable, and interpretable systems of AI-powered flood analysis, as well as their corresponding components, remains an area of active research and development. The integration of cutting-edge deep learning, hydrology, and real-time satellite systems is vital for accomplishing real-time flood prediction and effective impact mitigation. This framework will help address the escalating challenges posed by climate change and urban growth, enhancing our ability to predict, manage, and mitigate flood risks in a timely and effective manner.

REFERENCES

[1] X. Ruan, H. Sun, W. Shou, and J. Wang, "The impact of climate change and urbanization on compound flood risks in coastal areas: A comprehensive review of methods," *Applied Sciences*, vol. 14, no. 21, p. 10019, 2024.

[2] H. Tabari, "Climate change impact on flood and extreme precipitation increases with water availability," *Scientific reports*, vol. 10, no. 1, p. 13768, 2020.

[3] Y. Hirabayashi, R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae, "Global flood risk under climate change," *Nature climate change*, vol. 3, no. 9, pp. 816–821, 2013.

[4] United Nations Office for Disaster Risk Reduction (UNDRR), "Gar 2025 hazards: Floods," *Global Assessment Report on Disaster Risk Reduction*, 2025.

[5] L. Alfieri, L. Feyen, and G. Di Baldassarre, "Increasing flood risk under climate change: a pan-european assessment of the benefits of four adaptation strategies," *Climatic Change*, vol. 136, no. 3, pp. 507–521, 2016.

[6] L. Bibbò, G. Bilotta, G. M. Meduri, E. Genovese, and V. Barrile, "Flood risk forecasting: An innovative approach with machine learning and markov chains using lidar data," *Applied Sciences*, vol. 15, no. 13, p. 7563, 2025.

[7] M. Mareschal and B. L. Holian, "From fluid particles to physical particles: computing hydrodynamics," in *Microscopic simulations of complex hydrodynamic phenomena*. Springer, 1992, pp. 1–12.

[8] A. E. Jenifer, A. Aparna, N. Sudha, and A. Kumar, "Agrifloodnet: a dual patch cnn architecture for mapping flooded agricultural lands via bi-temporal multi-sensor images," *Geocarto International*, vol. 37, no. 26, pp. 13 618–13 637, 2022.

[9] N. I. Ulloa, S.-H. Yun, S.-H. Chiang, and R. Furuta, "Sentinel-1 spatiotemporal simulation using convolutional lstm for flood mapping," *Remote sensing*, vol. 14, no. 2, p. 246, 2022.

[10] I. Chamatidis, D. Istrati, and N. D. Lagaros, "Vision transformer for flood detection using satellite images from sentinel-1 and sentinel-2," *Water*, vol. 16, no. 12, p. 1670, 2024.

[11] S. Sharma and L. Chaudhari, "Vision transformers and multi-sensor earth observation," in *Deep Learning for Multi-Sensor Earth Observation*. Elsevier, 2025, pp. 201–210.

[12] N. K. Sharma and M. Saharia, "Deepsarflood: Rapid and automated sar-based flood inundation mapping using vision transformer-based deep ensembles with uncertainty estimates," *Science of Remote Sensing*, vol. 11, p. 100203, 2025.

[13] T. Saleh, S. Holail, X. Xiao, and G.-S. Xia, "High-precision flood detection and mapping via multi-temporal sar change analysis with semantic token-based transformer," *International Journal of Applied Earth Observation and Geoinformation*, vol. 131, p. 103991, 2024.

[14] Q. Poterek, A. Caretto, R. Braun, S. Clandillon, C. Huber, and P. Ceccato, "Interpolated flood surface (inflos), a rapid and operational tool to estimate flood depths from earth observation data for emergency management," *Remote Sensing*, vol. 17, no. 2, p. 329, 2025.

[15] T. Destefanis, S. Guliyeva, P. Boccardo, and V. Fissore, "Advancing flood detection and mapping: A review of earth observation services, 3d data integration, and ai-based techniques," *Remote Sensing*, vol. 17, no. 17, p. 2943, 2025.

[16] S. Yao, W. Du, N. Chen, C. Wang, and Z. Chen, "An integrated process-based framework for flood phase segmentation and assessment," *International Journal of Geographical Information Science*, vol. 37, no. 6, pp. 1315–1337, 2023.

[17] J. Savage, P. Uhe, O. Wing, C. Sampson, A. Smith, N. Lord, N. Addor, S. Hatchard, J. Hoch, J. Bates *et al.*, "A new framework for building global flood models for the present day and future climates," in *EGU General Assembly Conference Abstracts*, 2023, pp. EGU–15 051.

[18] R. Sadiq, Z. Akhtar, M. Imran, and F. Ofli, "Integrating remote sensing and social sensing for flood mapping," *Remote Sensing Applications: Society and Environment*, vol. 25, p. 100697, 2022.

[19] K. J. Beven, *Rainfall-runoff modelling: the primer*. John Wiley & Sons, 2012.

[20] B. Zhao, H. Sui, J. Liu, W. Shi, W. Wang, C. Xu, and J. Wang, "Flood inundation monitoring using multi-source satellite imagery: A knowledge transfer strategy for heterogeneous image change detection," *Remote Sensing of Environment*, vol. 314, p. 114373, 2024.

[21] I. Ahmad, R. Farooq, M. Ashraf, M. Waseem, and D. Shangguan, "Improving flood hazard susceptibility assessment by integrating hydrodynamic modeling with remote sensing and ensemble machine learning," *Natural Hazards*, pp. 1–30, 2025.

[22] A. Tarpanelli, A. C. Mondini, and S. Camici, "Effectiveness of sentinel-1 and sentinel-2 for flood detection assessment in europe," *Natural Hazards and Earth System Sciences*, vol. 22, no. 8, pp. 2473–2489, 2022.

[23] G. S. Santecchia, G. N. Revollo Sarmiento, S. A. Genchi, A. J. Vitale, and C. A. Delrieux, "Assessment of landsat-8 and sentinel-2 water indices: A case study in the southwest of the buenos aires province (argentina)," *Journal of Imaging*, vol. 9, no. 9, p. 186, 2023.

[24] S. Dhanabalan, S. Abdul Rahaman, and R. Jegankumar, "Flood monitoring using sentinel-1 sar data: A case study based on an event of 2018 and 2019 southern part of kerala," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 44, pp. 37–41, 2021.

[25] A. Toma, I. Șandric, and B.-A. Mihai, "Flooded area detection and mapping from sentinel-1 imagery. complementary approaches and comparative performance evaluation," *European Journal of Remote Sensing*, vol. 57, no. 1, p. 2414004, 2024.

[26] K. D. Bathe and N. S. Patil, "Flood detection and mapping: A critical review of methods, challenges and future prospects," *Journal of the Indian Society of Remote Sensing*, pp. 1–32, 2025.

[27] Y. Yuan, X. Meng, W. Sun, G. Yang, L. Wang, J. Peng, and Y. Wang, "Multi-resolution collaborative fusion of sar, multispectral and hyperspectral images for coastal wetlands mapping," *Remote Sensing*, vol. 14, no. 14, p. 3492, 2022.

[28] D. Bonafilia, B. Tellman, T. Anderson, and E. Issenberg, "Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 210–211.

[29] W. Tan, N. Qin, Y. Zhang, H. McGrath, M. Fortin, and J. Li, "A rapid high-resolution multi-sensory urban flood mapping framework via dem upscaling," *Remote Sensing of Environment*, vol. 301, p. 113956, 2024.

[30] M. Siddique and T. Ahmed, "Ccd-conv1d: A deep learning based coherent change detection technique to monitor and forecast floods using sentinel-1 images," *Remote Sensing Applications: Society and Environment*, vol. 37, p. 101440, 2025.

[31] R. Bentivoglio, E. Isufi, S. N. Jonkman, and R. Taormina, "Deep learning methods for flood mapping: a review of existing applications and future research directions," *Hydrology and Earth System Sciences Discussions*, vol. 2022, pp. 1–50, 2022.

[32] A. Jamali, S. K. Roy, L. H. Beni, B. Pradhan, J. Li, and P. Ghamisi, "Residual wave vision u-net for flood mapping using dual polarization sentinel-1 sar imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 127, p. 103662, 2024.

[33] Y. Mu, J. Sun, and J. He, "The combined focal cross entropy and dice loss function for segmentation of protein secondary structures from cryo-em 3d density maps," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2022, pp. 3454–3461.

[34] S. S. Kulkarni and A. Mahapatra, "Improving satellite flood image classification using attention-based cnn and transformer models." *International Journal of Advanced Computer Science & Applications*, vol. 16, no. 3, 2025.

[35] L. A. Kadiyala, O. Mermer, D. J. Samuel, Y. Sermet, and I. Demir, "The implementation of multimodal large language models for hydrological applications: A comparative study of gpt-4 vision, gemini, llava, and multimodal-gpt," *Hydrology*, vol. 11, no. 9, p. 148, 2024.

[36] I. Ahmed, M. Ahmad, G. Jeon, and A. Chehri, "An internet of things and ai-powered framework for long-term flood risk evaluation," *IEEE Internet of Things Journal*, vol. 11, no. 3, pp. 3812–3819, 2023.

[37] V. Santhi, I. Krishnamurthi, and N. Madhumitha, "Flood detection and segmentation using deep learning models," in *2024 International Conference on Smart Systems for Electrical, Electronics, Communication and Computer Engineering (ICSSEECC)*. IEEE, 2024, pp. 226–231.

[38] N. Mohamadiazar, A. Ebrahimian, and H. Hosseiny, "Integrating deep learning, satellite image processing, and spatial-temporal analysis for urban flood prediction," *Journal of Hydrology*, vol. 639, p. 131508, 2024.

[39] J. Donnelly, A. Daneshkhah, and S. Abolfathi, "Physics-informed neural networks as surrogate models of hydrodynamic simulators," *Science of the Total Environment*, vol. 912, p. 168814, 2024.

[40] Y. Bhattarai, S. Bista, R. Talchabhadel, S. Duwal, and S. Sharma, "Rapid prediction of urban flooding at street-scale using physics-informed machine learning-based surrogate modeling," *Total Environment Advances*, vol. 12, p. 200116, 2024.

[41] M. Fawakherji, J. Blay, M. Anokye, L. Hashemi-Beni, and J. Dorton, “Deepflood for inundated vegetation high-resolution dataset for accurate flood mapping and segmentation,” *Scientific Data*, vol. 12, no. 1, p. 271, 2025.

[42] B. Ghosh, S. Garg, M. Motagh, and S. Martinis, “Automatic flood detection from sentinel-1 data using a nested unet model and a nasa benchmark dataset,” *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 92, no. 1, pp. 1–18, 2024.

[43] V. Hertel, C. Chow, O. Wani, M. Wieland, and S. Martinis, “Probabilistic sar-based water segmentation with adapted bayesian convolutional neural network,” *Remote Sensing of Environment*, vol. 285, p. 113388, 2023.

[44] Y. Bai, W. Wu, Z. Yang, J. Yu, B. Zhao, X. Liu, H. Yang, E. Mas, and S. Koshimura, “Enhancement of detecting permanent water and temporary water in flood disasters by fusing sentinel-1 and sentinel-2 imagery using deep learning algorithms: Demonstration of sen1floods11 benchmark datasets,” *Remote Sensing*, vol. 13, no. 11, p. 2220, 2021.

[45] G. I. Drakonakis, G. Tsagkatakis, K. Fotiadou, and P. Tsakalides, “Ombrianet—supervised flood mapping via convolutional neural networks using multitemporal sentinel-1 and sentinel-2 data fusion,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 2341–2356, 2022.

[46] M. Wieland, S. Martinis, R. Kiefl, and V. Gstaiger, “Semantic segmentation of water bodies in very high-resolution satellite and aerial images,” *Remote Sensing of Environment*, vol. 287, p. 113452, 2023.

[47] S. Lv, L. Meng, D. Edwing, S. Xue, X. Geng, and X.-H. Yan, “High-performance segmentation for flood mapping of hisea-1 sar remote sensing images,” *Remote Sensing*, vol. 14, no. 21, p. 5504, 2022.

[48] M. Bereczky, M. Wieland, C. Krullikowski, S. Martinis, and S. Plank, “Sentinel-1-based water and flood mapping: Benchmarking convolutional neural networks against an operational rule-based processing chain,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 2023–2036, 2022.

[49] N. Notarangelo, C. Wirion, and F. van Winsen, “Sturm-flood: a curated dataset for deep learning-based flood extent mapping leveraging sentinel-1 and sentinel-2 imagery,” *Big Earth Data*, pp. 1–27, 2025.

[50] Y. Wang, Y. Shen, B. Salahshour, M. Cetin, K. Iftekharuddin, N. Tahvildari, G. Huang, D. K. Harris, K. Ampofo, and J. L. Goodall, “Urban flood extent segmentation and evaluation from real-world surveillance camera images using deep convolutional neural network,” *Environmental Modelling & Software*, vol. 173, p. 105939, 2024.

[51] Z. Situ, Q. Zhong, J. Zhang, S. Teng, X. Ge, Q. Zhou, and Z. Zhao, “Attention-based deep learning framework for urban flood damage and risk assessment with improved flood prediction and land use segmentation,” *International Journal of Disaster Risk Reduction*, vol. 116, p. 105165, 2025.

[52] J. Sakamoto, “Proposal of a flood damage road detection method based on deep learning and elevation data,” *Geomatics, Natural Hazards and Risk*, vol. 15, no. 1, p. 2375545, 2024.

[53] R. Pally and S. Samadi, “Application of image processing and convolutional neural networks for flood image classification and semantic segmentation,” *Environmental modelling & software*, vol. 148, p. 105285, 2022.

[54] J. Li, R. Cai, Y. Tan, H. Zhou, A.-M. Sadick, W. Shou, and X. Wang, “Automatic detection of actual water depth of urban floods from social media images,” *Measurement*, vol. 216, p. 112891, 2023.

[55] S. M. Gulzar, F. U. H. Mir, M. Rafiqui, and M. A. Tantray, “Damage assessment of residential constructions in post-flood scenarios: A case of 2014 kashmir floods.” *Environment, Development & Sustainability*, vol. 23, no. 3, 2021.

[56] M. S. Rahman, L. Di, E. Yu, L. Lin, and Z. Yu, "Remote sensing based rapid assessment of flood crop damage using novel disaster vegetation damage index (dvdi)," *International Journal of Disaster Risk Science*, vol. 12, no. 1, pp. 90–110, 2021.

[57] Y. Dang, L. Yang, and J. Song, "The construction of a crop flood damage assessment index to rapidly assess the extent of postdisaster impact," *Remote Sensing*, vol. 16, no. 9, p. 1527, 2024.

[58] Y. Kurihara, M. Miyamoto, and R. Sunakawa, "Flood direct damage assessment due to typhoon ulysses by satellite images," *International Journal of Disaster Risk Reduction*, vol. 118, p. 105067, 2025.

[59] M. Saharia, S. K. Jain, V. Prakash, H. Malik, O. Sreejith, and D. Joshi, "A district-level flood severity index for flood management in india," *Natural Hazards*, pp. 1–17, 2025.

[60] H. Hao and Y. Wang, "Leveraging multimodal social media data for rapid disaster damage assessment," *International Journal of Disaster Risk Reduction*, vol. 51, p. 101760, 2020.

[61] R. I. Ogie, S. James, A. Moore, T. Dilworth, M. Amirghasemi, and J. Whittaker, "Social media use in disaster recovery: A systematic literature review," *International Journal of Disaster Risk Reduction*, vol. 70, p. 102783, 2022.

[62] N. P. Shetty, Y. Bijalwan, P. Chaudhari, J. Shetty, and B. Muniyal, "Disaster assessment from social media using multimodal deep learning," *Multimedia Tools and Applications*, vol. 84, no. 18, pp. 18 829–18 854, 2025.

[63] A. Islam, S. Ghosh, and M. Sarkar, "Assessing livelihood vulnerability of rural communities in the wake of recurrent tropical flood hazards in india," *Natural Hazards*, vol. 121, no. 1, pp. 677–704, 2025.

[64] F. I. W. Rohmat, A. J. Löhr, F. Pratama, N. S. Burnama, and A. A. Kuntoro, "Quantifying time-dependent flood resilience index in a densely populated urban environment in manado, indonesia," *International Journal of Disaster Risk Reduction*, vol. 116, p. 105112, 2025.

[65] P. L. B. Ghazali, R. A. Ibrahim, M. Mamat, A. Sambas, Y. Hidyat *et al.*, "Price model of multiple-trigger flood bond with trigger indices of aggregate losses and maximum number of submerged houses," *International Journal of Disaster Risk Reduction*, vol. 116, p. 105156, 2025.

[66] J. E. Schubert, K. J. Mach, and B. F. Sanders, "National-scale flood hazard data unfit for urban risk management," *Earth's future*, vol. 12, no. 7, p. e2024EF004549, 2024.

[67] H. Chen, J. Song, O. Dietrich, C. Broni-Bediako, W. Xuan, J. Wang, X. Shao, Y. Wei, J. Xia, C. Lan *et al.*, "Bright: A globally distributed multimodal building damage assessment dataset with very-high-resolution for all-weather disaster response," *Earth System Science Data Discussions*, vol. 2025, pp. 1–51, 2025.

[68] S. Al Shafian, C. He, and D. Hu, "Damagescope: An integrated pipeline for building damage segmentation, geospatial mapping, and interactive web-based visualization," *Remote Sensing*, vol. 17, no. 13, p. 2267, 2025.

[69] S. T. Seydi, H. Rastiveis, B. Kalantar, A. A. Halin, and N. Ueda, "Bdd-net: An end-to-end multiscale residual cnn for earthquake-induced building damage detection," *Remote Sensing*, vol. 14, no. 9, p. 2214, 2022.

[70] R. Gupta, R. Hosfelt, S. Sajeev, N. Patel, B. Goodman, J. Doshi, E. Heim, H. Choset, and M. Gaston, "xbd: A dataset for assessing building damage from satellite imagery," *arXiv preprint arXiv:1911.09296*, 2019.

[71] C.-F. Liu, L. Huang, K. Yin, S. Brody, and A. Mostafavi, "Flooddamagecast: Building flood damage nowcasting with machine-learning and data augmentation," *International Journal of Disaster Risk Reduction*, vol. 114, p. 104971, 2024.