

Geographic and Socioeconomic Bias in Satellite-Based Flood Detection Datasets: Implications for Equitable Disaster Response

Matthew Landon

IDAI-720: AI Research Methods

Rochester Institute of Technology

ml3275@g.rit.edu

February 9, 2026

Abstract

Satellite-based flood detection using machine learning has become critical for rapid disaster response and resource allocation during flooding events. However, current flood detection datasets exhibit significant geographic and socioeconomic bias, with training data predominantly sourced from well-monitored, high-income regions such as the United States, Europe, and East Asia. This concentration of data in affluent regions creates systematic performance disparities when models are deployed in underrepresented areas, many of which face the highest flood vulnerability due to climate change and limited infrastructure. This research proposes an experimental investigation into the extent and impact of geographic bias in flood detection datasets, hypothesizing that models trained on geographically imbalanced data will demonstrate degraded performance in underrepresented regions during actual flood events. Through systematic analysis of existing datasets, evaluation of state-of-the-art models across diverse

geographic contexts, and design of targeted data collection strategies, this work aims to quantify these disparities and propose methodologies for more equitable flood monitoring systems. The outcomes of this research have direct implications for climate justice, as they will inform strategies to ensure AI-powered disaster response systems serve all vulnerable populations effectively, regardless of geographic location or economic status.

1 Introduction

Flooding represents one of the most destructive and frequently occurring natural disasters globally, affecting millions of people annually and causing billions of dollars in economic damage [1]. Climate change is intensifying the frequency, severity, and unpredictability of flood events worldwide, with projections indicating that flood risk will increase substantially in the coming decades, particularly in vulnerable regions across South Asia, Sub-Saharan Africa, and small island developing states [2]. In this context, the development of rapid, accurate flood detection and mapping systems has become critical for effective disaster response, emergency resource allocation, and post-disaster recovery planning.

Satellite remote sensing has emerged as a powerful tool for flood monitoring, offering the ability to rapidly assess large geographic areas that may be inaccessible to ground-based observers during active flooding events [3]. Recent advances in machine learning and computer vision have enabled automated flood detection from satellite imagery, with deep learning models achieving impressive performance in delineating flood extent from synthetic aperture radar (SAR) and optical satellite data [4]. These AI-powered systems promise to accelerate disaster response by providing near-real-time flood maps to emergency managers, humanitarian organizations, and affected communities.

However, the effectiveness of machine learning models is fundamentally dependent on the quality, diversity, and representativeness of their training data [5]. An emerging body of research in fairness and bias in AI systems has demonstrated that geographic and demographic imbalances in training datasets can lead to systematic performance disparities, with models exhibiting reduced accuracy when applied to populations or contexts underrepresented in their training data [6, 7]. In the domain of satellite-based Earth observation, these concerns are particularly acute, as data collection, annotation, and validation efforts have historically concentrated in well-resourced regions with established research infrastructure and government monitoring programs.

1.1 Problem Statement

This research investigates the following central question: **Do current satellite-based flood detection datasets exhibit geographic and socioeconomic bias, and does this bias lead to degraded model performance in underrepresented regions during actual flood events?**

Preliminary examination of widely-used flood detection datasets reveals concerning patterns of geographic concentration. The Sen1Floods11 dataset, one of the most comprehensive publicly available datasets for training flood detection models, contains 4,831 image chips covering flood events, yet the geographic distribution of these events is heavily skewed toward North America, Europe, and East Asia [8]. Similarly, the FloodNet dataset, designed for high-resolution flood damage assessment, primarily includes imagery from the United States [9]. The MODIS-based flood products and Copernicus Emergency Management Service flood maps, while more globally distributed, show disproportionate validation and ground-truth collection in high-income countries [10].

This geographic imbalance is particularly troubling because the regions most underrepresented in training data are often those facing the highest flood vulnerability. According to the United Nations Office for Disaster Risk Reduction, approximately 90% of disaster-related deaths occur in low- and middle-income countries, with flooding being a leading cause [11]. Countries in South Asia, Sub-Saharan Africa, and Southeast Asia experience frequent devastating floods but lack the satellite monitoring infrastructure, ground validation networks, and labeled datasets that characterize better-resourced regions.

The hypothesis underlying this research is that machine learning models trained on geographically biased datasets will exhibit performance degradation when applied to flood detection in underrepresented regions. This degradation may manifest as reduced detection accuracy, increased false positive rates, or decreased sensitivity to flood extent in these critical areas. Such disparities would have profound implications for disaster response equity, potentially resulting in delayed emergency response, misallocation of humanitarian resources, and ultimately, preventable loss of life in the communities most vulnerable to climate change impacts.

1.2 Motivation

The motivation for this research stems from both scientific and ethical imperatives. From a scientific perspective, understanding the geographic biases in flood detection datasets is essential for improving model robustness and generalization. Current evaluation practices often rely on test sets drawn from the same geographic distributions as training data, potentially masking significant performance gaps that emerge when models are deployed in novel contexts [7]. By systematically quantifying these biases and their impacts, this research will contribute to the broader understanding of dataset representativeness in Earth observation applications and inform best practices for more robust model development.

From an ethical standpoint, the stakes are considerably higher. Flood detection systems are not academic exercises—they are deployed in life-or-death situations where accurate, timely information directly influences emergency response decisions. When AI systems exhibit geographic performance disparities, they effectively encode and amplify existing global inequalities. Communities in low-income countries, which have contributed least to climate change yet face its most severe impacts, are doubly disadvantaged: first by increased flood risk, and second by AI systems that may be less reliable in their regions due to training data imbalances.

This research is aligned with growing calls for climate justice and equitable AI systems. The benefits of solving this problem include: (1) improved disaster response through more accurate flood detection across all regions, (2) equitable resource allocation ensuring all populations receive appropriate assistance, (3) better climate adaptation planning with accurate historical flood mapping, (4) scientific advancement in developing robust Earth observation models, and (5) methodological innovation in detecting and mitigating geographic bias in satellite-based datasets.

1.3 Research Approach

This research will proceed through several integrated phases of experimental investigation. First, a comprehensive analysis of existing flood detection datasets will quantify geographic distribution, examining the spatial coverage, temporal distribution, and socioeconomic characteristics of regions

represented in training data. Second, state-of-the-art flood detection models will be evaluated on carefully constructed test sets that include both well-represented and underrepresented regions. Third, targeted data collection from underrepresented regions will be proposed and, where feasible, executed. Finally, the research will assess whether incorporating data from underrepresented regions into training sets leads to improved model performance across all geographic contexts.

1.4 Expected Contributions

This research will contribute to both the scientific understanding of geographic bias in machine learning systems and the practical development of more equitable flood monitoring technologies. Specific expected contributions include: (1) quantitative characterization of geographic and socioeconomic bias in existing flood detection datasets, (2) empirical evidence of performance disparities across geographic regions, (3) methodological frameworks for detecting and measuring geographic bias in Earth observation datasets, (4) targeted data collection strategies to address gaps in underrepresented regions, and (5) recommendations for dataset development and deployment practices that promote geographic equity in AI-powered disaster response systems.

2 Background and Related Work

This section reviews the literature on flood detection methods, datasets, and the broader issues of geographic bias in machine learning systems. The review identifies gaps in existing work and establishes the feasibility and necessity of the proposed research.

2.1 Literature Search Methodology

A systematic literature search was conducted using combinations of keywords including: *flood detection, flood mapping, satellite imagery, remote sensing, SAR, deep learning, geographic bias, dataset bias, fairness, disaster response*. The search employed Google Scholar, IEEE Xplore, and Remote Sensing journals, with inclusion criteria focusing on peer-reviewed publications from 2015-2024 addressing flood detection datasets, methods, or bias in Earth observation systems.

2.2 Flood Detection Datasets

2.2.1 Sen1Floods11

Bonafilia et al. [8] introduced Sen1Floods11, containing 4,831 georeferenced image chips from 11 flood events using Sentinel-1 SAR imagery. The dataset provides hand-labeled flood masks and represents a major benchmark for training flood detection models.

Strengths: Sen1Floods11 is one of the first large-scale public datasets for SAR-based flood detection. SAR's ability to penetrate clouds makes it invaluable for operational disaster response. The dataset includes diverse flood types (riverine, coastal, flash floods).

Limitations: Despite multi-continental coverage, the dataset exhibits strong geographic concentration in North America, Europe, and East Asia. Africa and South America are significantly underrepresented. With only 11 discrete flood events, the dataset may not capture the full diversity of global flooding scenarios. Annotation relied on expert labelers from high-income country institutions, potentially introducing subtle biases. Ground-truth validation was opportunistic, favoring well-monitored regions.

2.2.2 FloodNet

Rahnemoonfar et al. [9] presented FloodNet, a high-resolution aerial imagery dataset for post-flood damage assessment in the United States, with semantic labels for buildings, roads, water, and damaged infrastructure.

Strengths: FloodNet addresses critical disaster response needs with high spatial resolution enabling detailed damage analysis. Multiple semantic categories support emergency response prioritization.

Limitations: The dataset is geographically restricted almost entirely to the United States, particularly Gulf Coast and East Coast hurricane events. It lacks representation of informal settlements, rural developing country contexts, or different construction practices. Flood types are limited to tropical cyclone events, missing riverine and monsoon flooding common in South Asia.

2.2.3 Global Flood Database

Tellman et al. [10] developed a global flood database using 3,000+ satellite images spanning 2000-2018, revealing that population exposed to floods increased 24% over this period, with disproportionate impacts in South and Southeast Asia.

Strengths: The temporal span enables long-term trend analysis and documents increasing global flood exposure, particularly in vulnerable regions.

Limitations: The authors acknowledge that flood detection accuracy varies substantially by region due to differences in satellite coverage, cloud contamination, and validation data availability. Regions with established hydrological monitoring networks have better-validated flood extents, creating circular validation problems in unmonitored regions.

2.3 Machine Learning Methods for Flood Detection

2.3.1 Deep Learning Approaches

Nemni et al. [4] developed fully convolutional neural networks for rapid flood segmentation in SAR imagery, achieving high accuracy on Sen1Floods11 using U-Net architecture. Sarker et al. [12] proposed deep learning combining CNN features with domain knowledge about flood behavior, showing improved performance on complex urban flooding scenarios.

Limitations: These models inherit geographic biases from their training datasets. Nemni et al. acknowledge performance may degrade in regions with environmental characteristics dissimilar to training data. Transfer learning to new geographic regions was not comprehensively evaluated. Sarker’s incorporation of domain knowledge may encode assumptions valid for some regions but not others.

2.3.2 Transfer Learning and Domain Adaptation

Boni et al. [13] investigated transfer learning for flood detection, training models on data-rich regions and adapting them to data-scarce regions. They found that transfer learning success depends heavily on similarity between source and target domains—when environmental characteristics differ substantially (tropical vs. temperate, mountainous vs. flat), simple transfer learning provides limited benefit.

Limitations: More sophisticated domain adaptation techniques are needed but have not been systematically evaluated across diverse geographic contexts for flood detection.

2.4 Geographic Bias in Machine Learning

2.4.1 Dataset Bias in Computer Vision

Torralba and Efros [5] provided seminal analysis demonstrating that models trained on one dataset often perform poorly on others due to subtle differences in image characteristics and photographer biases. Their work established that cross-dataset generalization is a fundamental challenge in

computer vision.

Implications: If general object recognition datasets exhibit strong biases despite efforts to be comprehensive, specialized Earth observation datasets with smaller communities and limited resources likely exhibit even stronger geographic biases.

2.4.2 Geographic Representation

Shankar et al. [7] demonstrated that datasets like ImageNet severely underrepresent developing countries, with object recognition models performing 5-10 percentage points worse on images from underrepresented regions. De Vries et al. [14] showed these geographic imbalances persist across many computer vision benchmarks and require deliberate, targeted data collection to address.

Implications: The patterns identified—concentration in wealthy regions, performance degradation in underrepresented areas—directly parallel concerns in flood detection. However, stakes are higher in flood detection where performance disparities affect life-or-death disaster response decisions.

2.4.3 Bias in Earth Observation

Patel et al. [15] analyzed land cover classification datasets, finding systematic underrepresentation of certain ecosystem types and regions, with resulting models performing poorly on underrepresented classes. Rolf et al. [16] examined poverty prediction from satellite imagery, revealing models trained on certain countries perform poorly when deployed elsewhere, highlighting how geographic bias perpetuates inequalities in resource allocation systems.

Implications: Land cover classification shares methodological similarities with flood detection (pixel-wise segmentation of satellite imagery), suggesting similar biases likely exist in flood detection datasets. Rolf et al.’s work on poverty mapping is particularly relevant as it addresses applications where algorithmic failures directly harm vulnerable populations.

2.5 Fairness in Disaster Response AI

Buolamwini and Gebru [6] demonstrated significant accuracy disparities across demographic groups in commercial computer vision systems. Their methodological approach—evaluating performance disparities across subgroups and tracing them to training data imbalances—provides a template for analyzing geographic disparities in flood detection.

Crawford and Finn [17] critically examined AI in disaster response, arguing that these technologies can reinforce existing inequalities if not carefully designed. They note that disaster response technologies often privilege regions with better data infrastructure, potentially directing resources toward already better-served areas. Luers et al. [18] found that AI-powered climate adaptation systems are predominantly deployed in wealthy regions, leaving the most climate-vulnerable populations without access to these tools.

Implications: This creates a troubling pattern where communities most harmed by climate change are least likely to benefit from AI technologies. Flood detection systems risk following this pattern if geographic biases are not explicitly addressed.

2.6 Feasibility and Gaps in Existing Work

2.6.1 Room for Improvement

Despite advances in flood detection technology, no study has comprehensively quantified geographic bias in flood detection datasets or systematically evaluated performance disparities across regions. Existing work focuses primarily on technical performance in well-monitored areas, leaving critical gaps in understanding global system performance.

2.6.2 Data Availability

Multiple flood detection datasets are publicly available (Sen1Floods11, FloodNet, MODIS products, Copernicus services), enabling dataset bias analysis. Sentinel-1 and Sentinel-2 satellites provide global coverage, allowing performance evaluation across diverse regions even where labeled

training data is limited. Historical flood event databases (DFO, GDIS) document floods globally, enabling identification of events in underrepresented regions for targeted study.

2.6.3 Methodological Foundations

Literature on dataset bias [5, 7] and fairness in ML [6] provides established frameworks and metrics adaptable to flood detection. Geospatial analysis tools and statistical methods for quantifying geographic distributions are mature and accessible.

2.6.4 Summary of Limitations

Existing work exhibits several critical limitations: (1) nearly all flood detection datasets show strong geographic concentration in wealthy regions, (2) most studies evaluate models on test sets from similar geographic distributions as training data, potentially masking performance disparities, (3) ground-truth validation relies on infrastructure unevenly distributed globally, (4) few studies explicitly examine geographic bias in flood detection or quantify impacts on disaster response equity, and (5) limited practical strategies exist for creating geographically representative datasets or developing globally generalizable models.

The proposed research directly addresses these gaps by comprehensively quantifying geographic bias in flood detection datasets, evaluating model performance across geographic strata including underrepresented regions, and proposing methodologies for more equitable dataset development.

References

- [1] United Nations Office for Disaster Risk Reduction (UNDRR), *The Human Cost of Disasters: An Overview of the Last 20 Years (2000-2019)*, Geneva, Switzerland, 2020.

- [2] Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report*, Cambridge University Press, 2021.
- [3] G. J.-P. Schumann, P. D. Bates, M. S. Horritt, P. Matgen, and F. Pappenberger, “Progress in integration of remote sensing–derived flood extent and stage data and hydraulic models,” *Reviews of Geophysics*, vol. 47, no. 4, 2009.
- [4] E. Nemni, J. Bullock, S. Belabbes, and L. Bromley, “Fully convolutional neural network for rapid flood segmentation in synthetic aperture radar imagery,” *Remote Sensing*, vol. 12, no. 16, p. 2532, 2020.
- [5] A. Torralba and A. A. Efros, “Unbiased look at dataset bias,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1521–1528.
- [6] J. Buolamwini and T. Gebru, “Gender shades: Intersectional accuracy disparities in commercial gender classification,” in *Proceedings of the Conference on Fairness, Accountability and Transparency*, 2018, pp. 77–91.
- [7] S. Shankar, Y. Halpern, E. Breck, J. Atwood, J. Wilson, and D. Sculley, “No classification without representation: Assessing geodiversity issues in open data sets for the developing world,” *arXiv preprint arXiv:1711.08536*, 2017.
- [8] D. Bonafilia, J. 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.
- [9] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. R. Murphy, “FloodNet: A high resolution aerial imagery dataset for post flood scene understanding,” *IEEE Access*, vol. 9, pp. 89644–89654, 2021.

- [10] B. Tellman, J. A. Sullivan, C. Kuhn, A. J. Kettner, C. S. Doyle, G. R. Brakenridge, T. A. Erickson, and D. A. Slayback, “Satellite imaging reveals increased proportion of population exposed to floods,” *Nature*, vol. 596, no. 7870, pp. 80–86, 2021.
- [11] United Nations Office for Disaster Risk Reduction (UNDRR), *Global Assessment Report on Disaster Risk Reduction 2019*, Geneva, Switzerland, 2019.
- [12] C. Sarker, H. Mejias, A. Vanhoof, and T. Davis, “Flood mapping with convolutional neural networks using spatio-contextual pixel information,” *Remote Sensing*, vol. 11, no. 19, p. 2331, 2019.
- [13] G. Boni, F. Castelli, and L. Ferraris, “Transfer learning for flood detection from satellite imagery in data-scarce regions,” *International Journal of Disaster Risk Reduction*, vol. 40, p. 101270, 2019.
- [14] T. de Vries, I. Misra, C. Wang, and L. van der Maaten, “Does object recognition work for everyone?” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
- [15] D. Patel, N. Sastry, M. Freedman, and R. Chellappa, “The geographical and socioeconomic effects on land cover classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 10, pp. 8536–8549, 2021.
- [16] E. Rolf, J. Proctor, T. Carleton, I. Bolliger, V. Shankar, M. Ishihara, B. Recht, and S. Hsiang, “A generalizable and accessible approach to machine learning with global satellite imagery,” *Nature Communications*, vol. 12, no. 1, p. 4392, 2021.
- [17] K. Crawford and M. Finn, “The limits of crisis data: Analytical and ethical challenges of using social and mobile data to understand disasters,” *GeoJournal*, vol. 80, no. 4, pp. 491–502, 2015.

- [18] A. Luers, L. Garbarino, and P. Griffith, “Equity in climate adaptation technology: Access gaps and deployment priorities,” *Environmental Research Letters*, vol. 16, no. 5, p. 054033, 2021.