Earth Observation for Disaster Mapping: Benchmarks, Methods, Challenges and Future Perspectives

Hongruixuan Chen†, Jian Song†, Weihao Xuan†, Junjue Wang†, Heli Qi, Zeqi Zhou, Pengyu Dai, Olivier Dietrich, Erika Gutierrez, Lars Bromly, Edoardo Nemni, Yafei Ou, Jie Zhao, Zhuo Zheng, Yonghao Xu, Ronny Hänsch, Wenzhe Jiao, Marco Chini, Claudio Persello, Junshi Xia, Shijian Lu, Lixin Wang, Zhe Zhu, Evan Shelhamer, Jocelyn Chanussot, Konrad Schindler, Naoto Yokoya

Abstract—Natural hazards cause hundreds of billions of dollars in losses annually, with impacts that disproportionately affect vulnerable populations. Earth observation (EO) provides comprehensive, repeatable measurements that are essential for assessing disaster impacts when ground-based systems fail. Over the past two decades, methods for post-disaster mapping impacts from EO data have evolved rapidly, from manual interpretation and spectral indices to deep learning and, most recently, foundation models. Yet this methodological diversity has outpaced systematic evaluation: most existing studies focus on individual hazards, single tasks, or specific method families, and lack reproducible cross-method comparison. In this paper, we provide a unified review and benchmark for EO-based disaster mapping. We define a task-driven taxonomy distinguishing disaster extent mapping from infrastructure damage mapping, and systematically review methods across both tasks, spanning traditional, deep learning, and foundation model paradigms. We then benchmark over 30 methods under a standardized protocol across 13 datasets covering diverse hazards, sensors, and operational settings. Our studies reveal that architectural complexity yields diminishing returns for extent mapping, that change detection architectures benefit damage mapping but not extent mapping, and that foundation models do not yet consistently outperform task-specific baselines, particularly on non-RGB modalities. Based on these findings, we identify key priorities for advancing operational disaster mapping. Code and benchmark resources are publicly available at here.

Index Terms—Earth observation, natural hazards, disaster mapping, deep learning, foundation models

†: *H. Chen, J. Song, W. Xuan, and J. Wang contribute this work equally. Contact: yokoya@k.u-tokyo.ac.jp; qschrx@gmail.com*
H. Chen, J. Song, W. Xuan, H. Qi, P. Dai, Y. Ou, J. Xia, and N. Yokoya are with RIKEN Center for Advanced Intelligence Project (AIP), Japan
W. Xuan, J. Wang, P. Dai, and N. Yokoya are with The University of Tokyo, Japan
Z. Zhou is with Brown University, United States
O. Dietrich and K. Schindler are with ETH Zurich, Switzerland
Z. Zheng is with Stanford University, United States
E. Gutierrez and L. Bromly are with United Nations Satellite Centre
E. Nemni is with Barcelona School of Economics, Spain
J. Zhao is with Technical University of Munich, Germany
Y. Xu is with Linköping University, Sweden
R. Hänsch is with German Aerospace Center (DLR), Germany
W. Jiao is with Texas A&M University, United States
M. Chini is with Luxembourg Institute of Science and Technology, Luxembourg
C. Persello is with University of Twente, Netherlands
S. Lu is with Nanyang Technological University, Singapore
L. Wang is with Indiana University Indianapolis, United States
Z. Zhu is with University of Connecticut, United States
E. Shelhamer is with The University of British Columbia and Vector Institute, Canada
J. Chanussot is with Université Grenoble Alpes INRIA, France

1 Introduction

Natural disasters claim tens of thousands of lives and cost the global economy hundreds of billions of dollars each year, and with impacts falling disproportionately on low- and middle-income countries least equipped to absorb them [54], [95], [446]. Climate change is intensifying this burden: the frequency and severity of hydro-meteorological extremes, including floods, storms, and wildfires, are rising [445], [494], compounding exposure for populations already facing poverty, ageing infrastructure, and constrained disaster preparedness.

Fig. 1 provides a global perspective spanning half a century (1974-2024) and highlights three patterns that motivate this review[1]. First, the growth in recorded disasters is dominated by hydro-meteorological hazards (e.g., floods and storms), consistent with an intensifying climate risk landscape. Second, impacts on human remain severe and hazard-dependent, spanning mortality, displacement, and livelihood disruption. Third, economic losses are deeply unequal: high-income countries bear the largest absolute losses, while low- and lower-middle-income countries suffer far greater losses relative to GDP, reflecting structural vulnerability and uneven capacity for risk reduction [155], [444]. These trends directly challenge the Sendai Framework for Disaster Risk Reduction 2015–2030 [442], whose Target C commits signatories to substantially reducing direct disaster

1. Trends in recorded disaster counts should be interpreted with caution. Improvements in reporting infrastructure and information flow over the past decades contribute to the apparent rise in event numbers, particularly in data-sparse regions. The geographic distribution in Fig. 1-(a) likely reflects reporting coverage as much as actual hazard exposure.

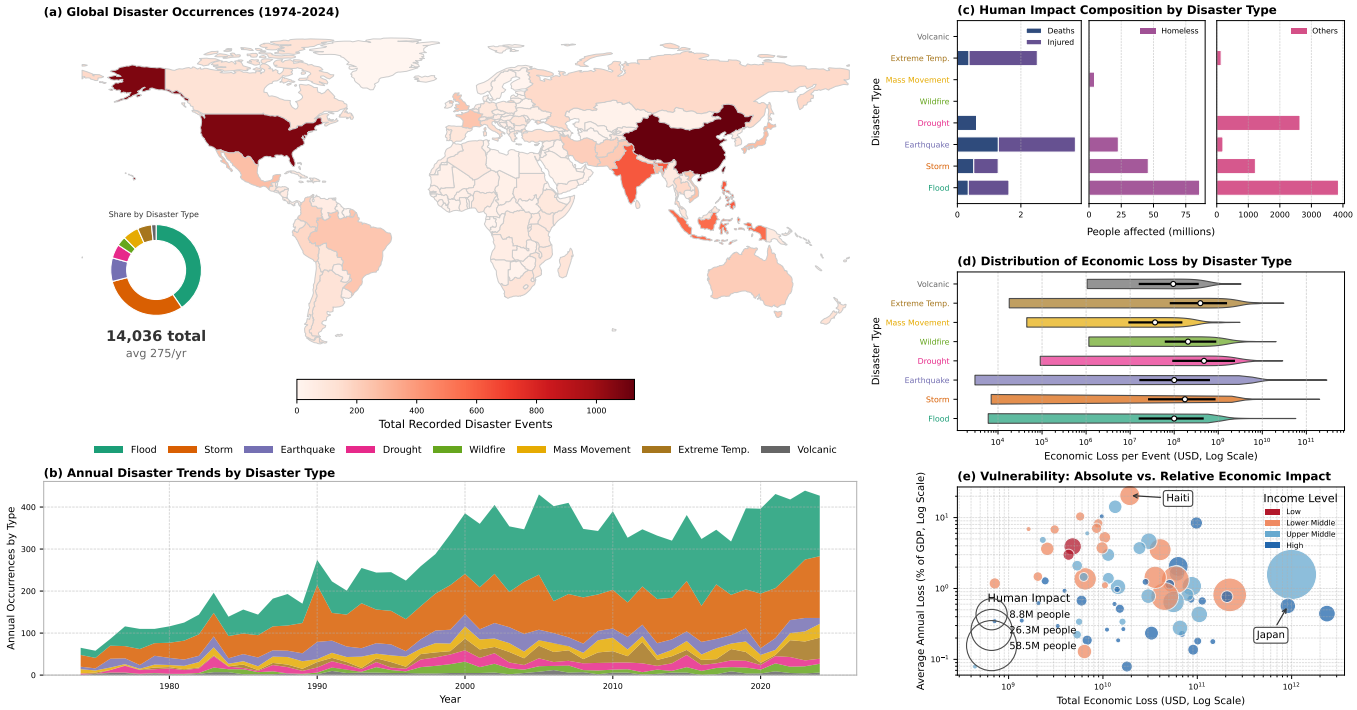

Fig. 1. Global natural disasters, impacts, and vulnerability (1974–2024). (a) Global distribution of disaster occurrences. (b) Annual trends by hazard type. (c) Human impacts across hazards. (d) Economic loss per event (log scale). (e) National vulnerability: total losses vs. annual loss as %GDP. *Data Sources:* EM-DAT [87], World Bank [493].

economic loss relative to GDP, and undermine progress toward the Sustainable Development Goals (SDGs), particularly poverty reduction (SDG 1), resilient infrastructure (SDG 11), and climate adaptation (SDG 13) [441], [443], [447]. In this context, rapid, accurate, and scalable post-event mapping is not merely a technical objective, but a critical enabler of timely response, recovery planning, and longer-term disaster risk reduction.

In the aftermath of a major disaster, ground-based information systems often fail: communication networks collapse, roads are blocked, and on-site assessment becomes slow or impossible. Earth observation (EO), here encompassing spaceborne, airborne, and Unmanned Aerial Vehicle (UAV) platforms, fills this gap by providing synoptic, repeatable, and objective measurements at scale without placing responders at risk. Over the past two decades, EO has undergone a data revolution, transitioning from sporadic acquisitions to a persistent, multi-modal global monitoring capability. This now includes sub-metre optical constellations (e.g., WorldView, Pléiades), all-weather SAR systems (e.g., Sentinel-1, TerraSAR-X, Capella), and high-revisit missions (e.g., Sentinel constellations, Planet). In addition, UAVs complement these with on-demand, centimetre-scale acquisitions. As Fig. 2 illustrates, the resulting mapping tasks span vastly different damage signatures, from landscape-scale flood inundation and burn scars to discrete structural failures in urban settings, across multiple sensor modalities.

The proliferation of EO data, coupled with advances in computing, has driven rapid methodological evolution in disaster mapping. Early approaches relied on pixel-level physical properties such as spectral indices and thresholding [74]. Classical machine learning then introduced handcrafted feature engineering to improve classification in specific contexts. The current paradigm is dominated by deep learning (DL): convolutional neural networks (CNNs) have automated feature extraction [571], Transformers and state-space models (e.g., Mamba) have enabled global context modeling [59], [230], and foundation models now target broad generalization across tasks and sensors [45], [168], [184], [463], [468].

This rapid progress, however, has outpaced systematic evaluation. New models are published at a rapid pace, often benchmarked on small or proprietary datasets with inconsistent protocols. While the broader EO community has begun to address this through standardized benchmarks such as PANGAEA [279], disaster mapping specifically still lacks a comparable large-scale, cross-method evaluation framework. Existing reviews in this domain have provided useful overviews but remain limited in scope: most focus on a single hazard type (e.g., floods [565] or earthquakes [99]), a single modality (e.g., SAR [127], [565]), or exclusively on deep learning methods [476], without bridging across these dimensions. Although some general EO benchmarks include individual disaster-relevant datasets such as flood or burn scar mapping [279], [420], they treat disaster as one application among many rather than systematically addressing the diversity of hazard types, the extent-damage task distinction, cross-modal operational settings, and the full methodological spectrum from classical approaches to foundation models. Key questions specific to this domain therefore remain open. Which methods generalize reliably across hazard types, sensors, geographic regions, and operational constraints? How do performance patterns change between SAR, optical, and multimodal settings? What trade-offs emerge among accuracy, computational cost, and deployment readiness? Without such domain-specific bench-

marks, selecting the right method for operational disaster response remains an ad hoc decision.

This review goes beyond existing surveys by providing not only a systematic taxonomy but also a large-scale quantitative benchmark of disaster mapping methods. We argue that such an empirical analysis is essential to move the field from methodological proliferation toward operational maturity. We first define a taxonomy of disaster mapping tasks, distinguishing between disaster extent mapping and infrastructure damage mapping. Subsequently, we benchmark over 30 methods, spanning traditional remote sensing techniques, CNNs, Transformers, and foundation models, under a unified evaluation protocol across 13 datasets covering diverse hazards, sensors, and mapping tasks. The primary contributions are:

- A systematic review and taxonomy from traditional to foundation model paradigms, accompanied by a consolidated resource of datasets and codebases to support reproducible research.
- The first large-scale benchmark dedicated to disaster mapping that evaluates over 30 methods spanning traditional, deep learning, and foundation model paradigms across distinct hazard types, mapping tasks, and sensor modalities under a unified protocol, providing an evidence-based guide for method selection.
- An empirically grounded roadmap identifying persistent challenges and priorities for developing disaster mapping solutions that are robust, scalable, and aligned with global resilience objectives.

2 Scope and Taxonomy of Disaster Mapping Tasks

This review covers major natural hazards, including earthquakes, tsunamis, floods, wildfires, severe storms, volcanic eruptions, and landslides, as well as anthropogenic disasters such as armed conflicts and large-scale explosions. The latter are included because their damage signatures in the built environment are functionally analogous to those of natural hazards and are addressed by the same class of methods. We focus exclusively on post-event mapping: methods that characterize the physical manifestations of an disaster event from EO data, namely its spatial extent and the resulting damage to built assets. Pre-event tasks such as susceptibility modelling [356] or risk assessment [7], during-event monitoring, and niche post-event applications such as population displacement estimation [265] or infrastructure accessibility analysis [214] are not covered. Similarly, upstream preprocessing steps essential for operational pipelines, such as atmospheric correction [454], [455], cloud and shadow screening [342], [579], [580], and image co-registration [112], are prerequisites for the mapping methods reviewed here but are not themselves covered.

Our literature search mainly covers peer-reviewed articles and major conference proceedings published between 2005 and 2025, drawn from leading venues in remote sensing (e.g., *RSE*, *ISPRS J P&RS*, *IEEE TGRS*) and computer vision (e.g., *CVPR*, *ICCV*). Search terms combined "disaster mapping", "remote sensing", and "Earth observation" with specific hazard types. The initial search returned approximately 25,000 records. Fig. 3 shows a twelve-fold increase in annual output over this period, with growth accelerating markedly after 2018, coinciding with the maturation of deep learning frameworks and the expanding availability of open satellite imagery. According to Fig. 3, flood-related studies account for the largest share (25%), followed by landslides (16%), and earthquakes (12%), broadly reflecting the global frequency and data availability for each hazard type.

Organizing this review by hazard type alone would obscure the fact that underlying algorithmic challenges often transcend specific disasters. Delineating a flood boundary is functionally similar to mapping a burned area or a lava flow: all are segmentation tasks over amorphous regions. Likewise, assessing building damage from an earthquake and from an armed conflict share the same objective. We therefore structure this review around a functional taxonomy of mapping tasks. Based on our survey, the large majority of EO-based post-disaster mapping efforts fall into two categories:

- **Disaster extent mapping** concerns the spatial footprint of a hazard's impact on the landscape, such as flood inundation, wildfire burned areas, or landslide deposits. The core task is semantic segmentation of the affected area.
- **Infrastructure damage mapping** concerns damage to the built environment, such as building collapse after earthquakes [5], [58], road and bridge damage [346], or structural destruction from conflict [97]. The core task is detection and classification of discrete damaged objects.

This task-driven taxonomy enables more direct comparison of methods, since a model's suitability depends primarily on the nature of the mapping target (landscape-scale extent versus discrete infrastructure) rather than the specific hazard that caused the damage. In practice, these two task families are not strictly disjoint. The increasing availability of global building footprint datasets means that per-building damage assessment can be approximated by performing landscape-scale semantic segmentation and intersecting the result with existing footprints [283], [292]. Nevertheless, the distinction remains methodologically meaningful: dedicated damage mapping architectures explicitly model object-level context, multi-level damage grading, and structural cues that a simple intersection cannot capture, particularly for partial damage categories or buildings absent from existing inventories.

It is also worth noting that disaster extent mapping can be viewed as a special case of abrupt land disturbance detection, a framing that connects it to continuous change monitoring research [581]. While a full treatment of continuous change detection is beyond our scope, the conceptual link is relevant. The same principles of baseline modelling and anomaly detection underpin both ecological disturbance monitoring and post-disaster extent mapping.

3 Progress in Data for Disaster Mapping Tasks

A review of methods is incomplete without examining the data that drive their development and evaluation. In the

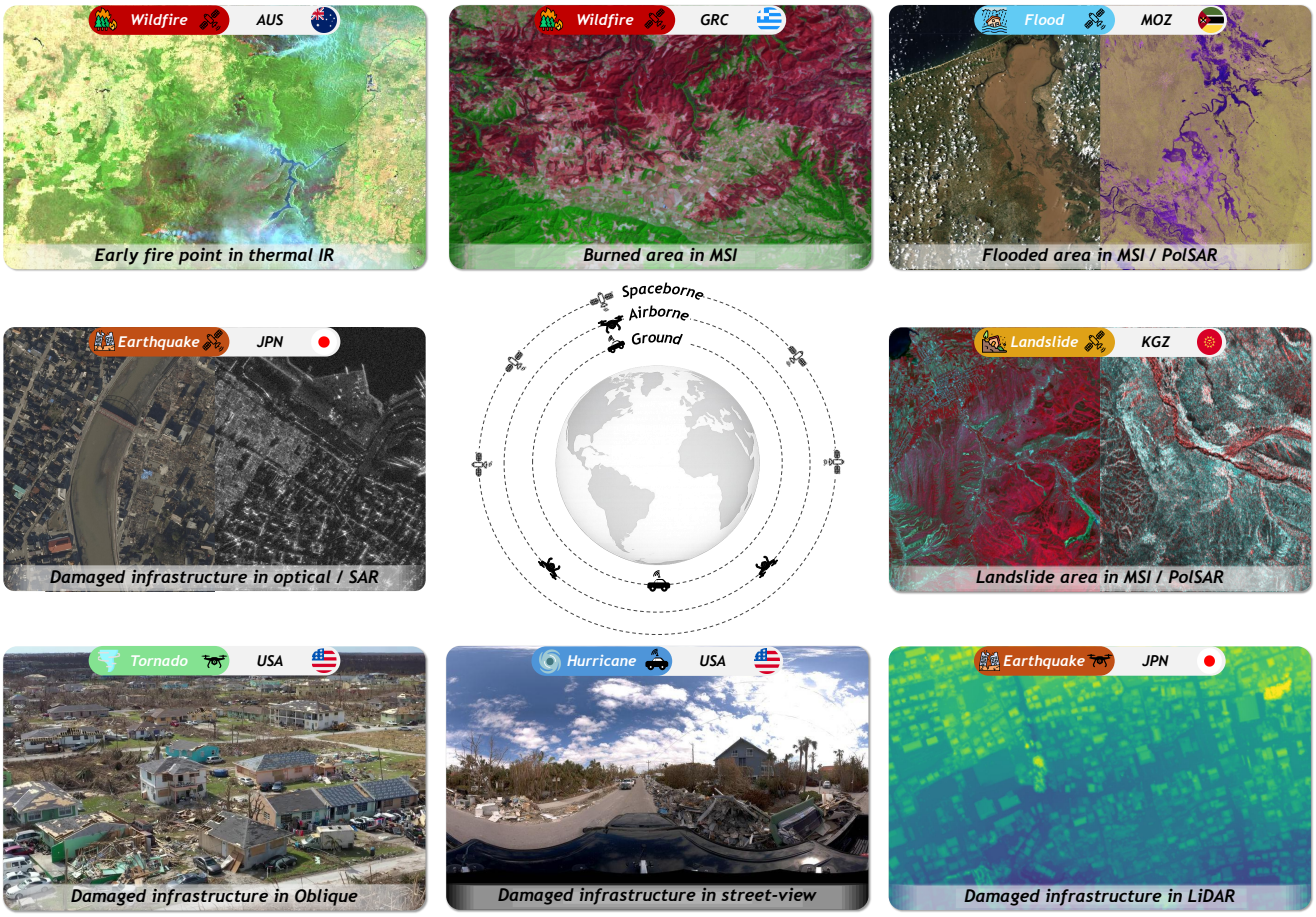

Fig. 2. The diversity of disaster mapping as observed by multi-platform, multi-modal Earth observation. The figure illustrates representative hazard types and their characteristic signatures across three observation scales: spaceborne, airborne, and ground-level. Disaster extent mapping tasks include early fire detection via thermal infrared anomalies, burned area delineation from multi-spectral imagery (MSI), and flood inundation mapping from MSI and polarimetric SAR. Infrastructure damage mapping tasks (middle and bottom rows) span nadir-view satellite observations in optical and SAR, oblique UAV photography, street-level panoramic imagery, and airborne LiDAR point clouds. Country codes indicate the event location. *Image sources:* Thermal IR and MSI wildfire imagery © ESA/Sentinel; MSI/PolSAR flood imagery © ESA/Sentinel; optical/SAR earthquake imagery is from [58]; MSI/PolSAR landslide imagery © ESA/Sentinel; oblique UAV tornado imagery is from [69]; street-view hurricane imagery is from [224]; LiDAR earthquake imagery is from [72].

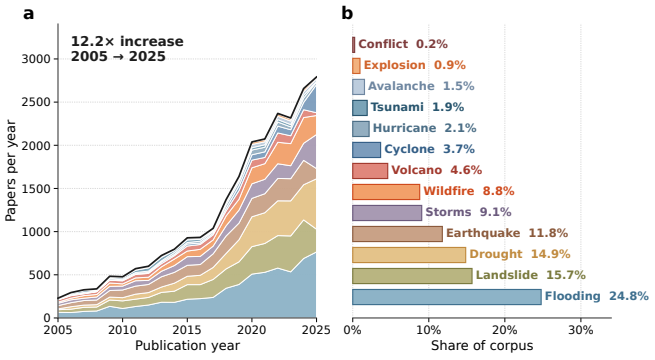

Fig. 3. Number of publications in the last two decades based on the Web of Science Core Collection database and obtained by searching keywords: “Earth observation”, “remote sensing”, “post-disaster mapping”, and different disaster types.

deep learning era, large-scale annotated datasets are central to both architectural innovation and performance assessment. Rather than attempting an exhaustive catalogue, we focus on the de facto benchmarks and dominant data sources that define current performance baselines for each task. The discussion follows the task taxonomy established in Section 2.

3.1 Disaster Extent Mapping

Disaster extent mapping aims to delineate the spatial footprint of hazard impacts on natural and semi-natural surfaces. We organize the dataset review by hazard category, as this aligns with how researchers typically search for benchmarks, while noting that many hazards that are geophysically distinct produce similar observable signatures in EO data. We present dedicated subsections for flood, wildfire, and mass-movement-related geomorphic change, which together dominate current benchmark development for disaster extent mapping. Hazards whose primary EO formulation is atmospheric monitoring [355], [559] or slow-onset environmental assessment [197], [297], [423], such as storms and droughts, are not reviewed as separate extent-mapping categories here because they do not typically produce a discrete post-event surface footprint comparable

TABLE 1
Representative open disaster extent mapping datasets. Within each hazard type, datasets are grouped by dominant modality or task regime and sorted by publication year.

Hazard type	Dataset	Modality	Sensor/Platform	GSD (m/px)	Temporal	Events	Tiles
Flood	GLOBALFLOODDATABASE [426]	Multispectral	MODIS	250	Multitemporal	913	12,719
	WORLDFLOODS [284]	Multispectral	Sentinel-2	≥10	Post-event	119	185,574
	MIDWESTERN USA [188]	Multispectral	Planet	3	Multi-temporal	1	500
	SEN1FLOODS11 [39]	SAR	Sentinel-1	10	Post-event	11	4,831
	UNOSAT [312]	SAR	Sentinel-1	10	Bi-temporal	N/A	58,128
	ETCI2021 [309]	SAR	Sentinel-1	5x20	Post-event	N/A	66,810
	S1GFLOODS [370]	SAR	Sentinel-1	10	Bi-temporal	46	5,360
	URBANSARFLOODS [567]	SAR	Sentinel-1	20	Bi-temporal	18	8,879
	SEN12-FLOOD [348]	Multispectral, SAR	Sentinel-1, Sentinel-2	≥10	Multi-temporal	N/A	336
	OMBRIA [100]	Multispectral, SAR	Sentinel-1, Sentinel-2	10	Bi-temporal	23	1,688
	GF-FLOODNET [560]	Multispectral, SAR	GF-2, GF-3	1.5–5	Bi-temporal	8	13,388
	CAU-FLOOD [159]	Multispectral, SAR	Sentinel-1, Sentinel-2	10	Bi-temporal	18	18,302
	FLOODPLANET [561]	Multispectral, SAR	PlanetScope, Sentinel-1, Sentinel-2	3–10	Post-event	19	366
	STURM-FLOOD [319]	Multispectral, SAR	Sentinel-1, Sentinel-2	10	Post-event	60	24,277
	MMFLOOD [304]	SAR, DEM	Sentinel-1, DEM	20	Post-event	95	8,522
	KURO SIWO [42]	SAR, DEM	Sentinel-1, SRTM	≥10	Bi-temporal	43	67,490
	DFC2024 [332]	Multispectral, SAR, DEM, land-cover map	Sentinel-1, Sentinel-2, Landsat series, DEM	10–90	Post-event	N/A	2,888
	DEEPFLOOD [111]	RGB, SAR	UAV, Sentinel-1	0.015–10	Post-event	4	20,593
Wildfire	MCD64A1 C6 [136]	Optical	MODIS Terra/Aqua	500	Multi-temporal	Global	N/A
	S2-WCD [449]	Multispectral	Sentinel-2 L1C	10	Bi-temporal	25	41 pairs
	FLOGA [377]	Multi-modal	Sentinel-2, MODIS	20	Bi-temporal	326	1,167
	MODIS MOD14 [200]	Thermal	MODIS Terra	1,000	Multi-temporal	Global	N/A
	VIIRS 375M AF [376]	Thermal	VIIRS S-NPP	375	Multi-temporal	Global	N/A
	HIMAWARI-8 AHI AF [514]	Thermal	Himawari-8 AHI	2,000	Multi-temporal	Asia–Pacific	N/A
	SENTINEL-3 SLSTR AF [513]	Thermal	Sentinel-3 SLSTR	∼1,000	Multi-temporal	Global	N/A
	LANDSAT-8 ACTIVE FIRE [92]	Optical, Thermal	Landsat-8 OLI/TIRS	30	During-event	Global	146,214
	WILDFIREDB [400]	Multi-modal	VIIRS, LANDFIRE	375	Multi-temporal	N/A	2,367,209
	NDWS [176]	Multi-modal	MODIS, GRIDMET, VIIRS, SRTM	1,000	Multi-temporal	N/A	18,545
	WILDFIRESPREADTS [131]	Multi-modal	VIIRS, MODIS, GRIDMET, GFS, SRTM	375	Multi-temporal	607	13,607
	SEASFIRE [202]	Multi-modal	ERA5, MODIS, CAMS, FireCCI, GFED, NOAA	∼27,500	Multi-temporal	Global	N/A
	WSTS+ [219]	Multi-modal	VIIRS, MODIS, GRIDMET, GFS, SRTM	375	Multi-temporal	1,005	24,462
	FIRE DETECTION 360 [29]	RGB (Omnidirectional)	UAV	N/A	During-event	N/A	150
	FLAME [383]	RGB, thermal	UAV	N/A	During-event	1	2,003 seg.
	BURNEDAREAUAV [359]	RGB	DJI Phantom 4 Pro	∼0.03	During-event	1	249
Landslide	LANDSLIDE4SENSE [134]	Multispectral, DEM	Sentinel-2, ALOS PALSAR	10	Post-event	4	4,844
	HR-GLDD [293]	Multispectral	PlanetScope	4	Post-event	13	1,758
	GVLM [558]	RGB	Google Earth	0.59	Bi-temporal	17	7,596
	CAS LANDSLIDE [516]	Optical	WorldView2/3, Sentinel-2, PlanetScope, SuperView-1, GF-1, Landsat, UAV	0.5–5	Post-event	9	20,865
	MRGSLD [104]	Multispectral, DEM	Sentinel-2, ALOS PALSAR	10	Post-event	21	5,995
	LANDREF [323]	Multispectral, SAR, DEM	Sentinel-1, Sentinel-2, TanDEM-X	10	Bi-temporal	28	5,995

to inundation, burned area, or mass-movement deposits[2]. Tables 1 collate key open-access benchmarks discussed in this subsection.

3.1.1 Surface Water and Inundation

Flood inundation mapping, encompassing riverine floods, flash floods, coastal storm surges, and tsunami inundation, benefits from the most mature data ecosystem among disaster extent tasks. This maturity reflects two factors: the availability of scalable open EO data (particularly Sentinel-1 and Sentinel-2), and the relatively distinct spectral and backscatter signature of open water, which enables consistent labelling across diverse regions.

2. This does not imply that EO data are unimportant for storms or droughts. Storm-related research often relies on meteorological satellites and radar for monitoring and nowcasting, while drought studies commonly use vegetation indices, soil moisture, and hydro-meteorological data cubes for continuous monitoring and forecasting. These formulations are adjacent to, but distinct from, the post-event spatial footprint mapping tasks reviewed in this subsection.

Early benchmarks were built on medium-resolution optical archives. WorldFloods [284], for example, provides 10 m Sentinel-2 imagery and labels for 119 flood events worldwide and remains a common reference for cross-region evaluation. Because major floods frequently coincide with cloud cover and adverse illumination, SAR-based benchmarks have become essential for rapid response mapping [462]. A growing number of datasets are built around Sentinel-1, spanning post-event collections [39], [298], [304] and paired pre/post-event resources [312], [370], [567]. Among these, URBANSARFLOOD [567] is notable for introducing an urban-oriented label taxonomy (*non-flooded*, *flooded open area*, *flooded urban area*), reflecting the distinct observational characteristics of inundation in built-up environments.

The current frontier has shifted from single-modality archives to analysis-ready multi-source data cubes that address observational gaps through two strategies. Multi-sensor fusion combines distinct EO modalities (typically optical and SAR) to ensure all-weather coverage. Phys-

ical and semantic augmentation incorporates static priors such as DEMs, hydrography, and land-cover maps to constrain water accumulation and distinguish transient flooding from permanent water bodies. Representative multi-source datasets include SEN12-FLOOD [348], OMBRIA [100], FLOODPLANET [561], CAU-FLOOD [159], and competition-style releases [332]. Several benchmarks also refine label semantics for cross-region portability. For instance, KUROSIWO [42] introduces a dedicated permanent water class to decouple long-standing water bodies from transient inundation. To bridge the resolution gap in complex urban scenarios, a small number of datasets leverage VHR optical or UAV imagery as high-fidelity validation sources [111], [347], complementing the dominant satellite-based baselines.

3.1.2 Burned/Disturbed Land Surface

Wildfire mapping differs from most other hazards in that its observable signature evolves through distinct phases, giving rise to three dataset regimes: (i) early-stage active fire detection based on thermal anomalies, (ii) burned area footprint mapping based on surface reflectance change, and (iii) short-horizon fire progression mapping[3] using multi-temporal data cubes that combine EO with environmental covariates.

For early-stage detection, the primary signal is intense thermal emission. Because fires are highly dynamic, temporal resolution is prioritized over spatial detail. Most benchmarks in this regime are anchored to global operational products from moderate-resolution sensors, including MODIS [200], VIIRS [376], and Himawari-8 [514], which provide daily monitoring through contextual brightness-temperature thresholding. Complementary thermal products from other missions further extend global coverage and detection sensitivity, e.g., Sentinel-3 SLSTR [513]. Landsat-8 extends detection to 30 m resolution by combining reflectance and thermal bands [92], while SEN2FIRE [515] integrates Sentinel-2 multispectral imagery with Sentinel-5P aerosol products. At finer scales, UAV-based benchmarks such as FLAME [383] support flame-level detection that coarse sensors cannot resolve.

Once the fire front passes, the task shifts to delineating burned scars from surface reflectance changes, with NIR and SWIR bands being particularly informative. Medium-resolution optical archives (Sentinel-2, Landsat) constitute the primary data source. Early datasets were limited to single events [266], [556], but several large-scale benchmarks with expert-annotated pixel-level masks have since emerged [282], [377], [449], and automatic global burned-area mapping algorithms have been developed using Sentinel-2 and active fire data [30]. Complementary UAV datasets and single-event studies also support method development in specific scenarios [359], [434], [450]. At global scale, long-running burned-area products [136] are frequently used as proxy ground truth for training and evaluation. When smoke or cloud cover degrades optical observations during active burning, SAR provides a practical alternative, with Sentinel-1 serving as the most accessible platform and longer-wavelength systems (e.g., ALOS-2) offering complementary sensitivity to vegetation-structure change [1]. However, SAR-based wildfire benchmarks remain less standardized than their optical counterparts, with most studies limited to small numbers of events [28], [266], [556], [557].

Since wildfire is a sustained, rapidly evolving process, an important task extends beyond scar delineation to forecasting short-horizon progression (e.g., next-day spread). This requires multi-temporal data cubes rather than single images. The Next Day Wildfire Spread (NDWS) dataset [176] is the foundational benchmark, providing temporally indexed samples over numerous U.S. fire events with multi-source covariates. Subsequent datasets have expanded toward richer time-series inputs and more explicit multi-modal conditioning [131], [219], [541], while analysis-ready datacube products consolidating climate, land-surface, and fire-product variables support next-day to seasonal-scale forecasting [202], [335].

3.1.3 Mass Movement Deposits and Geomorphic Change

This category covers hazards that displace and deposit material across the landscape, including landslides, volcanic eruptions (lava flows, pyroclastic deposits), and snow avalanches. Unlike water or fire, which produce distinct spectral or thermal anomalies, these hazards are identified primarily through geomorphic and textural disruption. The displaced material often shares a similar spectral signature with surrounding bare ground or snow, making detection dependent on topographic context and structural change rather than spectral contrast. Among the three, landslides have received the most attention in the EO community and are the only hazard type with established large-scale benchmarks.

Benchmark construction first matured around medium-resolution optical archives (Sentinel-2 class), where wide coverage enables multi-event curation but limits the minimum mappable landslide size [52]. LANDSLIDE4SENSE [134] is a milestone benchmark in this regime, providing globally distributed events with co-registered optical and topographic layers (DEM, slope). MRGSLD [104] further expands geographic scope, enabling evaluation of cross-region generalization. To capture smaller slope failures, several benchmarks leverage VHR optical imagery, with HR-GLDD [293] and GVLM [558] being prominent examples. While optical data remain dominant, recent resources emphasize paired optical-SAR availability. For instance, the Landslide Reference Data [323] curates masks with co-temporal Sentinel-1 and Sentinel-2 coverage, supporting benchmarking under all-weather constraints. Deformation-sensitive modalities such as InSAR time series [53], [477] and airborne LiDAR [182], [453] provide complementary evidence for slow-moving slopes and canopy-obscured failures, but these resources are typically released as event-specific collections rather than standardized open benchmarks, limiting their role in community-wide evaluation.

Volcanic eruptions and snow avalanches produce analogous deposit signatures in EO data (lava flows, pyroclastic

3. Although this review is centered on post-event mapping, we include a small number of next-day wildfire extent prediction studies as a boundary case because they are methodologically adjacent to EO-based extent mapping.

TABLE 2
Representative open infrastructure damage mapping datasets. Within each platform group, datasets are sorted by publication year.

Platform	Dataset	Modality	GSD (m/px)	Temporal	Disaster type	Events	Tiles	Target (Damage level)
Spaceborne	ABCD [120]	RGB	0.4	Bi-temporal	Tsunami	1	10,777	Building (2 grades)
	ADBD [65]	RGB	0.5	Post-event	Hurricane	1	3,600	Building (2 or 5 grades)
	xBD [151]	RGB	<0.8	Bi-temporal	Earthquake, tsunami, wildfire, flood, storm, volcano	19	11,034	Building (4 grades)
	SpaceNet-8 [178]	RGB	0.3–0.8	Bi-temporal	Flood	3	1,207	Building (2 grades), road (2 grades)
	HaitiBRD [244]	RGB	0.5	Bi-temporal	Earthquake	1	20	Building (4 grades), road (2 grades)
	Ida-BD [205]	RGB	0.5	Bi-temporal	Hurricane	1	87	Building (4 grades)
	QQB [416]	RGB, SAR	0.35	Bi-temporal	Earthquake	1	4,029	Building (2 grades)
	EBD [482]	RGB	0.3–0.5	Bi-temporal	Earthquake, flood, storm, volcano	12	18,215	Building (4 grades)
	Bright [58]	RGB, SAR	0.3–1.0	Bi-temporal	Earthquake, wildfire, flood, storm, volcano, armed conflict, explosion	14	4,246	Building (3 grades)
Airborne	DoriaNet [69]	RGB	N/A	Post-event	Hurricane	1	2,409	Building (5 grades)
	FloodNet [347]	RGB	0.015	Post-event	Flood	1	2,343	Building (2 grades), road (2 grades)
	ISBDA [578]	RGB	N/A	Post-event	Hurricane, tornado	6	1,030	Building (3 grades)
	Flood-3i [193]	RGB	0.05	Post-event	Flood	1	17,055	Building (2 grades), road (2 grades)
	RescueNet [346]	RGB	N/A	Post-event	Hurricane	1	4,494	Building (4 grades), road (2 grades)
	Crasar-U-Droids [277]	RGB	0.02–0.13	Post-event	Hurricane, tornado, volcano, wildfire, tower collapse	10	18,780	Building (4 grades)
Ground	CrisisMMD [13]	RGB, text	N/A	Post-event	Earthquake, hurricane, wildfire, flood	7	16,097	Building (3 grades), road (3 grades), bridge (3 grades)
	MediaEval-2018 [36]	RGB	0.3–30	Post-event	Hurricane, flood	3	12,763	Road (2 grades)
	CVIAN [224]	RGB	N/A	Post-event	Hurricane	1	4,121	Building (3 grades)
	Milton [532]	RGB	N/A	Bi-temporal	Hurricane	1	2,249	Building (3 grades)

deposits, debris fans, and avalanche runout zones) and are detected through similar geomorphic and textural change cues. However, both remain data-scarce relative to landslides. For volcanic mapping, most studies rely on event-specific optical or SAR acquisitions without standardized community benchmarks [51], [85]. Avalanche mapping has progressed through optical inventory construction and SAR-based debris detection [47], [102], [201], but existing datasets are regional and not yet consolidated into open, community-wide benchmarks. This absence of established benchmarks for both hazard types represents a notable gap that limits systematic method development and comparison.

3.2 Infrastructure Damage Mapping

While disaster extent mapping focuses on delineating the perimeter of an affected area, infrastructure damage mapping demands finer-grained analysis, typically at the level of individual buildings or critical facilities. Building damage mapping dominates this category, given its direct relevance to life-saving operations and emergency response [99], [127]. The datasets supporting this task are best understood by platform, each corresponding to a different spatial scale and assessment capability: satellite for wide-area coverage, airborne for high-fidelity detail, and ground-level for facade-visible damage. Table 2 summarizes current open-access benchmarks.

Satellite imagery is the most widely used source for rapid, large-area damage assessment. Optical VHR sensors are the natural starting point, as they depict scene reflectance in a human-interpretable way that supports object-level damage analysis. Early datasets were typically limited to single events [65], [186], [190], [228], [244], [430]. The release of xBD [151] marked a major milestone as the first large-scale, multi-hazard benchmark providing matched pre- and post-event VHR optical images with building footprints and multi-class damage labels. xBD catalysed rapid progress in deep learning for bi-temporal damage mapping, and subsequent datasets such as SpaceNet-8 [178] and EBD [482] have further broadened coverage. However, optical sensors are rendered ineffective by cloud cover, smoke, or darkness, conditions that are common in the immediate aftermath of disasters. This operational gap has driven growing interest in SAR, whose all-weather, day-night capability is well suited to rapid post-event assessment. At medium resolution, Sentinel-1 supports regional damage estimation [73], [97], [204], [287], while VHR SAR satellites (e.g., TerraSAR-X, COSMO-SkyMed, Capella) increasingly enable object-level analysis. Historically, high cost and restrictive licensing limited VHR SAR studies to isolated case studies [26], [64], [94], [211], [300], and optical data remained superior for precise building localization [5], [503]. The emergence of more open data policies from new-generation SAR providers, together with these

complementary strengths, has motivated the construction of multimodal benchmarks [4], [5], [58], [416]. The BRIGHT dataset [58] is the key benchmark in this category: the first open-access, globally distributed, multi-hazard dataset providing matched VHR optical-SAR pairs for building damage mapping.

Airborne platforms, particularly Unmanned Aerial Vehicles (UAVs), provide ultra-high-resolution imagery (<0.1 m) from closer vantage points. Crucially, oblique photography captures building facades, revealing structural damage such as wall cracks and partial collapses that are invisible from nadir satellite views [208], [278]. Event-specific datasets including FLOODNET [347] and RESCUENET [346] serve as key benchmarks, while recent multi-event collections [277], [333], [334], [578] signal a move toward broader coverage. Airborne LiDAR complements optical UAV data by providing precise 3D measurements that can directly quantify building height and volume changes, making it uniquely effective for identifying collapse types missed by 2D analysis [285]. However, high acquisition costs limit LiDAR to post-hoc fine-grained assessments rather than rapid response; its coverage is concentrated in high-income countries, and no standardized large-scale LiDAR benchmark currently exists [119], [153], [187], [220], [226], [458], [480], [575].

Ground-level data, primarily street-view imagery, provide detailed perspectives on facade and interior damage that are invisible from overhead platforms. This makes them highly complementary to satellite and airborne views, particularly for light-to-moderate damage assessment and long-term recovery monitoring [224], [507]. Data sources range from structured platforms such as Google Street View[4] and Mapillary[5] to unstructured social media imagery used for rapid real-time analysis [8], [13], [317]. However, existing benchmarks remain small-scale and typically cover single events in developed countries [224], [518], [532], [544].

A critical distinction from extent mapping is the need for multi-level damage classification. Rather than binary labelling (e.g., *flooded* vs. *not flooded*), infrastructure damage mapping requires categorizing structural condition across a graded scale (e.g., *no damage*, *minor damage*, *major damage*, *destroyed*). A significant challenge is the absence of a universally adopted standard. Multiple scales coexist, from engineering-focused systems such as the European Macroseismic Scale (EMS-98) [144] and the AIJ scale [320] to simplified 3-to-5-level schemes developed for EO benchmarks [5], [58], [151]. This diversity complicates cross-dataset comparison and model generalization. The annotation process itself introduces further inconsistency: damage interpretation depends strongly on sensor resolution and viewing angle. Manzini *et al.* showed that satellite-derived labels systematically under-report damage, particularly for *minor damage* categories, which are often invisible from a top-down perspective but clearly visible in UAV imagery, with inter-platform disagreement reaching 29% [278]. This highlights a fundamental challenge in establishing unified ground truth for multi-platform damage assessment.

It is worth noting that global building footprint datasets have reached a maturity that enables their routine use as ancillary inputs. Products from OpenStreetMap [321], Microsoft Building Footprints [294], Google Open Buildings [401], and Overture Maps [324] now cover large parts of the inhabited world, providing the geometric priors that underpin both footprint-intersection pipelines and object-level post-processing in damage mapping workflows.

3.3 Summarization

According to the preceding review, as well as the two dataset summary tables (Table 1 and Table 2), disaster extent mapping datasets are more often organized around hazard-specific observables and sensing regimes, whereas infrastructure damage datasets increasingly adopt hazard-spanning designs centered on a shared object-level damage target. Table 3 summarizes this benchmark landscape by linking each hazard to its primary task, dominant modalities, typical resolution regime, and benchmark maturity. The resulting picture is uneven: floods are supported by multiple large-scale open datasets, whereas volcanic deposits, avalanches, and armed conflict remain comparatively benchmark-scarce, limiting systematic method development and cross-method comparison for these hazards.

Across these benchmarks, dataset design is primarily driven by the effective mapping scale and the observability of the target, leading to distinct sensor regimes for the two task families. Disaster extent mapping is most commonly supported by medium-resolution archives (10–30 m), typically Sentinel-1/2 and Landsat, where wide coverage and frequent revisit enable scalable multi-event curation and the key requirement is separating a surface state rather than resolving fine geometry. Infrastructure damage mapping is predominantly benchmarked with VHR observations (often <1 m) from commercial satellites or airborne platforms, where object-level interpretability is essential. Recent damage benchmarks tend to be hazard-agnostic, encompassing earthquakes, storms, and fires within a single dataset, because reliable damage assessment depends on spatial resolution and annotation granularity rather than hazard-specific spectral signatures. These regimes are not rigid: VHR data are increasingly used for boundary-critical extent mapping in complex settings, while moderate-resolution data can support coarse damage screening, suggesting that resolution is best understood as a dataset design choice rather than a strict task-level constraint.

Despite these differences, benchmark development across both tasks shows three converging trends:

- From single-event case studies to multi-event, geographically diverse collections with standardized splits, enabling more meaningful evaluation under cross-region and cross-event domain shift.
- From single-sensor inputs to analysis-ready multi-source stacks that mitigate clear-sky limitations and contextual ambiguity. Optical-SAR pairing is now common for all-weather response, while many extent benchmarks treat imagery combined with terrain or thematic priors as a canonical input configuration.
- From static snapshots to temporally indexed data that include pre-event baselines or time-series stacks, enabling change-aware mapping and, in domains

4. https://www.google.com/streetview/
5. https://www.mapillary.com

TABLE 3
Typical EO sensor–task regimes across major disaster scenarios. The table summarizes dominant settings reported in the literature rather than strict rules; exceptions and hybrid configurations exist, especially in urban, multimodal, or high-resolution studies. Benchmark maturity is indicated as ● ● ● (multiple large-scale open benchmarks), ●● (some benchmarks with limited coverage), and ● (few standardized benchmarks).

Hazard	Primary observable	Primary task	Key modalities	Typical resolution	Typical temporal	Benchmark maturity
Flood	Water surface (specular reflection, NIR absorption)	Extent	SAR, multispectral	Medium (10–30 m)	Post / bi-temporal	● ● ●
Wildfire (active)	Thermal emission	Extent (detection)	Thermal IR	Coarse (375–1000 m)	Multi-temporal	● ● ●
Wildfire (burned area)	Surface reflectance change	Extent	Multispectral, SAR	Medium (10–30 m)	Bi-temporal	●●
Wildfire (spread)	Multi-source dynamics	Extent (prediction)	Multispectral, meteorology, DEM	Coarse (375–1000 m)	Multi-temporal	● ● ●
Landslide	Geomorphic disruption	Extent	Multispectral, DEM, SAR	Medium to VHR (0.5–10 m)	Post / bi-temporal	●●
Volcanic deposits	Surface deposition	Extent	Multispectral, SAR	Medium (10–30 m)	Post / bi-temporal	●
Avalanche	Snow deposit / surface roughness	Extent	SAR, multispectral	Medium (10–30 m)	Post / bi-temporal	●
Earthquake	Structural collapse	Damage	VHR optical, VHR SAR	VHR ($<$ 1 m)	Bi-temporal	●●
Tsunami	Coastal inundation + built damage	Damage + extent	VHR optical; SAR for inundation	VHR to medium	Bi-temporal / post	●
Storm / hurricane	Flooded surface + built damage	Damage + extent	VHR optical, SAR	VHR to medium	Bi-temporal / post-event	●●
Armed conflict	Structural destruction	Damage	VHR optical, VHR SAR	VHR ($<$ 1 m)	Bi-temporal	●

such as wildfire, short-horizon progression forecasting.

These evolving benchmarks have been instrumental in enabling the transition from manual interpretation to automated, data-driven disaster mapping, and have in turn driven the development of the approaches reviewed in the following sections.

4 Progress of Disaster Mapping Methodology

Methods for EO-based disaster mapping have evolved substantially over the past two decades. Early efforts relied on manual interpretation and sensor-specific, feature-engineered pipelines whose performance depended on carefully designed indices, texture descriptors, and heuristic rules. With the growth of benchmark datasets and computational resources, the field has shifted toward learning-based paradigms, culminating in end-to-end deep networks that jointly learn representations and task-specific outputs for segmentation, detection, and damage classification. We review methods along this trajectory.

4.1 Traditional Methods

Before the widespread adoption of deep learning, disaster mapping relied largely on a priori physical knowledge and expert experience. The most fundamental approach is visual interpretation, where trained analysts manually compare pre- and post-event imagery to identify disaster impacts. It remains a routine component of operational mapping, but is subject to inter-analyst variability and becomes resource-intensive at very large scales or under tight response windows [99]. This limitation drove the development of automated methods for both continuous disaster extent mapping and discrete infrastructure damage mapping. These methods are intrinsically tied to the sensor modality and, crucially, to the availability of pre-event data. Fig. 4 summarizes the core paradigms, workflows, and sensor-specific physical mechanisms underlying traditional disaster mapping.

4.1.1 *Bi-Temporal Methods*

The bi-temporal paradigm leverages both pre- and post-event acquisitions and is fundamentally a change detection task. Its core strength lies in the ability to directly measure physical differences in EO signals. Here, the meaning of "change" differs by task. For extent mapping, the objective is to isolate a surface-state transition (e.g., dry-to-wet, vegetated-to-burned), so pixel- or region-level change measures often suffice. For damage mapping, change must be attributed to specific assets (*e.g.*, buildings, roads) and interpreted as a damage state, which typically requires object localization, contextual constraints, and features sensitive to structural integrity.

For optical data, bi-temporal comparison identifies impacts through changes in spectral and textural properties. For disaster extent mapping, the standard approach exploits spectral contrasts that are diagnostic of specific surface-state transitions: water indices for flood inundation [18], [116], [194], [227], [291], [389], [511], burn-sensitive indices (e.g., dNBR) for wildfire severity [209], [295], [344], and vegetation-loss indicators for landslide delineation [82], [118], [528], [562]. While effective in clear-view conditions, these spectral approaches are sensitive to atmospheric contamination, shadow, and mixed land-cover, which has motivated a range of scene-specific refinements. For infrastructure damage, textural and contextual cues become central. Building collapse manifests as a transition from ordered, homogeneous textures to chaotic, heterogeneous patterns, and features capturing edge regularity, morphological structure, and shadow geometry have proven effective at distinguishing damaged from intact structures [71], [75], [177], [353], [366], [429], [431], [461], [553].

SAR offers all-weather, day-night imaging and provides complementary change signals rooted in microwave scattering physics. Bi-temporal SAR analyses are typically categorized into three types [127], [565]. Intensity-based methods quantify changes in backscatter amplitude. For flood mapping, open water acts as a specular reflector, producing a pronounced backscatter reduction [78], [137],

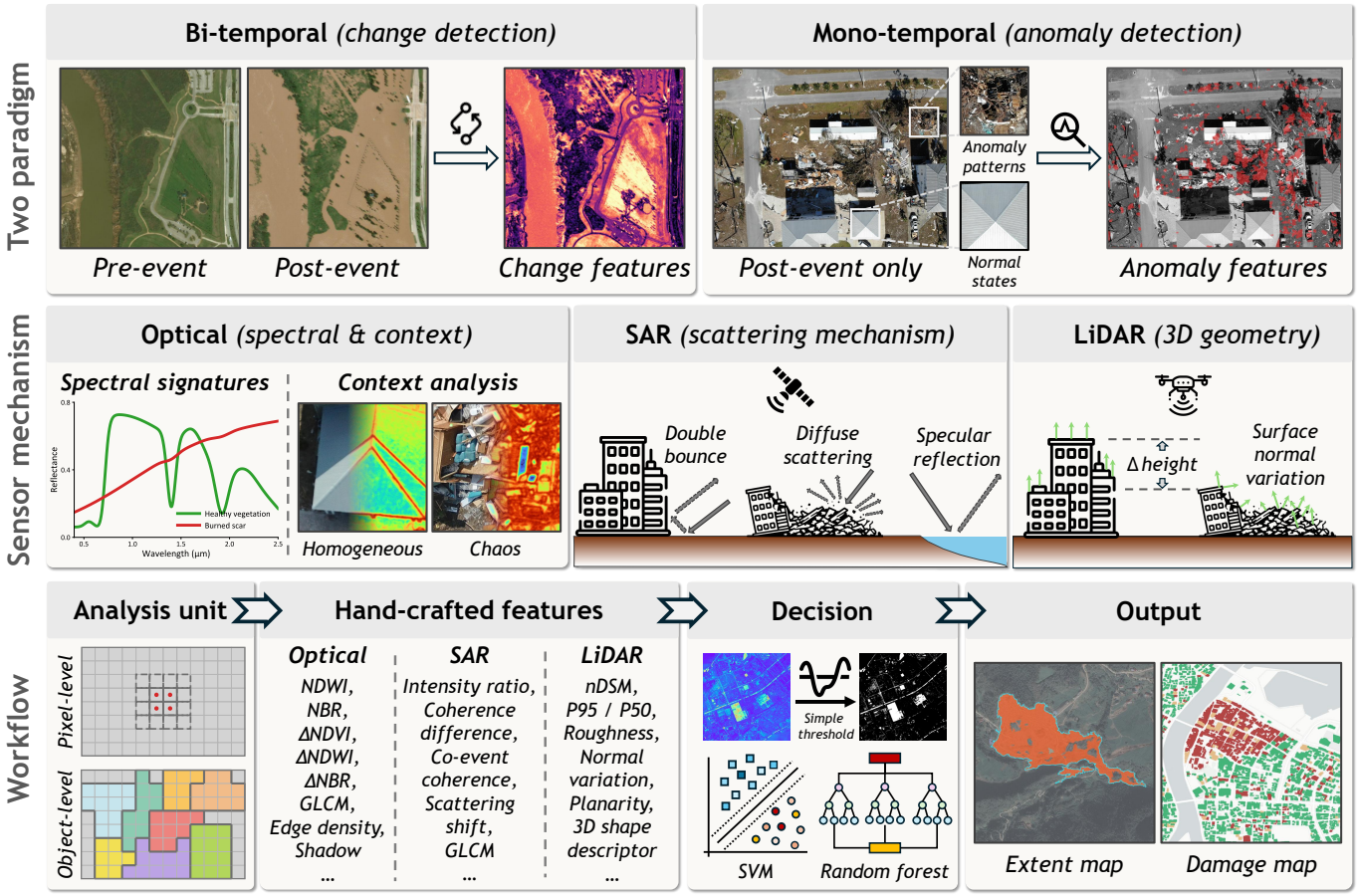

Fig. 4. Overview of the traditional paradigm for EO-based disaster mapping. Traditional pipelines can be grouped into bi-temporal change detection and mono-temporal anomaly detection. The middle panel summarizes the dominant sensor mechanisms exploited by hand-crafted designs, including optical, SAR, and LiDAR cues. The bottom panel illustrates the canonical workflow from analysis unit, to hand-crafted indicators, to rule-based or shallow ML decisions, producing either extent maps or damage maps. Note that this schematic highlights common insights rather than an exhaustive taxonomy.

[216], [254], [451]. For building damage, the key mechanism is the loss of the dihedral (double-bounce) scattering formed by vertical wall-ground structures, resulting in a characteristic backscatter decrease [16], [75], [96], [138], [229], [252], [286], [289], [290]. Wildfire and landslide responses are more ambiguous, as backscatter changes depend on sensor frequency, vegetation type, and terrain geometry [28], [173], [175], [396], [403], [556]. Coherence-based methods measure the phase correlation between acquisitions and are sensitive to sub-pixel surface disruption. Both floods and landslides cause strong decorrelation [409], [432]. For building damage, SAR coherence has proven particularly valuable for detecting moderate damage that remains invisible in intensity data alone [11], [17], [38], [63], [117], [167], [251], [288], [311]. This so-called Coherent Change Detection (CCD) framework has become a standard approach for rapid post-disaster damage assessment [425], [484], [486], [539], as exemplified by the widely-used (and patented) *Damage Proxy Maps* developed by NASA-JPL. To reduce false detections driven by noise, recent works have exploited long time series of coherence maps to establish robust pre-event baselines [11], [374], [375], [410], [530]. These large-scale, VHR-independent approaches have gained particular traction for monitoring damage in conflict zones, where they have been widely adopted by humanitarian organizations and newsrooms alike [374], [375]. Polarimetric methods (PolSAR) leverage shifts in scattering mechanisms: a transition from volume to surface scattering indicates vegetation removal caused by wildfires or landslides [48], [142], [240], [318], while a transition from double-bounce to diffuse scattering indicates structural collapse [66], [191], [399], [485], [522]. In practice, combining these complementary feature types consistently yields improved performance [1], [17], [128], [388], [557].

LiDAR provides the most direct means of assessing disaster impact through precise 3D elevation measurements. Pre- and post-event point clouds (or derived surface models) are co-registered and differenced to produce elevation-change maps. For mass movements, the resulting depletion and accumulation zones enable volumetric estimation of displaced material [40], [98], [436]. For wildfires, canopy height model differencing quantifies vertical structural loss [357], and for floods, LiDAR is primarily used to characterize post-event geomorphic change such as channel erosion and deposition [6], [535]. For infrastructure damage, the primary indicators are abrupt height loss (structural collapse) and positive anomalies (debris accumulation), with more advanced methods analyzing geometric properties

such as surface roughness and normal vector variance to detect subtler deformations [119], [158], [305], [368], [540].

To overcome the limitations of any single sensor, multi-modal methods fuse complementary data sources [76], [84]. In supervised settings, features from heterogeneous sources are concatenated and classified using methods such as Support Vector Machines (SVM) or Random Forests (RF) [4]. However, for rapid disaster response, unsupervised approaches are especially valuable because they do not require manual annotation. The central challenge is modality heterogeneity [60], which prevents direct pixel-level comparison between sensors. Three strategies have been developed to address this: independent classification of each modality followed by thematic map comparison [259], [465], [466], though this is prone to error propagation [495]; modality translation, which transforms one sensor's imagery to resemble the other through regression or physical simulation [46], [268], [418]; and feature transformation, which projects both modalities into a shared latent space where sensor-invariant relationships can be exploited [60], [417], [419], [433], [459], [464]

Moreover, regardless of the core algorithm, ancillary data is frequently integrated as a post-processing step to refine the results. GIS vector maps of building footprints or land cover are essential for defining the Region of Interest (ROI), allowing algorithms to mask out irrelevant changes (e.g., seasonal vegetation changes) and focus the analysis exclusively on potential risk areas like urban infrastructure or floodplains comparisons [121], [253].

4.1.2 Mono-Temporal Methods

When only post-event data are available, mapping relies on recognizing affected states from a single acquisition. For extent mapping, this is feasible when the hazard produces a directly observable signal in the relevant EO modality (e.g., thermal emission from active fires, strong water absorption in near-infrared bands). Damage mapping is inherently more ambiguous because many damage signatures are indirect and confounded by pre-existing urban heterogeneity. Mono-temporal damage methods therefore place greater emphasis on structural priors, seeking indicators such as abnormal texture, loss of geometric regularity, and height inconsistencies visible in the post-event scene alone.

For optical data, the approach depends on how distinctive the spectral signal is. In extent mapping, the hazard itself is often directly detectable. Active fires are identified through their strong thermal infrared emission using contextual algorithms applied to moderate-resolution sensors such as MODIS and VIIRS [200], [376]. Flood water is segmented by exploiting its strong absorption in near- and shortwave-infrared bands, typically through histogram-based thresholding [31], [116], [291], [378], [394]. For infrastructure damage, however, spectral information alone is insufficient because debris is spectrally similar to intact concrete and asphalt. Methods instead rely on textural and geometric anomalies: the transition from ordered roof patterns to chaotic, high-entropy textures serves as a primary indicator of collapse, often complemented by morphological and edge-based features that detect the loss of linear structure in building footprints [148], [174], [413], [523], [534].

For SAR, mono-temporal flood mapping exploits the specular reflection of smooth water surfaces, which appear as distinctly dark regions in the radar image; the task reduces to finding an optimal backscatter threshold to delineate the inundation boundary [161], [280], [358]. For infrastructure damage, the absence of a pre-event reference demands more complex reasoning. Polarimetric decomposition is central: structural collapse shifts the dominant scattering mechanism from double-bounce (intact wall-ground structures) to volume or diffuse scattering (rubble), providing a physically grounded damage indicator [233], [543]. Texture measures capturing spatial heterogeneity further correlate with damage severity. These features are applied either through comparison against simulated intact baselines from SAR simulators [27], [215] or as inputs to machine learning classifiers [126], [140], [496], with several studies demonstrating improved performance by fusing polarimetric and textural cues [23]–[25], [545], [568].

For LiDAR, post-event 3D point clouds enable geometric analysis without a pre-event reference. Landslide deposits can be identified by their characteristic rough ("hummocky") surface texture, which contrasts with surrounding smooth slopes, using roughness and slope variability metrics [41], [181]. For building damage, the primary indicator is violation of planarity: intact roofs exhibit consistent, near-horizontal surface normals, whereas collapsed structures produce clusters of abnormal normals and high local curvature [21], [217]. Surface roughness metrics further help distinguish chaotic debris from smooth, intact surfaces [187], [351], [352]. These geometric features are then classified through either rule-based thresholds [21] or machine learning classifiers [351]. Several studies also demonstrate the fusion of these various geometric cues for a more robust classification of damaged structures [187], [351].

As with the bi-temporal case, mono-temporal methods benefit from multi-modal fusion. Common configurations include combining post-event optical imagery with SAR [4], [75], [195], [411], [422], [467] or augmenting 2D imagery with LiDAR-derived 3D information [4], [153], [448], [458], [480]. In practice, hand-crafted features from each modality are typically integrated at the feature level for rule-based or shallow classifier decisions. Ancillary GIS data becomes even more critical in the mono-temporal setting. Without a pre-event reference, building footprint maps are often essential to localize the analysis and prevent confusion between debris and other spectrally or texturally similar surfaces [187], [352], [371], [534].

4.1.3 Analysis Units and Classifiers

As EO data shifted toward higher spatial resolutions and feature spaces grew more complex, simple pixel-level thresholds proved insufficient. Two complementary advances refined the decision-making process: Object-Based Image Analysis (OBIA), which operates on spatially coherent units rather than isolated pixels, and Machine Learning (ML) classifiers, which resolve nonlinear relationships in high-dimensional feature spaces.

With increasing spatial resolution, pixel-based methods often suffer from the "salt-and-pepper" effect, where individual pixels are misclassified due to local noise. OBIA addresses this by first segmenting the image into spectrally

homogeneous regions (superpixels) or using GIS-derived boundaries to define analysis units, and then computing features per object [37]. For disaster extent mapping, OBIA is used primarily to suppress spurious detections and enforce spatial consistency, with several studies incorporating hazard-specific segmentation rules (e.g., hydrographic constraints for flood boundaries [124], [198], [238], [276], [384], [437], [452], slope-guided delineation for landslides [237], [360], [412]) to recover more plausible boundaries. For infrastructure damage mapping, OBIA is particularly natural because damage is fundamentally an object-level attribute. Shape descriptors such as compactness, convexity, and rectangularity can assess whether a building's geometric integrity has been disrupted. However, OBIA trades pixel-level noise for a dependency on segmentation quality. When object boundaries are well defined, OBIA suppresses spurious detections and enforces spatial consistency. When they are not, for instance due to occlusion, shadow, or spectrally ambiguous debris, errors in the initial segmentation propagate into all downstream features [169].

While rule-based thresholding remains effective for spectrally distinctive targets such as open water or active fires, it cannot capture the complex decision boundaries required for building damage mapping or multi-modal feature fusion. To handle high-dimensional engineered features, traditional workflows adopted supervised ML classifiers, most commonly Support Vector Machines (SVM) [35], [70], [122], [153], [189], [199], [228], [326], [373], [386], [490], Random Forests and other ensemble methods [1], [15], [22], [30], [32], [103], [105], [153], [210], [250], [301], [326], [386], [450], and probabilistic models [242], [308], [344]. These classifiers serve as integrators, synthesizing the physical knowledge encoded in hand-crafted features into a final decision boundary.

4.1.4 Summary

The traditional paradigm for EO-based disaster mapping follows a common two-stage framework: (i) extracting hand-crafted indicators from EO data, and (ii) applying rule-based decisions or shallow classifiers to produce the final map. Early approaches often rely on unsupervised heuristics such as empirical thresholding, while later work increasingly adopts supervised pipelines feeding engineered features into conventional classifiers. Table 4 summarizes representative approaches by their practical attributes, allowing readers to identify which sensor-method combinations have proven effective for specific mapping targets.

A key insight from this part's literature review is that the distinction between extent mapping and damage mapping is particularly pronounced in the traditional paradigm. Traditional methods succeed when the target corresponds to a separable physical state in the chosen modality. This condition is frequently met for disaster extent mapping, where the goal is to delineate a surface-state footprint using relatively direct cues such as spectral contrast, scattering difference, or elevation change. Infrastructure damage mapping, by contrast, requires inferring the condition of discrete objects from indirect evidence: collapse rarely produces a unique spectral signature and instead manifests through changes in texture regularity, roof geometry, shadow patterns, and scattering mechanisms. As a result, traditional damage pipelines tend to be more object- and context-dependent, often requiring building footprint priors, whereas extent mapping pipelines are more threshold-driven.

More broadly, the traditional paradigm suffers from fundamental limitations that hinder operational scalability. Hand-crafted features are intrinsically tied to specific sensors, modalities, and hazard types, requiring substantial domain expertise to design and yielding models that generalize poorly. A workflow carefully tuned for one event and acquisition geometry often fails when transferred to a different setting. This case-specific nature, compounded by the limited scale of training data available for traditional classifiers, made it difficult to develop robust, broadly applicable solutions. These limitations directly motivated the shift toward the end-to-end learning paradigms of deep learning.

4.2 Deep Learning-Based Methods

The advent of deep learning, combined with the availability of large-scale, multi-disaster datasets, triggered a paradigm shift in disaster mapping. Unlike traditional approaches that relied on disparate, hand-crafted features for different targets, deep learning unified the field under a data-driven, end-to-end framework. Both infrastructure damage assessment and disaster extent mapping are now treated as pixel-wise semantic prediction problems, where deep networks learn hierarchical representations directly from raw EO data, treating sensor modality and temporal configuration as flexible inputs rather than rigid constraints. Fig. 5 provides an overview. Given this interconnected landscape, we organize this section along three axes: architectural evolution, fusion strategy, and operational considerations.

4.2.1 From Classification to Segmentation

In the initial phase, deep learning for disaster mapping was formulated as image classification. For discrete objects such as buildings, the approach was object-centric: individual building patches, cropped using ancillary footprint data (e.g., from OpenStreetMap), were fed into standard CNN classifiers (e.g., VGG, ResNet, DenseNet) to assign a single damage label per building [26], [43], [132], [206], [299], [315], [416], [427], [456], [527], [527]. While suitable for rapid assessment, this two-stage workflow was limited by the quality of ancillary footprints and lacked the spatial granularity to delineate damage extent within individual structures. For continuous hazard fields such as flood inundation or landslide extent, early studies adopted a sliding-window strategy, classifying one pixel at a time using its surrounding patch [28], [154]. Although this produced pixel-level outputs, it suffered from severe computational redundancy and limited contextual modelling.

The release of large-scale, densely annotated datasets [58], [151], [346], [347] enabled a shift from patch-based classification to end-to-end semantic segmentation, in which every pixel receives a label simultaneously. The dominant framework is the encoder-decoder architecture, with several design choices driven by the specific requirements of disaster mapping. To handle bi-temporal inputs, the encoder is commonly implemented as a Siamese network with shared or partially shared

TABLE 4
Representative traditional methods for EO-based disaster mapping. The table summarizes dominant methodological patterns rather than exhaustively listing all variants. Exceptions and hybrid pipelines exist.

Temporal	Primary Modality	Key Clue	Target	Analysis Unit	Decision	Reference
Bi-temporal	Multispectral	Spectral/water indices	Extent (flood)	Pixel	Heuristic change rules	[402]
		Burn indices	Extent (wildfire)	Pixel	Heuristic change rules	[295], [296]
		Vegetation/soil spectral change	Extent (landslide)	Pixel	Heuristic change rules	[118], [562]
	Optical VHR	Spectral values	Extent (landslide)	Pixel	Heuristic change rules	[237]
		Texture features	Damage (building)	Pixel	Statistical classifier	[353]
		Morphological features	Damage (building)	Object	Shallow ML (decision tree)	[461]
		Texture + spectral features	Damage (building)	Object	Shallow ML (SVM)	[228]
		Shadow change	Multi-level damage (building)	Object	Heuristic thresholding	[431]
	SAR	Intensity	Extent (flood)	Pixel	Heuristic change rules	[79], [261]
		Intensity, Coherence	Extent (flood)	Pixel	Heuristic fusion	[337]
		Polarimetric features	Extent (flood)	Pixel	Decomposition-based change analysis	[44]
		Intensity	Extent (wildfire)	Pixel	Difference, ratio	[173], [396]
		Coherence	Extent (wildfire)	Pixel	Decorrelation analysis	[1]
		Polarimetric features	Extent (wildfire)	Pixel	Decomposition analysis	[142]
		Intensity features	Extent (landslide)	Pixel	Expert interpretation	[303]
		Intensity change	Damage (building)	Pixel	Shallow ML (LDA)	[289]
		Coherence	Damage (building)	Object	Heuristic change rules	[167], [251]
		Polarimetric	Damage (building)	Object	Heuristic change rules	[66]
		Hybrid	Damage (building)	Object	Rule-based thresholding	[17]
	LiDAR	3D shape descriptor	Damage (building)	Object	Heuristic thresholding	[158]
		Building point drifts	Multi-level damage (building)	Object	Heuristic thresholding	[119]
		Hand-crafted geometric features	Multi-level damage (building)	Object	Shallow ML (RF)	[540]
	Multispectral–SAR	Cross-sensor structure consistency	Extent (flood)	Object	Graphical model	[417]
	Multispectral–SAR	Affinity matrix	Extent (flood, wildfire)	Pixel	Shallow MLs	[268]
	Multispectral–SAR	Intensity, coherence, spectral indices	Extent (burned area)	Pixel	Shallow ML (RF)	[1], [557]
	SAR, DEM, Land-cover	Intensity, topographic consistency	Extent (flood)	Pixel	Rule-based classification	[338]
	Optical, SAR, DEM	Spectral value, SAR intensity, SAR coherence, height	Multi-level damage (building)	Pixel	Shallow MLs	[4]
Mono-temporal	Thermal IR	Brightness temperature anomaly	Extent (active fire)	Pixel	Contextual thresholding	[200], [376]
	Multispectral	Post-fire spectral indices/features	Extent (wildfire)	Pixel	Thresholding/Feature selection, RF	[30], [450]
	Optical VHR	Texture, edge, and morphology features	Multi-level damage (building)	Object	Rule-based thresholding	[523]
	Optical VHR	Edge, gradient orientation	Damage (building)	Object	Rule-based thresholding	[534]
	SAR	Intensity features	Extent (flood)	Pixel	Heuristic thresholding	[239], [281]
		Intensity features	Damage (building)	Object	Shallow MLs	[140]
		Polarimetry and texture features	Multi-level damage (building)	Object	Heuristic thresholding	[568]
		Scattering mechanism cues	Damage (building)	Object	Expert interpretation	[27]
	LiDAR	Roof geometry	Damage (building)	Object	Heuristic thresholding	[21]
		Roof geometry, texture	Damage (building)	Object	Unsupervised clustering	[187]
	Multispectral, DEM	Spectral flood response	Extent (flood)	Pixel	Shallow ML (AdaBoost)	[81]
	LiDAR, UAV	Geometry + texture cues	Damage (building)	Object	Shallow MLs (SVM, RF)	[458]

weights, fusing features from pre- and post-event images to explicitly model change [2], [5], [57], [59], [147], [160], [316], [377], [421], [449], [471], [479], [563], [571], [574]. The decoder configuration reflects the output objective. For disaster extent mapping, a single-branch decoder typically suffices to produce a binary hazard mask [133], [255], [266], [282], [359], [393], [434], [521]. For infrastructure damage mapping, multi-task decoders are common, with one branch generating a building localization mask and another classifying damage grades, thereby reducing false positives in complex urban backgrounds [57], [59], [571], [574]. ChangeOS [571] exemplifies this design, coupling a partial Siamese encoder with a multi-task decoder that jointly localizes buildings and classifies damage within a unified framework.

The performance of encoder-decoder segmentation models depends critically on the feature extraction backbone. The field has followed the broader trajectory of computer vision: early work adopted CNN architectures such as UNet [5], [9], [132], [139], [151], [162], [225], [282], [392], [434], [505], [549], [550], [556], [566], ResNet [132], [497], [505], [556], [571], EfficientNet [77], [263], [385] and HRNet [55], [212], [246], [548], often enhanced with attention mechanisms to refine feature representations [5], [9], [88], [146], [218], [232], [246], [256], [377], [392], [470], [502], [541], [549], [550]. However, CNNs are constrained by local receptive fields, which limits their ability to connect spatially disjointed damage patterns across large scenes. This motivated the adoption of Transformer-based architectures (e.g., Swin Transformer [88], [89], [129], [263], [408], [491], [536], SegFormer [57], [230], [264]), which capture global spatial context through self-attention [57], [89], [230], [263]. More recently, state-space models such as Mamba [145] have emerged as an efficient alternative, offering global modelling capability with linear computational complexity. ChangeMamba [59] is a representative example, integrating spatio-temporal scanning mechanisms to quantify damage severity.

Decoder design has evolved in parallel. Multi-scale feature aggregation through pyramid pooling and skip connections has become standard practice for reconciling global context with boundary precision [135], [207], [271], [273], [302], [460]. Attention-based decoders have been adopted to capture long-range dependencies and cross-temporal correspondences [377], while Transformer decoders perform global feature modelling within the decoding stage [423]. For temporal sequences, Recurrent Neural Networks (RNNs) and their variants, e.g., LSTM and ConvLSTM, propagate evidence across time steps, a role increasingly filled by Mamba-based decoders that achieve

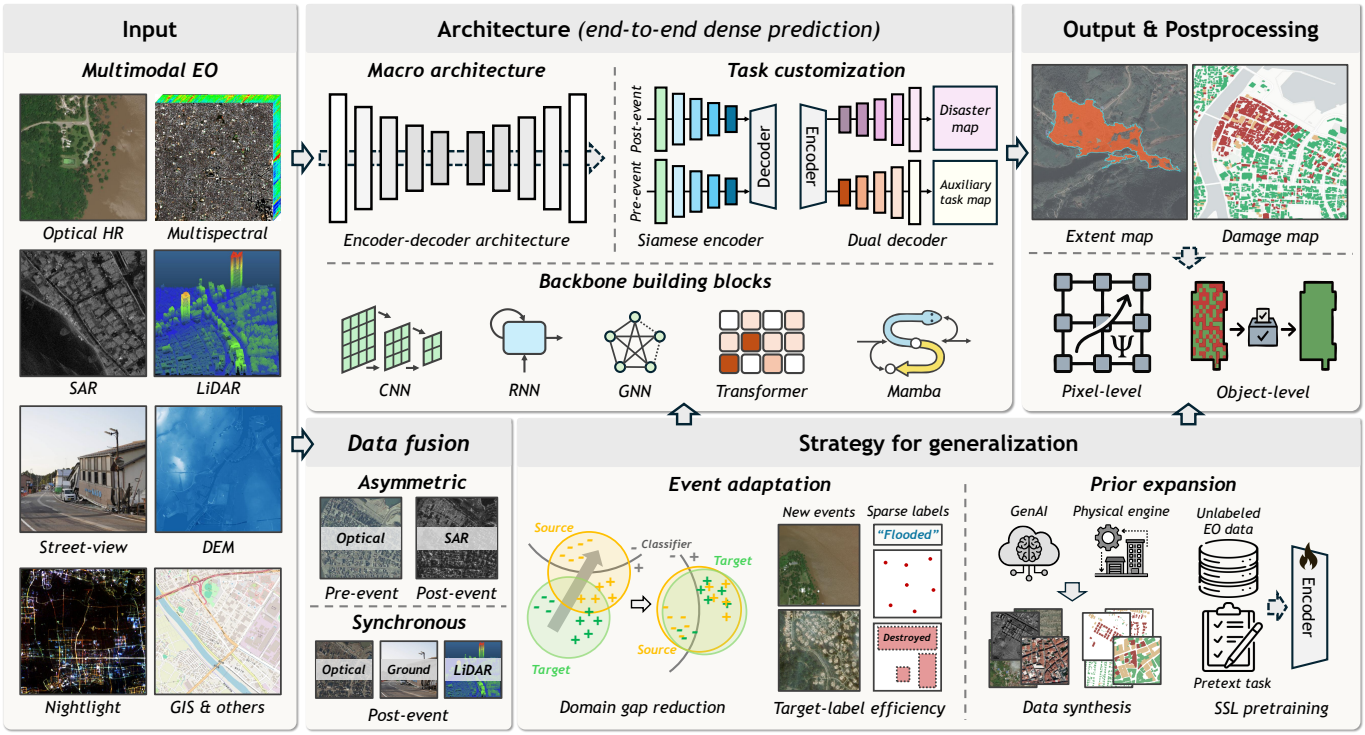

Fig. 5. Conceptual taxonomy of deep learning paradigms for EO-based disaster mapping. The figure summarizes a unified end-to-end dense prediction view. Inputs can be multimodal EO, extended with complementary viewpoints and priors. The central panel decomposes deep models into a macro architecture and interchangeable micro backbones as building blocks. The bottom panel highlights two core research frontiers: data fusion and generalization strategies. Outputs and postprocessing span both pixel-level segmentation maps and object-level products. This schematic is intended to distill dominant design patterns rather than enumerate all model variants.

comparable modelling with lower computational cost [59], [149]. Graph neural networks have also been explored to enforce neighbourhood consistency among detected objects [90], [180], [231], [379].

Beyond semantic segmentation, a subset of studies formulates building damage mapping as object detection or instance segmentation, localizing individual damaged structures with bounding boxes [512]. Architectures from the YOLO and R-CNN families have been adapted for this purpose [274], [481], [487], [537], [546]. For disaster extent mapping, this paradigm is generally unsuitable for large continuous fields but has found niche application in detecting discrete, small-scale slope failures [438], [501].

In addition to architecture, practical advances in training and post-processing have improved mapping quality. To address the pervasive class imbalance in disaster datasets (affected pixels are typically the minority), standard cross-entropy loss is commonly supplemented by focal loss [91], [471], [491], [536], dice loss [460], [471], [505], [536], [571] and Lovasz-softmax loss [57]–[59], [166]. For post-processing, probabilistic graphical models like Conditional Random Fields (CRFs) are used to enforce spatial smoothness at the pixel level [132], [329], [393], while object-level strategies such as majority voting within building footprints ensure that each structure receives a single, consistent damage label [58], [160], [542], [571]. For disaster extent mapping, rule-based filtering removes small spurious detections to produce cleaner hazard maps [28], [30], [377].

4.2.2 *From Unimodal to Multimodal*

Deep learning for disaster mapping initially advanced along unimodal tracks, shaped by data availability and label maturity in each modality. Optical imagery has been the dominant data source, driven by the availability of large-scale annotated datasets [151], [188], [282], [284], [359], [377], [558] that provided the labelled volume required to train deep models. The optical domain has consequently served as the primary testbed for architectural innovation, from CNNs to Transformers and state-space models [57], [59], [86], [147], [282], [377], [390], [392], [423], [449], [487], [571], [574]. SAR has attracted parallel interest for its weather- and daylight-independence and sensitivity to structural geometry, with deep learning applied to damage quantification [26], [203], [230], [325], [365], [529], burned area mapping [28], [266], [556], and flood delineation [39], [135], [185], [273], [470]. LiDAR-based deep learning remains comparatively scarce [6], [241], [509], [510], [535], with existing studies predominantly adapting 3D deep architectures [6], [241], [509], [510], [535], primarily due to the lack of large-scale 3D benchmark datasets.

The field has increasingly shifted toward multimodal fusion to mitigate sensor-specific failure modes and exploit complementary physical evidence. Deep learning is particularly well-suited to this challenge. Unlike traditional methods that struggled with heterogeneous data fusion, deep networks can learn nonlinear cross-modal mappings that bridge differences in imaging physics. Multimodal research in disaster mapping can be organized around two

operationally motivated paradigms: (i) asymmetric fusion across time, where pre- and post-event sensors differ, enabling rapid response with whatever data are available; and (ii) synchronous fusion of concurrent modalities and priors, enriching feature representations for more refined assessment.

The asymmetric paradigm recognizes that the sensors available before and after an event often differ. A widely studied configuration pairs pre-event optical imagery with post-event SAR, combining the semantic richness of optical data with the all-weather observability of radar. Progress has been uneven across tasks. For disaster extent mapping, particularly flood inundation, the relatively large spatial footprints and moderate resolution requirements have enabled scalable fusion pipelines, supported by established Sentinel-1/2 datasets [100], [159], [319], [348]. For infrastructure damage mapping, the requirements are more demanding: accurate damage assessment typically requires very high resolution imagery and object-level annotation, while cross-modal co-registration tolerances are tighter. Early studies consequently relied on private or small case-specific collections [5], [416], [548], limiting reproducibility. The release of the BRIGHT benchmark [58] marked an important milestone by providing an open, large-scale optical-SAR dataset for building damage mapping with community baselines. Architecturally, fusion strategies have evolved from simple input concatenation to designs that explicitly address cross-modal mismatch. Pseudo-Siamese encoders with modality-specific parameters allow each branch to accommodate the distinct statistical properties of optical and SAR inputs [5], [556], [557]. Cross-modal attention modules further improve robustness by conditionally aggregating features across modalities and spatial context, which is particularly valuable when pixel-wise correspondence is unreliable due to geometric differences between sensors [5], [159], [225], [538], [542], [554], [556], [582].

Alongside these supervised approaches, unsupervised methods have seen a resurgence, modernizing the classic paradigm with deep models. Generative Adversarial Networks (GANs) [143], notably CycleGAN [577], are used to translate imagery between domains to reduce modality heterogeneity [223], [269], [472], while autoencoder-based architectures map distinct modalities into a shared latent space [61], [247]–[249], [500]. These unsupervised methods have shown promise for disaster extent mapping, where targets are spatially homogeneous, but remain limited for infrastructure damage assessment where fine-grained discrimination is required.

Synchronous fusion assumes that multiple data sources are available within a similar post-event time window and aims to enrich the representation for more detailed mapping. This paradigm serves both tasks. For extent mapping, concurrent modalities help disambiguate complex surface states. For damage mapping, they provide evidence beyond rooftop appearance alone. The most common configuration fuses post-event optical and SAR data to reduce ambiguity in heterogeneous scenes, particularly in dense urban areas where debris can be spectrally similar to intact surfaces. Chen *et al*. [58] evaluated multiple architectures across several disaster events and found that optical-SAR fusion consistently outperformed optical-only baselines, even when optical observations were relatively clean. Beyond semantic complementarity, SAR provides a practical fallback when optical coverage is degraded by clouds or smoke [28], [221], [266], [372], [556], [557]. A second direction integrates explicit geometric information. For damage mapping, LiDAR-augmented frameworks add 3D cues that improve sensitivity to structural deformation, including story-level collapse patterns [245], [381], [457]. For extent mapping, elevation priors regularize physically plausible footprints, such as drainage-consistent inundation boundaries [49], [124], [306] or slope-constrained deposit extents [255], [436]. An emerging direction combines overhead observations with ground-level imagery to capture damage features that are weakly observable from nadir views. Social media and street-view imagery provide temporally dense, facade-level evidence that can complement satellite-derived maps, whether by validating and correcting omissions in flood products [267], [362], [367] or by enabling fine-grained structural assessment through cross-view feature alignment [224].

4.2.3 *From Single-Event Fitting to Cross-Event Generalization*

Early deep learning studies for disaster mapping were typically developed, trained, and evaluated on a single event or narrow region [315]. For the specific event studied, this is not inherently problematic since a model tailored to one disaster can produce highly accurate maps for that event. However, this paradigm makes reported performance gains difficult to interpret beyond the studied case, and offers no guarantee that the same model will perform adequately on a future event with different geographic, structural, or sensor characteristics. The release of multi-event benchmarks [58], [151], [558], [567] enabled a shift toward broader training and more reproducible comparison. However, strong benchmark performance does not necessarily translate into operational utility [58], [491]. In practice, models must be deployed on a new disaster immediately, using only knowledge from past events and with minimal or no new labelled data. This exposes a persistent generalization gap shared by both extent and damage mapping, though the dominant sources of domain shift differ. For extent mapping, shift is primarily driven by land-cover variability, seasonality, and sensor/acquisition differences. For damage mapping, visual manifestations vary drastically across hazard types, construction materials, building typologies, and viewing conditions.

Current research addresses this cross-event generalization gap through two complementary paradigms: (i) adapting to the target event at deployment time, and (ii) broadening prior knowledge before deployment so that less adaptation is needed. The first paradigm exploits the fact that, while labelled data for a new disaster are scarce, unlabelled post-event imagery is typically abundant within hours. Unsupervised domain adaptation methods [12], [179], [235], [236], [243], [405], [478] bridge the distribution gap between past training events and the incoming one by aligning source and target feature representations without requiring target labels. Semi-supervised [19], [234], [398], [482], [492] and weakly supervised approaches [14], [336], [339]–[341] further reduce the annotation barrier to cross-event transfer by substituting dense pixel-level labels with pseudo-labels,

TABLE 5
Summary of representative deep learning-based disaster mapping studies. Note that citations are categorized based on their primary methodological contribution. MM-A: asymmetric fusion across time; MM-S: synchronous post-event fusion within a time window. N/A: there is no relevant literature available.

Family	Methodology	Disaster Extent Mapping					Infrastructure Damage Mapping				
		Optical	SAR	LiDAR	MM-A	MM-S	Optical	SAR	LiDAR	MM-A	MM-S
Architecture	CNN	[92], [222]	[312], [566]	[107]	[159], [556]	[350], [393]	[571]	[26]	[241], [509]	[5]	[363], [381]
	RNN	[404]	[439], [555]	N/A	N/A	[109], [110]	N/A	[531]	N/A	[58]	N/A
	GNN	[90], [231]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	[180], [379]
	Transformer	[272], [377]	[370]	N/A	N/A	[423]	[57]	[230], [529]	N/A	N/A	N/A
	Mamba	[424]	[488], [489]	N/A	[391]	N/A	[59]	[166]	N/A	[542]	[166]
Learning strategy	Semi-supervised	[307], [349]	[234], [329]	N/A	N/A	N/A	[482], [492]	N/A	N/A	[58]	N/A
	Weakly supervised	[460]	[387]	N/A	N/A	[39]	[340]	N/A	N/A	N/A	N/A
	Self-supervised	[12], [449]	[519]	N/A	[113]	N/A	[330], [504]	N/A	N/A	N/A	N/A
	Unsupervised	[382]	[196], [520]	N/A	[247], [270]	N/A	[428]	[369]	N/A	[58]	N/A
	Transfer learning	[282], [558]	[123]	N/A	N/A	[499]	[125], [573]	N/A	N/A	[225], [542]	N/A
Generation	Data-driven synthesis	[114]	N/A	N/A	N/A	N/A	[569]	N/A	N/A	N/A	N/A
	Physics-based simulation	[535]	N/A	N/A	N/A	N/A	[406]	[106], [164]	N/A	N/A	N/A

point annotations, or image-level tags that can be collected rapidly for the new event. The practical relevance of these strategies for cross-event deployment was demonstrated in the 2025 IEEE GRSS Data Fusion Contest [58], [331], [503] and ESA Φ-Lab Challenge [101], where winning solutions showed that advanced pseudo-labelling can substantially improve generalization to unseen events [225], [542].

The second paradigm seeks to reduce the domain gap before a disaster occurs by exposing models to a broader space of disaster appearances, so that an unseen event is less likely to fall outside the learned distribution. Data synthesis is a prominent direction: generative models such as GANs [31], [33], [114], [143], [163], [327], [364], [405], [414], [428], [565] and diffusion models [165], [171], [314], [380], [569] can populate underrepresented corners of the disaster appearance space, generating realistic post-event imagery for hazard types, building typologies, or geographic contexts that are absent from existing archives. A notable example is Neural Disaster Simulation (NeDS) [569], which conditions synthetic post-event damage on pre-event imagery of the target area, effectively producing a site-specific training set before the event occurs. Physics-based simulation provides a complementary route, using 3D city models and physical engines to model structural failure under controlled conditions and generate labels for scenarios rarely captured in real observations [106], [164], [406], [407], [540]. Song *et al.* [406], [407] demonstrated successful transfer from such synthetic data to real unseen events including the 2023 Hawaii wildfire and the 2023 Turkey–Syria earthquakes. Self-supervised pretraining complements these data-centric strategies from the representation side: by learning general visual structures from large volumes of unlabelled EO imagery, it provides an initialization that is less brittle to domain shift and reduces the amount of event-specific supervision needed for effective cross-event transfer [330], [504], [526]. When scaled to massive and diverse corpora, this line of work converges with the foundation model paradigm discussed in Section 4.3.

4.2.4 Summary

The evolution of deep learning for disaster mapping can be characterized along three trajectories. First, architecturally, the field has moved from patch-based classification to pixel-level semantic segmentation, driven by successive adoption of CNNs, Transformers, and state-space models that progressively improve contextual modelling in complex post-disaster scenes. Second, in terms of input modality, the paradigm has shifted from single-sensor dependence to multimodal fusion. Modern frameworks bridge feature heterogeneity across optical, SAR, LiDAR, and ground-level sources, resolving semantic ambiguities while maintaining all-weather operational capability. Third, the field is moving beyond optimizing performance on single-event benchmarks toward designing models that generalize across hazards, regions, and sensors. This shift is reflected in evaluation protocols that explicitly simulate deployment conditions, including leave-event-out testing [58], [331], [491], few-shot adaptation [58], [469], and systematic studies of how synthetic and semi-supervised signals improve performance on unseen disasters [569].

Table 5 summarizes representative studies categorized by their primary technical contributions, spanning architecture families, learning strategies, and generation paradigms. This organization aims to assist readers in quickly identifying key references and grasping the fundamental ideas required to apply these advanced techniques to specific disaster mapping scenarios.

4.3 The Era of Foundation Model

The advances discussed above have pushed the boundaries of disaster mapping, yet a fundamental limitation persists: the paradigm remains largely task-specific, requiring separate training pipelines for different hazards, sensors, or regions. Foundation Models (FMs) represent a departure from this fragmented approach. Pre-trained on massive and diverse datasets, FMs learn general-purpose representations that can be adapted to diverse downstream tasks with minimal fine-tuning. Current research explores this frontier along two trajectories: (i) adapting general-purpose vision models to EO disaster tasks, and (ii) developing models pre-trained natively on EO data.

4.3.1 Adapting General Vision Foundation Models

The first trajectory leverages FMs from the computer vision domain, most notably the Segment Anything Model

(SAM) family [213], [354] and the DINO family [50], [322]. These models possess powerful, class-agnostic segmentation and feature extraction capabilities learned from billions of natural images. However, applying them directly to disaster mapping faces a significant domain gap between ground-level photographs and overhead EO data. To bridge this gap without full retraining, Parameter-Efficient Fine-Tuning (PEFT) strategies such as Low-Rank Adaptation (LoRA) [170] and visual prompt tuning [192] inject lightweight learnable modules into frozen backbones, steering the model toward EO-specific semantics while preserving its pre-trained knowledge.

For disaster mapping specifically, SAM has been adapted through visual prompting for building damage evaluation on pre-/post-event image pairs [564], through dual-encoder architectures that fuse terrain features for landslide mapping [525], and through LoRA-based tuning for SAR-based flood and avalanche delineation [130], [483]. Beyond task-specific adaptation, SAM has also been used as an external prior for deployment-time generalization: generating pseudo-labels of relevant structures that, combined with test-time adaptation, enable damage detection in unseen regions without target annotations [10], [552]. DINO-style models have received less attention in disaster contexts but show promise: their self-supervised embeddings can localize damage regions even in zero-shot settings [172], and recent work has explored DINOv2-based domain generalization for post-disaster infrastructure assessment [141]. Applications to extent mapping with DINO features remain sparse and largely limited to high-resolution UAV imagery [93].

Despite these encouraging results, general vision FMs are trained predominantly on RGB ground-level imagery and can struggle with EO-specific characteristics, particularly non-optical modalities. SAR backscatter, LiDAR geometry, and multispectral bands lie outside the training distribution of internet-scale pretraining, motivating the development of FMs pre-trained natively on EO data.

4.3.2 Training Native EO Foundation Models

The second trajectory develops FMs pre-trained directly on large archives of unlabelled EO imagery, learning representations that inherently capture the physics and spatial scales of satellite data. Pre-training generally follows two self-supervised paradigms [156], [551]. Generative approaches, typically based on Masked Image Modelling (MIM) [157], train the model to reconstruct masked portions of input imagery, forcing it to learn spatial and spectral structure. Representative models include SatMAE [83], SpectralGPT [168], RingMo [415], TerraMind [184], and the Prithvi family [183], [420]. Contrastive approaches instead learn invariant representations by distinguishing semantically equivalent views from dissimilar ones; models such as SeCo [533] leverage temporal augmentations, while the SkySense family [150], [498] extends DINO-like objectives to multimodal EO data by aligning optical, SAR, and meteorological inputs. A recent subclass departs from this fine-tunable backbone paradigm: embedding field models such as AlphaEarth Foundations [45] and TESSERA [115] release pre-computed annual global pixel embeddings at 10m resolution, which downstream users consume directly as input features rather than as initial weights for further training.

These EO-native FMs provide strong representations that reduce labelled-data requirements and improve transfer under domain shift, and disaster mapping is increasingly adopted as a key downstream evaluation task [20], [183], [435], [498]. Prithvi-EO-2.0 [420] exemplifies this trend: leveraging temporal and location embeddings within a global pre-training framework, it demonstrates adaptation across flood inundation, fire scar detection, and landslide mapping. SkySense family [498] provides complementary evidence for multimodal scalability, achieving consistent gains on a disaster mapping suite spanning both extent delineation and building damage mapping. Embedding field models occupy a different position. Their pre-computed, analysis-ready delivery makes them highly accessible as geospatial feature layers [108]. However, current annual global embeddings are temporally coarse for the event-dated pre/post observations required in disaster mapping. They are therefore unlikely to replace event-specific imagery in rapid mapping pipelines, but may serve as convenient pre-event references or contextual priors, which we return to in Section 6.

4.3.3 Summary

Foundation models are beginning to reshape disaster mapping by shifting the bottleneck from task-specific supervised training to transferable representation learning and label-efficient adaptation. As summarized in Table 6, progress follows two complementary directions. General vision FMs provide strong reusable components for rapid delineation, pseudo-labelling, and lightweight adaptation in time-critical settings, but their utility is bounded by domain and modality mismatch and is most reliable on RGB-like VHR optical imagery. Native EO FMs, pre-trained on large EO archives, offer representations that better respect the physics, scale, and temporal structure of EO data, providing stronger initialization under cross-event and cross-region domain shift. While most EO FMs are still validated primarily on general-purpose benchmarks, recent work has demonstrated broad transfer to multiple disaster-relevant tasks [420], [498], [508].

A core question motivating this review, however, remains open: how much operational value do foundation models actually provide for disaster mapping across hazards, sensors, and deployment regimes? Current evidence is fragmented, often limited to a narrow set of hazards, RGB-dominant inputs, or restricted evaluation protocols, and does not yet establish when FMs genuinely outperform well-tuned task-specific models or how robust they remain under the constraints that matter most in practice. Addressing this gap requires systematic, regime-aware benchmarking, which motivates the comparative evaluation presented in the following sections.

5 Benchmark Experiments

The preceding sections have traced the rapid expansion of disaster mapping methods, from classical feature engineering to foundation models. While this diversity offers a powerful toolkit, it also creates practical ambiguity for deployment. Models differ substantially in parameter count,

TABLE 6
Summary of foundation model-based approaches for disaster mapping.

Family	Base FM	Pretrain	Input	Adaptation	Usage	Reference
General Vision FMs	SAM	Natural RGB	Bi-temporal optical	Visual prompt	Building damage mapping	[564]
			Optical + DEM	Adapter	Landslide extent mapping	[525]
			Sentinel-1 SAR	Adapter	Avalanche extent mapping	[130]
			Bi-temporal optical	-	Generating pseudo-labels for building damage mapping	[10]
			Bi-temporal optical	-	Providing structural prior for building damage mapping	[552]
	DINOv2	Natural RGB	Street-view image	Fine-tune	Building damage mapping	[172]
			Post-event optical	Adapter	Building damage mapping	[141]
Native EO FMs	Prithvi-EO-1.0	Multispectral	HLS	Fine-tune	Flood/wildfire extent mapping	[183]
	Prithvi-EO-1.0	Multispectral	Sentinel-2	Fine-tune	Wildfire extent mapping	[395]
	Prithvi-EO-2.0	Multispectral	HLS	Fine-tune	Flood/wildfire/landslide extent mapping	[420]
	DoFA+	Optical VHR, multispectral, hyperspectral, SAR	Multimodal EO	Fine-tune	Flood/wildfire extent mapping	[508]
	SkySense++	Optical VHR, multispectral, SAR	Multimodal EO	Fine-tune	Disaster extent mapping and building damage mapping	[498]
	TerraMind	Multispectral, SAR, LULC map, DEM, vegetation index, location, text	Multimodal EO	Fine-tune	Flood/wildfire extent mapping	[184]
	AnySat	Optical VHR, multispectral, SAR	Multimodal EO	Fine-tune	Flood/wildfire extent mapping	[20]
	Galileo	Multispectral, SAR, nightlight, slope, elevation, LULC map, climate, weather, population, location	Multimodal EO	Fine-tune	Flood extent mapping	[435]

TABLE 7
Summary of the 13 benchmark datasets used in our evaluation, grouped by task family and hazard type.

Task	Dataset	Hazard	Modality	GSD	Temporal	Classes	Tiles	Evaluation protocol
Extent	CAUFLOOD [159]	Flood	Multispectral–SAR	10 m	Bi-temporal	2	18,302	Official split
	KUROSIWO [42]	Flood	SAR+DEM+Slope	$\geq$10 m	Bi-temporal	3	67,490	Official split
	URBANSARFLOODS [567]	Flood	SAR	20 m	Bi-temporal	3	8,879	Official split
	LANDSLIDE4SENSE [134]	Landslide	Multispectral+DEM	10 m	Post-event	2	4,844	Official split
	HR-GLDD [293]	Landslide	Multispectral	4 m	Post-event	2	1,758	Official split
	GVLM [558]	Landslide	RGB	0.59 m	Bi-temporal	2	7,596	Official split
	S2WCD [449]	Wildfire	Multispectral	10 m	Bi-temporal	2	41	Custom split
	SBAD [80]	Wildfire	Multispectral–SAR	10 m	Bi-temporal	5	73	Official split
	FLOGA [377]	Wildfire	Multispectral	20 m	Bi-temporal	2	1,167	Official split
	WILDFIRESPREADTS [131]	Wildfire (spread)	Multi-source time series	375 m	Multi-temporal	2	13,607	Official split
Damage	XBD [151]	Multi-hazard (6 types)	RGB	$<$0.8 m	Bi-temporal	5	11,034	Official split
	BRIGHT [58]	Multi-hazard (7 types)	RGB–SAR	0.3–1.0 m	Bi-temporal	5	4,246	Official split
	RESCUENET [346]	Hurricane	UAV RGB	$\sim$0.015 m	Post-event	10	4,494	Official split

data requirements, modality assumptions, and training paradigms, yet it remains unclear which choices are justified under realistic response constraints. Is a computation-heavy foundation model necessary for a structurally simple target such as open-area flood extent, or does a lightweight UNet already saturate performance? Do general-purpose models that promise zero- or few-shot transfer consistently outperform task-specific models trained with limited but domain-relevant labels? And how do these paradigms perform when forced to operate outside the RGB comfort zone, relying instead on SAR or bi-temporal cross-modal inputs?

To move beyond qualitative comparison, we design a systematic benchmarking framework and evaluate representative methods, from standard deep learning baselines to foundation model pipelines, across the two task categories defined earlier: infrastructure damage mapping and disaster extent mapping. Table 7 summarizes the 13 datasets used in our evaluation.

5.1 Benchmark Datasets

5.1.1 *Disaster Extent Mapping*

For disaster extent mapping, we select datasets covering three major hazard types. For floods, we employ three benchmarks spanning different complexities:

- URBANSARFLOOD [567]: Sentinel-1 SAR observations targeting the distinction between flooded open areas and flooded urban areas, where urban scattering effects introduce significant ambiguity.
- KUROSIWO [313]: A large-scale, multi-event SAR benchmark bundling Sentinel-1 with DEM-derived auxiliary layers, enabling assessment of how topographic constraints contribute to robust flood extent mapping.
- CAU-FLOOD [159]: Pre-event Sentinel-2 optical imagery paired with post-event Sentinel-1 SAR, serving as a benchmark for cross-modal change detection under adverse post-flood imaging conditions.

For landslides, we use a multi-resolution suite:

- LANDSLIDE4SENSE [134]: Sentinel-2 imagery fused with ALOS PALSAR DEM, representing the moderate-resolution regime for regional monitoring.
- HR-GLDD [293]: PlanetScope imagery offering finer spatial resolution and covering diverse geomorphological settings.
- GVLM [558]: VHR imagery with pre- and post-event optical pairs, enabling bi-temporal change-based landslide detection.

For wildfire mapping, we employ four datasets:

- S2WCD [449]: A bi-temporal Sentinel-2 dataset covering 25 wildfire events across Europe and Oceania, providing pre- and post-event multispectral image pairs with pixel-level burned area masks. We use S2WCD to evaluate change detection-based burned scar mapping.
- FLOGA [377]: A large-scale, expert-annotated dataset of 326 wildfire events with bi-temporal Sentinel-2 imagery and MODIS-derived auxiliary layers, providing a multi-event benchmark for burned area mapping at moderate resolution.
- SBAD [80]: 73 acquisitions of European forest wildfires from Sentinel-1 and Sentinel-2, with pixel-level annotations for both burned area delineation and five-level burn severity estimation. We use the binary burned/unburned task as a post-event mapping benchmark.
- WILDFIRESPREADTS [131]: A multi-temporal dataset covering 607 U.S. fire events, packaging VIIRS active fire detections with MODIS, GRIDMET meteorological, and SRTM terrain covariates into time-series cubes. We use this dataset to benchmark short-horizon fire spread prediction.

5.1.2 *Infrastructure Damage Mapping*

We employ three widely used open benchmarks that together represent complementary operational regimes: bi-temporal optical (optical–optical), bi-temporal multimodal (optical–SAR), and post-event-only UAV mapping.

- XBD [151]: The de facto standard for building damage mapping, spanning 19 disaster events globally with pre- and post-event VHR optical imagery. We use it to evaluate the standard bi-temporal optical setting.
- BRIGHT [58]: The first large-scale benchmark for optical-SAR building damage mapping, pairing pre-event optical imagery with post-event VHR SAR. This dataset evaluates the ability to bridge the modality gap between visual textures and radar backscatter under all-weather conditions.
- RESCUENET [346]: Post-event UAV imagery at ultra-high resolution with pixel-wise annotations covering buildings, roads, and other infrastructure classes. We use it to benchmark a post-event-only setting where pre-event references are unavailable.

5.2 Evaluation Metrics

We adopt standardized metrics tailored to each task. For extent mapping benchmarks, the primary target is the affected-region footprint. We report positive-class F_1 and Intersection over Union (IoU), which are more informative under class imbalance than Overall Accuracy (OA) alone. For URBANSARFLOODS, which distinguishes flooded open and urban areas, we report F_1 and IoU per subclass alongside OA and mean Intersection over Union (mIoU).

As for infrastructure damage mapping, we follow the protocols of the XVIEW2 Challenge [151] and the 2025 IEEE Data Fusion Contest [331] on XBD and BRIGHT datasets. Performance is quantified at two levels: localization F_1 (denoted as F_1^{loc}), measuring binary building detection accuracy, and damage classification F_1 (denoted as F_1^{dam}), the harmonic mean of per-class F1 scores computed on correctly localized pixels to decouple classification from detection errors. We additionally report OA and mIoU for holistic assessment. For RESCUENET, which involves multi-class segmentation of diverse infrastructure, OA and mIoU serve as the primary metrics.

5.3 Model Zoo for Disaster Mapping

To systematically assess the methodological landscape, we assemble a model zoo of over 30 representative methods, all evaluated under a unified training and evaluation protocol. These are organized into three groups: (i) general-purpose computer vision models, (ii) EO-specialized architectures designed for EO scene understanding, and (iii) foundation models pre-trained at scale on either natural images or EO data.

The first group comprises widely adopted architectures that serve as stable baselines for dense prediction. CNN and hybrid backbones include FCN [260], UNet [361], UNet++ [576], DeepLabV3+ [62], HRNet [473], and ConvNeXt [258]. Transformer-based models include Swin Transformer [257], SegFormer [506], and Mask2Former [68]. These establish a consistent reference for evaluating the marginal gains of more specialized designs.

The second group includes architectures designed for EO scene understanding. For single-temporal inputs, we evaluate UNetFormer [475], FarSeg [570], FarSeg++ [572], and RS3Mamba [275]. For bi-temporal settings, we include Siamese-style change detection architectures that explicitly model cross-temporal interactions: SiamCRNN [67], DSIFN [547], BIT-CD [56], ChangeOS [571], and ChangeMamba [59].

The third group evaluates foundation models along two trajectories. General vision FMs include SAM [213], SAM 2 [354], DINOv2 [322], DINOv3 [397], and CLIP [345]. Native EO FMs include SatMAE [83], SkySense [150], SpectralGPT [168], HyperSIGMA [468] and DINOv3 [397]. DINOv3 is located under both groups, as it provides both natural-image and EO-pretrained weights.

To support reproducibility and community adoption, we release an integrated toolbox[6] providing standardized dataset loaders, unified training and evaluation scripts, and ready-to-run implementations of all models. The repository is designed as a living benchmark that will be continuously updated with new datasets, architectures, and foundation model adaptations.

6. All code, model implementations, and evaluation protocols are publicly available at https://github.com/ChenHongruixuan/AnyDisasterMapping

TABLE 8
Comprehensive benchmark on CAUFLOOD, KUROSIWO, and URBANSARFLOODS. A **Family** column is added to group models into six broad categories for easier navigation. We report class-specific **IoU**, **OA**, **mF1**, and **mIoU** scores. Values denote the mean over 5 runs, with standard deviations in parentheses. The top three results are highlighted in **red**, orange, and yellow.

Family	Model	Venue	Encoder	Decoder	CAUFLOOD				KUROSIWO					URBANSARFLOODS				
					IoU_{flood}	OA	mF1	mIoU	IoU_{flood}	$IoU_{perm.}$	OA	mF1	mIoU	IoU_{open}	IoU_{urban}	OA	mF1	mIoU
Trad.	Random Guessing	-	-	-	4.27 (0.00)	81.86 (0.00)	49.07 (0.00)	42.99 (0.00)	3.70 (0.00)	1.88 (0.00)	70.25 (0.00)	31.36 (0.00)	25.46 (0.00)	0.67 (0.00)	0.01 (0.00)	97.26 (0.00)	33.32 (0.00)	32.65 (0.00)
	Index-Rule		-	-	33.99 (0.00)	73.19 (0.00)	66.16 (0.00)	51.44 (0.00)	39.18 (0.00)	46.37 (0.00)	92.70 (0.00)	72.00 (0.00)	59.49 (0.00)	30.95 (0.00)	0.48 (0.00)	98.45 (0.00)	49.14 (0.00)	43.29 (0.00)
	Random Forest		-	-	55.74 (0.00)	89.23 (0.00)	82.42 (0.00)	71.54 (0.00)	49.90 (0.00)	38.70 (0.00)	93.55 (0.00)	73.07 (0.00)	60.81 (0.00)	29.12 (0.00)	2.23 (0.00)	96.24 (0.00)	49.18 (0.00)	42.52 (0.00)
General Vision CNN	FCN-8s	$CVPR_{2015}$	VGG-16	-	80.52 (0.09)	96.96 (0.02)	93.72 (0.03)	88.52 (0.05)	63.47 (1.81)	61.98 (0.83)	97.12 (0.15)	84.27 (0.66)	74.26 (0.90)	58.89 (0.35)	11.34 (2.60)	99.40 (0.01)	64.70 (1.42)	56.54 (0.90)
	UNet	$MICCAI_{2015}$	-	-	**83.28** (0.15)	**97.43** (0.03)	**94.69** (0.05)	**90.17** (0.09)	64.47 (0.79)	**63.43** (0.76)	97.25 (0.04)	84.91 (0.28)	75.11 (0.38)	**67.33** (0.65)	13.25 (4.79)	**99.52** (0.02)	67.79 (2.47)	**60.03** (1.59)
	UNet++	TMI_{2019}	-	-	**83.15** (0.29)	**97.41** (0.05)	**94.64** (0.10)	**90.09** (0.17)	64.37 (0.52)	62.54 (0.11)	97.20 (0.02)	84.65 (0.11)	74.77 (0.15)	**67.25** (0.67)	8.59 (5.89)	**99.52** (0.01)	65.18 (3.29)	58.45 (1.80)
	DeepLabV3+	$ECCV_{2018}$	ResNet-50	ASPP	82.60 (0.44)	97.28 (0.11)	94.44 (0.16)	89.74 (0.28)	**65.34** (0.60)	63.04 (0.94)	**97.28** (0.07)	**85.03** (0.26)	**75.29** (0.35)	62.03 (0.92)	7.05 (1.72)	99.46 (0.01)	63.14 (1.16)	56.18 (0.79)
			ResNet-101	ASPP	82.60 (0.46)	97.30 (0.10)	94.45 (0.17)	89.75 (0.29)	**65.07** (1.43)	**63.62** (0.68)	**97.28** (0.07)	**85.10** (0.41)	**75.38** (0.56)	61.62 (1.00)	8.08 (1.36)	99.47 (0.01)	63.63 (0.96)	56.39 (0.73)
	HRNet	$TPAMI_{2021}$	HRNetV2-W18	-	82.71 (0.39)	97.34 (0.07)	94.50 (0.14)	89.84 (0.23)	64.17 (0.76)	63.28 (1.21)	97.25 (0.05)	84.79 (0.45)	74.96 (0.60)	61.84 (1.56)	6.98 (3.68)	99.47 (0.02)	63.01 (1.88)	56.10 (1.05)
			HRNetV2-W48	-	82.41 (0.29)	97.27 (0.05)	94.38 (0.10)	89.64 (0.17)	64.69 (1.63)	62.53 (1.38)	97.26 (0.10)	84.73 (0.67)	74.89 (0.90)	62.63 (1.13)	11.83 (1.19)	99.48 (0.01)	65.97 (0.35)	57.98 (0.04)
	ConvNeXt	$CVPR_{2022}$	ConvNeXt-T	UperNet	82.76 (0.13)	97.34 (0.02)	94.51 (0.04)	89.86 (0.07)	64.43 (0.74)	62.51 (1.89)	97.23 (0.10)	84.66 (0.60)	74.78 (0.81)	64.04 (0.72)	9.42 (2.10)	99.48 (0.01)	64.99 (1.17)	57.65 (0.70)
			ConvNeXt-S	UperNet	82.77 (0.29)	97.36 (0.05)	94.52 (0.10)	89.87 (0.18)	63.37 (1.10)	61.01 (3.34)	97.10 (0.09)	83.98 (1.03)	73.88 (1.33)	63.31 (0.79)	9.09 (2.22)	99.47 (0.02)	64.62 (1.33)	57.29 (0.85)
			ConvNeXt-B	UperNet	82.86 (0.32)	97.37 (0.06)	94.55 (0.11)	89.92 (0.19)	64.10 (1.38)	62.69 (1.86)	97.20 (0.15)	84.62 (0.78)	74.73 (1.05)	64.07 (0.73)	9.98 (1.49)	99.48 (0.01)	65.32 (0.97)	57.84 (0.71)
General Vision Transformer	Swin	$ICCV_{2021}$	Swin-T	UperNet	82.44 (0.09)	97.27 (0.02)	94.39 (0.04)	89.66 (0.06)	64.53 (0.60)	62.60 (1.25)	97.20 (0.04)	84.70 (0.20)	74.84 (0.26)	64.37 (0.79)	12.61 (3.04)	99.50 (0.01)	66.79 (1.54)	58.83 (0.97)
			Swin-S	UperNet	82.38 (0.16)	97.27 (0.03)	94.38 (0.06)	89.63 (0.10)	64.53 (1.25)	63.16 (0.87)	97.24 (0.07)	84.85 (0.39)	75.04 (0.53)	64.63 (0.61)	**15.35** (1.56)	99.49 (0.01)	**68.28** (0.82)	**59.82** (0.59)
			Swin-B	UperNet	82.37 (0.23)	97.27 (0.05)	94.37 (0.08)	89.62 (0.14)	64.69 (0.62)	62.89 (1.31)	97.21 (0.05)	84.82 (0.41)	74.99 (0.55)	64.48 (0.61)	**14.63** (1.92)	99.49 (0.01)	**67.88** (1.09)	59.54 (0.79)
	SegFormer	$NeurIPS_{2021}$	MiT-B0	MLP	82.57 (0.38)	97.32 (0.07)	94.45 (0.13)	89.75 (0.23)	64.11 (1.18)	61.20 (2.49)	97.19 (0.12)	84.25 (0.78)	74.25 (1.03)	62.46 (1.71)	8.66 (2.96)	99.45 (0.04)	64.14 (1.49)	56.85 (0.87)
			MiT-B1	MLP	82.59 (0.23)	97.32 (0.04)	94.45 (0.08)	89.76 (0.14)	63.58 (0.41)	62.59 (1.15)	97.20 (0.05)	84.47 (0.32)	74.53 (0.42)	63.93 (0.48)	9.61 (1.70)	99.48 (0.01)	65.08 (0.88)	57.67 (0.48)
			MiT-B2	MLP	82.38 (0.24)	97.28 (0.04)	94.38 (0.08)	89.63 (0.14)	64.06 (3.07)	63.26 (0.98)	97.25 (0.18)	84.75 (0.95)	74.92 (1.29)	64.01 (0.88)	9.78 (2.43)	99.48 (0.01)	65.18 (1.48)	57.76 (1.02)
			MiT-B3	MLP	82.62 (0.26)	97.33 (0.04)	94.46 (0.09)	89.78 (0.16)	**64.95** (1.41)	62.02 (1.94)	97.24 (0.10)	84.67 (0.83)	74.81 (1.10)	63.49 (0.14)	11.47 (1.38)	99.47 (0.01)	65.99 (0.71)	58.14 (0.42)
			MiT-B4	MLP	82.43 (0.24)	97.30 (0.04)	94.40 (0.09)	89.67 (0.14)	64.60 (1.26)	61.91 (0.81)	97.21 (0.09)	84.55 (0.51)	74.64 (0.69)	63.92 (0.28)	10.52 (1.04)	99.47 (0.01)	65.58 (0.58)	57.97 (0.37)
			MiT-B5	MLP	82.54 (0.14)	97.31 (0.03)	94.44 (0.05)	89.73 (0.09)	63.50 (1.93)	62.20 (0.74)	97.18 (0.10)	84.35 (0.66)	74.37 (0.87)	62.75 (1.17)	12.51 (3.00)	99.45 (0.04)	66.32 (1.46)	58.23 (0.81)
	Mask2Former	$CVPR_{2022}$	Swin-B	Mask2Former	82.74 (0.14)	97.34 (0.03)	94.51 (0.05)	89.85 (0.08)	63.65 (0.81)	61.51 (0.77)	97.05 (0.10)	84.19 (0.38)	74.14 (0.52)	62.54 (1.11)	**21.18** (1.17)	99.47 (0.01)	**70.54** (0.75)	**61.06** (0.71)
General Vision FM	SAM	$ICCV_{2023}$	ViT-B-16	DPT	81.96 (0.15)	97.21 (0.02)	94.23 (0.05)	89.38 (0.09)	63.46 (0.83)	62.12 (1.39)	97.12 (0.10)	84.30 (0.53)	74.29 (0.70)	61.36 (0.26)	9.26 (2.01)	99.46 (0.00)	64.23 (1.07)	56.69 (0.61)
			ViT-L-16	DPT	82.01 (0.15)	97.22 (0.03)	94.25 (0.05)	89.41 (0.09)	63.14 (1.14)	62.58 (1.03)	97.11 (0.09)	84.33 (0.39)	74.33 (0.52)	61.85 (0.24)	10.31 (2.02)	99.46 (0.01)	64.93 (1.10)	57.21 (0.70)
	SAM2	$ICLR_{2025}$	Hiera-S	UperNet	68.54 (0.10)	94.95 (0.03)	89.21 (0.04)	81.43 (0.06)	54.40 (1.81)	54.77 (0.86)	96.29 (0.15)	79.79 (0.71)	68.52 (0.86)	47.99 (0.80)	0.21 (0.25)	99.21 (0.02)	54.96 (0.34)	49.13 (0.31)
			Hiera-B+	UperNet	68.75 (0.15)	94.98 (0.04)	89.29 (0.06)	81.55 (0.09)	55.19 (1.21)	55.39 (0.95)	96.39 (0.06)	80.21 (0.54)	69.03 (0.66)	47.71 (0.41)	0.91 (0.49)	99.21 (0.02)	55.33 (0.34)	49.27 (0.21)
	CLIP	$ICML_{2021}$	ViT-B-16	DPT	79.41 (0.46)	96.82 (0.07)	93.34 (0.16)	87.89 (0.27)	59.62 (1.19)	58.73 (1.41)	96.73 (0.15)	82.38 (0.68)	71.76 (0.88)	54.84 (2.19)	4.46 (1.40)	99.34 (0.05)	59.67 (0.73)	52.88 (0.64)
			ViT-L-14	DPT	79.91 (0.99)	96.92 (0.17)	93.52 (0.36)	88.20 (0.59)	59.56 (0.80)	59.80 (1.07)	96.83 (0.06)	82.66 (0.39)	72.13 (0.49)	56.32 (0.80)	3.93 (3.04)	99.36 (0.02)	59.72 (1.76)	53.20 (0.95)
	DINOV2	$TMLR_{2024}$	ViT-B-14	DPT	82.16 (0.22)	97.24 (0.04)	94.30 (0.08)	89.50 (0.13)	62.88 (1.49)	60.77 (2.22)	97.04 (0.18)	83.79 (0.96)	73.62 (1.26)	59.85 (1.28)	10.51 (3.62)	99.43 (0.03)	64.48 (1.92)	56.59 (1.22)
			ViT-L-14	DPT	81.79 (0.28)	97.18 (0.05)	94.17 (0.10)	89.28 (0.17)	62.72 (2.11)	59.57 (1.92)	96.98 (0.10)	83.43 (0.50)	73.16 (0.67)	60.97 (0.53)	10.97 (1.98)	99.45 (0.01)	65.07 (1.14)	57.13 (0.77)
	DINOV3	$TMLR_{2026}$	ViT-B-16	DPT	81.49 (0.22)	97.13 (0.04)	94.06 (0.08)	89.10 (0.13)	63.11 (0.34)	62.06 (1.41)	97.11 (0.10)	84.20 (0.37)	74.16 (0.48)	59.41 (1.00)	7.55 (3.85)	99.42 (0.02)	62.70 (1.93)	55.46 (0.99)
			ViT-L-16	DPT	82.41 (0.12)	97.29 (0.02)	94.39 (0.04)	89.66 (0.07)	62.50 (0.76)	62.05 (1.89)	97.05 (0.13)	84.03 (0.61)	73.92 (0.81)	61.44 (1.11)	10.33 (4.46)	99.46 (0.01)	64.78 (2.58)	57.08 (1.73)
EO-Ori.	UNetFormer	$ISPRS_{2022}$	ResNet-18	UNetFormer	**82.96** (0.12)	**97.39** (0.02)	**94.58** (0.04)	**89.98** (0.07)	64.14 (0.52)	62.27 (1.03)	97.17 (0.06)	84.52 (0.37)	74.59 (0.50)	61.92 (0.80)	4.29 (1.39)	99.46 (0.01)	61.47 (0.84)	55.22 (0.50)
	FarSeg	$CVPR_{2020}$	ResNet-50	FarSeg	82.48 (0.17)	97.30 (0.03)	94.41 (0.06)	89.69 (0.10)	64.04 (0.49)	61.86 (0.50)	97.17 (0.04)	84.40 (0.19)	74.44 (0.25)	60.53 (0.83)	7.69 (1.30)	99.44 (0.01)	63.13 (0.94)	55.88 (0.69)
	FarSeg++	$TPAMI_{2023}$	MiT-B2	FarSeg++	82.84 (0.23)	97.37 (0.04)	94.54 (0.08)	89.91 (0.14)	64.68 (0.60)	62.79 (0.58)	97.24 (0.04)	84.80 (0.17)	74.97 (0.23)	64.13 (0.70)	14.21 (2.90)	99.48 (0.01)	67.56 (1.55)	59.27 (1.05)
	RS^3Mamba	$GRSL_{2024}$	VMamba-T	UNetFormer	82.24 (0.28)	97.25 (0.06)	94.33 (0.10)	89.54 (0.17)	64.68 (1.18)	**63.68** (0.66)	**97.30** (0.07)	**85.02** (0.45)	**75.28** (0.62)	62.46 (0.97)	11.24 (3.23)	99.45 (0.01)	65.57 (1.74)	57.72 (1.09)
Bi-temporal	SiamCRNN	$TGRS_{2020}$	ResNet-50	ConvLSTM	82.75 (0.11)	97.39 (0.02)	94.52 (0.04)	89.88 (0.07)	62.67 (0.81)	57.81 (1.12)	96.88 (0.11)	82.95 (0.48)	72.53 (0.62)	57.99 (1.18)	6.86 (2.52)	99.40 (0.02)	61.95 (1.59)	54.75 (1.02)
	DSIFN	$ISPRS_{2020}$	VGG-16	DSIFN	82.24 (1.03)	97.35 (0.15)	94.36 (0.35)	89.61 (0.60)	61.76 (1.20)	61.40 (0.90)	97.10 (0.05)	83.69 (0.33)	73.48 (0.43)	57.38 (1.01)	2.15 (2.46)	99.39 (0.03)	58.91 (1.81)	52.97 (1.15)
	BIT-CD	$TGRS_{2021}$	ResNet-50	BIT	78.35 (0.17)	96.58 (0.03)	92.94 (0.06)	87.22 (0.10)	62.69 (0.44)	60.09 (1.05)	96.94 (0.05)	83.55 (0.32)	73.29 (0.42)	50.01 (0.94)	1.12 (1.58)	99.27 (0.03)	56.16 (1.07)	50.13 (0.63)
	ChangeOS	RSE_{2021}	ResNet-50	FPN	81.50 (0.69)	97.27 (0.10)	94.11 (0.24)	89.20 (0.40)	62.53 (0.88)	58.49 (1.81)	96.90 (0.17)	83.10 (0.73)	72.73 (0.94)	51.44 (1.35)	9.39 (1.50)	99.29 (0.02)	61.57 (0.70)	53.38 (0.45)
	ChangeMamba	$TGRS_{2024}$	VMamba-T	STSS	82.03 (0.68)	97.28 (0.11)	94.27 (0.24)	89.46 (0.40)	63.53 (1.10)	58.79 (1.65)	96.93 (0.14)	83.43 (0.72)	73.16 (0.95)	62.01 (1.03)	12.81 (2.84)	99.44 (0.02)	66.29 (1.57)	58.09 (1.01)
Native EO FM	SatMAE	$NeurIPS_{2022}$	ViT-B-16	DPT	81.00 (0.19)	97.05 (0.03)	93.89 (0.07)	88.81 (0.11)	61.08 (2.94)	60.98 (0.88)	96.96 (0.15)	83.37 (0.89)	73.07 (1.16)	59.46 (0.77)	8.24 (4.28)	99.41 (0.01)	63.09 (2.26)	55.70 (1.26)
	SkySense	$CVPR_{2024}$	Swin-V2-H	DPT	81.09 (0.16)	97.03 (0.03)	93.91 (0.06)	88.84 (0.10)	63.50 (0.85)	61.74 (0.68)	97.07 (0.03)	84.20 (0.09)	74.16 (0.12)	64.05 (0.69)	14.38 (0.91)	99.48 (0.01)	67.65 (0.53)	59.30 (0.42)
	SpectralGPT	$TPAMI_{2024}$	ViT-B-8	DPT	82.73 (0.26)	97.34 (0.06)	94.50 (0.10)	89.84 (0.16)	63.49 (1.82)	62.34 (0.75)	97.18 (0.09)	84.38 (0.57)	74.41 (0.76)	60.72 (0.50)	10.33 (2.54)	99.45 (0.01)	64.64 (1.37)	56.83 (0.84)
	HyperSIGMA	$TPAMI_{2025}$	ViT	DPT	81.58 (0.49)	97.17 (0.07)	94.10 (0.17)	89.17 (0.28)	64.39 (0.80)	60.78 (0.87)	97.13 (0.07)	84.20 (0.17)	74.17 (0.22)	**65.12** (8.75)	11.76 (3.03)	**99.51** (0.13)	66.44 (2.30)	58.80 (2.91)
	DINOV3	$TMLR_{2026}$	ViT-L-16*	DPT	82.39 (0.09)	97.29 (0.02)	94.38 (0.03)	89.64 (0.05)	62.77 (0.56)	60.00 (1.08)	96.94 (0.05)	83.55 (0.34)	73.30 (0.45)	61.19 (1.30)	11.06 (3.97)	99.45 (0.01)	65.13 (2.17)	57.23 (1.43)

5.4 Benchmark Results

The benchmark results across over 30 methods, 13 datasets, and two task families reveal several cross-cutting findings that challenge prevailing assumptions about model selection for disaster mapping. Tables 8–11 present the full quantitative results for flood extent mapping, landslide mapping, wildfire mapping and building damage mapping, respectively.

Finding 1: No single architecture dominates across tasks and hazards. For bi-temporal building damage mapping, ChangeMamba achieves the best performance on both xBD (mIoU = 64.24%) and BRIGHT (mIoU = 67.93%), indicating that explicit cross-temporal interaction is beneficial when pre- and post-event imagery are available. On the post-event-only RESCUENET benchmark, DINOv2 (ViT-B-14) achieves the highest mIoU of 65.30%, suggesting that strong general-purpose visual representations can be more effective when temporal evidence is absent. For disaster extent mapping, the winners vary even within the same hazard family: UNet leads on CAUFLOOD, DeepLabV3+ with ResNet-101 achieves the best mIoU on KUROSIWO, and Mask2Former obtains the highest mIoU on URBANSARFLOODS. A similar pattern appears for landslide and wildfire mapping: SegFormer-B4 leads on LANDSLIDE4SENSE, FarSeg leads on HR-GLDD, Swin-T leads on GVLM, while Mask2Former achieves the highest scores on S2WCD and FLOGA. These results rule out a universal architectural recommendation and instead point to hazard-, modality-, and dataset-specific inductive biases.

Finding 2: Traditional methods reveal when disaster mapping is feature-separable. The traditional baselines are not merely lower bounds. They help identify tasks where the target is already well expressed by low-level spectral, backscatter, topographic, or temporal cues. Random Forest reaches 71.54% mIoU on CAUFLOOD, 60.81% on KUROSIWO, 65.45% IoU on GVLM, and 92.67% F1 on S2WCD. On S2WCD, even the hand-engineered index rule reaches 90.91% F1 and 83.33% IoU, close to many learned models. The limitation is equally clear: on URBANSARFLOODS, Random Forest obtains only 2.23% IoU for the urban-flood class, and on xBD, BRIGHT, and RESCUENET its mIoU stays well below the deep baselines. Where traditional methods are competitive, much of the signal needed for the task is already in the per-pixel feature

TABLE 9
Comprehensive benchmark on LANDSLIDE4SENSE, HR-GLDD, and GVLM. A **Family** column is added to group models into six broad categories for easier navigation. We report **OA**, **F1**, and **IoU** scores. Values denote the mean over 5 runs, with standard deviations in parentheses. The top three results in each metric are highlighted in **red**, orange, and yellow.

Family	Model	Venue	Encoder	Decoder	LANDSLIDE4SENSE			HR-GLDD			GVLM		
					OA	F1	IoU	OA	F1	IoU	OA	F1	IoU
Trad.	Random Guessing	-	-	-	95.88 (0.00)	2.08 (0.01)	1.05 (0.01)	81.19 (0.01)	10.50 (0.02)	5.54 (0.01)	87.54 (0.00)	6.67 (0.00)	3.45 (0.00)
	Index-Rule	-	-	-	97.97 (0.00)	48.83 (0.00)	32.30 (0.00)	87.69 (0.00)	58.74 (0.00)	41.58 (0.00)	95.08 (0.00)	71.10 (0.00)	55.16 (0.00)
	Random Forest	-	-	-	97.70 (0.00)	51.47 (0.00)	34.65 (0.00)	90.64 (0.00)	67.47 (0.00)	50.91 (0.00)	96.59 (0.00)	79.12 (0.00)	65.45 (0.00)
General Vision CNN	FCN-8s	$CVPR_{2015}$	VGG-16	-	98.58 (0.09)	62.29 (0.56)	45.23 (0.60)	93.26 (0.09)	68.53 (0.36)	52.12 (0.42)	98.33 (0.02)	88.12 (0.10)	78.76 (0.16)
	UNet	$MICCAI_{2015}$	-	-	**98.67** (0.03)	64.08 (1.48)	47.16 (1.62)	**94.10** (0.11)	**72.42** (0.47)	**56.76** (0.57)	98.46 (0.03)	88.95 (0.19)	80.10 (0.31)
	UNet++	TMI_{2019}	-	-	98.60 (0.07)	62.63 (2.93)	45.64 (3.03)	**94.07** (0.09)	72.25 (0.44)	56.55 (0.54)	98.39 (0.04)	88.47 (0.28)	79.32 (0.46)
	DeepLabV3+	$ECCV_{2018}$	ResNet-50	ASPP	98.56 (0.11)	62.27 (0.72)	45.22 (0.76)	93.45 (0.61)	70.51 (0.95)	54.45 (1.13)	98.41 (0.03)	88.58 (0.25)	79.50 (0.41)
			ResNet-101	ASPP	98.52 (0.12)	62.61 (2.49)	45.61 (2.63)	93.42 (0.38)	70.51 (1.06)	54.46 (1.26)	98.41 (0.03)	88.46 (0.19)	79.31 (0.30)
	HRNet	$TPAMI_{2021}$	HRNetV2-W18	-	98.56 (0.11)	59.89 (1.46)	42.76 (1.49)	93.60 (0.07)	68.90 (0.81)	52.56 (0.95)	98.38 (0.02)	88.37 (0.17)	79.17 (0.27)
			HRNetV2-W48	-	98.57 (0.09)	62.56 (1.04)	45.53 (1.09)	93.73 (0.11)	70.62 (0.38)	54.59 (0.45)	98.45 (0.01)	88.83 (0.11)	79.91 (0.17)
	ConvNeXt	$CVPR_{2022}$	ConvNeXt-T	UperNet	98.48 (0.17)	63.63 (1.61)	46.68 (1.75)	93.50 (0.15)	70.80 (0.86)	54.80 (1.02)	98.36 (0.04)	88.20 (0.17)	78.90 (0.27)
			ConvNeXt-S	UperNet	98.55 (0.21)	64.16 (2.40)	47.27 (2.61)	93.77 (0.19)	70.95 (0.64)	54.98 (0.77)	98.37 (0.04)	88.35 (0.21)	79.13 (0.34)
			ConvNeXt-B	UperNet	98.54 (0.15)	64.54 (1.33)	47.65 (1.45)	93.62 (0.18)	71.02 (0.36)	55.06 (0.43)	98.38 (0.03)	88.32 (0.16)	79.09 (0.26)
General Vision Transformer	Swin	$ICCV_{2021}$	Swin-T	UperNet	98.56 (0.07)	64.36 (0.69)	47.46 (0.75)	93.72 (0.07)	71.20 (0.45)	55.28 (0.54)	**98.53** (0.01)	**89.38** (0.05)	**80.79** (0.08)
			Swin-S	UperNet	98.57 (0.08)	64.37 (0.76)	47.46 (0.83)	93.70 (0.09)	70.77 (0.44)	54.76 (0.53)	98.50 (0.02)	**89.25** (0.11)	**80.58** (0.19)
			Swin-B	UperNet	98.58 (0.10)	64.34 (1.09)	47.44 (1.19)	93.76 (0.17)	71.25 (0.71)	55.35 (0.85)	**98.51** (0.03)	**89.32** (0.12)	**80.69** (0.19)
	SegFormer	$NeurIPS_{2021}$	MiT-B0	MLP	**98.68** (0.12)	**66.35** (1.18)	**49.65** (1.33)	93.39 (0.22)	68.99 (0.70)	52.66 (0.82)	98.44 (0.04)	88.83 (0.20)	79.90 (0.32)
			MiT-B1	MLP	98.64 (0.10)	66.06 (1.72)	49.34 (1.88)	93.48 (0.14)	70.03 (0.62)	53.88 (0.74)	98.46 (0.03)	88.88 (0.16)	79.99 (0.26)
			MiT-B2	MLP	98.62 (0.24)	**66.12** (2.60)	**49.43** (2.87)	93.58 (0.15)	70.05 (0.62)	53.91 (0.74)	98.47 (0.03)	89.00 (0.16)	80.19 (0.26)
			MiT-B3	MLP	98.59 (0.08)	65.30 (1.13)	48.49 (1.25)	93.55 (0.17)	69.99 (0.50)	53.84 (0.59)	98.45 (0.03)	88.97 (0.15)	80.13 (0.24)
			MiT-B4	MLP	**98.73** (0.05)	**67.13** (1.02)	**50.53** (1.16)	93.54 (0.17)	70.79 (0.71)	54.80 (0.85)	98.46 (0.02)	89.00 (0.12)	80.18 (0.20)
			MiT-B5	MLP	98.60 (0.13)	65.78 (0.40)	49.01 (0.45)	93.43 (0.10)	69.44 (0.43)	53.19 (0.50)	98.42 (0.03)	88.74 (0.20)	79.76 (0.32)
	Mask2Former	$CVPR_{2022}$	Swin-B	Mask2Former	98.55 (0.08)	62.84 (1.44)	45.83 (1.53)	94.05 (0.06)	**72.38** (0.14)	**56.71** (0.17)	98.48 (0.05)	89.15 (0.36)	80.42 (0.59)
General Vision FM	SAM	$ICCV_{2023}$	ViT-B-16	DPT	98.45 (0.09)	58.86 (1.64)	41.72 (1.65)	93.25 (0.18)	68.80 (0.65)	52.44 (0.75)	98.38 (0.03)	88.39 (0.16)	79.19 (0.26)
			ViT-L-16	DPT	98.40 (0.14)	57.91 (2.24)	40.79 (2.18)	92.96 (0.25)	67.84 (0.59)	51.34 (0.67)	98.38 (0.04)	88.50 (0.21)	79.37 (0.33)
	SAM2	$ICLR_{2025}$	Hiera-S	UperNet	98.57 (0.06)	61.56 (0.88)	44.47 (0.91)	93.34 (0.15)	68.50 (0.71)	52.10 (0.82)	98.21 (0.02)	87.03 (0.09)	77.04 (0.15)
			Hiera-B+	UperNet	98.48 (0.08)	59.86 (1.31)	42.73 (1.32)	93.06 (0.19)	67.95 (0.86)	51.46 (0.98)	98.23 (0.02)	87.27 (0.11)	77.41 (0.18)
	CLIP	$ICML_{2021}$	ViT-B-16	DPT	98.53 (0.10)	60.42 (3.58)	43.37 (3.60)	92.97 (0.24)	67.71 (1.08)	51.19 (1.24)	98.18 (0.06)	86.85 (0.34)	76.76 (0.53)
			ViT-L-14	DPT	98.49 (0.11)	58.08 (2.42)	40.96 (2.40)	92.37 (0.63)	65.56 (1.55)	48.78 (1.70)	98.34 (0.02)	88.12 (0.11)	78.76 (0.18)
	DINOV2	$TMLR_{2024}$	ViT-B-14	DPT	98.45 (0.13)	61.89 (1.26)	44.82 (1.31)	92.36 (0.27)	65.03 (1.63)	48.20 (1.77)	98.15 (0.11)	86.64 (0.96)	76.44 (1.49)
			ViT-L-14	DPT	98.53 (0.10)	62.48 (1.49)	45.44 (1.57)	93.51 (0.18)	70.24 (0.92)	54.13 (1.09)	98.14 (0.14)	86.64 (0.86)	76.44 (1.34)
	DINOV3	$TMLR_{2026}$	ViT-B-16	DPT	98.48 (0.08)	61.66 (0.78)	44.58 (0.81)	93.10 (0.10)	68.18 (0.40)	51.72 (0.47)	98.28 (0.07)	87.67 (0.40)	78.04 (0.62)
			ViT-L-16	DPT	98.51 (0.07)	59.08 (3.54)	41.99 (3.53)	93.12 (0.49)	68.40 (2.35)	52.01 (2.66)	98.29 (0.08)	87.71 (0.52)	78.11 (0.82)
EO-Ori.	UNetFormer	$ISPRS_{2022}$	ResNet-18	UNetFormer	98.55 (0.09)	63.60 (1.16)	46.64 (1.25)	93.32 (0.05)	66.73 (0.67)	50.07 (0.76)	98.34 (0.01)	88.05 (0.07)	78.65 (0.10)
	FarSeg	$CVPR_{2020}$	ResNet-50	FarSeg	98.56 (0.11)	61.39 (1.70)	44.30 (1.77)	**94.19** (0.04)	**72.58** (0.52)	**56.97** (0.65)	98.41 (0.02)	88.60 (0.14)	79.53 (0.23)
	FarSeg++	$TPAMI_{2023}$	MiT-B2	FarSeg++	98.66 (0.12)	65.20 (1.38)	48.38 (1.51)	93.64 (0.17)	70.46 (0.64)	54.40 (0.77)	98.31 (0.07)	87.89 (0.39)	78.40 (0.61)
	RS^3Mamba	$GRSL_{2024}$	VMamba-T	UNetFormer	98.54 (0.12)	63.20 (1.41)	46.21 (1.53)	93.76 (0.10)	71.40 (0.47)	55.52 (0.57)	98.33 (0.04)	88.08 (0.26)	78.70 (0.41)
Bi-temporal	SiamCRNN	$TGRS_{2020}$	ResNet-50	ConvLSTM	-	-	-	-	-	-	98.18 (0.07)	86.79 (0.39)	76.66 (0.60)
	DSIFN	$ISPRS_{2020}$	VGG-16	DSIFN	-	-	-	-	-	-	98.37 (0.02)	88.03 (0.29)	78.61 (0.46)
	BIT-CD	$TGRS_{2021}$	ResNet-50	BIT	-	-	-	-	-	-	98.03 (0.03)	85.81 (0.16)	75.14 (0.25)
	ChangeOS	RSE_{2021}	ResNet-50	FPN	-	-	-	-	-	-	98.18 (0.06)	86.81 (0.32)	76.70 (0.50)
	ChangeMamba	$TGRS_{2024}$	VMamba-T	STSS	-	-	-	-	-	-	**98.51** (0.05)	89.24 (0.35)	80.57 (0.57)
Native EO FM	SatMAE	$NeurIPS_{2022}$	ViT	DPT	98.50 (0.24)	62.54 (1.41)	45.51 (1.48)	93.17 (0.03)	67.95 (0.49)	51.46 (0.56)	98.30 (0.04)	87.80 (0.25)	78.26 (0.40)
	SkySense	$CVPR_{2024}$	ViT	DPT	98.59 (0.09)	65.38 (1.65)	48.58 (1.81)	93.58 (0.17)	69.83 (0.33)	53.64 (0.39)	98.44 (0.04)	88.87 (0.14)	79.97 (0.22)
	SpectralGPT	$TPAMI_{2024}$	ViT	DPT	98.49 (0.20)	62.92 (2.11)	45.93 (2.25)	93.33 (0.16)	68.59 (0.66)	52.20 (0.77)	98.41 (0.03)	88.48 (0.27)	79.34 (0.44)
	HyperSIGMA	$TPAMI_{2025}$	ViT	DPT	98.60 (0.02)	61.95 (1.38)	44.89 (1.45)	93.34 (0.19)	67.94 (1.10)	51.45 (1.26)	98.23 (0.05)	87.17 (0.31)	77.26 (0.49)
	DINOV3	$TMLR_{2026}$	ViT-L-16*	DPT	98.51 (0.10)	61.45 (1.23)	44.36 (1.28)	93.33 (0.13)	69.25 (0.86)	52.97 (1.00)	98.34 (0.03)	87.98 (0.24)	78.53 (0.38)

distribution; where they collapse, that distribution alone is no longer sufficient, and identifying what additional information is required (e.g., spatial context, temporal comparison, or finer semantics) becomes a per-task question that the benchmark itself cannot fully answer.

Finding 3: Explicit temporal modelling mainly benefits damage mapping. As reported in Table 11, the strongest change detection models, including ChangeMamba (mIoU = 64.24% on xBD, 67.93% on BRIGHT) and ChangeOS (62.59% on xBD), outperform the best early-fusion baseline on these benchmarks (Swin-B at 62.15% on xBD). The gain is not automatic across the family, however: SiamCRNN (61.50%), DSIFN (59.28%), and BIT-CD (56.28%) all trail Swin-B on xBD despite using two-stream formulations. In contrast, on flood and wildfire extent mapping, change detection models offer no systematic advantage and sometimes underperform early-fusion baselines that take the same two-timestamp input. Table 8 shows that on KUROSIWO, ChangeMamba (mIoU = 73.16%) falls below UNet (75.11%) and DeepLabV3+ R101 (75.38%). On wildfire, ChangeMamba is mid-pack on FLOGA (F_1 = 91.23%, behind Mask2Former at 92.18% and several CNNs). The split aligns with the extent-versus-damage distinction in our taxonomy: extent mapping asks whether each pixel belongs to a hazard-affected surface state, a question that early fusion of the two timestamps appears to answer adequately, whereas damage mapping requires judging how an object changed relative to its pre-event condition, which benefits from architectures that keep the two phases' representations explicitly comparable.

Finding 4: Foundation models do not consistently outperform task-specific baselines. Across all datasets, general

TABLE 10
Comprehensive benchmark on S2WCD, SBAD, FLOGA, and WILDFIRESPREADTS. A **Family** column is added to group models into six broad categories for easier navigation. We report **OA**, **F1** (or **mF1**), and **IoU** (or **mIoU**) scores. Values denote the mean over 5 runs, with standard deviations in parentheses. The top three results in each metric are highlighted in **red**, orange, and yellow.

Family	Model	Venue	Encoder	Decoder	S2WCD			SBAD			FLOGA			WILDFIRESPREADTS		
					OA	F1	IoU	OA	mF1	mIoU	OA	F1	IoU	OA	F1	IoU
Trad.	Random Guessing	-	-	-	54.59 (0.02)	34.40 (0.02)	20.76 (0.02)	86.79 (0.01)	19.29 (0.01)	17.75 (0.01)	89.59 (0.00)	5.27 (0.02)	2.71 (0.01)	99.58 (0.00)	0.18 (0.01)	0.09 (0.00)
	Index-Rule		-	-	93.91 (0.00)	90.91 (0.00)	83.33 (0.00)	26.18 (0.00)	10.90 (0.00)	6.59 (0.00)	97.91 (0.00)	72.80 (0.00)	57.24 (0.00)	98.54 (0.00)	26.02 (0.00)	14.95 (0.00)
	Random Forest		-	-	95.14 (0.00)	92.67 (0.00)	86.35 (0.00)	86.50 (0.00)	19.28 (0.00)	17.72 (0.00)	97.89 (0.00)	79.44 (0.00)	65.89 (0.00)	99.40 (0.00)	40.91 (0.00)	25.71 (0.00)
General Vision CNN	FCN-8s	$CVPR_{2015}$	VGG-16	-	92.26 (0.15)	88.24 (0.31)	78.95 (0.50)	**76.04** (0.64)	24.11 (3.87)	19.17 (2.25)	99.10 (0.09)	86.50 (1.00)	76.22 (1.54)	99.65 (0.02)	51.96 (0.41)	35.10 (0.37)
	UNet	$MICCAI_{2015}$	-	-	95.01 (0.25)	92.50 (0.35)	86.05 (0.60)	69.86 (3.27)	**28.94** (1.99)	**21.53** (0.92)	**99.45** (0.04)	**91.52** (0.59)	**84.36** (0.99)	**99.74** (0.01)	**60.36** (0.41)	**43.22** (0.42)
	UNet++	TMI_{2019}	-	-	94.90 (0.43)	92.32 (0.65)	85.74 (1.11)	72.77 (1.94)	**29.66** (1.25)	**22.36** (0.51)	99.30 (0.16)	89.58 (1.92)	81.17 (3.15)	**99.74** (0.01)	**60.33** (0.21)	**43.19** (0.21)
	DeepLabV3+	$ECCV_{2018}$	ResNet-50	ASPP	94.00 (0.93)	90.94 (1.28)	83.41 (2.15)	72.38 (6.39)	21.91 (2.86)	17.33 (0.97)	99.29 (0.08)	88.92 (1.06)	80.07 (1.71)	99.71 (0.01)	57.37 (0.47)	40.23 (0.46)
			ResNet-101	ASPP	92.97 (0.89)	89.05 (1.47)	80.28 (2.38)	72.70 (3.99)	21.72 (2.59)	17.23 (1.11)	99.31 (0.05)	89.04 (0.86)	80.26 (1.40)	99.72 (0.01)	57.56 (0.17)	40.41 (0.17)
	HRNet	$TPAMI_{2021}$	HRNetV2-W18	-	93.64 (0.94)	90.57 (1.31)	82.78 (2.17)	**75.53** (4.03)	20.50 (3.36)	17.08 (2.15)	99.41 (0.04)	90.93 (0.51)	83.37 (0.85)	99.71 (0.01)	57.22 (0.10)	40.08 (0.10)
			HRNetV2-W48	-	94.53 (0.73)	91.81 (1.00)	84.88 (1.68)	70.53 (4.71)	23.17 (2.43)	18.03 (1.25)	99.43 (0.04)	91.13 (0.59)	83.71 (1.00)	99.71 (0.01)	57.43 (0.14)	40.28 (0.13)
	ConvNeXt	$CVPR_{2022}$	ConvNeXt-T	UperNet	93.66 (1.34)	90.56 (1.74)	82.78 (2.89)	64.18 (16.12)	23.81 (3.39)	17.77 (3.54)	99.36 (0.17)	90.35 (2.16)	82.45 (3.51)	99.72 (0.01)	58.57 (0.20)	41.41 (0.20)
			ConvNeXt-S	UperNet	94.96 (0.34)	92.39 (0.45)	85.85 (0.77)	67.48 (7.19)	26.36 (3.12)	19.65 (1.92)	99.23 (0.15)	88.71 (1.88)	79.75 (3.04)	99.72 (0.01)	**58.72** (0.14)	**41.56** (0.14)
			ConvNeXt-B	UperNet	94.70 (0.23)	92.01 (0.33)	85.20 (0.56)	65.09 (7.87)	23.41 (4.29)	17.66 (2.67)	99.39 (0.04)	90.70 (0.42)	82.99 (0.71)	99.72 (0.01)	58.52 (0.20)	41.36 (0.20)
General Vision Transformer	Swin	$ICCV_{2021}$	Swin-T	UperNet	95.00 (0.23)	92.43 (0.30)	85.92 (0.52)	70.18 (4.78)	23.09 (1.11)	17.81 (0.49)	99.44 (0.01)	91.23 (0.19)	83.88 (0.33)	99.72 (0.01)	58.37 (0.44)	41.21 (0.44)
			Swin-S	UperNet	94.85 (0.22)	92.27 (0.28)	85.66 (0.48)	69.44 (2.97)	23.92 (2.65)	18.10 (1.61)	**99.45** (0.02)	91.42 (0.24)	84.20 (0.40)	**99.73** (0.01)	58.24 (0.16)	41.08 (0.16)
			Swin-B	UperNet	95.06 (0.19)	92.55 (0.23)	86.14 (0.40)	71.60 (2.79)	23.63 (2.47)	18.31 (1.35)	99.45 (0.03)	**91.55** (0.34)	**84.41** (0.57)	99.73 (0.01)	58.01 (0.69)	40.86 (0.68)
	SegFormer	$NeurIPS_{2021}$	MiT-B0	MLP	91.24 (1.01)	86.66 (1.42)	76.48 (2.24)	73.37 (1.26)	23.91 (1.84)	18.61 (1.06)	99.33 (0.01)	89.40 (0.20)	80.84 (0.33)	99.70 (0.00)	55.92 (0.13)	38.81 (0.12)
			MiT-B1	MLP	92.50 (0.56)	88.30 (0.88)	79.07 (1.41)	72.86 (2.13)	22.99 (3.24)	17.95 (2.03)	99.30 (0.03)	89.13 (0.53)	80.39 (0.86)	99.70 (0.01)	55.93 (0.40)	38.83 (0.39)
			MiT-B2	MLP	87.11 (2.63)	78.79 (5.27)	65.24 (7.07)	72.53 (2.07)	21.07 (1.70)	16.82 (1.03)	99.36 (0.06)	90.16 (0.77)	82.09 (1.28)	99.71 (0.00)	56.05 (0.22)	38.94 (0.21)
			MiT-B3	MLP	89.90 (1.69)	83.89 (2.76)	72.33 (4.11)	71.84 (1.20)	22.28 (1.75)	17.43 (0.89)	99.40 (0.02)	90.74 (0.22)	83.05 (0.38)	99.70 (0.01)	56.13 (0.23)	39.01 (0.22)
			MiT-B4	MLP	84.65 (2.82)	73.76 (5.60)	58.68 (7.29)	70.93 (2.53)	23.07 (2.05)	17.80 (1.07)	99.40 (0.02)	90.67 (0.29)	82.93 (0.49)	99.70 (0.01)	56.10 (0.23)	38.98 (0.22)
			MiT-B5	MLP	85.48 (1.40)	75.50 (2.86)	60.71 (3.78)	72.60 (0.52)	21.43 (1.01)	17.10 (0.59)	99.39 (0.03)	90.67 (0.37)	82.93 (0.62)	99.69 (0.01)	56.21 (0.27)	39.09 (0.26)
	Mask2Former	$CVPR_{2022}$	Swin-B	Mask2Former	**95.57** (0.36)	**93.30** (0.47)	**87.44** (0.82)	68.37 (8.30)	24.54 (4.54)	18.57 (3.37)	**99.49** (0.03)	**92.18** (0.43)	**85.50** (0.74)	99.72 (0.01)	56.47 (0.41)	39.34 (0.40)
General Vision FM	SAM	$ICCV_{2023}$	ViT-B-16	DPT	94.07 (0.50)	91.16 (0.63)	83.76 (1.05)	70.82 (4.62)	24.64 (4.03)	18.72 (1.80)	99.35 (0.04)	89.90 (0.58)	81.66 (0.95)	99.68 (0.01)	52.70 (0.57)	35.78 (0.52)
			ViT-L-16	DPT	94.80 (0.27)	92.11 (0.39)	85.37 (0.67)	71.96 (3.62)	24.42 (1.74)	18.72 (0.91)	99.33 (0.06)	89.76 (0.87)	81.43 (1.43)	99.68 (0.02)	52.37 (0.59)	35.47 (0.54)
	SAM2	$ICLR_{2025}$	Hiera-S	UperNet	94.78 (0.34)	92.16 (0.48)	85.46 (0.83)	67.48 (3.06)	25.89 (0.99)	19.13 (0.61)	99.18 (0.07)	87.89 (0.84)	78.41 (1.34)	99.71 (0.01)	57.00 (0.28)	39.86 (0.28)
			Hiera-B+	UperNet	95.00 (0.19)	92.46 (0.24)	85.98 (0.42)	67.46 (3.14)	26.56 (2.58)	19.62 (1.52)	99.14 (0.17)	87.53 (2.05)	77.87 (3.25)	99.72 (0.01)	57.32 (0.35)	40.18 (0.34)
	CLIP	$ICML_{2021}$	ViT-B-16	DPT	93.98 (0.54)	91.03 (0.67)	83.54 (1.12)	69.30 (5.19)	25.18 (2.55)	18.88 (1.40)	99.15 (0.21)	87.24 (2.72)	77.46 (4.22)	99.66 (0.01)	49.76 (1.43)	33.13 (1.27)
			ViT-L-14	DPT	**95.34** (0.20)	**92.89** (0.30)	**86.73** (0.51)	71.28 (3.57)	25.37 (3.80)	19.36 (2.04)	99.28 (0.11)	89.07 (1.38)	80.31 (2.21)	99.67 (0.01)	52.37 (0.80)	35.47 (0.73)
	DINOV2	$TMLR_{2024}$	ViT-B-14	DPT	94.90 (0.18)	92.25 (0.26)	85.62 (0.45)	69.27 (6.27)	24.90 (2.44)	18.80 (1.83)	99.25 (0.04)	88.58 (0.55)	79.51 (0.88)	99.69 (0.01)	53.54 (0.61)	36.56 (0.57)
			ViT-L-14	DPT	**95.24** (0.30)	**92.74** (0.42)	**86.46** (0.73)	71.78 (2.16)	27.50 (2.99)	20.59 (1.73)	99.26 (0.03)	88.73 (0.68)	79.75 (1.09)	99.68 (0.01)	53.62 (0.57)	36.64 (0.53)
	DINOV3	$TMLR_{2026}$	ViT-B-16	DPT	94.69 (0.31)	91.98 (0.38)	85.15 (0.65)	71.16 (2.97)	24.04 (2.05)	18.42 (1.07)	99.29 (0.09)	88.94 (1.34)	80.10 (2.16)	99.66 (0.01)	50.61 (0.66)	33.88 (0.59)
			ViT-L-16	DPT	94.85 (0.16)	92.23 (0.22)	85.59 (0.39)	71.50 (1.47)	24.66 (1.86)	18.87 (1.06)	99.34 (0.09)	89.94 (1.24)	81.74 (2.02)	99.66 (0.01)	50.70 (0.39)	33.96 (0.35)
EO-Ori.	UNetFormer	$ISPRS_{2022}$	ResNet-18	UNetFormer	94.02 (0.27)	91.16 (0.30)	83.75 (0.51)	70.07 (2.25)	19.59 (1.46)	15.99 (0.59)	99.35 (0.02)	89.79 (0.30)	81.47 (0.50)	99.72 (0.00)	58.23 (0.13)	41.07 (0.13)
	FarSeg	$CVPR_{2020}$	ResNet-50	FarSeg	91.17 (0.38)	86.91 (0.48)	76.86 (0.74)	73.87 (1.84)	22.91 (2.17)	18.00 (1.25)	99.18 (0.12)	87.24 (1.64)	77.39 (2.55)	99.71 (0.01)	57.15 (0.24)	40.01 (0.23)
	FarSeg++	$TPAMI_{2023}$	MiT-B2	FarSeg++	93.56 (0.52)	89.97 (0.79)	81.78 (1.29)	71.38 (1.63)	19.25 (0.83)	15.72 (0.43)	99.34 (0.05)	89.82 (0.66)	81.53 (1.09)	99.71 (0.00)	57.13 (0.26)	39.99 (0.25)
	RS^3Mamba	$GRSL_{2024}$	VMamba-T	UNetFormer	94.28 (0.34)	91.44 (0.50)	84.23 (0.85)	60.29 (9.56)	20.91 (4.09)	15.95 (3.01)	99.43 (0.05)	91.17 (0.62)	83.79 (1.04)	99.72 (0.01)	58.22 (0.14)	41.06 (0.14)
Bi-temporal	SiamCRNN	$TGRS_{2020}$	ResNet-50	ConvLSTM	94.67 (0.30)	92.04 (0.40)	85.26 (0.69)	70.02 (5.28)	24.30 (1.30)	18.37 (0.65)	99.39 (0.05)	90.46 (0.77)	82.59 (1.28)	-	-	-
	DSIFN	$ISPRS_{2020}$	VGG-16	DSIFN	90.27 (3.12)	85.51 (4.22)	74.88 (6.35)	54.01 (22.79)	16.53 (4.47)	12.73 (4.62)	99.21 (0.03)	87.75 (0.32)	78.17 (0.51)	-	-	-
	BIT-CD	$TGRS_{2021}$	ResNet-50	BIT	92.68 (0.68)	89.24 (0.84)	80.59 (1.36)	69.63 (8.50)	19.76 (2.70)	15.84 (0.94)	99.08 (0.21)	85.93 (2.66)	75.40 (3.98)	-	-	-
	ChangeOS	RSE_{2021}	ResNet-50	FPN	89.82 (1.11)	86.00 (1.24)	75.45 (1.91)	69.78 (2.32)	20.59 (0.88)	16.24 (0.54)	99.27 (0.19)	89.05 (2.48)	80.33 (3.99)	-	-	-
	ChangeMamba	$TGRS_{2024}$	VMamba-T	STSS	95.08 (0.20)	92.56 (0.29)	86.15 (0.50)	70.72 (2.80)	25.90 (2.02)	19.44 (1.43)	99.44 (0.04)	91.23 (0.67)	83.89 (1.13)	-	-	-
Native EO FM	SatMAE	$NeurIPS_{2022}$	ViT-B	DPT	94.55 (0.14)	91.73 (0.19)	84.73 (0.32)	69.37 (10.23)	27.26 (1.23)	20.55 (0.94)	99.28 (0.05)	88.88 (0.64)	79.99 (1.03)	99.66 (0.01)	50.66 (0.51)	33.93 (0.46)
	SkySense	$CVPR_{2024}$	Swin-V2-H	DPT	94.58 (0.20)	91.91 (0.25)	85.03 (0.43)	69.77 (2.73)	28.07 (1.36)	20.81 (0.82)	99.43 (0.04)	91.10 (0.46)	83.65 (0.78)	99.73 (0.01)	58.32 (0.47)	41.16 (0.47)
			ViT-H	DPT	94.81 (0.23)	92.23 (0.29)	85.58 (0.50)	69.20 (4.91)	26.45 (0.90)	19.77 (0.76)	99.42 (0.05)	91.00 (0.79)	83.50 (1.34)	-	-	-
	SpectralGPT	$TPAMI_{2024}$	ViT	DPT	87.56 (4.40)	82.83 (4.85)	70.92 (6.85)	64.39 (15.72)	19.64 (4.19)	15.45 (3.49)	99.09 (0.07)	85.90 (0.96)	75.30 (1.48)	-	-	-
	HyperSIGMA	$TPAMI_{2025}$	ViT	DPT	94.63 (0.36)	91.91 (0.44)	85.03 (0.74)	73.35 (4.42)	26.17 (0.68)	20.52 (0.90)	99.36 (0.05)	90.11 (0.73)	82.01 (1.21)	-	-	-
	DINOV3	$TMLR_{2026}$	ViT-L-16*	DPT	94.64 (0.45)	91.77 (0.75)	84.80 (1.27)	**74.63** (0.99)	**28.32** (2.21)	**21.43** (1.31)	99.24 (0.11)	88.57 (1.40)	79.50 (2.26)	99.64 (0.01)	50.00 (0.70)	33.33 (0.62)

vision FMs generally trail behind well-tuned segmentation models. The gap is moderate on VHR optical tasks but widens sharply on non-RGB data: as shown in Table 8, on URBANSARFLOODS, SAM2 (Hiera-B+) achieves an urban flood IoU of only 0.91%, compared to 14.63% for Swin-B and 21.18% for Mask2Former. SAM2 also underperforms SAM on several benchmarks. Among general vision FMs, DINOv2 is the most consistent, ranking first on RESCUENET (mIoU = 65.30%, Table 11) and third on S2WCD (F_1 = 92.74%, Table 10). CLIP is competitive on wildfire benchmarks, with CLIP (ViT-L-14) ranking second on S2WCD (F_1 = 92.89%), but remains weak on URBANSARFLOODS (mIoU = 53.20%). EO-native FMs show a task-specific failure mode rather than uniform underperformance. On wildfire extent mapping, SatMAE on S2WCD (F_1 = 91.73%) and SkySense on FLOGA (91.10%) sit within 1 F1 point of UNet, and HyperSIGMA reaches the third-highest open-flood IoU on URBANSARFLOODS (65.12%). The picture changes on infrastructure damage mapping: on RESCUENET (Table 11), SkySense (mIoU = 53.27%), SpectralGPT (23.56%), and HyperSIGMA (33.77%) fall well below DINOv2-B (65.30%); on xBD, all four EO-native FMs cluster below 56% mIoU, behind ChangeMamba (64.24%). DINOv3, evaluated with both natural-image and EO-pretrained weights, shows no consistent direction of transfer between the two variants across our benchmarks.

The underperformance of FMs in our benchmark is not unique to disaster mapping. Recent evaluations across the geospatial domain [3], [279] and other specialized fields [517] have consistently shown that, under standard supervised settings with sufficient labels, well-tuned task-specific models match or outperform FMs, with FM advantages emerging primarily in low-data regimes. Our results are consistent with this pattern. General vision FMs have never encountered SAR, multispectral, or temporal change data during pre-training. EO-native FMs, while exposed to EO data, are typically trained with reconstruction objectives on general land-cover scenes rather than disaster-specific semantics. Both categories can also suffer from over-parameterization, adding redundant representations rather

TABLE 11
Models on three infrastructure damage mapping benchmark datasets. A **Family** column is added to indicate the broad methodological category of each model for easier navigation. Values denote the mean over 5 runs, with standard deviations in parentheses. The top three highest results in each dataset are highlighted in **red**, **orange**, and **yellow**.

Family	Models	Venue	Encoder	Decoder	xBD				BRIGHT				RESCUENET	
					F_1^{loc}	F_1^{clf}	OA	mIoU	F_1^{loc}	F_1^{clf}	OA	mIoU	OA	mIoU
Trad.	Random Guessing	-	-	-	4.01 (0.00)	0.58 (0.00)	91.17 (0.00)	18.65 (0.00)	14.12 (0.00)	2.79 (0.00)	75.16 (0.00)	20.67 (0.00)	33.29 (0.00)	5.52 (0.00)
	Random Forest	-	-	-	42.01 (0.00)	51.41 (0.00)	83.88 (0.00)	27.70 (0.00)	59.29 (0.00)	63.39 (0.00)	81.79 (0.00)	38.59 (0.00)	53.40 (0.00)	24.62 (0.00)
General Vision CNN	FCN-8s	$CVPR_{2015}$	VGG-16	-	83.06 (0.10)	67.23 (1.31)	97.30 (0.04)	57.87 (0.31)	87.35 (0.15)	69.56 (0.32)	95.24 (0.07)	63.53 (0.29)	81.43 (0.81)	53.28 (0.46)
	UNet	$MICCAI_{2015}$	-	-	83.46 (0.37)	66.50 (0.83)	97.07 (0.04)	56.92 (0.18)	88.68 (0.15)	70.85 (1.27)	95.63 (0.04)	64.86 (0.53)	74.81 (0.87)	40.17 (1.10)
	UNet++	TMI_{2019}	-	-	82.26 (0.11)	63.98 (0.66)	96.92 (0.05)	55.40 (0.39)	87.70 (0.22)	70.75 (0.28)	95.35 (0.06)	63.97 (0.23)	76.18 (0.79)	37.82 (0.71)
	DeepLabV3+	$ECCV_{2018}$	ResNet-50	ASPP	78.70 (1.04)	64.40 (1.40)	97.05 (0.14)	56.14 (0.87)	86.96 (0.69)	70.55 (1.64)	95.37 (0.12)	64.36 (0.81)	80.66 (1.08)	52.62 (1.08)
			ResNet-101	ASPP	79.41 (1.65)	64.29 (1.79)	97.13 (0.09)	56.48 (0.86)	85.77 (1.37)	68.73 (1.26)	95.14 (0.21)	63.18 (0.88)	80.56 (0.56)	52.42 (1.70)
	HRNet	$TPAMI_{2021}$	HRNetV2-W18	-	84.29 (0.35)	69.88 (0.68)	97.41 (0.06)	59.60 (0.41)	88.63 (0.17)	71.09 (0.68)	95.62 (0.09)	65.19 (0.43)	79.90 (0.16)	52.12 (0.37)
			HRNetV2-W48	-	86.14 (0.15)	**72.17** (0.59)	97.67 (0.04)	61.88 (0.27)	89.80 (0.21)	**72.71** (0.46)	95.96 (0.05)	66.69 (0.11)	81.62 (0.90)	56.65 (0.43)
	ConvNeXt	$CVPR_{2022}$	ConvNeXt-T	UperNet	85.62 (0.10)	71.52 (0.68)	97.67 (0.04)	61.65 (0.33)	89.65 (0.13)	70.89 (0.36)	95.93 (0.04)	65.80 (0.26)	83.59 (0.25)	60.29 (0.92)
			ConvNeXt-S	UperNet	85.78 (0.23)	71.14 (0.89)	97.68 (0.05)	61.48 (0.35)	89.78 (0.10)	72.07 (1.15)	95.95 (0.07)	66.36 (0.73)	83.39 (0.52)	59.79 (0.87)
			ConvNeXt-B	UperNet	85.91 (0.42)	71.67 (0.49)	97.73 (0.03)	62.01 (0.19)	89.82 (0.08)	71.56 (0.51)	95.98 (0.03)	66.35 (0.27)	83.59 (0.92)	60.53 (0.62)
General Vision Transformer	Swin	$ICCV_{2021}$	Swin-T	UperNet	85.90 (0.19)	**72.26** (0.92)	97.68 (0.04)	61.87 (0.55)	89.85 (0.10)	72.00 (0.41)	96.00 (0.03)	66.38 (0.35)	82.52 (0.62)	57.72 (0.61)
			Swin-S	UperNet	85.85 (0.16)	70.93 (1.29)	97.70 (0.02)	61.52 (0.37)	89.96 (0.06)	71.33 (0.74)	96.00 (0.01)	66.24 (0.37)	83.49 (0.56)	59.87 (0.23)
			Swin-B	UperNet	86.07 (0.07)	72.13 (0.84)	97.72 (0.05)	62.15 (0.31)	90.02 (0.08)	71.88 (0.36)	96.04 (0.05)	66.55 (0.31)	83.95 (0.41)	59.83 (0.47)
	SegFormer	$NeurIPS_{2021}$	MiT-B0	MLP	84.52 (0.16)	70.98 (0.83)	97.46 (0.05)	60.02 (0.18)	88.45 (0.12)	70.80 (0.87)	95.59 (0.06)	64.94 (0.49)	82.74 (0.56)	57.34 (0.33)
			MiT-B1	MLP	85.59 (0.16)	71.07 (0.61)	97.64 (0.03)	61.20 (0.33)	89.27 (0.17)	70.69 (0.87)	95.80 (0.05)	65.35 (0.37)	83.72 (0.66)	59.23 (1.13)
			MiT-B2	MLP	85.88 (0.22)	71.54 (0.89)	97.72 (0.05)	61.92 (0.23)	89.79 (0.13)	71.12 (0.69)	95.96 (0.05)	65.89 (0.43)	83.35 (0.54)	59.71 (0.95)
			MiT-B3	MLP	85.80 (0.20)	69.95 (1.23)	97.73 (0.05)	61.48 (0.39)	89.83 (0.17)	71.30 (1.39)	95.97 (0.04)	66.03 (0.73)	84.19 (0.48)	61.25 (0.93)
			MiT-B4	MLP	86.05 (0.08)	70.59 (0.40)	97.72 (0.02)	61.56 (0.19)	89.86 (0.13)	71.77 (1.32)	96.00 (0.04)	66.30 (0.68)	84.33 (0.20)	61.14 (0.51)
			MiT-B5	MLP	86.03 (0.15)	70.33 (0.75)	97.74 (0.03)	61.50 (0.19)	89.68 (0.13)	72.05 (0.43)	95.96 (0.04)	66.38 (0.14)	84.94 (0.13)	62.01 (0.56)
	Mask2Former	$CVPR_{2022}$	Swin-B	Mask2Former	**87.29** (0.14)	67.15 (0.82)	97.84 (0.03)	60.91 (0.62)	**90.45** (0.07)	70.72 (0.85)	96.06 (0.05)	65.52 (0.38)	81.20 (1.63)	55.27 (0.95)
General Vision FM	SAM	$ICCV_{2023}$	ViT-B-16	DPT	85.46 (0.19)	70.25 (0.78)	97.62 (0.04)	60.48 (0.45)	89.53 (0.26)	71.26 (1.22)	95.89 (0.05)	65.75 (0.48)	83.54 (0.63)	58.81 (0.69)
			ViT-L-16	DPT	86.29 (0.26)	71.10 (0.72)	97.77 (0.03)	61.78 (0.29)	90.39 (0.11)	71.20 (1.02)	96.09 (0.03)	66.29 (0.36)	83.72 (0.70)	60.27 (0.71)
	SAM2	$ICLR_{2025}$	Hiera-S	UperNet	78.56 (0.10)	71.28 (0.38)	96.65 (0.04)	56.02 (0.24)	84.28 (0.14)	70.78 (1.41)	94.52 (0.01)	62.80 (0.63)	82.83 (0.66)	57.47 (0.34)
			Hiera-B+	UperNet	79.20 (0.12)	71.55 (0.65)	96.71 (0.03)	56.41 (0.15)	84.41 (0.14)	70.05 (0.89)	94.53 (0.01)	62.47 (0.41)	84.36 (0.85)	59.61 (0.78)
	CLIP	$ICML_{2021}$	ViT-B-16	DPT	85.49 (0.09)	69.73 (0.62)	97.67 (0.04)	60.70 (0.41)	89.74 (0.13)	70.82 (0.59)	95.94 (0.03)	65.60 (0.45)	85.18 (0.56)	62.06 (0.77)
			ViT-L-14	DPT	86.38 (0.08)	70.89 (1.13)	97.81 (0.04)	61.98 (0.16)	90.31 (0.09)	71.17 (0.99)	**96.13** (0.04)	66.31 (0.63)	**86.09** (0.27)	**63.60** (0.66)
	DINOV2	$TMLR_{2024}$	ViT-B-14	DPT	86.47 (0.20)	71.38 (1.11)	97.78 (0.02)	61.98 (0.38)	90.32 (0.37)	**72.29** (0.78)	96.11 (0.07)	**66.91** (0.50)	**87.03** (0.20)	**65.30** (0.51)
			ViT-L-14	DPT	86.87 (0.13)	71.42 (0.69)	**97.87** (0.05)	**62.57** (0.59)	**90.70** (0.10)	72.20 (0.52)	**96.22** (0.07)	**67.05** (0.34)	85.66 (0.48)	62.88 (1.07)
	DINOV3	$TMLR_{2026}$	ViT-B-16	DPT	86.02 (0.14)	71.11 (0.41)	97.76 (0.03)	61.59 (0.39)	89.58 (0.11)	69.75 (1.50)	95.90 (0.03)	65.26 (0.65)	85.52 (0.37)	63.36 (0.39)
			ViT-L-16	DPT	86.72 (0.23)	71.80 (0.67)	**97.86** (0.05)	62.48 (0.35)	90.28 (0.23)	71.13 (0.67)	96.11 (0.09)	66.39 (0.53)	**86.37** (0.57)	**64.84** (0.67)
EO-Ori.	UNetFormer	$ISPRS_{2022}$	ResNet-18	UNetFormer	83.59 (0.25)	69.30 (0.66)	97.23 (0.07)	58.34 (0.55)	88.12 (0.07)	68.99 (1.02)	95.49 (0.08)	63.97 (0.52)	80.67 (0.66)	53.09 (0.71)
	FarSeg	$CVPR_{2020}$	ResNet-50	FarSeg	85.17 (0.15)	69.52 (0.58)	97.64 (0.02)	60.60 (0.34)	89.14 (0.28)	69.96 (0.68)	95.75 (0.11)	64.87 (0.52)	84.11 (0.72)	60.08 (0.47)
	FarSeg++	$TPAMI_{2023}$	MiT-B2	FarSeg++	86.13 (0.21)	72.17 (0.90)	97.74 (0.08)	62.31 (0.66)	90.00 (0.08)	71.89 (0.63)	96.04 (0.04)	66.35 (0.36)	83.45 (0.59)	59.63 (0.77)
	RS^3Mamba	$GRSL_{2024}$	VMamba-T	UNetFormer	83.59 (0.22)	67.81 (0.64)	97.20 (0.06)	57.73 (0.43)	88.45 (0.25)	70.00 (0.37)	95.55 (0.05)	64.30 (0.43)	79.70 (0.54)	50.27 (1.86)
Bi-temporal	SiamCRNN	$TGRS_{2020}$	ResNet-50	ConvLSTM	86.42 (0.14)	70.23 (0.84)	97.70 (0.02)	61.50 (0.38)	88.80 (0.13)	67.06 (1.27)	95.46 (0.04)	63.14 (0.45)	-	-
	DSIFN	$ISPRS_{2020}$	VGG-16	DSIFN	85.07 (0.16)	67.28 (0.54)	97.49 (0.03)	59.28 (0.24)	86.64 (1.01)	67.24 (0.31)	95.43 (0.19)	63.34 (0.45)	-	-
	BIT-CD	$TGRS_{2021}$	ResNet-50	BIT	81.29 (1.44)	67.12 (2.09)	96.96 (0.18)	56.28 (1.54)	87.53 (0.19)	71.40 (0.55)	95.32 (0.07)	64.63 (0.43)	-	-
	ChangeOS	RSE_{2021}	ResNet-50	FPN	**86.97** (0.17)	71.34 (0.52)	97.82 (0.04)	**62.59** (0.28)	90.14 (0.10)	70.19 (0.41)	95.97 (0.06)	65.62 (0.08)	-	-
	ChangeMamba	$TGRS_{2024}$	VMamba-T	STSS	**87.94** (0.19)	**73.58** (0.46)	**97.96** (0.03)	**64.24** (0.25)	**91.09** (0.08)	**73.38** (0.56)	**96.25** (0.05)	**67.93** (0.39)	-	-
Native EO FM	SatMAE	$NeurIPS_{2022}$	ViT-B-16	DPT	79.81 (0.29)	64.77 (1.16)	96.92 (0.07)	55.13 (0.37)	86.34 (0.20)	68.49 (1.30)	95.06 (0.06)	62.64 (0.56)	84.34 (0.15)	56.00 (0.78)
	SkySense	$CVPR_{2024}$	Swin-V2-H	UperNet	80.89 (6.61)	64.45 (10.41)	96.63 (1.50)	54.53 (8.90)	88.76 (0.12)	70.80 (0.67)	95.63 (0.06)	64.81 (0.42)	80.79 (0.74)	53.27 (1.82)
	SpectralGPT	$TPAMI_{2024}$	ViT-B-8	UperNet	79.35 (0.36)	65.69 (0.20)	96.75 (0.10)	54.95 (0.26)	83.11 (0.11)	69.67 (1.09)	93.99 (0.09)	60.38 (0.24)	63.46 (1.26)	23.56 (0.72)
	HyperSIGMA	$TPAMI_{2025}$	ViT	FPN	79.49 (0.60)	65.90 (0.45)	96.82 (0.11)	55.34 (0.36)	82.86 (0.19)	68.16 (0.50)	93.87 (0.04)	59.76 (0.12)	74.76 (0.34)	33.77 (0.62)
	DINOV3	$TMLR_{2025}$	ViT-L-16	DPT	86.06 (0.24)	70.76 (0.83)	97.77 (0.04)	61.61 (0.34)	89.59 (0.28)	71.65 (0.58)	95.91 (0.10)	66.09 (0.51)	85.99 (0.20)	63.03 (0.47)

than new abstractions [152]. Disaster mapping additionally demands generalization to unseen events, hazard-specific damage grading, and cross-modal transfer, which together make it a demanding testbed for the FM paradigm.

Finding 5: Computational cost scales poorly with accuracy gains. Fig. 6 reveals a stark disconnect between model complexity and mapping performance. Lightweight architectures such as SegFormer-B0 (3.7M parameters, 147 FPS) and UNetFormer (11.7M parameters, 146 FPS) achieve competitive accuracy across multiple datasets while running at over 140 frames per second on a single GPU. At the other end of the spectrum, SkySense (672M parameters, 29 FPS) and DINOv2-L (345M parameters, 24 FPS) consume 10–50× more parameters and run 5–6× slower, yet offer no systematic accuracy advantage. Among change detection models, SiamCRNN (21M parameters, 118 FPS) delivers the best balance of accuracy and throughput, whereas ChangeMamba, despite achieving the highest damage mapping scores, operates at only 35 FPS due to its scanning mechanism. The practical implication is clear: for operational disaster response, where latency and hardware constraints are real, a well-chosen lightweight model can deliver comparable or superior mapping quality at a fraction of the computational cost. This is especially relevant for extent mapping, where our benchmark shows that accuracy differences among architectures are small to begin with. For damage mapping, where richer architectures do provide measurable gains, the cost-accuracy tradeoff is more nuanced but still favours mid-range models (30–60M parameters) over the largest FMs.

Finding 6: Difficulty concentrates in ambiguous and minority disaster states. Performance varies sharply across the benchmark, but the hardest cases are not defined simply by hazard type or spatial resolution. Two datasets expose this most clearly. On URBANSARFLOODS, the best urban-flood IoU is only 21.18% with Mask2Former, compared with 67.33% for open-water flooding on the same benchmark. This gap reflects the difficulty of separating flooded urban surfaces from layover, shadow, double-bounce scattering, and other built-up SAR responses in 20 m imagery. On

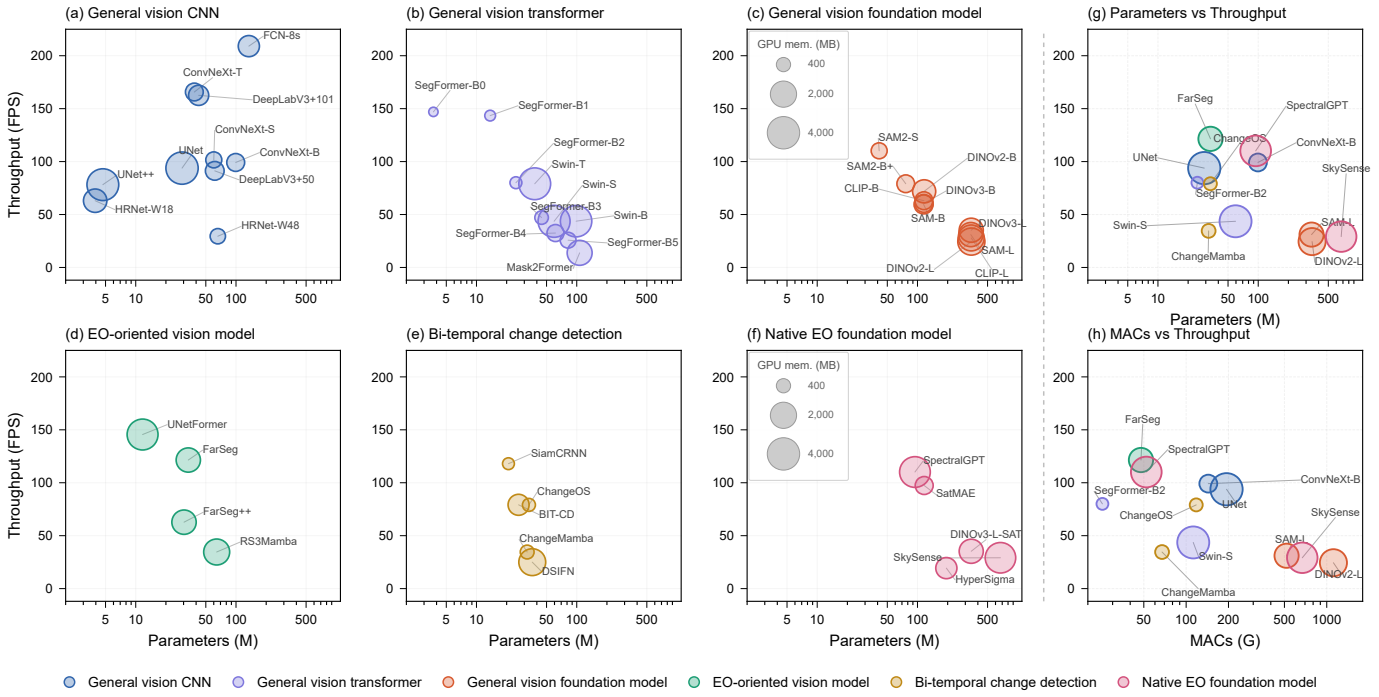

Fig. 6. Efficiency comparison of different models. Models are grouped into six families of Table 8 to Table 11 in panels (a–f), and representative models are compared across families in panels (g,h). The x-axis shows the number of parameters in (a–g) and MACs in (h), the y-axis shows throughput (FPS), and bubble size denotes GPU memory usage. Memory is measured using bi-temporal inputs with the size of 512×512×3.

SBAD, all models struggle with multi-class burned-area severity mapping: the best result is only 22.36% mIoU with UNet++, despite high overall accuracy for even random guessing. This indicates that OA is dominated by majority classes and masks poor discrimination of minority severity classes. Compared with binary burned-area mapping on FLOGA, SBAD requires finer semantic separation among subtly different damage states. Together, these results show that the main remaining bottlenecks are ambiguous target definitions, class imbalance, and weakly separable disaster states, rather than model capacity alone.

5.5 Case Studies

To assess whether benchmark performance translates to real-world utility, we conducted zero-shot evaluations on recent disaster events not represented in any datasets.

5.5.1 The 2025 Mozambique Flood

In late December 2025, heavy rainfall across southern Africa caused the Limpopo River to overflow, triggering severe flooding in Gaza Province, Mozambique. The Chokwe district was among the hardest hit, with over 390,000 people displaced across the province [440]. This event provides a particularly demanding test case: it is located in the Global South, a region severely underrepresented in existing flood benchmarks. We applied representative models from each paradigm to bi-temporal Sentinel-1 SAR imagery covering the Chibabel area near Chokwe, evaluating flood extent delineation in a zero-shot setting.

As shown in Fig. 7, all models successfully delineate the large-scale inundation footprint, but performance diverges sharply on the distinction between transient flooding and permanent water. UNetFormer achieves the highest flood IoU (86.28%) and overall accuracy (92.53%), while SiamCRNN produces the most balanced results across both classes (mIoU = 71.95%, permanent water IoU = 52.71%). The most revealing failure is UNet++: despite strong flood detection (IoU_f = 83.96%), it nearly completely fails on permanent water (IoU_w = 6.04%), effectively classifying all water as flood. SAM, the strongest general vision FM in this test (mIoU = 68.08%), outperforms SegFormer and HyperSIGMA but trails both SiamCRNN and UNetFormer, consistent with the benchmark finding that FMs remain competitive but not dominant on SAR-based tasks. Notably, the change detection architecture SiamCRNN provides the best permanent water discrimination, suggesting that explicit temporal comparison helps disentangle transient from persistent water, an advantage that direct segmentation models lack.

5.5.2 The 2025 Ōita Fire

On November 18, 2025, a large-scale fire struck Saganoseki, Oita City, Japan, destroying 187 buildings over approximately 48,900 m^2, fueled by strong northwesterly winds during an exceptionally dry autumn. The affected area is characterized by high building density and a severely aging population (72.2% over 65 in the Tanaka district), making rapid infrastructure assessment critical for evacuation and resource allocation. We applied representative models from each paradigm to a pre-event optical / post-event SAR input pair without manual co-registration, simulating the constraints of real-time response.

As shown in Fig. 8, the performance hierarchy from controlled benchmarks partially transfers: ChangeMamba

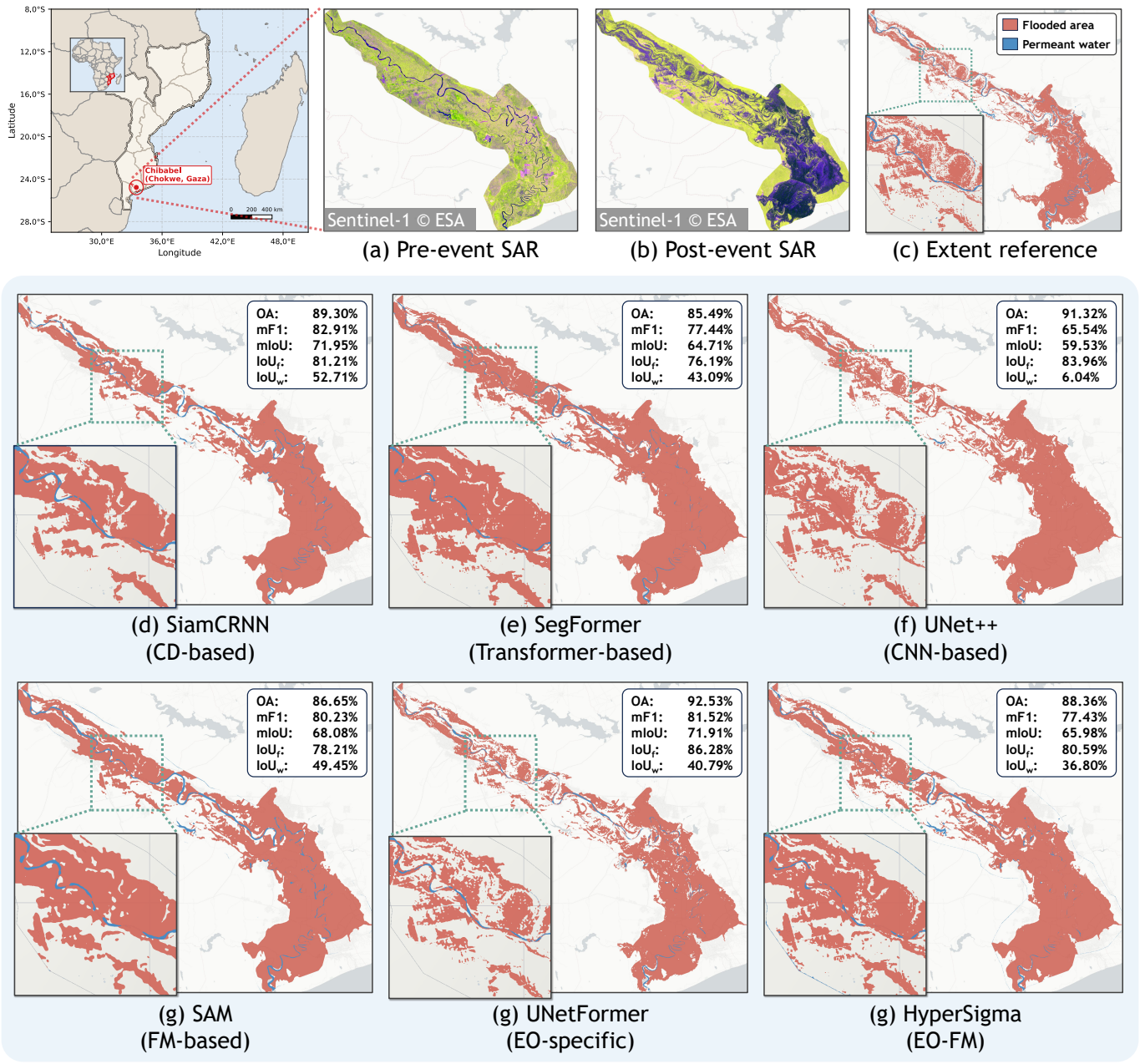

Fig. 7. Visual comparison of disaster extent mapping results for the 2025 Mozambique Flood. (a) Post-event SAR imagery; (b) Post-event SAR imagery; (c) Reference map for flood extent; (d)-(g) Predictions from representative deep learning families trained on KuroSiwo. In the visualizations, flooded areas are shown in salmon and permanent water bodies in blue.

produces the most spatially coherent damage map, followed by Swin-T and FarSeg++. However, two deployment-specific challenges emerge that are not captured by standard benchmark metrics. First, all models exhibit high false positive rates in industrial zones, where metallic structures and dense machinery produce volume scattering patterns similar to debris, leading to systematic misclassification. Second, geometric offsets between the pre-event optical and post-event SAR imagery, left uncorrected to simulate operational urgency, introduce spatial errors that degrade localization accuracy. Foundation models including SAM and DINOv3 perform notably worse than task-specific architectures in this setting, consistent with their weaker benchmark performance on cross-modal tasks but amplified by the distribution shift of an entirely unseen event and geography.

6 Open Challenges and Future Perspectives

The challenges identified by our benchmark span the full disaster mapping pipeline, from the data that models are trained on, through the preprocessing and alignment of inputs, to the models themselves and the format of their outputs. We organize the discussion accordingly.

6.1 Data: Coverage, Modality, and Annotation

Current benchmarks are geographically concentrated in North America, Europe, and East Asia, leaving model per-

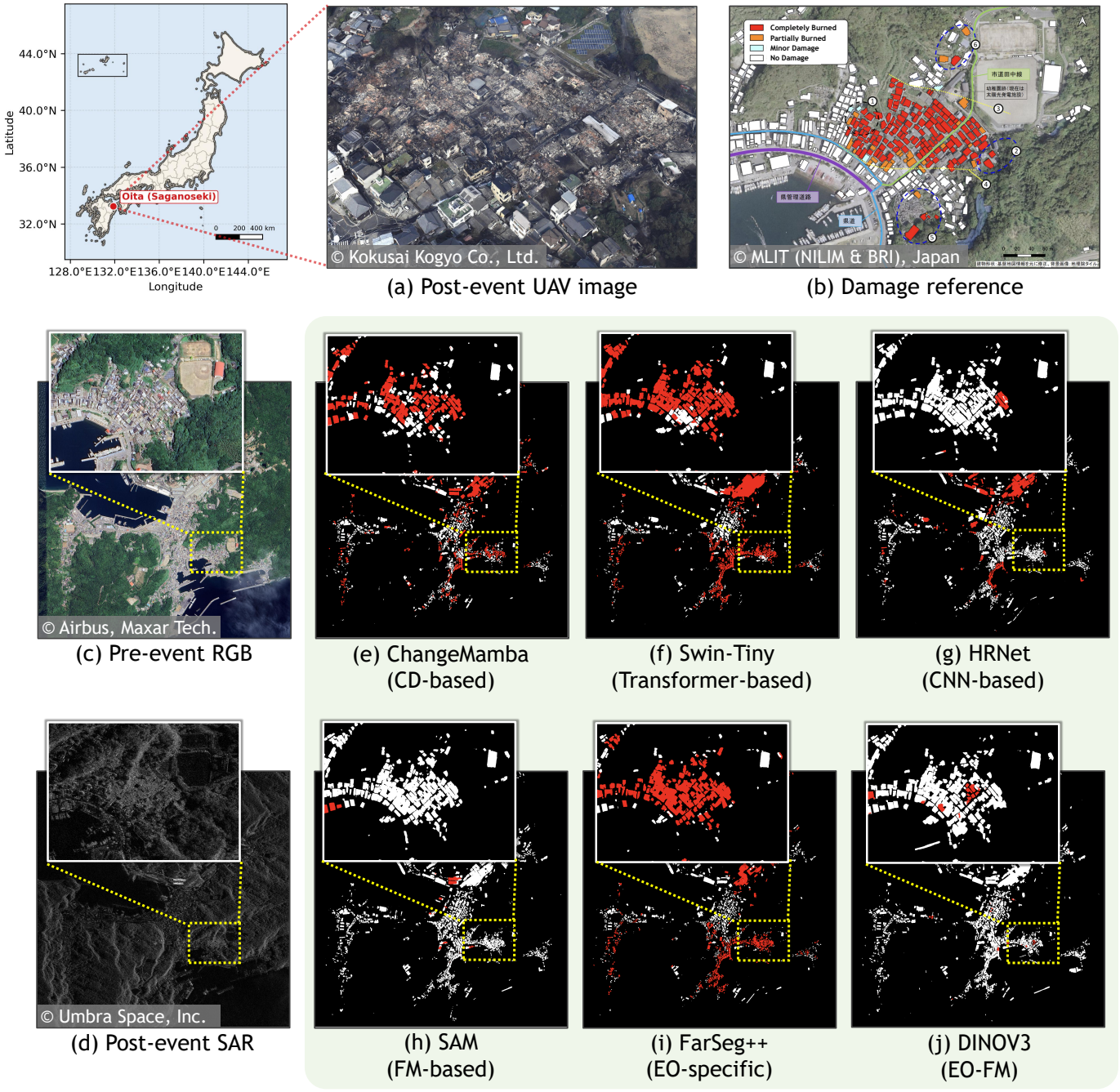

Fig. 8. Visual comparison of building damage mapping results for the 2025 Ōita Fire. (a) Post-event UAV imagery with oblique view; (b) Damage reference map based on ground truth surveys; (c) Pre-event optical image; (d) Post-event SAR image. (e)-(j) Predictions from representative deep learning families trained on BRIGHT. In the visualization, red represents the combined class of completely and partially destroyed buildings, white denotes intact structures, and black is the background. Panel (a) is reproduced from the Bois Disaster Information Service operated by Kokusai Kogyo Co., Ltd. (image ID 23; acquired on November 19, 2025, at 11:55 JST), available at https://bois-free.bousai.genavis.jp/diarsweb, with permission. Panel (b) is adapted from [310], with minor modifications by the authors.

formance in the Global South, where disaster vulnerability is highest, largely unvalidated. Modality coverage is similarly incomplete. While optical and medium-resolution SAR datasets are relatively mature, most LiDAR, InSAR, and VHR SAR resources remain confined to isolated case studies without community benchmarks. This matters in practice: all models in our benchmark fail on urban flood mapping from Sentinel-1, a limitation rooted partly in the absence of VHR SAR training data that could resolve urban scattering ambiguity. Volcanic deposits, avalanches, and armed conflict face analogous data scarcity, as noted in Section 3.

Beyond coverage, annotation quality limits what models can learn. For infrastructure damage, multiple grading scales coexist (EMS-98, AIJ, and various simplified schemes), and cross-platform label disagreement can reach 29% [278], particularly for moderate damage categories that are visible from UAV but not from satellite. For extent mapping, reference labels from different products (e.g.,

MODIS-based fire products vs. expert-delineated Sentinel-2 masks) introduce systematic biases that propagate into training. Labelling procedures are also poorly documented: published datasets rarely specify whether annotations come from a single interpreter or from multi-annotator review with consensus, and inter-annotator agreement statistics are almost never reported. Given disagreement rates of this magnitude, performance gaps of a few percentage points on current benchmarks may fall within the uncertainty of the reference labels themselves, a caveat that applies to our own results and argues for routine significance testing. Investing in protocols that surface and resolve ambiguity, and in standards for documenting labelling procedures, may yield greater returns than scaling label counts.

6.2 Input Reliability Under Operational Constraints

Beyond benchmark coverage and annotation quality, real-world deployment also depends critically on the quality and stability of the inputs themselves. Before any model can be applied, disaster mapping requires a reliable pre-event reference and, for multi-modal pipelines, accurate alignment between heterogeneous inputs. Both remain practical bottlenecks.

For bi-temporal pipelines, obtaining a usable pre-event optical reference is often the first operational bottleneck. Cloud cover, smoke, and seasonal mismatch frequently make the most recent pre-event acquisition unsuitable. A common workaround is to search historical archives for a cleaner scene, but this is time-consuming and can introduce temporal gaps that confound disaster-induced change with normal land-cover or phenological variation. A more systematic alternative is to construct cloud-reduced pre-event references from multi-temporal image archives, for example through compositing or related time-series aggregation strategies, rather than relying on a single acquisition [343]. Recent embedding field models such as AlphaEarth [45] and TESSERA [115] suggest a further step in this direction. Rather than selecting one cloud-free image, these models learn analysis-ready representations that integrate spatial, temporal, and cross-sensor context into a embedding. Such multi-source embeddings could potentially serve as pre-event reference priors for downstream disaster mapping, reducing reliance on manual archive search and on any single optical acquisition. However, whether these learned representations can effectively complement, or in some settings replace, conventional pre-event imagery in bi-temporal disaster mapping pipelines remains an open question.

The second part is geometric consistency across sensors. For disaster extent mapping at moderate resolution, residual co-registration error is often tolerable and automated optical-to-SAR registration can be adequate in many cases. For fine-grained damage mapping, however, alignment requirements are substantially tighter: even small offsets can move predictions between adjacent buildings or between damaged and undamaged structures. Our 2025 Ōita Fire case study illustrates this effect. When geometric offsets between pre-event optical and post-event SAR imagery are left uncorrected to reflect operational urgency, all tested models exhibit spatial errors. The problem is especially acute in dense urban areas, where layover, shadow, tall structures, and viewing geometry make cross-sensor registration more fragile [58], [524].

The third aspect, which is less frequently discussed, is pre-event source sensitivity. Our additional experiments in Appendix B.5 show that replacing the pre-event optical provider leads to severe degradation in damage classification, even when building localization remains relatively stable. This suggests that the learned cross-modal correspondence is at least partly tied to the source distribution seen during training. When an unseen provider is introduced, shifts in radiometry, viewing geometry, colour balance, or preprocessing can weaken the cues used for damage grading. Taken together, these observations suggest that source harmonization, source-diverse training, and more robust baseline construction deserve attention alongside model architecture design.

6.3 Generalization: From Benchmark to Deployment

Our benchmark shows that no single architecture dominates across datasets, and our out-of-distribution case studies on the Mozambique flood and the Saganoseki fire further show that strong benchmark performance does not necessarily translate into operational reliability. The dominant sources of domain shift also appear to be task-dependent: for extent mapping, land-cover variability, seasonality, and sensor differences are often central, whereas for damage mapping, construction typology, hazard mechanism, and viewing geometry can play a larger role. Current evaluation protocols often do not test these shifts explicitly. More realistic assessment will require standardized leave-event-out and cross-region splits, together with few-shot and zero-shot evaluation regimes [58], [569], to better reflect deployment conditions.

Generalization is not only geographic and event-specific. It also depends on the input regime under which models are evaluated. Our spectral ablation in Appendix B.3 shows that architectural comparisons can be incomplete when conducted under a single input configuration. This indicates that model rankings can invert depending on the spectral information provided. Likewise, our topographic ablation in Appendix B.2 shows that adding DEM or slope does not reliably improve performance and can even reduce it, challenging the assumption that physically meaningful auxiliary inputs are uniformly beneficial for flood extent mapping. Taken together, these results argue for evaluation protocols that vary not only geographic and temporal splits, but also input configurations.

More broadly, deployment readiness should be understood as robustness to both distribution shift and input variation, rather than as performance on a fixed benchmark configuration alone. From a methodological perspective, domain adaptation and synthetic data generation are promising directions for improving cross-event robustness [406], [573], but their effectiveness for operational post-disaster mapping remains insufficiently validated at scale.

6.4 Models: Scale, Inductive Bias, and Efficiency

Model development for post-disaster mapping is not simply a question of choosing a stronger backbone. The central issue is how to balance model scale, inductive bias, and

deployment cost under multi-modal and non-RGB settings. In our benchmark experiments, neither general vision nor EO-native foundation models consistently outperform well-tuned task-specific baselines, especially on non-RGB modalities. Similar patterns have also been reported in geospatial and other specialized domains under standard supervised settings with sufficient labels [3], [279], [517].

One notable finding is that EO-native foundation models on average do not outperform general vision foundation models. A plausible explanation is that general vision models benefit from much larger and more diverse pre-training corpora, while many EO-native models still rely heavily on reconstruction-oriented objectives that may not align well with downstream semantic discrimination in disaster settings. More broadly, our results indicate that pre-training scale and domain origin are both weaker predictors of downstream performance than is often assumed.

Efficiency is part of the same model-selection problem. Our profiling analysis shows that large models require roughly 10–50$\times$ more parameters than SegFormer-B0, but do not show a consistent accuracy advantage in our benchmark. For time-critical deployment, this weakens the case for simply scaling models up. Lightweight architectures remain competitive when latency, memory, and engineering overhead matter.

Progress will likely depend less on adopting ever larger generic models and more on designing representations that better match disaster mapping itself. Promising directions include self-supervised pre-training on disaster-relevant and multi-modal corpora, objectives better aligned with change and damage understanding, and synthetic data generation to cover rare hazard types, viewing conditions, and sensor configurations [407], [569].

6.5 Outputs: Beyond Maps to Operational Decision Support

Most post-disaster mapping methods and benchmarks are formulated as semantic segmentation problems. While this is convenient for model development, a segmentation map is rarely the final product needed in real response settings. Responders require actionable information: which buildings are likely destroyed, which road segments may be impassable, which facilities are affected, and where intervention should be prioritized. In this sense, pixel-level maps are better viewed as an intermediate representation than as the endpoint of operational inference.

Bridging this gap requires moving not only from pixels to objects, but from maps to structured decisions. Object-level outputs, such as per-building damage grades or per-road status labels, are an important intermediate layer and are currently often approximated through heuristic post-processing, such as majority voting within building footprints [58], [571]. However, these heuristics depend on the quality and completeness of external inventories and remain brittle when footprints are missing, misaligned, or outdated. More explicit instance-level architectures, together with benchmarks that provide per-object annotations and metrics, would better align model development with operational reporting requirements.

Yet even object-level labels are not sufficient on their own. In practice, responder-facing systems should translate geospatial predictions into concise, interpretable outputs, such as counts of severely affected buildings, estimates of displaced populations requiring immediate shelter, candidate blocked routes, facility-level status summaries, and priority-ranked areas for follow-up. Natural-language situation summaries may be particularly valuable as an interface layer for non-technical stakeholders [474]. The goal is not to discard maps, but to make them easier to interpret and act upon under time pressure.

Evaluation should evolve accordingly. Aggregate segmentation scores alone are insufficient to characterize output usefulness. What matters is whether the system supports clear operational judgments under realistic constraints, for example by controlling false alarms, preserving high-priority objects, and producing summaries that are understandable to non-specialist users. More broadly, benchmarks should evolve from static leaderboards toward regime-aware evaluation that considers not only accuracy, but also the form, interpretability, and decision utility of the outputs delivered to end users.

6.6 Future Perspective

Taken together, the challenges above point to a shared implication: no single architectural advance will close the gap between benchmark accuracy and operational reliability. The next stage of EO-based disaster mapping should therefore be organized around the construction of disaster intelligence systems (Fig. 9) that coordinate multiple specialist models over multi-modal observations, rather than around the pursuit of a single universal backbone.

Four properties should characterize such systems. First, they must integrate heterogeneous observations together with documented pre-event references constructed from archives rather than from any single acquisition. Second, a reasoning layer, such as an Large Language Model (LLM)-based orchestrator, may coordinate these specialist modules: fusing multi-modal evidence, reconciling conflicting predictions, selecting tools appropriate to each hazard, and continuing to adapt as new event types and sensors emerge. Third, outputs should be delivered in forms responders can act on rather than pixel maps alone. Fourth, the system must be trustworthy. Each product should carry calibrated uncertainty, transparent provenance from raw observation to reported decision, and explicit documentation of its failure modes, so that users can judge where to rely on the output and where further verification is needed.

Realizing these properties is not only a modeling problem; it will also require datasets documented with the transparency expected of scientific measurements, benchmarks that evaluate the system as a whole under realistic operational conditions, and sustained dialogue with the responders whose decisions these systems are designed to support. The measure of progress should shift accordingly, from leaderboard accuracy on isolated tasks to the operational dependability of the evidence delivered during a response.

7 Conclusion

This article has provided a unified review and benchmark of EO-based disaster mapping, spanning traditional methods,

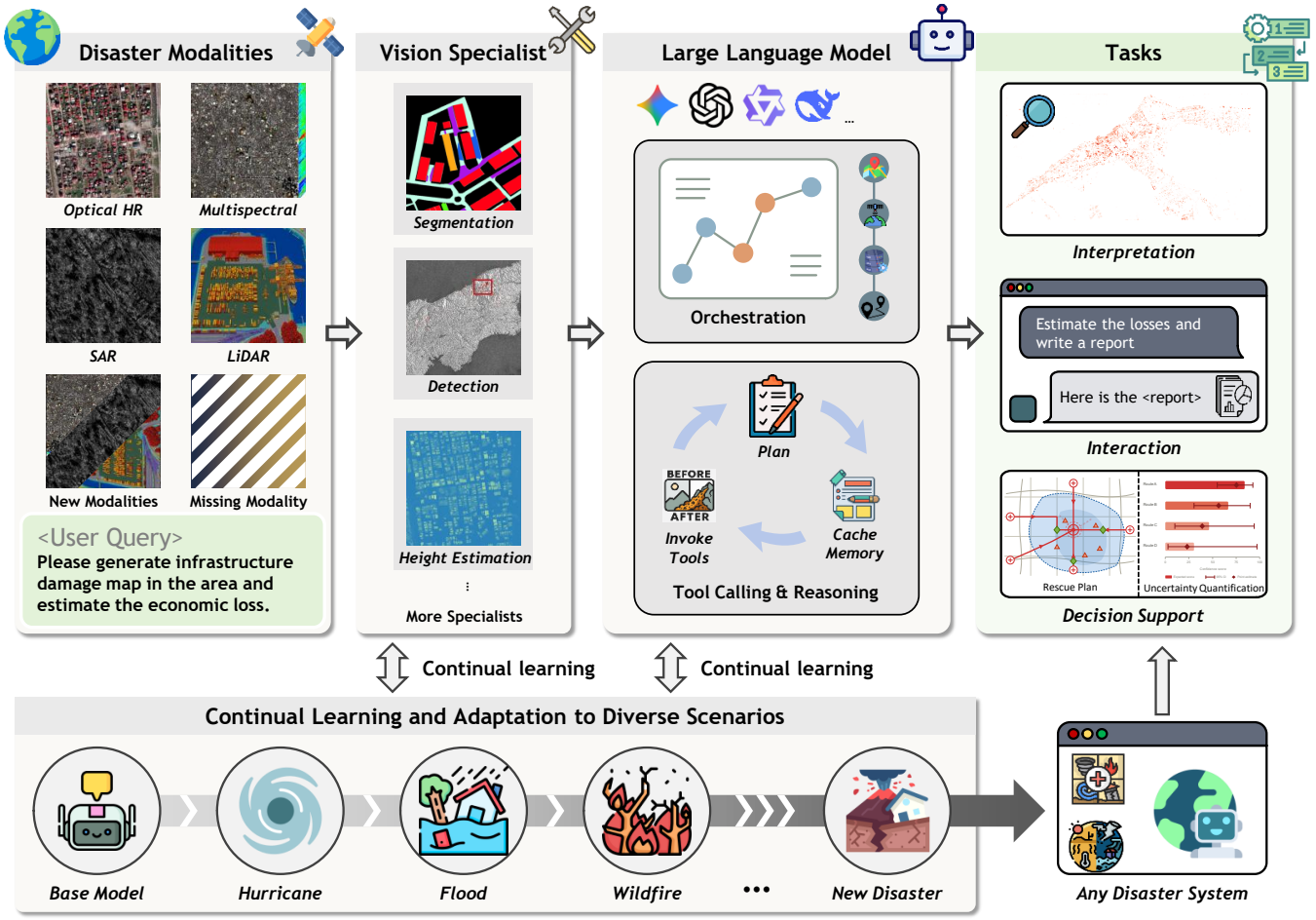

Fig. 9. A conceptual overview of next-generation disaster intelligent response systems. Multi-modal disaster observations are processed by specialist vision modules and coordinated by a large language model that supports orchestration, tool use, and reasoning, enabling downstream tasks such as interpretation, interaction, and decision support, while continually adapting across disaster types.

deep learning architectures, and foundation models across both disaster extent mapping and infrastructure damage mapping. We defined a task-driven taxonomy grounded in the observation that algorithmic challenges are shaped more by the nature of the mapping target than by the specific hazard. We systematically reviewed methods along three paradigm shifts: from hand-crafted features to learned representations, from single-sensor to multimodal fusion, and from single-event training to cross-event generalization. We compiled and organized benchmark datasets across floods, wildfires, landslides, and building damage, identifying critical gaps in geographic coverage, modality diversity, and annotation standardization.

Our benchmark of over 30 methods across 13 datasets yields several findings that challenge prevailing assumptions. Architectural complexity yields diminishing returns for extent mapping, where simple encoder-decoder models frequently match or outperform Transformers and state-space models. Change detection architectures provide clear advantages for damage mapping but not for extent mapping, empirically validating the task distinction at the core of our taxonomy. Foundation models, despite their promise of generality, do not yet consistently outperform well-tuned task-specific baselines, with the gap widening on non-RGB modalities and cross-modal settings. Out-of-distribution deployment confirms that benchmark performance does not guarantee operational reliability. Together, these results argue for regime-aware model selection that considers the mapping target, sensor configuration, temporal availability, and computational constraints, rather than the pursuit of a single universal architecture.

To support reproducibility and future research, we release an open-source evaluation toolbox with standardized data loaders, training protocols, and implementations of all benchmarked models. We hope this resource, together with the empirical findings and identified challenges, helps the community move toward disaster mapping systems that are dependable in practice, not only competitive on leaderboards.

Appendix A
Details of Experimental Setup

All experiments are conducted on NVIDIA A100 GPUs using PyTorch[7] [328]. We evaluate over 30 models across 13 benchmark datasets spanning four hazard domains: infrastructure damage (xBD, BRIGHT, RESCUENET), flood

7. https://pytorch.org/

(CAUFLOOD, KUROSIWO, URBANSARFLOODS), wildfire (FIRESPREAD, FLOGA, S2WCD, Satellite Burned Area), and landslide (GVLM, HR-GLDD, LANDSLIDE4SENSE). To ensure fair comparison, all models within each domain share the same training protocol, with domain-specific choices reflecting the distinct characteristics of each task. Full configuration files are available in our code repository[8].

Input modality and channel handling. The 14 datasets span a wide range of input configurations, from 3-channel RGB (RescueNet) to 200-channel multi-temporal stacks (FireSpread). For bi-temporal datasets, single-branch segmentation models receive all channels concatenated along the channel dimension, while Siamese-style change detection models receive pre- and post-event inputs through separate encoder branches.

Optimization. All models use the AdamW optimizer [262]. The learning rate schedule varies by domain: no scheduling for infrastructure damage, cosine annealing for flood, warm-up polynomial decay for wildfire, and step decay for landslide. The default learning rate is 1×10^{-4}, with model-specific adjustments for architectures that require different scales (e.g., 5×10^{-5} for Mask2Former, 5×10^{-6} for SpectralGPT on infrastructure tasks). For foundation models with pre-trained encoders, we apply a reduced encoder learning rate (typically 0.05–0.1$\times$ the base rate) to preserve pre-trained representations while allowing decoder adaptation.

Loss functions. We tried several combinations of loss functions for each type of task and then selected the optimal setting. Specifically, infrastructure damage mapping uses a combination of cross-entropy and Lovasz-softmax loss [34] to handle the multi-class imbalance inherent in damage grading. Wildfire mapping combines cross-entropy with dice loss. Flood and landslide mapping use cross-entropy alone. Mask2Former uses its built-in loss across all domains.

Data augmentation and normalization. Training augmentations include random horizontal and vertical flips, 90-degree rotations, and random cropping to domain-specific tile sizes (128–512 pixels depending on the dataset's native resolution). For infrastructure damage datasets, we additionally apply a class-balanced smart crop to mitigate the dominance of undamaged background pixels. FireSpread, which uses 200-channel temporal stacks, applies no augmentation. Normalization follows dataset-specific statistics: ImageNet statistics for RGB optical data, and dataset-level channel-wise means and standard deviations for multi-spectral and SAR inputs. Change detection models with heterogeneous pre- and post-event modalities use separate normalization per temporal phase.

Model-specific accommodations. Several models require fixed input sizes that override dataset defaults: SpectralGPT operates at 128×128, SatMAE at 224×224 for wildfire, and DINOv2/CLIP inputs are resized to multiples of their patch size (14 or 16 pixels). Batch sizes are adjusted per model to fit within GPU memory, ranging from 4 (large foundation models) to 32 (lightweight models on small-tile datasets). Inference uses sliding-window prediction with 50% overlap.

8. https://github.com/ChenHongruixuan/AnyDisasterMapping

TABLE 12
Ablation Study on Input Size for SegFormer-B0 on LANDSLIDE4SENSE Dataset.

Input Size	F_1	Pre.	Rec.	IoU	OA
128×128	64.99	65.01	64.96	48.14	98.67
256×256	67.71	64.96	70.72	51.19	98.72
384×384	68.24	68.79	67.71	51.79	98.81
512×512	69.56	67.06	72.24	53.30	98.79
640×640	68.91	67.00	70.93	52.56	98.78

Evaluation. All metrics are computed on the official test splits of each dataset. For datasets without predefined splits, we use the partitions established by prior work. No test-time augmentation is applied to any model, ensuring that reported results reflect single-pass inference performance.

Traditional baselines. In addition to deep models, we evaluate three families of non-deep baselines. *Random guessing* predicts each valid pixel by sampling from the empirical class prior of the corresponding training split. *Index-rule baselines* use hazard-specific physically motivated indices and threshold rules, with thresholds selected on the training/validation data. Because the datasets differ substantially in modality and label space, these rules are dataset-specific rather than uniform across domains: flood baselines rely on SAR backscatter, coherence, optical water/vegetation cues, and simple temporal change rules; landslide baselines use spectral disturbance indicators, brightness/vegetation suppression, and terrain-related priors; wildfire baselines use standard burn-related indices and change measures (e.g., NBR-, NDMI-, MIRBI-, or BAIS2-type cues), while FIRESPREAD uses a rule-based spread-risk score combining recent fire observations, neighborhood propagation, weather, fuel, and topography. *Random-forest baselines* are trained on class-balanced sampled pixels using handcrafted features tailored to each dataset, including raw channels, ratios and differences, temporal change features, local mean/std/range/gradient statistics, and selected GLCM texture descriptors. For heterogeneous pre-/post-event inputs, modality-specific features are extracted first and then concatenated. All hyperparameters, feature definitions, and threshold grids are provided in the code repository.

APPENDIX B
SUPPLEMENTARY EXPERIMENTS

In addition to the principal benchmark comparisons presented in the main text, we conducted a series of supplementary experiments to examine method behavior from complementary perspectives. The results summarized in the following tables and figures, while informative, fall outside the central narrative of the paper. We include them here for completeness and to assist readers who wish to reproduce or extend our analyses.

B.1 Effect of Input Resolution on Landslide Mapping

At the native 10 m resolution of Sentinel-2, individual landslides span only tens to hundreds of pixels, and their boundaries can be ambiguous when mixed with spectrally similar

TABLE 13
Ablation on additional topographic inputs on KuroSiwo. Baseline columns report the SAR-only results. The remaining columns show absolute performance changes after adding DEM, Slope, or DEM+Slope. Colored arrows indicate whether the score increases, decreases, or remains unchanged relative to the baseline.

Family	Model	Baseline				+DEM Δ				+Slope Δ				+DEM+Slope Δ			
		mF1	mIoU	Perm.	Flood	mF1	mIoU	Perm.	Flood	mF1	mIoU	Perm.	Flood	mF1	mIoU	Perm.	Flood
Seg.	UNet	84.19	74.14	62.30	62.86	↑ 0.26	↑ 0.35	↑ 0.69	↑ 0.32	↑ 1.32	↑ 1.80	↑ 2.63	↑ 2.55	↓ 5.81	↓ 7.12	↓ 15.10	↓ 5.83
	UNet++	84.88	75.06	63.72	64.12	↓ 0.71	↓ 0.95	↓ 0.36	↓ 2.36	↓ 0.36	↓ 0.48	↓ 0.59	↓ 0.84	↓ 0.12	↓ 0.16	↑ 0.21	↓ 0.68
	DeepLabV3+ (R50)	84.32	74.33	61.02	64.66	↑ 0.69	↑ 0.91	↑ 2.73	↓ 0.09	↑ 0.52	↑ 0.70	↑ 2.27	↓ 0.29	↑ 0.18	↑ 0.23	↑ 2.17	↓ 1.50
	HRNet-W48	84.49	74.55	61.76	64.48	↑ 0.21	↑ 0.28	↑ 0.83	→ 0.00	↑ 0.36	↑ 0.50	↑ 0.12	↑ 1.28	↓ 0.70	↓ 0.93	↓ 0.69	↓ 1.98
	SegFormer-B4	84.53	74.59	62.13	64.34	↓ 0.45	↓ 0.59	↓ 0.42	↓ 1.31	↓ 0.05	↓ 0.05	↓ 0.11	↓ 0.07	↓ 0.46	↓ 0.61	→ 0.00	↓ 1.75
	Swin-T	84.16	74.10	61.73	63.27	↓ 0.60	↓ 0.80	↓ 0.93	↓ 1.35	↑ 0.27	↑ 0.36	↑ 0.69	↑ 0.36	↓ 0.63	↓ 0.85	↓ 0.87	↓ 1.50
	SAM-ViT-L	82.95	72.51	57.99	62.58	↑ 0.17	↑ 0.20	↑ 2.41	↓ 1.83	↑ 1.30	↑ 1.70	↑ 4.57	↑ 0.30	↑ 0.46	↑ 0.58	↑ 2.80	↓ 1.15
	CLIP-ViT-L/14	82.06	71.34	58.18	59.04	↓ 2.07	↓ 2.57	↓ 3.73	↓ 3.76	↓ 0.11	↓ 0.14	↑ 0.09	↓ 0.51	↓ 1.61	↓ 2.02	↓ 2.63	↓ 3.20
	DINOv2-ViT-L/14	82.53	71.93	59.66	59.29	↓ 0.25	↓ 0.30	↓ 1.37	↑ 0.40	↑ 0.49	↑ 0.65	↑ 0.57	↑ 1.27	↑ 0.28	↑ 0.39	↓ 0.86	↑ 1.92
	DINOv3-ViT-L/16	83.03	72.59	59.35	61.46	↑ 0.48	↑ 0.64	↑ 1.38	↑ 0.42	↑ 0.63	↑ 0.84	↑ 1.27	↑ 1.09	↑ 0.52	↑ 0.69	↑ 1.17	↑ 0.78
Bi-temporal	SiamCRNN	83.59	73.37	59.29	63.57	↓ 0.40	↓ 0.53	↓ 0.49	↓ 1.00	↓ 0.19	↓ 0.25	↓ 0.28	↓ 0.47	↓ 0.06	↓ 0.10	↑ 1.14	↓ 1.37
	DSIFN	83.28	72.95	59.20	62.45	↓ 0.79	↓ 1.01	↓ 2.27	↓ 0.66	↓ 0.04	↓ 0.05	↑ 0.34	↓ 0.50	↓ 0.27	↓ 0.36	↑ 1.49	↓ 2.55
	BIT-CD	83.53	73.27	59.48	63.21	↓ 0.50	↓ 0.66	↑ 0.28	↓ 2.25	↓ 2.00	↓ 2.57	↑ 0.10	↓ 7.49	↓ 0.28	↓ 0.38	↑ 1.19	↓ 2.29
	ChangeOS	83.56	73.34	58.64	64.13	→ 0.00	↓ 0.04	↑ 1.38	↓ 1.38	↑ 0.48	↑ 0.61	↑ 1.65	↑ 0.19	↑ 0.66	↑ 0.85	↑ 2.67	↓ 0.16
	ChangeMamba	83.18	72.80	59.28	62.10	↓ 0.37	↓ 0.47	↓ 0.40	↓ 1.06	↑ 0.84	↑ 1.09	↑ 3.72	↓ 0.50	↓ 0.35	↓ 0.45	↑ 0.14	↓ 1.51

bare soil or shadow. A straightforward strategy to improve boundary delineation is to upsample the input patches to a larger pixel grid, effectively increasing the number of pixels representing the same geographic extent and providing the network with finer spatial detail. We investigate this strategy by training SegFormer-B0 on Landslide4Sense, upsampling each patch from the native 128×128 grid to sizes up to 640×640 while keeping the geographic footprint fixed.

As shown in Table 12, performance improves consistently as the input is upsampled from 128×128 (IoU $= 48.14\%$) to 512×512 (IoU $= 53.30\%$), a gain of over 5 percentage points. The effective pixel spacing decreases from 10 m to approximately 2.5 m, allowing the network to better resolve landslide edges and internal heterogeneity. Both precision and recall benefit: higher resolution helps reject false positives by revealing fine-grained texture differences between landslide deposits and spectrally similar bare ground, and recover true positives by sharpening subtle boundaries that are blurred at the native resolution. However, further upsampling to 640×640 ($\sim$2 m effective spacing) provides no additional benefit (IoU $= 52.56\%$), indicating that the information content of the original 10 m imagery is exhausted beyond a $4\times$ upsampling factor and that additional interpolated pixels introduce no new discriminative detail.

This finding has a practical implication: for landslide mapping with Sentinel-2 imagery, upsampling to 512×512 ($\sim 4\times$ the native resolution) offers the best accuracy–cost trade-off. It also suggests that, for this task, resolution enhancement through simple upsampling can yield accuracy gains comparable to or larger than architectural upgrades, complementing the main benchmark observation (Table 9) that lightweight models already perform competitively at the native resolution.

B.2 Impact of Topographic Priors on Flood Extent Mapping

Topographic variables such as elevation and terrain slope provide physically meaningful context for flood mapping, and datasets such as KuroSiwo therefore include DEM-derived layers alongside SAR imagery [42]. However, their empirical utility for deep learning models in this benchmark setting remains unclear. To examine this, we train a broad set of models on KuroSiwo under four input configurations: SAR only (baseline), SAR+DEM, SAR+Slope, and SAR+DEM+Slope. The results are summarized in Table 13.

Overall, topographic augmentation is not uniformly beneficial in this setting. Adding DEM decreases mIoU in 10 of 15 models, with an average change of approximately -0.37, while adding both DEM and Slope also lowers mIoU in 10 of 15 models and yields the largest average drop (approximately -0.68). Among the three augmented settings, SAR+Slope shows the most favorable aggregate trend, improving mIoU in 9 of 15 models with a modest average gain (approximately $+0.31$). These results suggest that auxiliary topographic channels should not be assumed to improve flood extent mapping by default.

A clearer pattern emerges in the class-specific scores. DEM-based augmentation tends to hurt flood detection more consistently than permanent-water discrimination: the flood score decreases in 11 of 15 models with +DEM and in 13 of 15 models with +DEM+Slope. By contrast, adding Slope more often benefits the permanent-water class, improving its score in 12 of 15 models. This suggests that static terrain cues may be more helpful for separating persistent water bodies from surrounding land than for delineating transient floodwater, whose extent depends on factors beyond topography alone.

The results also do not support a simple "more topographic context is better" conclusion. Although SAR+DEM+Slope provides the richest auxiliary input, it produces the worst average mIoU among the three augmented settings and introduces several large drops for individual models (e.g., UNet: -7.12 mIoU). This may indicate that some architectures struggle to exploit correlated auxiliary channels effectively. However, the present ablation does not isolate whether the issue is caused by model capacity, optimization difficulty, feature redundancy, or preprocessing mismatch.

Model-family trends are also mixed. Among the five bi-temporal architectures, adding DEM reduces mIoU for all models, while DEM+Slope reduces mIoU for four of five and Slope for three of five. This pattern is consistent with the possibility that static topographic channels are less compatible with architectures designed to emphasize

TABLE 14
Impact of input spectral bands on burned area mapping (FLOGA dataset). RGB uses 3 visible bands only; Sen2 uses all 9 available Sentinel-2 bands including NIR and SWIR. ΔF1 shows the absolute change from RGB to Sen2. Models are grouped by family as defined in Section 5.

Family	Model	RGB (3 bands)		Sentinel-2 (9 bands)		ΔF_1
		F_1	IoU	F_1	IoU	
Gen. CNN	FCN-8s	58.75	41.60	86.51	76.22	+27.76
	UNet	84.03	72.46	91.99	85.17	+7.96
	UNet++	85.32	74.40	90.50	82.65	+5.18
	DeepLabV3+ (R50)	88.07	78.68	90.27	82.26	+2.20
	HRNet-W18	88.48	79.34	91.04	83.55	+2.56
	ConvNeXt-T	86.25	75.83	91.78	84.80	+5.53
Gen. Trans.	Swin-B	87.11	77.17	91.96	85.12	+4.85
	SegFormer-B2	88.38	79.18	89.20	80.50	+0.82
	SegFormer-B5	87.36	77.55	90.84	83.22	+3.48
	Mask2Former	89.72	81.35	92.19	85.51	+2.47
Gen. FM	SAM (ViT-L)	87.05	77.06	88.63	79.58	+1.58
	SAM2 (Hiera-B+)	82.67	70.46	85.29	74.35	+2.62
	CLIP (ViT-L)	89.16	80.44	90.32	82.34	+1.16
	DINOv2 (ViT-B)	87.69	78.08	87.70	78.09	+0.01
	DINOv2 (ViT-L)	88.51	79.39	89.43	80.89	+0.92
EO	FarSeg++	90.59	82.80	90.88	83.28	+0.29
	RS^{3}Mamba	88.42	79.25	91.37	84.11	+2.95
Bi-temp.	SiamCRNN	87.68	78.07	90.43	82.54	+2.75
	ChangeOS	87.04	77.05	91.06	83.59	+4.02
	ChangeMamba	72.60	56.99	92.22	85.56	+19.62
EO FM	SatMAE	78.37	64.44	88.14	78.79	+9.77
	SkySense	77.22	62.89	91.55	84.42	+14.33
	SpectralGPT	71.34	55.44	85.89	75.26	+14.55
	HyperSIGMA	84.35	72.93	89.45	80.92	+5.10
	DINOv3 (EO)	83.99	72.39	89.69	81.31	+5.70

explicit temporal modeling, although this mechanism is not directly tested here. Pre-trained ViT-based models are likewise heterogeneous: SAM-ViT-L and DINOv3-ViT-L/16 improve under all three topographic settings, DINOv2-ViT-L/14 shows mixed behavior, and CLIP-ViT-L/14 degrades substantially under DEM-based augmentation. Pretraining therefore does not, by itself, guarantee effective use of auxiliary topographic inputs.

Taken together, these results suggest that DEM and slope should not be treated as universally helpful auxiliary inputs for flood extent mapping on KuroSiwo. In this ablation, Slope provides the most favorable aggregate trade-off, but the gains are modest and architecture-dependent. More broadly, these findings should not be interpreted as evidence that topographic priors are intrinsically unhelpful for flood mapping. Rather, their utility may depend on the spatial resolution and alignment of the topographic data, the sensor resolution, and the fusion strategy used by the model. It is therefore plausible that topographic priors would become more beneficial in higher-resolution settings, where finer-scale terrain structure is better aligned with inundation boundaries, but this remains to be verified empirically.

B.3 How Well Do Different Model Families Exploit multispectral Information?

NIR and SWIR bands carry diagnostic information for burned area mapping is well established. The more relevant question for this review is whether models pre-trained on RGB imagery can effectively exploit these additional bands when made available, or whether their learned representations remain anchored to the visible spectrum. We test this on FLOGA by comparing all models under two input configurations: RGB only (3 visible bands) and full Sentinel-2 (9 bands including NIR and SWIR). Table 14 reports the results.

The response to additional spectral bands varies dramatically across model families, revealing more about representational capacity than about burned area mapping itself. Three patterns emerge.

General vision foundation models pre-trained on RGB are largely unable to exploit multispectral inputs. DINOv2 (ViT-B) gains effectively nothing ($\Delta F_1 = +0.01$), and CLIP (ViT-L) gains only 1.16 points. Despite fine-tuning, their feature extractors appear to map unfamiliar spectral channels to redundant representations, effectively ignoring the additional information. This suggests that RGB-scale pre-training creates a representational bottleneck that lightweight adaptation cannot overcome for non-RGB modalities.

EO-native foundation models, by contrast, show substantial gains (ΔF_1 of +9.8 to +14.6 for SatMAE, SkySense, and SpectralGPT). This is one of the few settings in our entire benchmark where EO-specific pre-training confers a clear, measurable advantage over general vision pre-training. The result indicates that exposure to multispectral data during pre-training does produce representations capable of leveraging NIR and SWIR information, even if this advantage does not manifest on RGB-only tasks.

Standard deep learning architectures without pre-training occupy a middle ground but reveal an important pattern. Most CNNs and Transformers improve by 2–8 points, consistent with their ability to learn spectral features from scratch during task-specific training. The most striking case is ChangeMamba, which moves from the weakest model on RGB ($F_1 = 72.60\%$) to the strongest on Sentinel-2 ($F_1 = 92.22\%$, $\Delta F_1 = +19.6$). Its temporal scanning mechanism, largely ineffective on RGB colour differences, becomes highly discriminative when operating on the pronounced NIR/SWIR contrast between pre- and post-fire vegetation states. This demonstrates that architectural evaluations can be fundamentally misleading when conducted on a single input configuration: a model judged inferior on RGB may be superior under the operationally relevant multispectral setting.

The practical implication is twofold. For burned area mapping, input configuration matters more than architectural choice: a simple UNet with Sentinel-2 input ($F_1 = 91.99\%$) outperforms every general vision FM on RGB regardless of parameter count. For the foundation model community, the RGB lock-in effect observed here poses a concrete challenge: if pre-trained representations cannot absorb non-RGB channels, the promise of universal feature extractors remains unfulfilled for multispectral EO applications.

B.4 Investigation of Global Building Footprint Products for Building Damage Mapping

A practical question for operational damage mapping is whether existing global building footprint products can substitute for, or complement, the building localization branch of dedicated damage models. If these products provide sufficiently accurate footprints, damage assessment could be

TABLE 15
Agreement between global building footprint products and BRIGHT annotations. Three evaluation settings are reported: *All tiles* treats missing coverage as negative predictions; *Covered tiles* restricts evaluation to tiles where the product has non-empty predictions; *Joint coverage* evaluates only the 168 tiles (4.0% of the dataset) where all three products have predictions. All metrics are pixel-level binary agreement with BRIGHT labels (damage level ≥ 1 binarized as building).

Product	Tiles	Cov.(%)	All tiles ($n = 4{,}246$)				Covered tiles				Joint coverage ($n = 168$)			
			Pre.	Rec.	F_1	IoU	Pre.	Rec.	F_1	IoU	Pre.	Rec.	F_1	IoU
Overture Maps	3,135	73.83	75.24	55.14	63.64	46.67	75.24	57.52	65.20	48.37	55.84	58.27	57.03	39.89
Microsoft Building Footprints	2,287	53.86	74.08	46.55	57.18	40.03	74.08	58.78	65.55	48.75	67.72	42.04	51.88	35.03
Google Open Buildings	1,068	25.15	68.90	7.27	13.14	7.03	68.90	35.79	47.11	30.81	65.25	52.69	58.30	41.14

simplified to semantic segmentation followed by footprint intersection, as discussed in Section 2. We evaluate this question in two steps: first, we measure the spatial agreement between three widely used products and expert annotations; second, we test whether these footprints can serve as pseudo labels for training damage mapping models.

B.4.1 Spatial Agreement with Expert Annotations

We compare Overture Maps[9], Microsoft Building Footprints[10], and Google Open Buildings[11] against the BRIGHT dataset [58] by rasterizing their polygons onto the BRIGHT label grid and computing pixel-level binary agreement (treating all damage levels as building-positive).

As shown in Table 15, coverage varies dramatically: Overture provides predictions for 73.83% of BRIGHT tiles, Microsoft for 53.86%, and Google for only 25.15%. The intersection where all three products have non-empty predictions covers a mere 4.0% of the dataset. On the full dataset (counting missing coverage as all-negative), Overture achieves the highest IoU (46.67%), followed by Microsoft (40.03%) and Google (7.03%). Google's low score is driven primarily by limited geographic coverage rather than poor prediction quality: on tiles where each product actually provides predictions, Overture and Microsoft achieve comparable IoU (48.37% and 48.75%), while Google reaches 30.81%. On the joint-coverage subset, Google slightly outperforms the other two, suggesting that where it does provide data, its footprints are competitive.

Across all products and settings, recall rate is consistently lower than precision rate, indicating that the primary failure mode is omission rather than false detection, a pattern likely exacerbated in post-disaster settings where damaged structures no longer match the intact-building signatures these products were trained on.

The event-level breakdown in Table 16 reveals sharp geographic disparities. All three products perform best in regions with dense, well-mapped urban infrastructure (e.g., Turkey-Earthquake, Beirut-Explosion). Performance degrades in the Global South and in areas with informal construction. Most critically, several products have zero coverage for entire events: Microsoft provides no footprints for La Palma, Libya, or Morocco; Google covers none of the events in Turkey, Ukraine, Morocco, or six other locations. A damage mapping pipeline relying on any single product would therefore have complete blind spots for certain disasters.

9. https://overturemaps.org/
10. https://github.com/microsoft/GlobalMLBuildingFootprints/
11. https://sites.research.google/gr/open-buildings/

Pixel-level agreement is a useful geometric check, but in practice historical footprints are mainly used to provide building locations. For that purpose, exact boundary alignment is less important than whether a building is covered at all; missing a building is more consequential than a small contour offset. We therefore further evaluate the products at the building-instance level using the aligned instance annotations of BRIGHT. Table 17 reports the resulting GT-centric building coverage.

The instance-level analysis sharpens the picture from the raster comparison. Overture remains the strongest overall product, achieving an instance recall of 73.48% at a 10% coverage threshold, compared with 64.55% for Microsoft and 16.18% for Google. Microsoft has the highest polygon-level precision (82.55%), indicating cleaner predictions but lower coverage than Overture. Google again appears competitive only where it has actual support, but its sparse event coverage causes its whole-dataset building recall to collapse. This result is important for disaster mapping: the main limitation of these products is not only imperfect boundary alignment, but also systematic omission of buildings and entire uncovered regions.

B.4.2 Training with Footprint-Derived Pseudo Labels

The agreement analysis above measures footprint quality as a geometric prior. A more demanding test is whether these footprints can replace expert annotations entirely during model training. To evaluate this, we trained ChangeMamba on BRIGHT using Overture polygons as pseudo labels for the localization branch, while retaining the original damage grade annotations for classification. Table 18 compares the resulting model against the same architecture trained with expert labels.

The pseudo-label model achieves a localization F1 of 70.84%, a drop of 20 points from the expert-trained baseline (90.99%). Damage classification degrades further: F_1^{clf} falls from 75.47% to 46.16%, with the sharpest decline on the "destroyed" class (IoU = 17.56%). This disproportionate failure on severe damage has a specific cause beyond general label noise: footprint products are constructed from more recent available imagery, which often postdates the disaster. Buildings that were completely destroyed may therefore be absent from or misrepresented in the polygons, introducing systematic temporal misalignment between pseudo labels and the actual pre-disaster building inventory. Less severe damage categories, where buildings remain largely intact in post-disaster imagery, are less affected.

TABLE 16
Event-level building footprint agreement on BRIGHT, evaluated on tiles where each product has non-empty predictions. Coverage shows the number of covered tiles and percentage of the event's total tiles. Events are sorted by total sample count. "–" indicates zero coverage.

Event	Tiles	Overture Maps					Microsoft Building Footprints					Google Open Buildings				
		Cov.	Pre.	Rec.	F_1	IoU	Cov.	Pre.	Rec.	F_1	IoU	Cov.	Pre.	Rec.	F_1	IoU
Turkey-Earthquake	1,114	1,105 (99.2%)	76.67	58.43	66.32	49.61	1,109 (99.6%)	75.86	61.90	**68.17**	**51.72**	–	–	–	–	–
La Palma-Volcano	933	666 (71.4%)	87.63	56.29	**68.54**	**52.14**	–	–	–	–	–	614 (65.8%)	85.33	26.34	40.25	25.19
Morocco-Earthquake	567	194 (34.2%)	61.57	32.61	**42.64**	**27.09**	–	–	–	–	–	–	–	–	–	–
Ukraine-Conflict	513	448 (87.3%)	69.39	61.23	65.06	48.21	452 (88.1%)	76.81	63.66	**69.62**	**53.39**	–	–	–	–	–
Mexico-Hurricane	212	179 (84.4%)	68.92	64.13	**66.44**	**49.75**	111 (52.4%)	71.82	53.19	61.12	44.01	180 (84.9%)	81.62	47.37	59.95	42.81
Beirut-Explosion	133	125 (94.0%)	73.85	51.84	60.92	43.80	133 (100%)	74.17	51.90	**61.07**	**43.95**	–	–	–	–	–
Myanmar-Hurricane	126	12 (9.5%)	75.80	9.37	16.68	9.10	78 (61.9%)	45.79	36.09	**40.36**	**25.28**	–	–	–	–	–
Libya-Flood	124	111 (89.5%)	87.31	63.30	**73.39**	**57.97**	–	–	–	–	–	–	–	–	–	–
Congo-Volcano	123	118 (95.9%)	34.89	53.97	42.38	26.89	58 (47.2%)	36.17	8.55	13.83	7.43	103 (83.7%)	43.05	57.58	**49.26**	**32.68**
Bata-Explosion	107	4 (3.7%)	29.51	12.39	17.45	9.56	94 (87.9%)	51.17	56.83	**53.85**	**36.85**	98 (91.6%)	42.52	47.96	45.08	29.10
Noto-Earthquake	79	48 (60.8%)	77.31	65.05	**70.65**	**54.62**	50 (63.3%)	58.24	44.89	50.70	33.96	–	–	–	–	–
Marshall-Wildfire	77	59 (76.6%)	71.35	66.38	68.78	52.41	75 (97.4%)	70.53	78.00	**74.07**	**58.82**	–	–	–	–	–
Haiti-Earthquake	73	5 (6.8%)	58.14	9.81	16.78	9.16	69 (94.5%)	62.48	29.15	39.76	24.81	73 (100%)	56.80	52.10	**54.35**	**37.31**
Hawaii-Wildfire	65	61 (93.8%)	67.34	38.87	**49.29**	**32.70**	58 (89.2%)	46.48	23.78	31.47	18.67	–	–	–	–	–

TABLE 17
Building instance-level coverage of global footprint products on BRIGHT. We report GT-centric instance coverage using aligned building-instance annotations. "Prediction Precision (Any)" is the fraction of predicted polygons overlapping at least one annotated building. "Instance Recall @ 0.1" indicates that a building is counted as covered when at least 10% of its annotated footprint is overlapped by a product polygon.

Product	Pred. Prec. (Any)↑	Inst. Rec. (Any)↑	Inst. Rec.@0.1↑	Inst. Rec.@0.5↑	Area Rec.@0.1↑
Overture	79.46	77.26	73.48	46.54	80.96
Microsoft	82.55	68.76	64.55	42.50	73.33
Google	70.60	17.89	16.18	10.14	10.18

TABLE 18
Impact of replacing expert annotations with Overture footprint pseudo labels for training ChangeMamba on BRIGHT. Per-class IoU is reported for the pseudo-label model.

Training labels	F_1^{loc}	F_1^{clf}	OA	mIoU
Expert	90.99	75.47	96.31	68.92
Overture pseudo labels	70.84	46.16	92.24	48.83

Per-class IoU			
Background	Intact	Damaged	Destroyed
92.45	51.54	33.78	17.56

B.4.3 Summary

Taken together, these experiments show that global building footprint products are useful but insufficient for operational damage mapping. Although their footprints could potentially serve as geometric priors for post-processing (e.g., majority voting within footprints) to improve spatial consistency, this remains to be validated experimentally. Among the three products evaluated, Overture Maps currently offers the best coverage-quality trade-off. However, even the best product achieves a localization F1 of only 65% on its covered tiles, compared to 89–91% from dedicated models, and using footprints as training labels leads to substantial degradation, particularly for *destroyed* classes.

Moreover, the event-level analysis reveals that no single product guarantees coverage for all disasters. Microsoft and Google each have zero footprints for multiple events, making any pipeline built on a single product vulnerable to complete blind spots. Closing these gaps would require not only broader geographic coverage in disaster-prone regions of the Global South, but also temporally aligned footprint products that reflect the pre-disaster building inventory. As our pseudo-label experiment shows, footprints derived from post-disaster imagery systematically omit destroyed buildings, which is the primary cause of the disproportionate degradation on severe damage classes.

B.5 Sensitivity to Pre-event Optical Source Variability

In cross-modal damage mapping, the pre-event optical image serves as the reference against which post-event SAR observations are interpreted. The model must learn to distinguish genuine damage-induced changes in the SAR signal from the inherent appearance gap between optical and SAR modalities. Variations in the pre-event optical source, including spatial resolution, radiometric characteristics, orthorectification quality, and viewing or illumination conditions, may introduce source-specific discrepancies that are unrelated to actual damage. To show the effect of this sensitivity, we evaluate ChangeMamba and Swin-Small on the 2025 Saganoseki fire case study using pre-event optical imagery from three different providers: Maxar, GSI, and Bing Maps.

The qualitative results in Fig. 10 reveal a clear pattern.When Maxar imagery is used, both models produce damage maps that are broadly consistent with the reference labels in Fig. 10-(l). When GSI or Bing imagery is substituted in Fig. 10-(b)–(c), damage classification degrades substantially in both ChangeMamba and Swin-Small, as shown in Fig. 10-(d)–(i). Building localization, however, remains largely accurate across all three sources. This suggests that building localization is less sensitive to the choice of pre-event optical source, whereas damage grading is

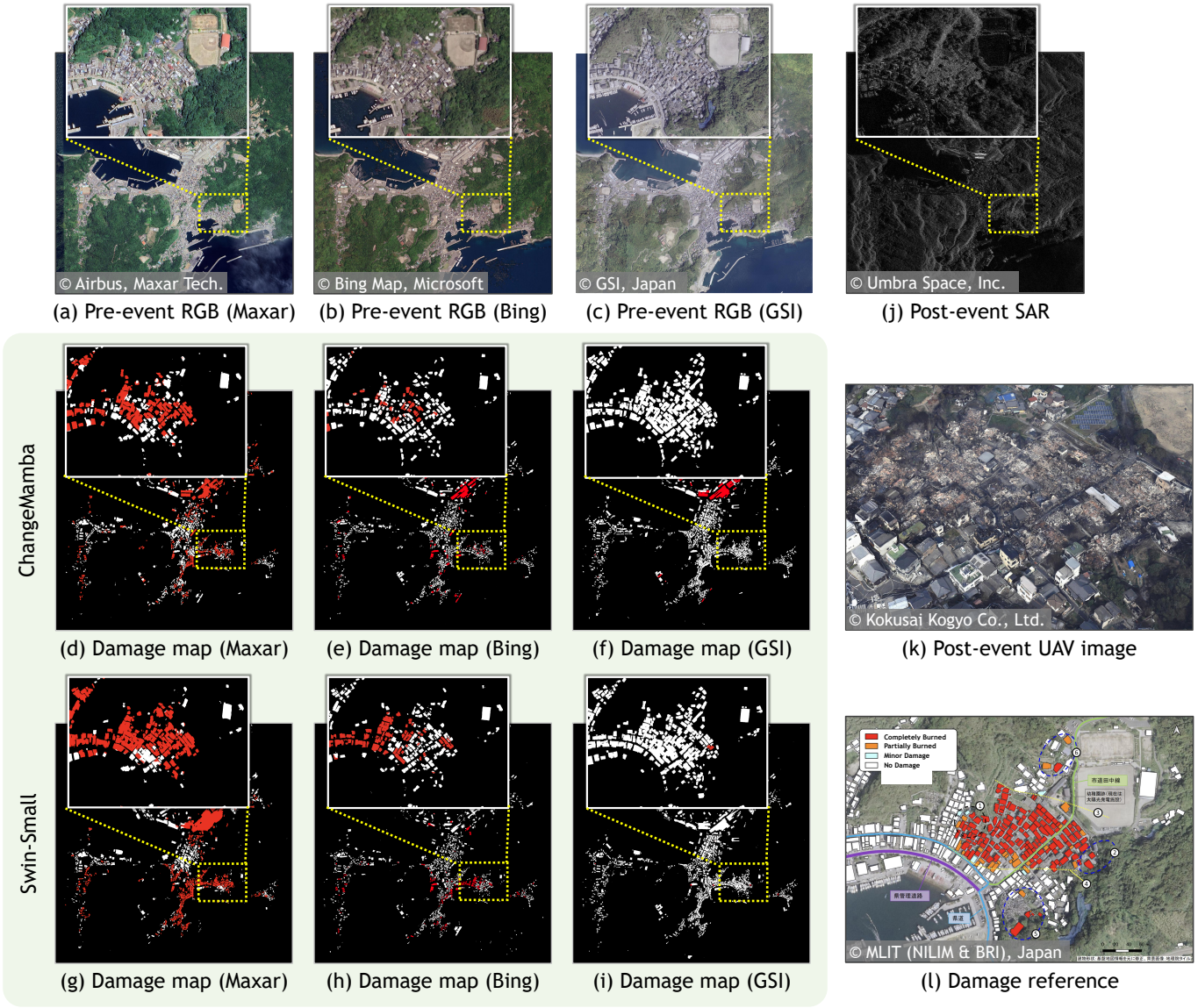

Fig. 10. Qualitative comparison of cross-modal damage mapping results under different pre-event optical sources for the 2025 Saganoseki fire case. (a)–(c) Pre-event optical imagery from Maxar, Bing Maps, and GSI, respectively; (j) post-event SAR image. (d)–(f) Damage maps predicted by ChangeMamba and (g)–(i) those predicted by Swin-Small using each pre-event source. (k) Post-event UAV image and (l) damage reference.

more dependent on the learned cross-modal correspondence between pre-event optical appearance and post-event SAR response. When an unfamiliar optical source is introduced, the model's learned mapping from optical appearance to expected SAR response is disrupted, and the subtle cross-modal cues that distinguish damage levels are overwhelmed by source-induced distribution shift.

This finding has direct operational implications. In rapid response, the pre-event optical reference is often sourced from whatever archive or basemap is available, which may differ from the imagery used during training. Our results indicate that pre-event source mismatch can be a substantial source of error, in some cases comparable to or larger than the differences between model architectures. Mitigating this sensitivity through source-diverse training augmentation, optical preprocessing harmonization, or explicit cross-source domain adaptation is an important priority for operational cross-modal damage mapping.

Acknowledgment

This work was supported by the Council for Science, Technology and Innovation (CSTI), the Cross-ministerial Strategic Innovation Pro-motion Program (SIP), Development of a Resilient Smart Network System against Natural Disasters (Funding agency: NIED); the JSPS, KAKENHI under Grant Number 22H03609, 24KJ0652, 25K03145, 26K21262, and 26K21244; JST, FOREST under Grant Number JPMJFR206S; Microsoft Research Asia; Next Generation AI Research Center of The University of Tokyo; Young Researchers Exchange Programme between Japan and Switzerland under the Japanese-Swiss Science and Technology Programme; and RIKEN Incentive Research 2026.

References

[1] Saygin Abdikan, Caglar Bayik, Aliihsan Sekertekin, Filiz Bektas Balcik, Sadra Karimzadeh, Masashi Matsuoka, and Fusun

Balik Sanli. Burned area detection using multi-sensor sar, optical, and thermal data in mediterranean pine forest. *Forests*, 13(2):347, 2022.

[2] Sara El Amrani Abou El Assad. *Flood Detection with a Deep Learning Approach Using Optical and SAR Satellite Data*. PhD thesis, MS thesis, 2019.

[3] Pierre Adorni, Minh-Tan Pham, Stéphane May, and Sébastien Lefèvre. Towards efficient benchmarking of foundation models in remote sensing: A capabilities encoding approach. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3087–3097, 2025.

[4] Bruno Adriano, Junshi Xia, Gerald Baier, Naoto Yokoya, and Shunichi Koshimura. Multi-source data fusion based on ensemble learning for rapid building damage mapping during the 2018 sulawesi earthquake and tsunami in palu, indonesia. *Remote Sensing*, 11(7), 2019.

[5] Bruno Adriano, Naoto Yokoya, Junshi Xia, Hiroyuki Miura, Wen Liu, Masashi Matsuoka, and Shunichi Koshimura. Learning from multimodal and multitemporal earth observation data for building damage mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175:132–143, 2021.

[6] Bruno Adriano, Naoto Yokoya, Kazuki Yamanoi, and Satoru Oishi. Predicting flood inundation depth based-on machine learning and numerical simulation. In *CEUR Workshop Proceedings*, volume 3207, pages 58–64. CEUR-WS. org, 2022.

[7] J. C. J. H. Aerts, W. J. Botzen, K. C. Clarke, S. L. Cutter, J. W. Hall, B. Merz, E. Michel-Kerjan, J. Mysiak, S. Surminski, and H. Kunreuther. Integrating human behaviour dynamics into flood disaster risk assessment. *Nature Climate Change*, 8(3):193–199, 2018.

[8] Kashif Ahmad, Konstantin Pogorelov, Michael Riegler, Olga Ostroukhova, Pål Halvorsen, Nicola Conci, and Rozenn Dahyot. Automatic detection of passable roads after floods in remote sensed and social media data. *Signal Processing: Image Communication*, 74:110–118, 2019.

[9] Seyed Ali Ahmadi, Ali Mohammadzadeh, Naoto Yokoya, and Arsalan Ghorbanian. Bd-skunet: Selective-kernel unets for building damage assessment in high-resolution satellite images. *Remote Sensing*, 16(1), 2024.

[10] Kyeongjin Ahn, Sungwon Han, Sungwon Park, Jihee Kim, Sangyoon Park, and Meeyoung Cha. Generalizable disaster damage assessment via change detection with vision foundation model. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'25/IAAI'25/EAAI'25. AAAI Press, 2025.

[11] Eleanor A. Ainscoe, Rohini Swaminathan, Lin Way, Sirio Modugno, Shi Tong Chin, Niroj Panta, Thierry Crevoisier, and Sang-Ho Yun. Earthquake damage mapped more comprehensively and accurately by radar satellites than optical imagery. *Communications Earth & Environment*, 6(1), August 2025.

[12] Peri Akiva, Matthew Purri, Kristin Dana, Beth Tellman, and Tyler Anderson. H2o-net: Self-supervised flood segmentation via adversarial domain adaptation and label refinement. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 111–122, 2021.

[13] Firoj Alam, Ferda Ofli, and Muhammad Imran. Crisismmd: Multimodal twitter datasets from natural disasters. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1), Jun. 2018.

[14] Muhammad Usman Ali, Waqas Sultani, and Mohsen Ali. Destruction from sky: Weakly supervised approach for destruction detection in satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:115–124, 2020.

[15] Sk Ajim Ali, Farhana Parvin, Quoc Bao Pham, Khaled Mohamed Khedher, Mahro Dehbozorgi, Yasin Wahid Rabby, Duong Tran Anh, and Duc Hiep Nguyen. An ensemble random forest tree with svm, ann, nbt, and lmt for landslide susceptibility mapping in the rangit river watershed, india. *Natural Hazards*, 113(3):1601–1633, 2022.

[16] Liqiang An, Jingfa Zhang, Lixia Gong, and Qiang Li. Integration of sar image and vulnerability data for building damage degree estimation. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 4263–4266, 2016.

[17] Gustavo A. Arciniegas, Wietske Bijker, Norman Kerle, and Valentyn A. Tolpekin. Coherence- and Amplitude-Based Analysis of Seismogenic Damage in Bam, Iran, Using ENVISAT ASAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1571–1581, 2007.

[18] RJ Aroma, K Raimond, VV Estrela, and MA De Jesus. A coastal band spectral combination for water body extraction using landsat 8 images. *International Journal of Environmental Science and Technology*, 21(2):1767–1784, 2024.

[19] Francesco Asaro, Gianluca Murdaca, and Claudio Maria Prati. Learning deep models from weak labels for water surface segmentation in sar images. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 6048–6051. IEEE, 2021.

[20] Guillaume Astruc, Nicolas Gonthier, Clément Mallet, and Loic Landrieu. Anysat: One earth observation model for many resolutions, scales, and modalities. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19530–19540, 2025.

[21] Colin Axel and Jan A. N. van Aardt. Building damage assessment using airborne lidar. *Journal of Applied Remote Sensing*, 11(4):046024, 2017.

[22] Shadi Sadat Baghermanesh, Shabnam Jabari, and Heather McGrath. Urban flood detection using sentinel1-a images. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 527–530. IEEE, 2021.

[23] Yanbing Bai, Bruno Adriano, Erick Mas, Hideomi Gokon, and Shunichi Koshimura. Object-based building damage assessment methodology using only post event alos-2/palsar-2 dual polarimetric sar intensity images. *Journal of Disaster Research*, 12(2):259–271, 2017.

[24] Yanbing Bai, Bruno Adriano, Erick Mas, and Shunichi Koshimura. Building damage assessment in the 2015 gorkha, nepal, earthquake using only post-event dual polarization synthetic aperture radar imagery. *Earthquake Spectra*, 33(1_suppl):185–195, 2017.

[25] Yanbing Bai, Bruno Adriano, Erick Mas, and Shunichi Koshimura. Machine learning based building damage mapping from the alos-2/palsar-2 sar imagery: Case study of 2016 kumamoto earthquake. *Journal of Disaster Research*, 12(sp):646–655, 2017.

[26] Yanbing Bai, Chang Gao, Sameer Singh, Magaly Koch, Bruno Adriano, Erick Mas, and Shunichi Koshimura. A framework of rapid regional tsunami damage recognition from post-event terrasar-x imagery using deep neural networks. *IEEE Geoscience and Remote Sensing Letters*, 15(1):43–47, 2018.

[27] Timo Balz and Mingsheng Liao. Building-damage detection using post-seismic high-resolution sar satellite data. *International Journal of Remote Sensing*, 31(13):3369–3391, 2010.

[28] Yifang Ban, Puzhao Zhang, Andrea Nascetti, Alexandre R Bevington, and Michael A Wulder. Near real-time wildfire progression monitoring with Sentinel-1 SAR time series and deep learning. *Scientific reports*, 10(1):1322, 2020.

[29] Panagiotis Barmpoutis, Tania Stathaki, Kosmas Dimitropoulos, and Nikos Grammalidis. Early fire detection based on aerial 360-degree sensors, deep convolution neural networks and exploitation of fire dynamic textures. *Remote Sensing*, 12(19):3177, 2020.

[30] Aitor Bastarrika, Armando Rodriguez-Montellano, Ekhi Roteta, Stijn Hantson, Magí Franquesa, Leyre Torre, Jon Gonzalez-Ibarzabal, Karmele Artano, Pilar Martinez-Blanco, Amaia Mesanza, et al. An automatic procedure for mapping burned areas globally using sentinel-2 and viirs/modis active fires in google earth engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 218:232–245, 2024.

[31] Kavita Devanand Bathe and Nita Sanjay Patil. Flood detection and mapping: A critical review of methods, challenges and future prospects. *Journal of the Indian Society of Remote Sensing*, pages 1–32, 2025.

[32] Mariana Belgiu and Lucian Drăguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114:24–31, 2016.

[33] Roberto Bentivoglio, Elvin Isufi, Sebastian Nicolaas Jonkman, and Riccardo Taormina. Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrology and Earth System Sciences Discussions*, 2022:1–50, 2022.

[34] Maxim Berman, Amal Rannen Triki, and Matthew B. Blaschko. The lovasz-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks.

In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4413–4421, 2018.

[35] María Bermúdez, Luis Cea, and Jerónimo Puertas. A rapid flood inundation model for hazard mapping based on least squares support vector machine regression. *Journal of Flood Risk Management*, 12:e12522, 2019.

[36] Benjamin Bischke, Patrick Helber, Zhengyu Zhao, Jens A. Bruijn, and Damian Borth. The multimedia satellite task at mediaeval 2018. In *MediaEval Benchmarking Initiative for Multimedia Evaluation*, 2018.

[37] T. Blaschke. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1):2–16, 2010.

[38] Ali Darvishi Boloorani, Mehdi Darvishi, Qihao Weng, and Xiangtong Liu. Post-war urban damage mapping using insar: The case of mosul city in iraq. *ISPRS International Journal of Geo-Information*, 10(3):140, March 2021.

[39] Derrick Bonafilia, Beth Tellman, Tyler Anderson, and Erica 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*, pages 210–211, 2020.

[40] David A Bonneau, Paul-Mark DiFrancesco, and D Jean Hutchinson. A method for vegetation extraction in mountainous terrain for rockfall simulation. *Remote Sensing of Environment*, 251:112098, 2020.

[41] Luigi Borrelli, Massimo Conforti, and Michele Mercuri. Lidar and UAV system data to analyse recent morphological changes of a small drainage basin. *ISPRS International Journal of Geo-Information*, 8(12):536, 2019.

[42] Nikolaos Ioannis Bountos, Maria Sdraka, Angelos Zavras, Ilektra Karasante, Andreas Karavias, Themistocles Herekakis, Angeliki Thanasou, Dimitrios Michail, and Ioannis Papoutsis. Kuro siwo: 33 billion m^2 under the water. a global multi-temporal satellite dataset for rapid flood mapping. *arXiv preprint arXiv:2311.12056*, 2023.

[43] Abdullah M. Braik and Maria Koliou. Automated building damage assessment and large-scale mapping by integrating satellite imagery, gis, and deep learning. *Computer-Aided Civil and Infrastructure Engineering*, 39(15):2389–2404, 2024.

[44] Brian Brisco, Andreas Schmitt, Kevin Murnaghan, Shannon Kaya, and Achim Roth. SAR polarimetric change detection for flooded vegetation. *International Journal of Digital Earth*, 6(2):103–114, 2013.

[45] Christopher F. Brown, Michal R. Kazmierski, Valerie J. Pasquarella, William J. Rucklidge, Masha Samsikova, Chenhui Zhang, Evan Shelhamer, Estefania Lahera, Olivia Wiles, Simon Ilyushchenko, Noel Gorelick, Lihui Lydia Zhang, Sophia Alj, Emily Schechter, Sean Askay, Oliver Guinan, Rebecca Moore, Alexis Boukouvalas, and Pushmeet Kohli. Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data, 2025.

[46] Dominik Brunner, Guido Lemoine, and Lorenzo Bruzzone. Earthquake damage assessment of buildings using vhr optical and sar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 48(5):2403–2420, 2010.

[47] Yves Bühler, Elisabeth D Hafner, Benjamin Zweifel, Mathias Zesiger, and Holger Heisig. Where are the avalanches? rapid spot6 satellite data acquisition to map an extreme avalanche period over the swiss alps. *The Cryosphere*, 13(12):3225–3238, 2019.

[48] Jiehua Cai, Lu Zhang, Jie Dong, Changcheng Wang, and Mingsheng Liao. Polarimetric sar pixel offset tracking for large-gradient landslide displacement mapping. *International Journal of Applied Earth Observation and Geoinformation*, 112:102867, 2022.

[49] Han Cao, Hong Zhang, Chao Wang, and Bo Zhang. Operational flood detection using sentinel-1 sar data over large areas. *Water*, 11(4):786, 2019.

[50] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jegou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9630–9640, 2021.

[51] Brett B. Carr, Einat Lev, Theresa Sawi, Kristen A. Bennett, Christopher S. Edwards, S. Adam Soule, Silvia Vallejo Vargas, and Gayatri Indah Marliyani. Mapping and classification of volcanic deposits using multi-sensor unoccupied aerial systems. *Remote Sensing of Environment*, 264:112581, 2021.

[52] Nicola Casagli, Emanuele Intrieri, Veronica Tofani, Giovanni Gigli, and Federico Raspini. Landslide detection, monitoring and prediction with remote-sensing techniques. *Nature Reviews Earth & Environment*, 4(1):51–64, 2023.

[53] Edinson Andrés Solarte Casanova, Diego Reale, Eugenio Sansosti, and Gianfranco Fornaro. Linear and planar projections of dinsar multi-orbit displacement measurements: General formulation and application to landslides. *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[54] Centre for Research on the Epidemiology of Disasters. 2023 Disasters in Numbers. Technical report, CRED, UCLouvain, Brussels, 2024.

[55] Fang Chen, Yao Sun, Lei Wang, Ning Wang, Huichen Zhao, and Bo Yu. Hrtbda: a network for post-disaster building damage assessment based on remote sensing images. *International Journal of Digital Earth*, 17(1):2418880, 2024.

[56] Hao Chen, Zipeng Qi, and Zhenwei Shi. Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

[57] Hongruixuan Chen, Edoardo Nemni, Sofia Vallecorsa, Xi Li, Chen Wu, and Lars Bromley. Dual-tasks siamese transformer framework for building damage assessment. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 1600–1603, 2022.

[58] Hongruixuan Chen, Jian Song, Olivier Dietrich, Clifford Broni-Bediako, Weihao Xuan, Junjue Wang, Xinlei Shao, Yimin Wei, Junshi Xia, Cuiling 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*, 17(11):6217–6253, 2025.

[59] Hongruixuan Chen, Jian Song, Chengxi Han, Junshi Xia, and Naoto Yokoya. Changemamba: Remote sensing change detection with spatiotemporal state space model. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–20, 2024.

[60] Hongruixuan Chen, Naoto Yokoya, and Marco Chini. Fourier domain structural relationship analysis for unsupervised multimodal change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 198:99–114, 2023.

[61] Hongruixuan Chen, Naoto Yokoya, Chen Wu, and Bo Du. Unsupervised Multimodal Change Detection Based on Structural Relationship Graph Representation Learning. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–18, 2022.

[62] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 833–851, Cham, 2018. Springer International Publishing.

[63] Lize Chen, Xuhui Shen, Shunying Hong, and Ye Shao. Estimation of the building damages in Yushu earthquake based on ALOS SAR data. In *MIPPR 2011: Remote Sensing Image Processing, Geographic Information Systems, and Other Applications*, volume 8006, page 800626, 2011.

[64] Qihao Chen, Hui Yang, Linlin Li, and Xiuguo Liu. A novel statistical texture feature for sar building damage assessment in different polarization modes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:154–165, 2020.

[65] Sean Andrew Chen, Andrew Escay, Christopher Haberland, Tessa Schneider, Valentina Staneva, and Youngjun Choe. Benchmark dataset for automatic damaged building detection from post-hurricane remotely sensed imagery. *arXiv preprint arXiv:1812.05581*, 2018.

[66] Si-Wei Chen and Motoyuki Sato. Tsunami Damage Investigation of Built-Up Areas Using Multitemporal Spaceborne Full Polarimetric SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4):1985–1997, 2013.

[67] Wei Chen, Yang Li, Weifeng Xue, Himan Shahabi, Shaojun Li, Haoyuan Hong, Xiaojing Wang, Huiyuan Bian, Shuai Zhang, Biswajeet Pradhan, et al. Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. *Science of The Total Environment*, 701:134979, 2020.

[68] Bowen Cheng, Ishan Misra, Alexander G. Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1280–1289, 2022.

[69] C. Cheng, A. Behzadan, and A. Noshadravan. DoriaNET: A visual dataset from hurricane dorian for post-disaster building damage assessment. DesignSafe-CI, 2021. Version 2.

[70] Gong Cheng, Junwei Han, Peicheng Zhou, and Lei Guo. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 98:119–132, 2014.

[71] Anne-Lise Chesnel, Renaud Binet, and Lucien Wald. Object oriented assessment of damage due to natural disaster using very high resolution images. In *2007 IEEE International Geoscience and Remote Sensing Symposium*, pages 3736–3739, 2007.

[72] T. Chiba. Post-kumamoto earthquake (16 april 2016) rupture lidar scan, 2018. Airborne LiDAR survey by Air Asia Survey Co., Ltd.; accessed 2026-04-23.

[73] M. Chini, C. Bignami, S. Stramondo, and N. Pierdicca. Uplift and subsidence due to the 26 december 2004 indonesian earthquake detected by sar data. *International Journal of Remote Sensing*, 29(13):3891–3910, 2008.

[74] M. Chini, F. R. Cinti, and S. Stramondo. Co-seismic surface effects from very high resolution panchromatic images: the case of the 2005 kashmir (pakistan) earthquake. *Natural Hazards and Earth System Sciences*, 11(3):931–943, 2011.

[75] Marco Chini, Nazzareno Pierdicca, and William J. Emery. Exploiting sar and vhr optical images to quantify damage caused by the 2003 bam earthquake. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1):145–152, 2009.

[76] Marco Chini, Alessandro Piscini, Francesca Romana Cinti, Stefania Amici, Rosa Nappi, and Paolo Marco DeMartini. The 2011 tohoku (japan) tsunami inundation and liquefaction investigated through optical, thermal, and sar data. *IEEE Geoscience and Remote Sensing Letters*, 10(2):347–351, 2013.

[77] P Chitra et al. Flood detection from satellite images using self-attention empowered efficientnetv2 mechanism. In *2023 7th international conference on computing, communication, control and automation (ICCUBEA)*, pages 1–4. IEEE, 2023.

[78] Miles A Clement, Chris G Kilsby, and Philip Moore. Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management*, 11(2):152–168, 2018.

[79] Miles A Clement, Chris G Kilsby, and Philip Moore. Multi-temporal synthetic aperture radar flood mapping using change detection. *Journal of Flood Risk Management*, 11(2):152–168, 2018.

[80] Luca Colomba, Alessandro Farasin, Simone Monaco, Salvatore Greco, Paolo Garza, Daniele Apiletti, Elena Baralis, and Tania Cerquitelli. A dataset for burned area delineation and severity estimation from satellite imagery. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, page 3893–3897, New York, NY, USA, 2022. Association for Computing Machinery.

[81] Brian Coltin, Scott McMichael, Trey Smith, and Terrence Fong. Automatic boosted flood mapping from satellite data. *International Journal of Remote Sensing*, 37(5):993–1015, 2016.

[82] Rosa Coluzzi, Angela Perrone, Caterina Samela, Vito Imbrenda, Salvatore Manfreda, Letizia Pace, and Maria Lanfredi. Rapid landslide detection from free optical satellite imagery using a robust change detection technique. *Scientific Reports*, 15(1):4697, 2025.

[83] Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David B. Lobell, and Stefano Ermon. Satmae: pre-training transformers for temporal and multi-spectral satellite imagery. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, 2022.

[84] Austin J. Cooner, Yang Shao, and James B. Campbell. Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 haiti earthquake. *Remote Sensing*, 8(10), 2016.

[85] Caroline Montagnino Corona, Gro B. M. Pedersen, Jakob Sigurdsson, and Margaux Heude. Automated mapping of lava: Classifcation of the lava flow field from the 2021 fagradalsfjall eruption. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 3821–3825, 2024.

[86] Jordi Cortes-Andres, Miguel-Angel Fernandez-Torres, and Gustau Camps-Valls. Deep learning with noisy labels for spatio-temporal drought detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[87] CRED and UNDRR. EM-DAT: The international disaster database. https://public.emdat.be/, 2025. Custom public-table export, accessed 2026-04-06.

[88] Liangyi Cui, Xin Jing, Yu Wang, Yixuan Huan, Yang Xu, and Qiangqiang Zhang. Improved swin transformer-based semantic segmentation of postearthquake dense buildings in urban areas using remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:369–385, 2023.

[89] Yifan Da, Zhiyuan Ji, and Yongsheng Zhou. Building damage assessment based on siamese hierarchical transformer framework. *Mathematics*, 10(11), 2022.

[90] Muhammad Umair Danish, Madhushan Buwaneswaran, Tehara Fonseka, and Katarina Grolinger. Graph attention convolutional u-net: A semantic segmentation model for identifying flooded areas. In *IECON 2024-50th Annual Conference of the IEEE Industrial Electronics Society*, pages 1–6. IEEE, 2024.

[91] Atanu Das, SM Abrar Rajin, Goh Kah Ong Michael, Shuvodip Biswas, Nabibun Billah, and Riasat Khan. Dual-attention resunet with masked focal-tversky loss for robust sar-based flood mapping. *IEEE Access*, 13:201460–201477, 2025.

[92] Gabriel Henrique de Almeida Pereira, Andre Minoro Fusioka, Bogdan Tomoyuki Nassu, and Rodrigo Minetto. Active fire detection in landsat-8 imagery: A large-scale dataset and a deep-learning study. *ISPRS Journal of Photogrammetry and Remote Sensing*, 178:171–186, 2021.

[93] Dibyabha Deb and Ujjwal Verma. Leveraging self-supervised features for efficient flooded region identification in uav aerial images. *arXiv preprint arXiv:2507.04915*, 2025.

[94] Rob J. Dekker. High-resolution radar damage assessment after the earthquake in haiti on 12 january 2010. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(4):960–970, 2011.

[95] Damien Delforge, Valentin Wathelet, Regina Below, Camila Lanfredi Sofia, Manon Tonnelier, Joris A. F. van Loenhout, and Niko Speybroeck. EM-DAT: the Emergency Events Database. *International Journal of Disaster Risk Reduction*, 124:105509, 2025.

[96] Fabio Dell'Acqua, Christian Bignami, Marco Chini, Gianni Lisini, Diego Aldo Polli, and Salvatore Stramondo. Earthquake damages rapid mapping by satellite remote sensing data: L'aquila april 6th, 2009 event. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(4):935–943, 2011.

[97] O. Dietrich, T. Peters, V. Sainte Fare Garnot, et al. An open-source tool for mapping war destruction at scale in Ukraine using Sentinel-1 time series. *Communications Earth & Environment*, 6:215, 2025.

[98] Mingtao Ding, Tao Huang, Hao Zheng, and Guohui Yang. Respective influence of vertical mountain differentiation on debris flow occurrence in the upper min river, china. *Scientific reports*, 10(1):11689, 2020.

[99] Laigen Dong and Jie Shan. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, 84:85–99, 2013.

[100] Georgios I Drakonakis, Grigorios Tsagkatakis, Konstantina Fotiadou, and Panagiotis 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*, 15:2341–2356, 2022.

[101] Patrick Ebel, Mounia El Baz, Junjue Wang, Weihao Xuan, Heli Qi, Zhuo Zheng, Naoto Yokoya, Junghwan Park, Jaewan Park, Arthur Elskens, Eléonore Charles, Iacopo Modica, Zachary Foltz, Philippe Bally, Christian Bossung, Marco Chini, Nicolas Longépé, and Gabriele Meoni. Artificial intelligence for earthquake response: Outcomes and insights from a global spaceborne rapid mapping challenge. *IEEE Geoscience and Remote Sensing Magazine*, pages 2–19, 2026.

[102] Markus Eckerstorfer, Yves Bühler, Regula Frauenfelder, and Eirik Malnes. Remote sensing of snow avalanches: Recent advances, potential, and limitations. *Cold Regions Science and Technology*, 121:126–140, 2016.

[103] Maelaynayn El Baida, Mohamed Hosni, Farid Boushaba, and Mimoun Chourak. A systematic literature review on classification machine learning for urban flood hazard mapping. *Water Resources Management*, 38(15):5823–5864, 2024.

[104] Ghislain Franck Emani, Weiya Xu, Kanon Guédet Guédé, Firdawus Ssemugga Nattabi, and Olive Mekontchou Yemele. Advancing global landslide segmentation: A coupled multispectral attention and data augmentation approach using the novel mrgsld dataset. *Earth Systems and Environment*, pages 1–22, 2025.

[105] M Esfandiari, S Jabari, H McGrath, and D Coleman. Flood mapping using random forest and identifying the essential conditioning factors; a case study in fredericton, new brunswick, canada. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:609–615, 2020.

[106] Yudai Ezaki, Chia Yee Ho, Bruno Adriano, Erick Mas, and Shunichi Koshimura. Evaluation of simulated sar images for building damage classification. *IEEE Geoscience and Remote Sensing Letters*, pages 1–1, 2024.

[107] Chengyong Fang, Xuanmei Fan, Hao Zhong, Luigi Lombardo, Hakan Tanyas, and Xin Wang. A novel historical landslide detection approach based on lidar and lightweight attention u-net. *Remote Sensing*, 14(17):4357, 2022.

[108] Heng Fang, Adam J Stewart, Isaac Corley, Xiao Xiang Zhu, and Hossein Azizpour. Earth embeddings as products: Taxonomy, ecosystem, and standardized access. *arXiv preprint arXiv:2601.13134*, 2026.

[109] Kuai Fang, Ming Pan, and Chaopeng Shen. The value of smap for long-term soil moisture estimation with the help of deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, 57(4):2221–2233, 2018.

[110] Kuai Fang, Chaopeng Shen, Daniel Kifer, and Xiao Yang. Prolongation of smap to spatiotemporally seamless coverage of continental us using a deep learning neural network. *Geophysical Research Letters*, 44(21):11–030, 2017.

[111] Mulham Fawakherji, Jeffrey Blay, Matilda Anokye, Leila Hashemi-Beni, and Jennifer Dorton. Deepflood for inundated vegetation high-resolution dataset for accurate flood mapping and segmentation. *Scientific Data*, 12(1):271, 2025.

[112] Ruitao Feng, Huanfeng Shen, Jianjun Bai, and Xinghua Li. Advances and opportunities in remote sensing image geometric registration: A systematic review of state-of-the-art approaches and future research directions. *IEEE Geoscience and Remote Sensing Magazine*, 9(4):120–142, 2021.

[113] Wenqing Feng, Fangli Guan, Chenhao Sun, and Wei Xu. Cross-modal change detection flood extraction based on self-supervised contrastive pre-training. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10:75–82, 2024.

[114] Xiao Feng, Juan Du, Minghua Wu, Bo Chai, Fasheng Miao, and Yang Wang. Potential of synthetic images in landslide segmentation in data-poor scenario: a framework combining gan and transformer models. *Landslides*, 21(9):2211–2226, September 2024.

[115] Zhengpeng Feng, Clement Atzberger, Sadiq Jaffer, Jovana Knezevic, Silja Sormunen, Robin Young, Madeline C Lisaius, Markus Immitzer, Toby Jackson, James Ball, et al. Tessera: Temporal embeddings of surface spectra for earth representation and analysis. *arXiv preprint arXiv:2506.20380*, 2025.

[116] Gudina L Feyisa, Henrik Meilby, Rasmus Fensholt, and Simon R Proud. Automated water extraction index: A new technique for surface water mapping using landsat imagery. *Remote sensing of environment*, 140:23–35, 2014.

[117] E. J. Fielding, M. Talebian, P. A. Rosen, H. Nazari, J. A. Jackson, M. Ghorashi, and R. Walker. Surface ruptures and building damage of the 2003 Bam, Iran, earthquake mapped by satellite synthetic aperture radar interferometric correlation. *Journal of Geophysical Research: Solid Earth*, 110(B3), March 2005.

[118] Federica Fiorucci, Francesca Ardizzone, Alessandro Cesare Mondini, Alessia Viero, and Fausto Guzzetti. Visual interpretation of stereoscopic ndvi satellite images to map rainfall-induced landslides. *Landslides*, 16(1):165–174, 2019.

[119] Fatemeh Foroughnia, Valentina Macchiarulo, Luis Berg, Matthew DeJong, Pietro Milillo, Kenneth W. Hudnut, Kenneth Gavin, and Giorgia Giardina. Quantitative assessment of earthquake-induced building damage at regional scale using LiDAR data. *International Journal of Disaster Risk Reduction*, 106:104403, 2024.

[120] Aito Fujita, Ken Sakurada, Tomoyuki Imaizumi, Riho Ito, Shuhei Hikosaka, and Ryosuke Nakamura. Damage detection from aerial images via convolutional neural networks. In *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, pages 5–8, 2017.

[121] Paolo Gamba, Fabio Dell'Acqua, and Giovanna Trianni. Rapid damage detection in the bam area using multitemporal sar and exploiting ancillary data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1582–1589, 2007.

[122] M Ganjirad and MR Delavar. Flood risk mapping using random forest and support vector machine. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10:201–208, 2023.

[123] Shubhika Garg, Ben Feinstein, Shahar Timnat, Vishal Batchu, Gideon Dror, Adi Gerzi Rosenthal, and Varun Gulshan. Cross-modal distillation for flood extent mapping. *Environmental Data Science*, 2:e37, 2023.

[124] Irene Garousi-Nejad, David G Tarboton, Mahyar Aboutalebi, and Alfonso F Torres-Rua. Terrain analysis enhancements to the height above nearest drainage flood inundation mapping method. *Water Resources Research*, 55(10):7983–8009, 2019.

[125] Jiayi Ge, Hong Tang, Naisen Yang, and Yijiang Hu. Rapid identification of damaged buildings using incremental learning with transferred data from historical natural disaster cases. *ISPRS Journal of Photogrammetry and Remote Sensing*, 195:105–128, 2023.

[126] Pinglan Ge, Hideomi Gokon, and Kimiro Meguro. Building damage assessment using intensity sar data with different incidence angles and longtime interval. *Journal of Disaster Research*, 14(3):456–465, 2019.

[127] Pinglan Ge, Hideomi Gokon, and Kimiro Meguro. A review on synthetic aperture radar-based building damage assessment in disasters. *Remote Sensing of Environment*, 240:111693, 2020.

[128] Pinglan Ge, Hideomi Gokon, Kimiro Meguro, and Shunichi Koshimura. Study on the intensity and coherence information of high-resolution alos-2 sar images for rapid massive landslide mapping at a pixel level. *Remote Sensing*, 11(23), 2019.

[129] Xiaosan Ge, Lin Zhou, and Di Meng. Ddnet: disaster damage detection for buildings based on dual-temporal joint attention network. *Scientific Reports*, 15(1):42513, November 2025.

[130] Riccardo Gelato, Carlo Sgaravatti, Jakob Grahn, Giacomo Boracchi, and Filippo Maria Bianchi. Promptable foundation models for sar remote sensing: Adapting the segment anything model for snow avalanche segmentation, 2026.

[131] Sebastian Gerard, Yu Zhao, and Josephine Sullivan. Wildfirespreadts: A dataset of multi-modal time series for wildfire spread prediction. *Advances in Neural Information Processing Systems*, 36:74515–74529, 2023.

[132] Saman Ghaffarian, Norman Kerle, Edoardo Pasolli, and Jamal Jokar Arsanjani. Post-disaster building database updating using automated deep learning: An integration of pre-disaster openstreetmap and multi-temporal satellite data. *Remote Sensing*, 11(20), 2019.

[133] Nafiseh Ghasemian Sorboni, Jinfei Wang, and Mohammad Reza Najafi. Urban flood mapping using sentinel-1 and radarsat constellation mission image and convolutional siamese network. *Natural Hazards*, 120(6):5711–5742, 2024.

[134] Omid Ghorbanzadeh, Yonghao Xu, Pedram Ghamisi, Michael Kopp, and David Kreil. Landslide4sense: Reference benchmark data and deep learning models for landslide detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022.

[135] BINAYAK Ghosh, Shagun Garg, and Mahdi Motagh. Automatic flood detection from sentinel-1 data using deep learning architectures. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:201–208, 2022.

[136] Louis Giglio, Luigi Boschetti, David P Roy, Michael L Humber, and Christopher O Justice. The collection 6 modis burned area mapping algorithm and product. *Remote sensing of environment*, 217:72–85, 2018.

[137] Laura Giustarini, Renaud Hostache, Patrick Matgen, Guy J-P Schumann, Paul D Bates, and David C Mason. A change detection approach to flood mapping in urban areas using terrasar-x. *IEEE transactions on Geoscience and Remote Sensing*, 51(4):2417–2430, 2012.

[138] Hideomi Gokon, Shunichi Koshimura, and Kimiro Meguro. Towards a damage assessment in a tsunami affected area using l-band and x-band sar data. In *2017 Joint Urban Remote Sensing Event (JURSE)*, pages 1–4, 2017.

[139] Masoomeh Gomroki, Mahdi Hasanlou, Jocelyn Chanussot, and Danfeng Hong. Unet-gcvit: a unet-based framework with global context vision transformer blocks for building damage detection. *International Journal of Remote Sensing*, 46(6):2587–2610, 2025.

[140] Lixia Gong, Chao Wang, Fan Wu, Jingfa Zhang, Hong Zhang, and Qiang Li. Earthquake-induced building damage detection with post-event sub-meter vhr terrasar-x staring spotlight imagery. *Remote Sensing*, 8(11), 2016.

[141] Ziyang Gong, Zhixiang Wei, Di Wang, Xiaoxing Hu, Xianzheng Ma, Hongruixuan Chen, Yuru Jia, Yupeng Deng, Zhenming Ji, Xiangwei Zhu, Xue Yang, Naoto Yokoya, Jing Zhang, Bo Du,

Junchi Yan, and Liangpei Zhang. Crossearth: Geospatial vision foundation model for domain generalizable remote sensing semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–18, 2025.

[142] David G Goodenough, Hao Chen, Ashlin Richardson, Shane Cloude, Wen Hong, and Yang Li. Mapping fire scars using radarsat-2 polarimetric sar data. *Canadian Journal of Remote Sensing*, 37(5):500–509, 2011.

[143] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, page 2672–2680, Cambridge, MA, USA, 2014. MIT Press.

[144] Gottfried Grünthal. European macroseismic scale 1998 (ems-98). 1998.

[145] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024.

[146] Zhihao Guan, Xinyu Miao, Yunjie Mu, Quan Sun, Qiaolin Ye, and Demin Gao. Forest fire segmentation from aerial imagery data using an improved instance segmentation model. *Remote Sensing*, 14(13):3159, 2022.

[147] Haonan Guo, Xin Su, Chen Wu, Bo Du, and Liangpei Zhang. Saan: Similarity-aware attention flow network for change detection with vhr remote sensing images. *IEEE Transactions on Image Processing*, 33:2599–2613, 2024.

[148] HuaDong Guo, LinLin Lu, JianWen Ma, Martino Pesaresi, and FangYan Yuan. An improved automatic detection method for earthquake-collapsed buildings from ADS40 image. *Chinese Science Bulletin*, 54(18):3303–3307, 2009.

[149] Xin Guo, Xudong Kang, Zihao Wang, Qiong Wu, Puhong Duan, Bin Yang, and Shutao Li. Flowmamba: Building damage assessment via optics flow-based state space model. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2025.

[150] Xin Guo, Jiangwei Lao, Bo Dang, Yingying Zhang, Lei Yu, Lixiang Ru, Liheng Zhong, Ziyuan Huang, Kang Wu, Dingxiang Hu, Huimei He, Jian Wang, Jingdong Chen, Ming Yang, Yongjun Zhang, and Yansheng Li. Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 27662–27673, 2024.

[151] Ritwik Gupta, Bryce Goodman, Nirav Patel, Ricky Hosfelt, Sandra Sajeev, Eric Heim, Jigar Doshi, Keane Lucas, Howie Choset, and Matthew Gaston. Creating xBD: A Dataset for Assessing Building Damage from Satellite Imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[152] Leonard Hackel, Tom Burgert, and Begüm Demir. How much of a model do we need? redundancy and slimmability in remote sensing foundation models, 2026.

[153] Masoud Hajeb, Sadra Karimzadeh, and Masashi Matsuoka. Sar and lidar datasets for building damage evaluation based on support vector machine and random forest algorithms—a case study of kumamoto earthquake, japan. *Applied Sciences*, 10(24), 2020.

[154] Wahyu Luqmanul Hakim, Fatemeh Rezaie, Arip Syaripudin Nur, Mahdi Panahi, Khabat Khosravi, Chang-Wook Lee, and Saro Lee. Convolutional neural network (cnn) with metaheuristic optimization algorithms for landslide susceptibility mapping in icheon, south korea. *Journal of environmental management*, 305:114367, 2022.

[155] Stephane Hallegatte, Adrien Vogt-Schilb, Mook Bangalore, and Julie Rozenberg. *Unbreakable: Building the Resilience of the Poor in the Face of Natural Disasters*. Climate Change and Development Series. World Bank, Washington, DC, 2017.

[156] Juefei He, Yuexing Peng, Wei Li, Junchuan Yu, Daqing Ge, and Wei Xiang. Mrife: A mask-recovering and interactive-feature-enhancing semantic segmentation network for relicf landslide detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[157] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15979–15988, 2022.

[158] Meizhang He, Qing Zhu, Zhiqiang Du, Han Hu, Yulin Ding, and Min Chen. A 3D Shape Descriptor Based on Contour Clusters for Damaged Roof Detection Using Airborne LiDAR Point Clouds. *Remote Sensing*, 8(3), 2016.

[159] Xiaoning He, Shuangcheng Zhang, Bowei Xue, Tong Zhao, and Tong Wu. Cross-modal change detection flood extraction based on convolutional neural network. *International Journal of Applied Earth Observation and Geoinformation*, 117:103197, 2023.

[160] Yongjun He, Jinfei Wang, Chunhua Liao, Xin Zhou, and Bo Shan. Ms4d-net: Multitask-based semi-supervised semantic segmentation framework with perturbed dual mean teachers for building damage assessment from high-resolution remote sensing imagery. *Remote Sensing*, 15(2), 2023.

[161] J-B Henry, P Chastanet, K Fellah, and Y-L Desnos. Envisat multi-polarized asar data for flood mapping. *International Journal of Remote Sensing*, 27(10):1921–1929, 2006.

[162] Victor Hertel, Omar Wani, Christian Geiß, Marc Wieland, and Hannes Taubenböck. Bayesiammtl: Uncertainty-aware multitask learning for post-disaster building damage assessment. *International Journal of Applied Earth Observation and Geoinformation*, 143:104759, 2025.

[163] Chia Yee Ho, Bruno Adriano, Gerald Baier, Erick Mas, Sesa Wiguna, Magaly Koch, and Shunichi Koshimura. Integrating GAN-Generated SAR and Optical Imagery for Building Damage Mapping. *Remote Sensing*, 18(1), 2026.

[164] Chia Yee Ho, Erick Mas, Bruno Adriano, and Shunichi Koshimura. Exploring the feasibility of ray tracing sar simulation on building damage assessment. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:1046–1059, 2024.

[165] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[166] Yu-Hsuan Ho and Ali Mostafavi. Multimodal mamba with multitask learning for building flood damage assessment using synthetic aperture radar remote sensing imagery. *Computer-Aided Civil and Infrastructure Engineering*, 40(26):4401–4424, 2025.

[167] J. Hoffmann. Mapping damage during the bam (iran) earthquake using interferometric coherence. *International Journal of Remote Sensing*, 28(6):1199–1216, 2007.

[168] Danfeng Hong, Bing Zhang, Xuyang Li, Yuxuan Li, Chenyu Li, Jing Yao, Naoto Yokoya, Hao Li, Pedram Ghamisi, Xiuping Jia, Antonio Plaza, Paolo Gamba, Jon Atli Benediktsson, and Jocelyn Chanussot. Spectralgpt: Spectral remote sensing foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5227–5244, 2024.

[169] Mohammad D. Hossain and Dongmei Chen. Segmentation for object-based image analysis (obia): A review of algorithms and challenges from remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150:115–134, 2019.

[170] Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

[171] Xiyang Hu and Maryam Rahnemoonfar. Flood simulation: Integrating uas imagery and ai-generated data with diffusion model. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 565–568. IEEE, 2024.

[172] Huili Huang, Andrew Zhang, Danrong Zhang, Max Mahdi Roozbahani, and James David Frost. Daseg: A domain-adaptive segmentation pipeline using vision foundation models—earthquake damage detection use case. *Remote Sensing*, 17(16), 2025.

[173] Shengli Huang and Florian Siegert. Backscatter change on fire scars in siberian boreal forests in envisat asar wide-swath images. *IEEE Geoscience and remote sensing letters*, 3(1):154–158, 2006.

[174] Xin Huang and Liangpei Zhang. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(1):161–172, 2012.

[175] Oldrich Hungr. Some methods of landslide hazard intensity mapping. In *Landslide risk assessment*, pages 215–226. Routledge, 2018.

[176] Fantine Huot, R Lily Hu, Nita Goyal, Tharun Sankar, Matthias Ihme, and Yi-Fan Chen. Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022.

[177] Charles K. Huyck, Beverley J. Adams, Sungbin Cho, Hung-Chi Chung, and Ronald T. Eguchi. Towards rapid citywide damage mapping using neighborhood edge dissimilarities in very high-resolution optical satellite imagery—application to the 2003 bam, iran, earthquake. *Earthquake Spectra*, 21(1_suppl):255–266, 2005.

[178] Ronny Hänsch, Jacob Arndt, Dalton Lunga, Matthew Gibb, Tyler Pedelose, Arnold Boedihardjo, Desiree Petrie, and Todd M. Bacastow. Spacenet 8 - the detection of flooded roads and buildings. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1471–1479, 2022.

[179] Kazi Aminul Islam, Mohammad Shahab Uddin, Chiman Kwan, and Jiang Li. Flood detection using multi-modal and multi-temporal images: A comparative study. *Remote Sensing*, 12(15):2455, 2020.

[180] Ali Ismail and Mariette Awad. Towards cross-disaster building damage detection with graph convolutional networks. In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 223–226, 2022.

[181] Michel Jaboyedoff, Antonio Abellán, Dario Carrea, Marc-Henri Derron, Battista Matasci, and Clément Michoud. Mapping and monitoring of landslides using lidar. In *Natural Hazards*, pages 397–420. CRC Press, 2018.

[182] Michel Jaboyedoff, Thierry Oppikofer, Antonio Abellán, Marc-Henri Derron, Alex Loye, Richard Metzger, and Andrea Pedrazzini. Use of lidar in landslide investigations: a review. *Natural hazards*, 61(1):5–28, 2012.

[183] Johannes Jakubik, Sujit Roy, C. E. Phillips, Paolo Fraccaro, Denys Godwin, Bianca Zadrozny, Daniela Szwarcman, Carlos Gomes, Gabby Nyirjesy, Blair Edwards, Daiki Kimura, Naomi Simumba, Linsong Chu, S. Karthik Mukkavilli, Devyani Lambhate, Kamal Das, Ranjini Bangalore, Dario Oliveira, Michal Muszynski, Kumar Ankur, Muthukumaran Ramasubramanian, Iksha Gurung, Sam Khallaghi, Hanxi (Steve) Li, Michael Cecil, Maryam Ahmadi, Fatemeh Kordi, Hamed Alemohammad, Manil Maskey, Raghu Ganti, Kommy Weldemariam, and Rahul Ramachandran. Foundation Models for Generalist Geospatial Artificial Intelligence. *Preprint Available on arxiv:2310.18660*, October 2023.

[184] Johannes Jakubik, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, Nikolaos Dionelis, Valerio Marsocci, Niklas Kopp, et al. Terramind: Large-scale generative multimodality for earth observation. *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

[185] Ali Jamali, Swalpa Kumar Roy, Leila Hashemi Beni, Biswajeet Pradhan, Jonathan Li, and Pedram Ghamisi. Residual wave vision u-net for flood mapping using dual polarization sentinel-1 sar imagery. *International Journal of Applied Earth Observation and Geoinformation*, 127:103662, 2024.

[186] Milad Janalipour and Ali Mohammadzadeh. Building damage detection using object-based image analysis and anfis from high-resolution image (case study: Bam earthquake, iran). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5):1937–1945, 2016.

[187] Milad Janalipour and Ali Mohammadzadeh. A novel and automatic framework for producing building damage map using post-event LiDAR data. *International Journal of Disaster Risk Reduction*, 39:101238, 2019.

[188] Youngsun Jang, Dongyoun Kim, Chulwoo Pack, and Kwanghee Won. A novel dataset for flood detection robust to seasonal changes in satellite imagery. In *Proceedings of the International Conference on Research in Adaptive and Convergent Systems*, pages 177–184, 2024.

[189] Mustafa Neamah Jebur, Biswajeet Pradhan, and Mahyat Shafapour Tehrany. Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (lidar) data at catchment scale. *Remote Sensing of Environment*, 152:150–165, 2014.

[190] Min Ji, Lanfa Liu, and Manfred Buchroithner. Identifying collapsed buildings using post-earthquake satellite imagery and convolutional neural networks: A case study of the 2010 haiti earthquake. *Remote Sensing*, 10(11), 2018.

[191] Yaqi Ji, Josaphat Tetuko Sri Sumantyo, Ming Yam Chua, and Mirza Muhammad Waqar. Earthquake/tsunami damage level mapping of urban areas using full polarimetric sar data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(7):2296–2309, 2018.

[192] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European Conference on Computer Vision (ECCV)*, 2022.

[193] Yuyu Jia, Junyu Gao, Wei Huang, Yuan Yuan, and Qi Wang. Holistic mutual representation enhancement for few-shot remote sensing segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–13, 2023.

[194] Wei Jiang, Yuan Ni, Zhiguo Pang, Xiaotao Li, Hongrun Ju, Guojin He, Juan Lv, Kun Yang, June Fu, and Xiangdong Qin. An effective water body extraction method with new water index for sentinel-2 imagery. *Water*, 13(12):1647, 2021.

[195] Xiao Jiang, You He, Gang Li, Yu Liu, and Xiao-Ping Zhang. Building damage detection via superpixel-based belief fusion of space-borne sar and optical images. *IEEE Sensors Journal*, 20(4):2008–2022, 2020.

[196] Xin Jiang, Shijing Liang, Xinyue He, Alan D. Ziegler, Peirong Lin, Ming Pan, Dashan Wang, Junyu Zou, Dalei Hao, Ganquan Mao, Yelu Zeng, Jie Yin, Lian Feng, Chiyuan Miao, Eric F. Wood, and Zhenzhong Zeng. Rapid and large-scale mapping of flood inundation via integrating spaceborne synthetic aperture radar imagery with unsupervised deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 178:36–50, 2021.

[197] Wenzhe Jiao, Lixin Wang, and Matthew F. McCabe. Multi-sensor remote sensing for drought characterization: current status, opportunities and a roadmap for the future. *Remote Sensing of Environment*, 256:112313, 2021.

[198] J Michael Johnson, Dinuke Munasinghe, Damilola Eyelade, and Sagy Cohen. An integrated evaluation of the national water model (nwm)–height above nearest drainage (hand) flood mapping methodology. *Natural Hazards and Earth System Sciences*, 19(11):2405–2420, 2019.

[199] Rabiul Islam Jony, Alan Woodley, Aishvarya Raj, and Dimitri Perrin. Ensemble classification technique for water detection in satellite images. In *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–8. IEEE, 2018.

[200] CO Justice, Louis Giglio, S Korontzi, J Owens, JT Morisette, D Roy, J Descloitres, Samuel Alleaume, F Petitcolin, and Y Kaufman. The modis fire products. *Remote sensing of Environment*, 83(1-2):244–262, 2002.

[201] Anna Karas, Fatima Karbou, Sophie Giffard-Roisin, Philippe Durand, and Nicolas Eckert. Automatic color detection-based method applied to sentinel-1 sar images for snow avalanche debris monitoring. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2021.

[202] Ilektra Karasante, Lazaro Alonso, Ioannis Prapas, Akanksha Ahuja, Nuno Carvalhais, and Ioannis Papoutsis. Seasfire cube-a multivariate dataset for global wildfire modeling. *Scientific Data*, 12(1):368, 2025.

[203] Sadra Karimzadeh, Mohammad Ghasemi, Masashi Matsuoka, Koichi Yagi, and Abdullah Can Zulfikar. A deep learning model for road damage detection after an earthquake based on synthetic aperture radar (sar) and field datasets. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:5753–5765, 2022.

[204] Sadra Karimzadeh and Masashi Mastuoka. Building damage assessment using multisensor dual-polarized synthetic aperture radar data for the 2016 m 6.2 amatrice earthquake, italy. *Remote Sensing*, 9(4), 2017.

[205] Navjot Kaur, Cheng-Chun Lee, Ali Mostafavi, and Ali Mahdavi-Amiri. Large-scale building damage assessment using a novel hierarchical transformer architecture on satellite images. *Computer-Aided Civil and Infrastructure Engineering*, 38(15):2072–2091, 2023.

[206] Swapandeep Kaur, Sheifali Gupta, Swati Singh, Vinh Truong Hoang, Sultan Almakdi, Turki Alelyani, and Asadullah Shaikh. Transfer learning-based automatic hurricane damage detection using satellite images. *Electronics*, 11(9), 2022.

[207] Arush Kaushal, Ashok Kumar Gupta, and Vivek Kumar Sehgal. A semantic segmentation framework with unet-pyramid for landslide prediction using remote sensing data. *Scientific Reports*, 14(1):30071, 2024.

[208] Norman Kerle, Francesco Nex, Markus Gerke, Diogo Duarte, and Anand Vetrivel. Uav-based structural damage mapping: A review. *ISPRS International Journal of Geo-Information*, 9(1), 2020.

[209] Carl H Key and Nathan C Benson. Landscape assessment (la). *In: Lutes, Duncan C.; Keane, Robert E.; Caratti, John F.; Key, Carl H.; Benson, Nathan C.; Sutherland, Steve; Gangi, Larry J. 2006. FIREMON: Fire effects monitoring and inventory system. Gen. Tech. Rep. RMRS-*

GTR-164-CD. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. LA-1-55, 164, 2006.

[210] Hyun Il Kim and Byung Hyun Kim. Flood hazard rating prediction for urban areas using random forest and lstm. *KSCE Journal of Civil Engineering*, 24(12):3884–3896, 2020.

[211] Minhwa Kim, Sang-Eun Park, and Seung-Jae Lee. Detection of damaged buildings using temporal sar data with different observation modes. *Remote Sensing*, 15(2), 2023.

[212] Wanyub Kim, Shinhyeon Cho, Junhyuk Jeong, Yeji Kim, Hyun Ok Kim, and Minha Choi. Evaluation of deep learning-based water bodies and flooded area detection with nanosatellites: The planetscope satellite imageries and hrnet model. *Korean Journal of Remote Sensing*, 40(5_1):617–627, 2024.

[213] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment Anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3992–4003, 2023.

[214] E. E. Koks, J. Rozenberg, C. Zorn, M. Tariverdi, M. Vousdoukas, S. A. Fraser, J. W. Hall, and S. Hallegatte. A global multi-hazard risk analysis of road and railway infrastructure assets. *Nature Communications*, 10(1):2677, 2019.

[215] Silvia Kuny and Karsten Schulz. Debris detection in sar imagery using statistics of simulated texture. In *2014 8th IAPR Workshop on Pattern Reconition in Remote Sensing*, pages 1–4, 2014.

[216] Youngjoo Kwak, Sang-ho Yun, and Yoichi Iwami. A new approach for rapid urban flood mapping using alos-2/palsar-2 in 2015 kinu river flood, japan. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1880–1883. IEEE, 2017.

[217] Richard C. Labiak, Jan A. N. van Aardt, Dmitry Bespalov, Darryl Eychner, Erin Wirch, and Hans-Peter Bischof. Automated method for detection and quantification of building damage and debris using post-disaster lidar data. In Monte D. Turner and Gary W. Kamerman, editors, *Laser Radar Technology and Applications XVI*, volume 8037, page 80370F. International Society for Optics and Photonics, SPIE, 2011.

[218] Umut Lagap, Saman Ghaffarian, Sophie Gelinas-Gagne, Jasmin Jilma, Zhiyu Liu, and Zhiyuan Luo. Towards reliable deep learning for post-disaster damage assessment: An xai-based evaluation. *International Journal of Disaster Risk Reduction*, 130:105839, 2025.

[219] Saad Lahrichi, Jake Bova, Jesse Johnson, and Jordan Malof. Advancing time series wildfire spread prediction: Modeling improvements and the wsts+ benchmark. *arXiv preprint arXiv:2502.12003*, 2025.

[220] Nhut Le and Maryam Rahnemoonfar. 3d semantic segmentation for post-disaster assessment. In *IGARSS 2025 - 2025 IEEE International Geoscience and Remote Sensing Symposium*, pages 6488–6492, 2025.

[221] Hyunho Lee and Wenwen Li. A spatially masked adaptive gated network for multimodal post-flood water extent mapping using sar and incomplete multispectral data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 232:492–508, 2026.

[222] Tao Lei, Yuxiao Zhang, Zhiyong Lv, Shuying Li, Shigang Liu, and Asoke K. Nandi. Landslide inventory mapping from bitemporal images using deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 16(6):982–986, 2019.

[223] Chengxi Li, Gang Li, Zhuoyue Wang, Xueqian Wang, and Pramod K. Varshney. Comic: An unsupervised change detection method for heterogeneous remote sensing images based on copula mixtures and cycle-consistent adversarial networks. *Information Fusion*, 106:102240, 2024.

[224] Hao Li, Fabian Deuser, Wenping Yin, Xuanshu Luo, Paul Walther, Gengchen Mai, Wei Huang, and Martin Werner. Cross-view geolocalization and disaster mapping with street-view and VHR satellite imagery: A case study of Hurricane IAN. *ISPRS Journal of Photogrammetry and Remote Sensing*, 220:841–854, 2025.

[225] Jiepan Li, He Huang, Yu Sheng, Yujun Guo, and Wei He. Building-guided pseudo-label learning for cross-modal building damage mapping. In *IGARSS 2025 - 2025 IEEE International Geoscience and Remote Sensing Symposium*, pages 228–232, 2025.

[226] ManChun Li, Liang Cheng, JianYa Gong, YongXue Liu, ZhenJie Chen, FeiXue Li, Gang Chen, Dong Chen, and XiaoGang Song. Post-earthquake assessment of building damage degree using LiDAR data and imagery. *Science in China Series E: Technological Sciences*, 51(2):133–143, 2008.

[227] Na Li, Arnaud Martin, and Rémi Estival. An automatic water detection approach based on dempster-shafer theory for multi-spectral images. In *2017 20th International Conference on Information Fusion (Fusion)*, pages 1–8. IEEE, 2017.

[228] Peijun Li, Haiqing Xu, and Jiancong Guo. Urban building damage detection from very high resolution imagery using ocsvm and spatial features. *International Journal of Remote Sensing*, 31(13):3393–3409, 2010.

[229] Qiang Li, Lixia Gong, and Jingfa Zhang. A correlation change detection method integrating pca and multi- texture features of sar image for building damage detection. *European Journal of Remote Sensing*, 52(1):435–447, 2019.

[230] Tianyang Li, Chao Wang, Hong Zhang, Fan Wu, and Xiaohan Zheng. DDFormer: A Dual-Domain Transformer for Building Damage Detection Using High-Resolution SAR Imagery. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023.

[231] Weiming Li, Yibin Fu, Shuaishuai Fan, Mingrui Xin, and Hongyang Bai. Dci-pgcn: Dual-channel interaction portable graph convolutional network for landslide detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–16, 2023.

[232] Wenmei Li, Jiaqi Wu, Huaihuai Chen, Yu Wang, Yan Jia, and Guan Gui. Unet combined with attention mechanism method for extracting flood submerged range. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:6588–6597, 2022.

[233] Xinwu Li, Huadong Guo, Lu Zhang, Xiao Chen, and Lei Liang. A new approach to collapsed building extraction using radarsat-2 polarimetric sar imagery. *IEEE Geoscience and Remote Sensing Letters*, 9(4):677–681, 2012.

[234] Yu Li, Sandro Martinis, and Marc Wieland. Urban flood mapping with an active self-learning convolutional neural network based on terrasar-x intensity and interferometric coherence. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152:178–191, 2019.

[235] Yundong Li, Wei Hu, Hongguang Li, Han Dong, Baochang Zhang, and Qing Tian. Aligning discriminative and representative features: An unsupervised domain adaptation method for building damage assessment. *IEEE Transactions on Image Processing*, 29:6110–6122, 2020.

[236] Yundong Li, Chen Lin, Hongguang Li, Wei Hu, Han Dong, and Yi Liu. Unsupervised domain adaptation with self-attention for post-disaster building damage detection. *Neurocomputing*, 415:27–39, 2020.

[237] Zhongbin Li, Wenzhong Shi, Soe W Myint, Ping Lu, and Qunming Wang. Semi-automated landslide inventory mapping from bitemporal aerial photographs using change detection and level set method. *Remote Sensing of Environment*, 175:215–230, 2016.

[238] Zhouyayan Li, Felipe Quintero Duque, Trevor Grout, Bradford Bates, and Ibrahim Demir. Comparative analysis of performance and mechanisms of flood inundation map generation using height above nearest drainage. *Environmental Modelling & Software*, 159:105565, 2023.

[239] Jiayong Liang and Desheng Liu. A local thresholding approach to flood water delineation using sentinel-1 sar imagery. *ISPRS journal of photogrammetry and remote sensing*, 159:53–62, 2020.

[240] Rubing Liang, Keren Dai, Juan M Lopez-Sanchez, Yakun Han, Xianlin Shi, and Qiang Xu. Landslide mapping from post-event single-temporal polarimetric SAR image by a deep learning method exploiting a morphological model. *Remote Sensing of Environment*, 328:114904, 2025.

[241] Yijun Liao, Mohammad Ebrahim Mohammadi, and Richard L. Wood. Deep learning classification of 2d orthomosaic images and 3d point clouds for post-event structural damage assessment. *Drones*, 4(2), 2020.

[242] Joongbin Lim and Kyoo-seock Lee. Flood mapping using multi-source remotely sensed data and logistic regression in the heterogeneous mountainous regions in north korea. *Remote Sensing*, 10(7):1036, 2018.

[243] Chen Lin, Yundong Li, Yi Liu, Xiang Wang, and Shuo Geng. Building damage assessment from post-hurricane imageries using unsupervised domain adaptation with enhanced feature discrimination. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–10, 2022.

[244] Szu-Yun Lin, Marisa Edocia, Fang Jung Tsai, Lien an Chen, and Wen-Ni Kuo. Haitibrd: A labeled satellite imagery dataset for building and road damage assessment of the 2010 haiti earthquake, 2023.

[245] Chang Liu, Linlin Ge, and Ting Bai. Rapid large-scale building damage level classification after earthquakes using deep learning with lidar and satellite optical data. *International Journal of Digital Earth*, 17(1):2441934, 2024.

[246] Chang Liu, Samad M.E. Sepasgozar, Qi Zhang, and Linlin Ge. A novel attention-based deep learning method for post-disaster building damage classification. *Expert Systems with Applications*, 202:117268, 2022.

[247] Jia Liu, Maoguo Gong, Kai Qin, and Puzhao Zhang. A Deep Convolutional Coupling Network for Change Detection Based on Heterogeneous Optical and Radar Images. *IEEE Transactions on Neural Networks and Learning Systems*, 29(3):545–559, 2018.

[248] Tongfei Liu, Jianjian Xu, Tao Lei, Yingbo Wang, Xiaogang Du, Weichuan Zhang, Zhiyong Lv, and Maoguo Gong. AEKAN: Exploring Superpixel-Based AutoEncoder Kolmogorov-Arnold Network for Unsupervised Multimodal Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–14, 2025.

[249] Tongfei Liu, Mingyang Zhang, Maoguo Gong, Qingfu Zhang, Fenlong Jiang, Hanhong Zheng, and Di Lu. Commonality feature representation learning for unsupervised multimodal change detection. *IEEE Transactions on Image Processing*, 34:1219–1233, 2025.

[250] Wangcai Liu, Yi Zhang, Yiwen Liang, Pingping Sun, Yuanxi Li, Xiaojun Su, Aijie Wang, and Xingmin Meng. Landslide risk assessment using a combined approach based on insar and random forest. *Remote Sensing*, 14(9):2131, 2022.

[251] Wen Liu and Fumio Yamazaki. Extraction of collapsed buildings in the 2016 kumamoto earthquake using multi-temporal palsar-2 data. *Journal of Disaster Research*, 12(2):241–250, 2017.

[252] Wen Liu, Fumio Yamazaki, Hideomi Gokon, and Shun ichi Koshimura. Extraction of tsunami-flooded areas and damaged buildings in the 2011 tohoku-oki earthquake from terrasar-x intensity images. *Earthquake Spectra*, 29(1_suppl):183–200, 2013.

[253] Wen Liu, Fumio Yamazaki, Hideomi Gokon, and Shunichi Koshimura. Extraction of damaged buildings due to the 2011 tohoku, japan earthquake tsunami. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pages 4038–4041, 2012.

[254] Xiaoyi Liu, Jiancheng Li, Hichem Sahli, Yu Meng, and Qingqing Huang. Improving unsupervised flood detection with spatio-temporal context on hj-1b ccd data. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 4402–4405. IEEE, 2016.

[255] Xinran Liu, Yuexing Peng, Zili Lu, Wei Li, Junchuan Yu, Daqing Ge, and Wei Xiang. Feature-fusion segmentation network for landslide detection using high-resolution remote sensing images and digital elevation model data. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–14, 2023.

[256] Yunlong Liu, Kai Zhang, Chunan Guan, Shanxin Zhang, Hong Li, Wenbo Wan, and Jiande Sun. Building change detection in earthquake: A multiscale interaction network with offset calibration and a dataset. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–17, 2024.

[257] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9992–10002, 2021.

[258] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11966–11976, 2022.

[259] Zhun Ga Liu, Gregoire Mercier, Jean Dezert, and Quan Pan. Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning. *IEEE Geoscience and Remote Sensing Letters*, 11(1):168–172, 2014.

[260] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[261] Stephanie Long, Temilola E Fatoyinbo, and Frederick Policelli. Flood extent mapping for namibia using change detection and thresholding with sar. *Environmental Research Letters*, 9(3):035002, 2014.

[262] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

[263] Wen Lu, Lu Wei, and Minh Nguyen. Bitemporal attention transformer for building change detection and building damage assessment. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:4917–4935, 2024.

[264] Xiaoqiang Lu, Tong Gou, Zhongjian Huang, Yuting Yang, Licheng Jiao, Lingling Li, Xu Liu, and Fang Liu. Multimodal segformer for flood rapid mapping with sentinel-2 data. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 790–793. IEEE, 2024.

[265] Xin Lu, Linus Bengtsson, and Petter Holme. Predictability of population displacement after the 2010 haiti earthquake. *Proceedings of the National Academy of Sciences*, 109(29):11576–11581, 2012.

[266] Harrison Luft, Calogero Schillaci, Guido Ceccherini, Diana Vieira, and Aldo Lipani. Deep learning based burnt area mapping using sentinel 1 for the santa cruz mountains lightning complex (czu) and creek fires 2020. *Fire*, 5(5):163, 2022.

[267] Huanzhang Luo, Jingjuan Liao, and Guozhuang Shen. Combining remote sensing and social media data for flood mapping: a case study in linhai, zhejiang province, china. *Journal of Applied Remote Sensing*, 17(2):024507–024507, 2023.

[268] Luigi Tommaso Luppino, Filippo Maria Bianchi, Gabriele Moser, and Stian Normann Anfinsen. Unsupervised image regression for heterogeneous change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):9960–9975, 2019.

[269] Luigi Tommaso Luppino, Mads Adrian Hansen, Michael Kampffmeyer, Filippo Maria Bianchi, Gabriele Moser, Robert Jenssen, and Stian Normann Anfinsen. Code-aligned autoencoders for unsupervised change detection in multimodal remote sensing images. *IEEE Transactions on Neural Networks and Learning*, 35(1):60–72, 2024.

[270] Luigi Tommaso Luppino, Michael Kampffmeyer, Filippo Maria Bianchi, Gabriele Moser, Sebastiano Bruno Serpico, Robert Jenssen, and Stian Normann Anfinsen. Deep image translation with an affinity-based change prior for unsupervised multimodal change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–22, 2022.

[271] Liang Lv, Tao Chen, Jie Dou, and Antonio Plaza. A hybrid ensemble-based deep-learning framework for landslide susceptibility mapping. *International Journal of Applied Earth Observation and Geoinformation*, 108:102713, 2022.

[272] Pengyuan Lv, Lusha Ma, Qiaomin Li, and Fang Du. Shapeformer: A shape-enhanced vision transformer model for optical remote sensing image landslide detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:2681–2689, 2023.

[273] Suna Lv, Lingsheng Meng, Deanna Edwing, Sihan Xue, Xupu Geng, and Xiao-Hai Yan. High-performance segmentation for flood mapping of hisea-1 sar remote sensing images. *Remote Sensing*, 14(21):5504, 2022.

[274] Haojie Ma, Yalan Liu, Yuhuan Ren, and Jingxian Yu. Detection of collapsed buildings in post-earthquake remote sensing images based on the improved yolov3. *Remote Sensing*, 12(1), 2020.

[275] Xianping Ma, Xiaokang Zhang, and Man-On Pun. Rs3mamba: Visual state space model for remote sensing image semantic segmentation. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024.

[276] Giorgos Mallinis, Ioannis Z Gitas, Vassileios Giannakopoulos, Fotis Maris, and Maria Tsakiri-Strati. An object-based approach for flood area delineation in a transboundary area using ENVISAT ASAR and LANDSAT TM data. *International Journal of Digital Earth*, 6(sup2):124–136, 2013.

[277] Thomas Manzini, Priyankari Perali, Raisa Karnik, and Robin Murphy. CRASAR-U-DROIDs: A Large Scale Benchmark Dataset for Building Alignment and Damage Assessment in Georectified sUAS Imagery. *arXiv preprint arXiv:2407.17673*, 2024.

[278] Thomas Manzini, Priyankari Perali, Jayesh Tripathi, and Robin R. Murphy. Now you see it, Now you don't: Damage Label Agreement in Drone & Satellite Post-Disaster Imagery. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, page 1998–2008, New York, NY, USA, 2025.

[279] Valerio Marsocci, Yuru Jia, Georges Le Bellier, David Kerekes, Liang Zeng, Sebastian Hafner, Sebastian Gerard, Eric Brune, Ritu Yadav, Ali Shibli, Heng Fang, Yifang Ban, Maarten Vergauwen, Nicolas Audebert, and Andrea Nascetti. Pangaea: Assessing geospatial foundation models capabilities through a global and inclusive benchmark. *IEEE Geoscience and Remote Sensing Magazine*, 14(1):245–285, 2026.

[280] Sandro Martinis, Jens Kersten, and André Twele. A fully automated terrasar-x based flood service. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104:203–212, 2015.

[281] Sandro Martinis, André Twele, and Stefan Voigt. Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution terrasar-x data. *Natural Hazards and Earth System Sciences*, 9(2):303–314, 2009.

[282] VS Martins, DP Roy, H Huang, L Boschetti, HK Zhang, and L Yan. Deep learning high resolution burned area mapping by transfer learning from landsat-8 to planetscope. *Remote Sensing of Environment*, 280:113203, 2022.

[283] Gonzalo Mateo-Garcia, Joshua Veitch-Michaelis, Cormac Purcell, Nicolas Longepe, Sam Reid, Alice Anlind, Freddie Cremer, Tanya Sherwood, and Richard Sherwood. Global Flood Extent Segmentation in Optical Satellite Images. *Scientific Reports*, 13:20316, 2023.

[284] Gonzalo Mateo-Garcia, Joshua Veitch-Michaelis, Lewis Smith, Silviu Vlad Oprea, Guy Schumann, Yarin Gal, Atılım Güneş Baydin, and Dietmar Backes. Towards global flood mapping onboard low cost satellites with machine learning. *Scientific reports*, 11(1):7249, 2021.

[285] L. Matikainen, H. Kaartinen, and J. Hyyppä. Classification tree based building detection from laser scanner and aerial image data. In *Laser Scanning 2007 and SilviLaser 2007*, volume XXXVI-3/W52 of *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Archives)*, pages 280–287, Espoo, Finland, September 2007.

[286] Masashi Matsuoka, Kei Horie, Hiroshi Ohkura, and Fumio Yamazaki. Characteristics of remote sensing images for the 2004 niigata-ken chuetsu earthquake. In *3 rd International Workshop on Remote Sensing for Post-Disaster Response, Chiba, Japan*, 2005.

[287] Masashi Matsuoka and Nobuoto Nojima. Building damage estimation by integration of seismic intensity information and satellite l-band sar imagery. *Remote Sensing*, 2(9):2111–2126, 2010.

[288] Masashi Matsuoka and Fumio Yamazaki. Use of interferometric satellite SAR for earthquake damage detection. *Sat*, 2:z1, 2000.

[289] Masashi Matsuoka and Fumio Yamazaki. Use of satellite sar intensity imagery for detecting building areas damaged due to earthquakes. *Earthquake Spectra*, 20(3):975–994, 2004.

[290] Masashi Matsuoka and Fumio Yamazaki. Building damage mapping of the 2003 bam, iran, earthquake using envisat/asar intensity imagery. *Earthquake Spectra*, 21(1_suppl):285–294, 2005.

[291] Stuart K McFeeters. The use of the normalized difference water index (ndwi) in the delineation of open water features. *International journal of remote sensing*, 17(7):1425–1432, 1996.

[292] James McRae, Max Abramowitz, Moses Schulhof, and Christopher Hill. Advancing Wildfire Damage Assessment with Aerial Thermal Remote Sensing and AI: Applications to the 2025 Eaton and Palisades Fires. *Remote Sensing*, 17(24):3962, 2025.

[293] Sansar Raj Meena, Lorenzo Nava, Kushanav Bhuyan, Silvia Puliero, Lucas Pedrosa Soares, Helen Cristina Dias, Mario Floris, and Filippo Catani. Hr-gldd: a globally distributed dataset using generalized deep learning (dl) for rapid landslide mapping on high-resolution (hr) satellite imagery. *Earth System Science Data*, 15(7):3283–3298, 2023.

[294] Microsoft. Global ml building footprints. https://github.com/microsoft/GlobalMLBuildingFootprints, 2026. Dataset repository; accessed 2026-03-26.

[295] Jay D Miller, Eric E Knapp, Carl H Key, Carl N Skinner, Clint J Isbell, R Max Creasy, and Joseph W Sherlock. Calibration and validation of the relative differenced normalized burn ratio (rdnbr) to three measures of fire severity in the sierra nevada and klamath mountains, california, usa. *Remote Sensing of Environment*, 113(3):645–656, 2009.

[296] Jay D. Miller and Andrea E. Thode. Quantifying burn severity in a heterogeneous landscape with a relative version of the delta normalized burn ratio (dnbr). *Remote Sensing of Environment*, 109(1):66–80, 2007.

[297] Christoph Minixhofer, Mark Swan, Calum McMeekin, and Pavlos Andreadis. Droughted: A dataset and methodology for drought forecasting spanning multiple climate zones. In *Tackling Climate Change with Machine Learning: Workshop at ICML 2021*, 2021.

[298] Amit Misra, Kevin White, Simone Fobi Nsutezo, William Straka III, and Juan Lavista. Mapping global floods with 10 years of satellite radar data. *Nature Communications*, 16(1):5762, 2025.

[299] Hiroyuki Miura, Tomohiro Aridome, and Masashi Matsuoka. Deep learning-based identification of collapsed, non-collapsed and blue tarp-covered buildings from post-disaster aerial images. *Remote Sensing*, 12(12), 2020.

[300] Hiroyuki Miura, Saburo Midorikawa, and Masashi Matsuoka. Building damage assessment using high-resolution satellite sar images of the 2010 haiti earthquake. *Earthquake Spectra*, 32(1):591–610, 2016.

[301] William Mobley, Antonia Sebastian, Russell Blessing, Wesley E Highfield, Laura Stearns, and Samuel D Brody. Quantification of continuous flood hazard using random forest classification and flood insurance claims at large spatial scales: a pilot study in southeast texas. *Natural Hazards and Earth System Sciences*, 21(2):807–822, 2021.

[302] Alessandro C Mondini, Fausto Guzzetti, and Massimo Melillo. Deep learning forecast of rainfall-induced shallow landslides. *Nature communications*, 14(1):2466, 2023.

[303] Alessandro C. Mondini, Michele Santangelo, Margherita Rocchetti, Enrica Rossetto, Andrea Manconi, and Oriol Monserrat. Sentinel-1 sar amplitude imagery for rapid landslide detection. *Remote Sensing*, 11(7), 2019.

[304] Fabio Montello, Edoardo Arnaudo, and Claudio Rossi. Mmflood: A multimodal dataset for flood delineation from satellite imagery. *IEEE Access*, 10:96774–96787, 2022.

[305] L. Moya, F. Yamazaki, W. Liu, and T. Chiba. Calculation of coseismic displacement from lidar data in the 2016 kumamoto, japan, earthquake. *Natural Hazards and Earth System Sciences*, 17(1):143–156, 2017.

[306] Marios Mpakratsas, Anastasia Moumtzidou, Ilias Gialampoukidis, Stefanos Vrochidis, and Ioannis Kompatsiaris. A deep neural network slope reduction model on sentinel-1 images for water mask extraction. In *Proceedings of the 40th Asian Conference on Remote Sensing (ACRS 2019), Daejeon, Korea*, pages 14–18, 2019.

[307] Manish Nadella, Garapati Venkata Krishna Rayalu, Menta Sai Akshay, SK Eswar Sudhan, VV Sajith Variyar, V Sowmya, and Ramesh Sivanpillai. Semi supervised flood damage detection using satellite images. In *International Conference on Computing, Intelligence and Data Analytics*, pages 163–174. Springer, 2024.

[308] Arpita Nandi, Arpita Mandal, Matthew Wilson, and David Smith. Flood hazard mapping in jamaica using principal component analysis and logistic regression. *Environmental Earth Sciences*, 75(6):465, 2016.

[309] NASA IMPACT and IEEE GRSS. Etci 2021 flood detection competition. https://nasa-impact.github.io/etci2021/, 2021. Online benchmark, accessed Jan. 2026.

[310] National Institute for Land and Infrastructure Management and Building Research Institute. Preliminary report on the building damage survey of the large-scale fire in saganoseki, oita city, oita prefecture, occurred on november 18, 2025. https://www.kenken.go.jp/japanese/contents/topics/2025/R7_12_9_oita.pdf, December 2025. Accessed 2026-04-10.

[311] Ryo Natsuaki, Hiroto Nagai, Naoya Tomii, and Takeo Tadono. Sensitivity and limitation in damage detection for individual buildings using insar coherence—a case study in 2016 kumamoto earthquakes. *Remote Sensing*, 10(2), 2018.

[312] Edoardo Nemni, Joseph Bullock, Samir Belabbes, and Lars Bromley. Fully convolutional neural network for rapid flood segmentation in synthetic aperture radar imagery. *Remote Sensing*, 12(16):2532, 2020.

[313] Edoardo Nemni, Joseph Bullock, Samir Belabbes, and Lars Bromley. Fully convolutional neural network for rapid flood segmentation in synthetic aperture radar imagery. *Remote Sensing*, 12(16):2532, 2020.

[314] Sun Han Neo, Sachith Seneviratne, Herath Mudiyanselage Viraj Vidura Herath, Abhishek Saha, Sanka Rasnayaka, and Lucy Amanda Marshall. Flood-ldm: Generalizable latent diffusion models for rapid and accurate zero-shot high-resolution flood mapping. *arXiv preprint arXiv:2511.14033*, 2025.

[315] Francesco Nex, Diogo Duarte, Fabio Giulio Tonolo, and Norman Kerle. Structural building damage detection with deep learning: Assessment of a state-of-the-art cnn in operational conditions. *Remote Sensing*, 11(23), 2019.

[316] Charles WW Ng, Tianli Pan, and Peifeng Ma. A novel deep learning model for landslide mapping using cross-resolution change technique. *ISPRS Journal of Photogrammetry and Remote Sensing*, 229:254–269, 2025.

[317] Dat T. Nguyen, Ferda Ofli, Muhammad Imran, and Prasenjit Mitra. Damage assessment from social media imagery data during disasters. In *Proceedings of the 2017 IEEE/ACM International*

Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17, page 569–576, 2017.

[318] Chaoyang Niu, Haobo Zhang, Wei Liu, Runsheng Li, and Tao Hu. Using a fully polarimetric sar to detect landslide in complex surroundings: Case study of 2015 shenzhen landslide. *ISPRS Journal of Photogrammetry and Remote Sensing*, 174:56–67, 2021.

[319] Nicla Notarangelo, Charlotte Wirion, and Frankwin van Winsen. Sturm-flood: a curated dataset for deep learning-based flood extent mapping leveraging sentinel-1 and sentinel-2 imagery. *Big Earth Data*, pages 1–27, 2025.

[320] Shigeyuki Okada and Nobuo Takai. Classifications of structural types and damage patterns of buildings for earthquake field investigation. In *Proceedings of the 12th world conference on earthquake engineering, Auckland, New Zealand*, volume 30, 2000.

[321] OpenStreetMap contributors. Planet dump retrieved from https://planet.openstreetmap.org. https://www.openstreetmap.org, 2026. Data file; accessed 2026-03-26.

[322] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024.

[323] Aiym Orynbaikyzy, Frauke Albrecht, Wei Yao, Mahdi Motagh, Wandi Wang, Sandro Martinis, and Simon Plank. Landslide mapping with deep learning: the role of pre-/post-event sar features and multi-sensor data fusion. *GIScience & Remote Sensing*, 62(1):2502214, 2025.

[324] Overture Maps Foundation. Overture maps. https://overturemaps.org, 2026. Accessed 2026-03-26.

[325] Lei Pang, Fengli Zhang, Lu Li, Qiqi Huang, Yanan Jiao, and Yun Shao. Assessing buildings damage from multi-temporal sar images fusion using semantic change detection. In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 6292–6295, 2022.

[326] Yongqi Pang, Yudong Li, Zhongke Feng, Zemin Feng, Ziyu Zhao, Shilin Chen, and Hanyue Zhang. Forest fire occurrence prediction in china based on machine learning methods. *Remote Sensing*, 14(21):5546, 2022.

[327] Minsoo Park, Dai Quoc Tran, Daekyo Jung, and Seunghee Park. Wildfire-detection method using densenet and cyclegan data augmentation-based remote camera imagery. *Remote Sensing*, 12(22):3715, 2020.

[328] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. *PyTorch: an imperative style, high-performance deep learning library*. Curran Associates Inc., Red Hook, NY, USA, 2019.

[329] Sayak Paul and Siddha Ganju. Flood segmentation on sentinel-1 sar imagery with semi-supervised learning. *arXiv preprint arXiv:2107.08369*, 2021.

[330] Bo Peng, Qunying Huang, and Jinmeng Rao. Spatiotemporal contrastive representation learning for building damage classification. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 8562–8565, 2021.

[331] Claudio Persello, Saurabh Prasad, Ujjwal Verma, Gemine Vivone, Hongruixuan Chen, Junshi Xia, Jian Song, Clifford Broni-Bediako, Olivier Dietrich, Konrad Schindler, and Naoto Yokoya. 2025 ieee grss data fusion contest: All-weather land cover and building damage mapping. *IEEE Geoscience and Remote Sensing Magazine*, 13(2):388–392, 2025.

[332] Claudio Persello, Saurabh Prasad, Gemine Vivone, Vincent Lonjou, Frederic Bretar, Raquel Rodriguez-Suquet, Pauline Guntzburger, Vincent Poulain, Jacqueline LE Moigne, Benjamin Smith, et al. 2024 ieee grss data fusion contest: Rapid flood mapping [technical committees]. *IEEE geoscience and remote sensing magazine*, 12(2):109–112, 2024.

[333] Yalong Pi, Nipun D. Nath, and Amir H. Behzadan. Convolutional neural networks for object detection in aerial imagery for disaster response and recovery. *Advanced Engineering Informatics*, 43:101009, 2020.

[334] Yalong Pi, Nipun D. Nath, and Amir H. Behzadan. Detection and semantic segmentation of disaster damage in uav footage. *Journal of Computing in Civil Engineering*, 35(2):04020063, 2021.

[335] Ioannis Prapas, Spyros Kondylatos, and Ioannis Papoutsis. Firecube: A daily datacube for the modeling and analysis of wildfires in greece. *URL: https://doi. org/10.5281/zenodo*, 6475592, 2022.

[336] Maria Presa-Reyes, Yudong Tao, Shu-Ching Chen, and Mei-Ling Shyu. Deep learning with weak supervision for disaster scene description in low-altitude imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–10, 2022.

[337] Luca Pulvirenti, Marco Chini, Nazzareno Pierdicca, and Giorgio Boni. Use of sar data for detecting floodwater in urban and agricultural areas: The role of the interferometric coherence. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3):1532–1544, 2016.

[338] Luca Pulvirenti, Nazzareno Pierdicca, Marco Chini, and Leila Guerriero. An algorithm for operational flood mapping from Synthetic Aperture Radar (SAR) data using fuzzy logic. *Natural Hazards and Earth System Sciences*, 11(2):529–540, 2011.

[339] Wenfan Qiao, Li Shen, Jicheng Wang, Xiaotian Yang, and Zhilin Li. A weakly supervised semantic segmentation approach for damaged building extraction from postearthquake high-resolution remote-sensing images. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023.

[340] Wenfan Qiao, Li Shen, Wei Wang, and Zhilin Li. A weakly supervised bitemporal scene change detection approach for pixel-level building damage assessment using pre- and post-disaster high-resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–23, 2024.

[341] Wenfan Qiao, Li Shen, Qi Wen, Quan Wen, Shiyang Tang, and Zhilin Li. Revolutionizing building damage detection: A novel weakly supervised approach using high-resolution remote sensing images. *International Journal of Digital Earth*, 17(1):2298245, 2024.

[342] Shi Qiu, Zhe Zhu, and Binbin He. Fmask 4.0: Improved cloud and cloud shadow detection in landsats 4–8 and sentinel-2 imagery. *Remote Sensing of Environment*, 231:111205, 2019.

[343] Shi Qiu, Zhe Zhu, Pontus Olofsson, Curtis E. Woodcock, and Suming Jin. Evaluation of landsat image compositing algorithms. *Remote Sensing of Environment*, 285:113375, 2023.

[344] Carmen Quintano, Alfonso Fernandez-Manso, and Dar A Roberts. Burn severity mapping from landsat mesma fraction images and land surface temperature. *Remote sensing of environment*, 190:83–95, 2017.

[345] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR, 18–24 Jul 2021.

[346] Maryam Rahnemoonfar, Tashnim Chowdhury, and Robin Murphy. RescueNet: A High Resolution UAV Semantic Segmentation Dataset for Natural Disaster Damage Assessment. *Scientific Data*, 10(1):913, 2023.

[347] Maryam Rahnemoonfar, Tashnim Chowdhury, Argho Sarkar, Debvrat Varshney, Masoud Yari, and Robin Roberson Murphy. FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding. *IEEE Access*, 9:89644–89654, 2021.

[348] Clément Rambour, Nicolas Audebert, E Koeniguer, Bertrand Le Saux, M Crucianu, and Mihai Datcu. Flood detection in time series of optical and sar images. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:1343–1346, 2020.

[349] Shuhao Ran, Gang Ma, Jiangzhou Mei, Fudong Chi, Wei Zhou, and Yonghong Weng. Semi-lm: A reliable multi-perspective consistency semi-supervised learning framework for landslide mapping from remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[350] Dmitry Rashkovetsky, Florian Mauracher, Martin Langer, and Michael Schmitt. Wildfire detection from multisensor satellite imagery using deep semantic segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:7001–7016, 2021.

[351] H. Rastiveis, F. Eslamizade, and E. Hosseini-Zirdoo. Building Damage Assessment After Earthquake Using Post-event LiDAR

Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1/W5:595–600, 2015.

[352] H. Rastiveis, E. Hosseini-Zirdoo, and F. Eslamizade. Automatic blocked roads assessment after earthquake using high resolution satellite imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1/W5:601–605, 2015.

[353] Ellen M. Rathje, Melba Crawford, Kyuseok Woo, and Amy Neuenschwander. Damage patterns from satellite images of the 2003 bam, iran, earthquake. *Earthquake Spectra*, 21(1_suppl):295–307, 2005.

[354] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollar, and Christoph Feichtenhofer. SAM 2: Segment Anything in Images and Videos. In *The Thirteenth International Conference on Learning Representations*, 2025.

[355] Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, Rachel Prudden, Amol Mandhane, Aidan Clark, Andrew Brock, Karen Simonyan, Raia Hadsell, Niall Robinson, Ellen Clancy, Alberto Arribas, and Shakir Mohamed. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677, 2021.

[356] Sufia Rehman, Mehebub Sahana, Haoyuan Hong, Haroon Sajjad, and Baharin Bin Ahmed. A systematic review on approaches and methods used for flood vulnerability assessment: framework for future research. *Natural Hazards*, 96(2):975–998, 2019.

[357] Sean Reilly, Matthew L Clark, Lisa Patrick Bentley, Corbin Matley, Elise Piazza, and Imma Oliveras Menor. The potential of multispectral imagery and 3d point clouds from unoccupied aerial systems (uas) for monitoring forest structure and the impacts of wildfire in mediterranean-climate forests. *Remote Sensing*, 13(19):3810, 2021.

[358] JH Reksten, A-B Salberg, and R Solberg. Flood detection in norway based on sentinel-1 sar imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:349–355, 2019.

[359] Tiago FR Ribeiro, Fernando Silva, José Moreira, and Rogerio Luis de C Costa. Burned area semantic segmentation: A novel dataset and evaluation using convolutional networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 202:565–580, 2023.

[360] Benjamin Aubrey Robson, Tobias Bolch, Shelley MacDonell, Daniel Hölbling, Philipp Rastner, and Nicole Schaffer. Automated detection of rock glaciers using deep learning and object-based image analysis. *Remote sensing of environment*, 250:112033, 2020.

[361] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[362] Julian F Rosser, Didier G Leibovici, and Margaret J Jackson. Rapid flood inundation mapping using social media, remote sensing and topographic data. *Natural Hazards*, 87(1):103–120, 2017.

[363] Tim G. J. Rudner, Marc Rußwurm, Jakub Fil, Ramona Pelich, Benjamin Bischke, Veronika Kopačková, and Piotr Biliński. Multi3Net: Segmenting Flooded Buildings via Fusion of Multiresolution, Multisensor, and Multitemporal Satellite Imagery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):702–709, Jul. 2019.

[364] Xue Rui, Yang Cao, Xin Yuan, Yu Kang, and Weiguo Song. Disastergan: Generative adversarial networks for remote sensing disaster image generation. *Remote Sensing*, 13(21), 2021.

[365] Luigi Russo, Deodato Tapete, Silvia Liberata Ullo, and Paolo Gamba. A deep learning framework for building damage assessment using vhr sar and geospatial data: demonstration on the 2023 turkiye earthquake. *arXiv preprint arXiv:2506.22338*, 2025.

[366] Samira Saadi, Khadidja Abbes, and Baya Nait Amara. Shadow buildings analysis for seismic damage assesment. In *2024 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*, pages 123–128, 2024.

[367] Rizwan Sadiq, Zainab Akhtar, Muhammad Imran, and Ferda Ofli. Integrating remote sensing and social sensing for flood mapping. *Remote Sensing Applications: Society and Environment*, 25:100697, 2022.

[368] Lucia Saganeiti, Federico Amato, Gabriele Nolè, Marco Vona, and Beniamino Murgante. Early estimation of ground displacements and building damage after seismic events using sar and lidar data: The case of the amatrice earthquake in central italy, on 24th august 2016. *International Journal of Disaster Risk Reduction*, 51:101924, 2020.

[369] Sudipan Saha, Francesca Bovolo, and Lorenzo Bruzzone. Building change detection in vhr sar images via unsupervised deep transcoding. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3):1917–1929, 2021.

[370] Tamer Saleh, Xingxing Weng, Shimaa Holail, Chen Hao, and Gui-Song Xia. Dam-net: Flood detection from sar imagery using differential attention metric-based vision transformers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 212:440–453, 2024.

[371] F Samadzadegan, H Rastiveisi, and WG Viii. Automatic detection and classification of damaged buildings, using high resolution satellite imagery and vector data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(Part B8):415–420, 2008.

[372] Jacob Sanderson, Hua Mao, Mohammed AM Abdullah, Raid Rafi Omar Al-Nima, and Wai Lok Woo. Optimal fusion of multispectral optical and sar images for flood inundation mapping through explainable deep learning. *Information*, 14(12):660, 2023.

[373] Younes Oulad Sayad, Hajar Mousannif, and Hassan Al Moatassime. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire safety journal*, 104:130–146, 2019.

[374] Corey Scher and Jamon Van Den Hoek. Active insar monitoring of building damage in gaza during the israel-hamas war, 2025.

[375] Corey Scher and Jamon Van Den Hoek. Nationwide conflict damage mapping with interferometric synthetic aperture radar: A study of the 2022 russia–ukraine conflict. *Science of Remote Sensing*, 11:100217, June 2025.

[376] Wilfrid Schroeder, Patricia Oliva, Louis Giglio, and Ivan A Csiszar. The new viirs 375 m active fire detection data product: Algorithm description and initial assessment. *Remote Sensing of Environment*, 143:85–96, 2014.

[377] Maria Sdraka, Alkinoos Dimakos, Alexandros Malounis, Zisoula Ntasiou, Konstantinos Karantzalos, Dimitrios Michail, and Ioannis Papoutsis. Floga: A machine-learning-ready dataset, a benchmark, and a novel deep learning model for burnt area mapping with sentinel-2. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:7801–7824, 2024.

[378] Aliihsan Sekertekin. A survey on global thresholding methods for mapping open water body using sentinel-2 satellite imagery and normalized difference water index. *Archives of Computational Methods in Engineering*, 28(3), 2021.

[379] Sivasakthy Selvakumaran, Iain Rolland, Luke Cullen, Rob Davis, Joshua Macabuag, Charbel Abou Chakra, Nanor Karageozian, Amir Gilani, Christian Geiβ, Miguel Bravo-Haro, and Andrea Marinoni. Improving operational use of post-disaster damage assessment for urban search and rescue by integrated graph-based multimodal remote sensing data analysis. *Progress in Disaster Science*, 25:100404, 2025.

[380] Minseok Seo, Jinwook Jung, and Dong-Geol Choi. Improved flood insights: Diffusion-based sar-to-eo image translation. *Remote Sensing*, 17(13):2260, 2025.

[381] Seyd Teymoor Seydi, Heidar Rastiveis, Bahareh Kalantar, Alfian Abdul Halin, and Naonori Ueda. Bdd-net: An end-to-end multiscale residual cnn for earthquake-induced building damage detection. *Remote Sensing*, 14(9), 2022.

[382] Hejar Shahabi, Maryam Rahimzad, Sepideh Tavakkoli Piralilou, Omid Ghorbanzadeh, Saied Homayouni, Thomas Blaschke, Samsung Lim, and Pedram Ghamisi. Unsupervised Deep Learning for Landslide Detection from Multispectral Sentinel-2 Imagery. *Remote Sensing*, 13(22), 2021.

[383] Alireza Shamsoshoara, Fatemeh Afghah, Abolfazl Razi, Liming Zheng, Peter Z Fulé, and Erik Blasch. Aerial imagery pile burn detection using deep learning: The FLAME dataset. *Computer Networks*, 193:108001, 2021.

[384] Wei Shan, Jiawen Liu, and Ying Guo. Rapid flood mapping and disaster assessment based on gee platform: Case study of a rainstorm from july to august 2024 in liaoning province, china. *Water*, 17(16):2416, 2025.

[385] Jinyuan Shao, Lina Tang, Ming Liu, Guofan Shao, Lang Sun, and Quanyi Qiu. Bdd-net: A general protocol for mapping buildings damaged by a wide range of disasters based on satellite imagery. *Remote Sensing*, 12(10), 2020.

[386] Yakui Shao, Zhongke Feng, Linhao Sun, Xuanhan Yang, Yudong Li, Bo Xu, and Yuan Chen. Mapping china's forest fire risks with machine learning. *Forests*, 13(6):856, 2022.

[387] Nirdesh Kumar Sharma and Manabendra Saharia. Deepsarflood: Rapid and automated sar-based flood inundation mapping using vision transformer-based deep ensembles with uncertainty estimates. *Science of Remote Sensing*, 11:100203, 2025.

[388] Ram C. Sharma, Ryutaro Tateishi, Keitarou Hara, Hoan Thanh Nguyen, Saeid Gharechelou, and Luong Viet Nguyen. Earthquake damage visualization (edv) technique for the rapid detection of earthquake-induced damages using sar data. *Sensors*, 17(2), 2017.

[389] Li Shen and Changchun Li. Water body extraction from landsat etm+ imagery using adaboost algorithm. In *2010 18th International Conference on Geoinformatics*, pages 1–4. IEEE, 2010.

[390] Runping Shen, Anqi Huang, Bolun Li, and Jia Guo. Construction of a drought monitoring model using deep learning based on multi-source remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 79:48–57, 2019.

[391] Yu Shen, Shuang Yao, Zhenkai Qiang, and Guanxiang Pei. Sd-mamba: A lightweight synthetic-decompression network for cross-modal flood change detection. *International Journal of Applied Earth Observation and Geoinformation*, 136:104409, 2025.

[392] Yu Shen, Sijie Zhu, Taojiannan Yang, Chen Chen, Delu Pan, Jianyu Chen, Liang Xiao, and Qian Du. Bdanet: Multiscale convolutional neural network with cross-directional attention for building damage assessment from satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

[393] Wenzhong Shi, Min Zhang, Hongfei Ke, Xin Fang, Zhao Zhan, and Shanxiong Chen. Landslide recognition by deep convolutional neural network and change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):4654–4672, 2020.

[394] Ranjay Shrestha, Liping Di, Genong Yu, Yuanzheng Shao, Lingjung Kang, and Bei Zhang. Detection of flood and its impact on crops using ndvi-corn case. In *2013 Second international conference on agro-geoinformatics (Agro-geoinformatics)*, pages 200–204. IEEE, 2013.

[395] Priyanka Shukla, Dericks Praise Shukla, Sina Klampt, and Daiki Kimura. Finetuning geospatial foundation model for wildfire scars in himalayan region. In *IGARSS 2025 - 2025 IEEE International Geoscience and Remote Sensing Symposium*, pages 3153–3157, 2025.

[396] F Siegert and G Ruecker. Use of multitemporal ers-2 sar images for identification of burned scars in south-east asian tropical rainforest. *International Journal of Remote Sensing*, 21(4):831–837, 2000.

[397] Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. DINOv3, 2025.

[398] Deepank Kumar Singh and Vedhus Hoskere. Post disaster damage assessment using ultra-high-resolution aerial imagery with semi-supervised transformers. *Sensors*, 23(19), 2023.

[399] Gulab Singh, Yoshio Yamaguchi, Wolfgang-Martin Boerner, and Sang-Eun Park. Monitoring of the March 11, 2011, Off-Tohoku 9.0 Earthquake With Super-Tsunami Disaster by Implementing Fully Polarimetric High-Resolution POLSAR Techniques. *Proceedings of the IEEE*, 101(3):831–846, 2013.

[400] Samriddhi Singla, Tina Diao, Ayan Mukhopadhyay, Ahmed Eldawy, Ross Shachter, and Mykel Kochenderfer. Wildfiredb: A spatio-temporal dataset combining wildfire occurrence with relevant covariates. In *34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, 2020.

[401] Wojciech Sirko, Sergii Kashubin, Marvin Ritter, Abigail Annkah, Yasser Salah Eddine Bouchareb, Yann Dauphin, Daniel Keysers, Maxim Neumann, Moustapha Cisse, and John A. Quinn. Continental-scale building detection from high resolution satellite imagery. *arXiv preprint arXiv:2107.12283*, 2021.

[402] Ramesh Sivanpillai, Kevin M Jacobs, Chloe M Mattilio, and Ela V Piskorski. Rapid flood inundation mapping by differencing water indices from pre-and post-flood landsat images. *Frontiers of Earth Science*, 15(1):1–11, 2021.

[403] Lorenzo Solari, Silvia Bianchini, Rachele Franceschini, Anna Barra, Oriol Monserrat, Patrick Thuegaz, Davide Bertolo, Michele Crosetto, and Filippo Catani. Satellite interferometric data for landslide intensity evaluation in mountainous regions. *International Journal of Applied Earth Observation and Geoinformation*, 87:102028, 2020.

[404] Amruta Sonavale, Midhun Chakkaravarthy, Surampudi Srinivasa Rao, Hishamuddin Bin M Salleh, and Jagannath Jadhav. Automatic semantic segmentation and classification of remote sensing image data for flood detection using novel lstm neural network. *SN Computer Science*, 5(8):992, 2024.

[405] Jian Song, Bruno Adriano, and Naoto Yokoya. Disaster detection from sar images with different off-nadir angles using unsupervised image translation. In *CDCEO@ IJCAI*, pages 14–20, 2022.

[406] Jian Song, Hongruixuan Chen, Weihao Xuan, Junshi Xia, and Naoto Yokoya. Synrs3d: a synthetic dataset for global 3d semantic understanding from monocular remote sensing imagery. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2024. Curran Associates Inc.

[407] Jian Song, Hongruixuan Chen, and Naoto Yokoya. Syntheworld: A large-scale synthetic dataset for land cover mapping and building change detection. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 8272–8281, 2024.

[408] Wenke Song, Mingfu Guan, and Dapeng Yu. Swinflood: A hybrid cnn-swin transformer model for rapid spatiotemporal flood simulation. *Journal of Hydrology*, 660:133280, 2025.

[409] Yuqi Song, Xie Hu, Xuguo Shi, Yifei Cui, Chao Zhou, and Yueren Xu. Hydrological proxy derived from insar coherence in landslide characterization. *Remote Sensing of Environment*, 322:114712, 2025.

[410] Oliver L. Stephenson, Tobias Kohne, Eric Zhan, Brent E. Cahill, Sang-Ho Yun, Zachary E. Ross, and Mark Simons. Deep learning-based damage mapping with InSAR coherence time series. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–17, 2022.

[411] S. Stramondo, C. Bignami, M. Chini, N. Pierdicca, and A. Tertulliani. Satellite radar and optical remote sensing for earthquake damage detection: results from different case studies. *International Journal of Remote Sensing*, 27(20):4433–4447, 2006.

[412] André Stumpf and Norman Kerle. Object-oriented mapping of landslides using random forests. *Remote sensing of environment*, 115(10):2564–2577, 2011.

[413] MITGT Sugiyama and Hitoshi Saji Keiichi Abe. Detection of earthquake damaged areas from aerial photographs by using color and edge information. In *Proceedings of the 5th Asian Conference on Computer Vision, Melbourne, Australia*, pages 23–25, 2002.

[414] Chongjing Sun, Dongping Ming, Lu Xu, Shizhe Xie, Ran Liu, and Xiao Ling. A hybrid damaged building sample generation method based on cross-scale fusion generative model for destroyed building detection after earthquake. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–15, 2025.

[415] Xian Sun, Peijin Wang, Wanxuan Lu, Zicong Zhu, Xiaonan Lu, Qibin He, Junxi Li, Xuee Rong, Zhujun Yang, Hao Chang, Qinglin He, Guang Yang, Ruiping Wang, Jiwen Lu, and Kun Fu. Ringmo: A remote sensing foundation model with masked image modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–22, 2023.

[416] Yao Sun, Yi Wang, and Michael Eineder. Quickquakebuildings: Post-earthquake sar-optical dataset for quick damaged-building detection. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024.

[417] Yuli Sun, Lin Lei, Dongdong Guan, and Gangyao Kuang. Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images. *IEEE Transactions on Image Processing*, 30:6277–6291, 2021.

[418] Yuli Sun, Lin Lei, Dongdong Guan, Ming Li, and Gangyao Kuang. Sparse-constrained adaptive structure consistency-based unsupervised image regression for heterogeneous remote-sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

[419] Yuli Sun, Lin Lei, Xiao Li, Hao Sun, and Gangyao Kuang. Nonlocal patch similarity based heterogeneous remote sensing change detection. *Pattern Recognition*, 109:107598, 2021.

[420] Daniela Szwarcman, Sujit Roy, Paolo Fraccaro, orsteinn Elí Gíslason, Benedikt Blumenstiel, Rinki Ghosal, Pedro Henrique de Oliveira, Joao Lucas de Sousa Almeida, Rocco Sedona, Yanghui Kang, Srija Chakraborty, Sizhe Wang, Carlos Gomes, Ankur Kumar, Vishal Gaur, Myscon Truong, Denys Godwin, Sam Khallaghi, Hyunho Lee, Chia-Yu Hsu, Ata Akbari Asanjan, Besart

Mujeci, Disha Shidham, Rufai Omowunmi Balogun, Venkatesh Kolluru, Trevor Keenan, Paulo Arevalo, Wenwen Li, Hamed Alemohammad, Pontus Olofsson, Timothy Mayer, Christopher Hain, Robert Kennedy, Bianca Zadrozny, David Bell, Gabriele Cavallaro, Campbell Watson, Manil Maskey, Rahul Ramachandran, and Juan Bernabe Moreno. Prithvi-eo-2.0: A versatile multitemporal foundation model for earth observation applications. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–20, 2026.

[421] Sahand Tahermanesh, Ali Mohammadzadeh, Amin Mohsenifar, and Armin Moghimi. Siscnet: A novel siamese inception-based network with spatial and channel attention for flood detection in sentinel-1 imagery. *Remote Sensing Applications: Society and Environment*, 38:101571, 2025.

[422] Nopphawan Tamkuan and Masahiko Nagai. Fusion of multi-temporal interferometric coherence and optical image data for the 2016 kumamoto earthquake damage assessment. *ISPRS International Journal of Geo-Information*, 6(7), 2017.

[423] Xuwei Tan, Qian Zhao, Yanlan Liu, and Xueru Zhang. Droughtset: Understanding drought through spatial-temporal learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 28440–28448, 2025.

[424] Xiaochuan Tang, Zhong Lu, Xuanmei Fan, Xiaochuang Yan, Xiaojun Yuan, Dongfen Li, Huailiang Li, Hongjun Li, Sansar Raj Meena, Alessandro Novellino, et al. Mamba for landslide detection: A lightweight model for mapping landslides with very high-resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[425] Cheryl W. J. Tay, Sang-Ho Yun, Shi Tong Chin, Alok Bhardwaj, Jungkyo Jung, and Emma M. Hill. Rapid flood and damage mapping using synthetic aperture radar in response to typhoon Hagibis, Japan. *Scientific Data*, 7(1), March 2020.

[426] Beth Tellman, Jonathan A Sullivan, Catherine Kuhn, Albert J Kettner, Colin S Doyle, G Robert Brakenridge, Tyler A Erickson, and Daniel A Slayback. Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, 596(7870):80–86, 2021.

[427] E. Tennant, S. F. Jenkins, V. Miller, R. Robertson, B. Wen, S.-H. Yun, and B. Taisne. Automating tephra fall building damage assessment using deep learning. *Natural Hazards and Earth System Sciences*, 24(12):4585–4608, 2024.

[428] Sofia Tilon, Francesco Nex, Norman Kerle, and George Vosselman. Post-disaster building damage detection from earth observation imagery using unsupervised and transferable anomaly detecting generative adversarial networks. *Remote Sensing*, 12(24), 2020.

[429] Daniel Tomowski, Sascha Klonus, Manfred Ehlers, Ulrich Michel, and Peter Reinartz. Change visualization through a texture-based analysis approach for disaster applications. In W. Wagner and B. Székely, editors, *ISPRS Technical Commission VII Symposium – 100 Years ISPRS: Advancing Remote Sensing Science*, volume XXXVIII-7A of *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Archives)*, pages 263–268, Vienna, Austria, July 2010. ISPRS.

[430] Xiaohua Tong, Zhonghua Hong, Shijie Liu, Xue Zhang, Huan Xie, Zhengyuan Li, Sonlin Yang, Weian Wang, and Feng Bao. Building-damage detection using pre- and post-seismic high-resolution satellite stereo imagery: A case study of the may 2008 wenchuan earthquake. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68:13–27, 2012.

[431] Xiaohua Tong, Xiaofei Lin, Tiantian Feng, Huan Xie, Shijie Liu, Zhonghua Hong, and Peng Chen. Use of shadows for detection of earthquake-induced collapsed buildings in high-resolution satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 79:53–67, 2013.

[432] Xiaopeng Tong and David Schmidt. Active movement of the cascade landslide complex in washington from a coherence-based insar time series method. *Remote Sensing of Environment*, 186:405–415, 2016.

[433] Redha Touati, Max Mignotte, and Mohamed Dahmane. Multimodal change detection in remote sensing images using an unsupervised pixel pairwise-based markov random field model. *IEEE Transactions on Image Processing*, 29:757–767, 2020.

[434] Dai Quoc Tran, Minsoo Park, Daekyo Jung, and Seunghee Park. Damage-map estimation using UAV images and deep learning algorithms for disaster management system. *Remote Sensing*, 12(24):4169, 2020.

[435] Gabriel Tseng, Anthony Fuller, Marlena Reil, Henry Herzog, Patrick Beukema, Favyen Bastani, James R Green, Evan Shelhamer, Hannah Kerner, and David Rolnick. Galileo: Learning global & local features of many remote sensing modalities. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 60280–60300. PMLR, 13–19 Jul 2025.

[436] Ken Tsutsui, Shuichi Rokugawa, Hideaki Nakagawa, Sanae Miyazaki, Chin-Tung Cheng, Takashi Shiraishi, and Shiun-Der Yang. Detection and volume estimation of large-scale landslides based on elevation-change analysis using dems extracted from high-resolution satellite stereo imagery. *IEEE transactions on geoscience and remote sensing*, 45(6):1681–1696, 2007.

[437] Kabir Uddin, Mir A Matin, and Franz J Meyer. Operational flood mapping using multi-temporal sentinel-1 sar images: A case study from bangladesh. *Remote Sensing*, 11(13):1581, 2019.

[438] Silvia Liberata Ullo, Amrita Mohan, Alessandro Sebastianelli, Shaik Ejaz Ahamed, Basant Kumar, Ramji Dwivedi, and Ganesh R Sinha. A new mask r-cnn-based method for improved landslide detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:3799–3810, 2021.

[439] Noel Ivan Ulloa, Sang-Ho Yun, Shou-Hao Chiang, and Ryoichi Furuta. Sentinel-1 spatiotemporal simulation using convolutional lstm for flood mapping. *Remote sensing*, 14(2):246, 2022.

[440] UNHCR. Severe floods displace nearly 400,000 in Mozambique, heightening protection risks. https://www.unhcr.org/news/briefing-notes/severe-floods-displace-nearly-400-000-mozambique-heightening-protection-risks, February 2026. Accessed: 2026-04-10.

[441] United Nations Office for Disaster Risk Reduction. Disaster Risk Reduction in the 2030 Agenda for Sustainable Development. Technical report, United Nations International Strategy for Disaster Reduction (UNISDR), Geneva, 2015.

[442] United Nations Office for Disaster Risk Reduction. Sendai Framework for Disaster Risk Reduction 2015–2030. Technical report, United Nations Office for Disaster Risk Reduction (UNDRR), Geneva, 2015.

[443] United Nations Office for Disaster Risk Reduction. Implementing the Sendai Framework to Achieve the Sustainable Development Goals. Technical report, United Nations International Strategy for Disaster Reduction (UNISDR), Geneva, 2016.

[444] United Nations Office for Disaster Risk Reduction. Global assessment report on disaster risk reduction 2022: Our world at risk — transforming governance for a resilient future. Technical report, UNDRR, Geneva, 2022.

[445] United Nations Office for Disaster Risk Reduction. Global Assessment Report on Disaster Risk Reduction 2023: Mapping Resilience for the Sustainable Development Goals. Technical report, United Nations Office for Disaster Risk Reduction (UNDRR), Geneva, 2023.

[446] United Nations Office for Disaster Risk Reduction. Global Assessment Report on Disaster Risk Reduction 2025: Resilience Pays – Financing and Investing for our Future. Technical report, United Nations Office for Disaster Risk Reduction (UNDRR), Geneva, 2025.

[447] United Nations Secretary-General. Implementation of the Sendai Framework for Disaster Risk Reduction 2015–2030: Report of the Secretary-General. Technical report, United Nations Office for Disaster Risk Reduction (UNDRR), Geneva, 2025.

[448] Serkan Ural, Ejaz Hussain, KyoHyouk Kim, Chiung-Shiuan Fu, and Jie Shan. Building extraction and rubble mapping for city port-au-prince post-2010 earthquake with geoeye-1 imagery and lidar data. *Photogrammetric Engineering & Remote Sensing*, 77(10):1011–1023, 2011.

[449] Dimitris Valsamis, Alexandros Oikonomidis, Chrysoula Chatzichristaki, Anastasia Moumtzidou, Ilias Gialampoukidis, Stefanos Vrochidis, and Ioannis Kompatsiaris. Towards advanced wildfire analysis: A siamese network-based change detection approach through self-supervised learning. In *2024 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–7. IEEE, 2024.

[450] Linh Nguyen Van, Vinh Ngoc Tran, Giang V Nguyen, Minho Yeon, May Thi-Tuyet Do, and Giha Lee. Enhancing wildfire mapping accuracy using mono-temporal sentinel-2 data: A novel

approach through qualitative and quantitative feature selection with explainable ai. *Ecological Informatics*, 81:102601, 2024.

[451] Venkata Sai Krishna Vanama and Yalamanchili S Rao. Change detection based flood mapping of 2015 flood event of chennai city using sentinel-1 sar images. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 9729–9732. IEEE, 2019.

[452] T Veljanovski, P Lamovec, P Pehani, and K Oštir. Comparison of three techniques for detection of flooded areas on envisat and radarsat-2 satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2011.

[453] Guido Ventura, Giuseppe Vilardo, Carlo Terranova, and Eliana Bellucci Sessa. Tracking and evolution of complex active landslides by multi-temporal airborne lidar data: The montaguto landslide (southern italy). *Remote Sensing of Environment*, 115(12):3237–3248, 2011.

[454] E.F. Vermote, D. Tanre, J.L. Deuze, M. Herman, and J.-J. Morcette. Second simulation of the satellite signal in the solar spectrum, 6s: an overview. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3):675–686, 1997.

[455] Eric Vermote, Christopher Justice, Martin Claverie, and Belen Franch. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185:46–56, 2016.

[456] Anand Vetrivel, Markus Gerke, Norman Kerle, Francesco Nex, and George Vosselman. Disaster damage detection through synergistic use of deep learning and 3d point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140:45–59, 2018.

[457] Anand Vetrivel, Markus Gerke, Norman Kerle, Francesco Nex, and George Vosselman. Disaster damage detection through synergistic use of deep learning and 3d point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140:45–59, 2018. Geospatial Computer Vision.

[458] Anand Vetrivel, Markus Gerke, Norman Kerle, and George Vosselman. Identification of damage in buildings based on gaps in 3D point clouds from very high resolution oblique airborne images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105:61–78, 2015.

[459] Michele Volpi, Gustau Camps-Valls, and Devis Tuia. Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, 107:50–63, 2015.

[460] Jirapa Vongkusolkit, Bo Peng, Meiliu Wu, Qunying Huang, and Christian G Andresen. Near real-time flood mapping with weakly supervised machine learning. *Remote Sensing*, 15(13):3263, 2023.

[461] T. T. Vu and Y. Ban. Context-based mapping of damaged buildings from high-resolution optical satellite images. *International Journal of Remote Sensing*, 31(13):3411–3425, 2010.

[462] Wolfgang Wagner, Bernhard Bauer-Marschallinger, Florian Roth, Tobias Raiger-Stachl, Christoph Reimer, Niall McCormick, Patrick Matgen, Marco Chini, Yu Li, Sandro Martinis, Marc Wieland, Franziska Kraft, Davide Festa, Muhammed Hassaan, Mark Edwin Tupas, Jie Zhao, Michaela Seewald, Michael Riffler, Luca Molini, Richard Kidd, Christian Briese, and Peter Salamon. The fully-automatic sentinel-1 global flood monitoring service: Scientific challenges and future directions. *Remote Sensing of Environment*, 333:115108, 2026.

[463] Leonard Waldmann, Ando Shah, Yi Wang, Nils Lehmann, Adam J. Stewart, Zhitong Xiong, Xiao Xiang Zhu, Stefan Bauer, and John Chuang. Panopticon: Advancing any-sensor foundation models for earth observation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2195–2205, 2025.

[464] L. Wan, T. Zhang, and H. J. You. Multi-sensor remote sensing image change detection based on sorted histograms. *International Journal of Remote Sensing*, 39(11):3753–3775, 2018.

[465] Ling Wan, Yuming Xiang, and Hongjian You. A Post-Classification Comparison Method for SAR and Optical Images Change Detection. *IEEE Geoscience and Remote Sensing Letters*, 16(7):1026–1030, 2019.

[466] Ling Wan, Yuming Xiang, and Hongjian You. An object-based hierarchical compound classification method for change detection in heterogeneous optical and SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):9941–9959, 2019.

[467] Chao Wang, Yan Zhang, Tao Xie, Lin Guo, Shishi Chen, Junyong Li, and Fan Shi. A detection method for collapsed buildings combining post-earthquake high-resolution optical and synthetic aperture radar images. *Remote Sensing*, 14(5), 2022.

[468] Di Wang, Meiqi Hu, Yao Jin, Yuchun Miao, Jiaqi Yang, Yichu Xu, Xiaolei Qin, Jiaqi Ma, Lingyu Sun, Chenxing Li, Chuan Fu, Hongruixuan Chen, Chengxi Han, Naoto Yokoya, Jing Zhang, Minqiang Xu, Lin Liu, Lefei Zhang, Chen Wu, Bo Du, Dacheng Tao, and Liangpei Zhang. Hypersigma: Hyperspectral intelligence comprehension foundation model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(8):6427–6444, 2025.

[469] Di Wang, Guorui Ma, Xiao Wang, Ronghao Yang, and Yongxian Zhang. Few-shot change detection in optical and sar remote sensing images for disaster response. *International Journal of Applied Earth Observation and Geoinformation*, 146:105100, 2026.

[470] Fajing Wang and Xu Feng. Flood change detection model based on an improved u-net network and multi-head attention mechanism. *Scientific Reports*, 15(1):3295, 2025.

[471] Jiaqi Wang, Haonan Guo, Xin Su, Li Zheng, and Qiangqiang Yuan. Pcdasnet: Position-constrained differential attention siamese network for building damage assessment. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–18, 2024.

[472] Jin-Ju Wang, Nicolas Dobigeon, Marie Chabert, Ding-Cheng Wang, Ting-Zhu Huang, and Jie Huang. Cd-gan: A robust fusion-based generative adversarial network for unsupervised remote sensing change detection with heterogeneous sensors. *Information Fusion*, 107:102313, 2024.

[473] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao. Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3349–3364, 2021.

[474] Junjue Wang, Weihao Xuan, Heli Qi, Zhihao Liu, Kunyi Liu, Yuhan Wu, Hongruixuan Chen, Jian Song, Junshi Xia, Zhuo Zheng, et al. Disasterm3: A remote sensing vision-language dataset for disaster damage assessment and response. *arXiv preprint arXiv:2505.21089*, 2025.

[475] Libo Wang, Rui Li, Ce Zhang, Shenghui Fang, Chenxi Duan, Xiaoliang Meng, and Peter M. Atkinson. Unetformer: A unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 190:196–214, 2022.

[476] Lili Wang, Jidong Wu, Youtian Yang, Rumei Tang, and Ru Ya. Deep learning models for hazard-damaged building detection using remote sensing datasets: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:15301–15318, 2024.

[477] Teng Wang, Yunjia Wang, Feng Zhao, Guangqian Zou, Sen Du, Leixin Zhang, Kewei Zhang, Nianbin Zhang, Zhanguo Ma, and José Fernández. In-dmu: Modeling uncertainty in interferometric sar-based deformation monitoring. *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[478] Xiang Wang, Yundong Li, Chen Lin, Yi Liu, and Shuo Geng. Building damage detection based on multi-source adversarial domain adaptation. *Journal of Applied Remote Sensing*, 15(3):036503, 2021.

[479] Xin Wang, Xuanmei Fan, Qiang Xu, and Peijun Du. Change detection-based co-seismic landslide mapping through extended morphological profiles and ensemble strategy. *ISPRS Journal of Photogrammetry and Remote Sensing*, 187:225–239, 2022.

[480] Xue Wang and Peijun Li. Extraction of urban building damage using spectral, height and corner information from vhr satellite images and airborne lidar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159:322–336, 2020.

[481] Yunlong Wang, Wenqing Feng, Kun Jiang, Qianchun Li, Ruipeng Lv, and Jihui Tu. Real-time damaged building region detection based on improved yolov5s and embedded system from uav images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:4205–4217, 2023.

[482] Zeyu Wang, Chuyi Wu, Feng Zhang, and Junshi Xia. Constructing an extensible building damage dataset via semi-supervised fine-tuning across 12 natural disasters. *Journal of Remote Sensing*, 5:0733, 2025.

[483] Ziming Wang and Ce Zhang. Adapting the segment anything model for sar-based flood detection using parameter-efficient

fine-tuning techniques. In *EGU General Assembly Conference Abstracts*, pages EGU25–3524, 2025.

[484] Prosper Washaya, Timo Balz, and Bahaa Mohamadi. Coherence change-detection with Sentinel-1 for natural and anthropogenic disaster monitoring in urban areas. *Remote Sensing*, 10(7):1026, June 2018.

[485] Manabu Watanabe, Takeshi Motohka, Yousuke Miyagi, Chinatsu Yonezawa, and Masanobu Shimada. Analysis of Urban Areas Affected by the 2011 Off the Pacific Coast of Tohoku Earthquake and Tsunami With L-Band SAR Full-Polarimetric Mode. *IEEE Geoscience and Remote Sensing Letters*, 9(3):472–476, 2012.

[486] Manabu Watanabe, Rajesh Bahadur Thapa, Tsuneo Ohsumi, Hiroyuki Fujiwara, Chinatsu Yonezawa, Naoya Tomii, and Sinichi Suzuki. Detection of damaged urban areas using interferometric SAR coherence change with PALSAR-2. *Earth, Planets and Space*, 68(1), July 2016.

[487] Ethan Weber and Hassan Kané. Building disaster damage assessment in satellite imagery with multi-temporal fusion. *arXiv preprint arXiv:2004.05525*, 2020.

[488] Yuxin Wei and Bin Liu. Cmfs-unet: A mamba-unet model for flood mapping in bitemporal and dual-polariziation sar imagery. In *2025 Photonics & Electromagnetics Research Symposium-Spring (PIERS-Spring)*, pages 1–7. IEEE, 2025.

[489] Yuxin Wei and Bin Liu. Mamba2-driven cnn for efficient flood mapping in sar imagery: A visual state space duality approach. In *IGARSS 2025 - 2025 IEEE International Geoscience and Remote Sensing Symposium*, pages 7124–7127, 2025.

[490] Marc Wieland, Wen Liu, and Fumio Yamazaki. Learning change from synthetic aperture radar images: Performance evaluation of a support vector machine to detect earthquake and tsunami-induced changes. *Remote Sensing*, 8(10), 2016.

[491] Sesa Wiguna, Bruno Adriano, Erick Mas, and Shunichi Koshimura. Evaluation of deep learning models for building damage mapping in emergency response settings. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:5651–5667, 2024.

[492] Sesa Wiguna, Bruno Adriano, Ruben Vescovo, Erick Mas, Ayumu Mizutani, and Shunichi Koshimura. Building damage mapping of the 2024 noto peninsula earthquake, japan, using semi-supervised learning and vhr optical imagery. *IEEE Geoscience and Remote Sensing Letters*, 21:1–5, 2024.

[493] World Bank. World development indicators: Gdp (current us dollars). https://data.worldbank.org/indicator/NY.GDP.MKTP.CD, 2026. Indicator code: NY.GDP.MKTP.CD, accessed 2026-04-06.

[494] World Meteorological Organization and United Nations Office for Disaster Risk Reduction. Global Status of Multi-Hazard Early Warning Systems 2023. Technical report, WMO and UNDRR, Geneva, 2023.

[495] Chen Wu, Bo Du, Xiaohui Cui, and Liangpei Zhang. A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion. *Remote Sensing of Environment*, 199:241–255, 2017.

[496] Fan Wu, Lixia Gong, Chao Wang, Hong Zhang, Bo Zhang, and Lei Xie. Signature analysis of building damage with terrasar-x new staring spotlight mode data. *IEEE Geoscience and Remote Sensing Letters*, 13(11):1696–1700, 2016.

[497] Han Wu, Huina Song, Jianhua Huang, Hua Zhong, Ronghui Zhan, Xuyang Teng, Zhaoyang Qiu, Meilin He, and Jiayi Cao. Flood detection in dual-polarization sar images based on multi-scale deeplab model. *Remote Sensing*, 14(20):5181, 2022.

[498] Kang Wu, Yingying Zhang, Lixiang Ru, Bo Dang, Jiangwei Lao, Lei Yu, Junwei Luo, Zifan Zhu, Yue Sun, Jiahao Zhang, Qi Zhu, Jian Wang, Ming Yang, Jingdong Chen, Yongjun Zhang, and Yansheng Li. A semantic-enhanced multi-modal remote sensing foundation model for earth observation. *Nature Machine Intelligence*, 7(8):1235–1249, August 2025.

[499] Lei Wu, Rui Liu, Nengpan Ju, Ao Zhang, Jingsong Gou, Guolei He, and Yuzhu Lei. Landslide mapping based on a hybrid cnn-transformer network and deep transfer learning using remote sensing images with topographic and spectral features. *International Journal of Applied Earth Observation and Geoinformation*, 126:103612, 2024.

[500] Yue Wu, Jiaheng Li, Yongzhe Yuan, A K Qin, Qi Guang Miao, and Mao Guo Gong. Commonality Autoencoder: Learning Common Features for Change Detection From Heterogeneous Images. *IEEE Transactions on Neural Networks and Learning Systems*, 33(9):4257–4270, 2022.

[501] Zherong Wu, Peifeng Ma, Yi Zheng, Feng Gu, Lin Liu, and Hui Lin. Automatic detection and classification of land subsidence in deltaic metropolitan areas using distributed scatterer insar and oriented r-cnn. *Remote Sensing of Environment*, 290:113545, 2023.

[502] Haobin Xia, Jianjun Wu, Jiaqi Yao, Hong Zhu, Adu Gong, Jianhua Yang, Liuru Hu, and Fan Mo. A deep learning application for building damage assessment using ultra-high-resolution remote sensing imagery in turkey earthquake. *International Journal of Disaster Risk Science*, 14(6):947–962, December 2023.

[503] Junshi Xia, Hongruixuan Chen, Clifford Broni-Bediako, Yimin Wei, Jian Song, and Naoto Yokoya. OpenEarthMap-SAR: A benchmark synthetic aperture radar dataset for global high-resolution land cover mapping. *arXiv preprint arXiv:2501.10891*, 2025.

[504] Zaishuo Xia, Zelin Li, Yanbing Bai, Jinze Yu, and Bruno Adriano. Self-supervised learning for building damage assessment from large-scale xbd satellite imagery benchmark datasets. In Christine Strauss, Alfredo Cuzzocrea, Gabriele Kotsis, A. Min Tjoa, and Ismail Khalil, editors, *Database and Expert Systems Applications*, pages 373–386, Cham, 2022. Springer International Publishing.

[505] Yiming Xiao and Ali Mostafavi. Damagecat: A deep learning transformer framework for typology-based post-disaster building damage categorization. *International Journal of Disaster Risk Reduction*, 128:105704, 2025.

[506] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. Segformer: simple and efficient design for semantic segmentation with transformers. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA, 2021. Curran Associates Inc.

[507] Ziyao Xing, Shuai Yang, Xuli Zan, Xinrui Dong, Yu Yao, Zhe Liu, and Xiaodong Zhang. Flood vulnerability assessment of urban buildings based on integrating high-resolution remote sensing and street view images. *Sustainable Cities and Society*, 92:104467, 2023.

[508] Zhitong Xiong, Yi Wang, Fahong Zhang, Adam J Stewart, Joëlle Hanna, Damian Borth, Ioannis Papoutsis, Bertrand Le Saux, Gustau Camps-Valls, and Xiao Xiang Zhu. Neural plasticity-inspired multimodal foundation model for earth observation. *arXiv preprint arXiv:2403.15356*, 2024.

[509] Haoyi Xiu, Xin Liu, Weimin Wang, Kyoung-Sook Kim, Takayuki Shinohara, Qiong Chang, and Masashi Matsuoka. Ds-net: A dedicated approach for collapsed building detection from post-event airborne point clouds. *International Journal of Applied Earth Observation and Geoinformation*, 116:103150, 2023.

[510] Haoyi Xiu, Takayuki Shinohara, Masashi Matsuoka, Munenari Inoguchi, Ken Kawabe, and Kei Horie. Collapsed building detection using 3d point clouds and deep learning. *Remote Sensing*, 12(24), 2020.

[511] Hanqiu Xu. Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery. *International journal of remote sensing*, 27(14):3025–3033, 2006.

[512] Joseph Z Xu, Wenhan Lu, Zebo Li, Pranav Khaitan, and Valeriya Zaytseva. Building damage detection in satellite imagery using convolutional neural networks. *arXiv preprint arXiv:1910.06444*, 2019.

[513] Weidong Xu, Martin J Wooster, Jiangping He, and Tianran Zhang. First study of sentinel-3 slstr active fire detection and frp retrieval: Night-time algorithm enhancements and global intercomparison to modis and viirs af products. *Remote Sensing of Environment*, 248:111947, 2020.

[514] Weidong Xu, Martin J Wooster, Takayuki Kaneko, Jiangping He, Tianran Zhang, and Daniel Fisher. Major advances in geostationary fire radiative power (frp) retrieval over asia and australia stemming from use of himarawi-8 ahi. *Remote Sensing of Environment*, 193:138–149, 2017.

[515] Yonghao Xu, Amanda Berg, and Leif Haglund. Sen2fire: A challenging benchmark dataset for wildfire detection using sentinel data. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 239–243, 2024.

[516] Yulin Xu, Chaojun Ouyang, Qingsong Xu, Dongpo Wang, Bo Zhao, and Yutao Luo. Cas landslide dataset: A large-scale and

multisensor dataset for deep learning-based landslide detection. *Scientific Data*, 11(1):12, 2024.

[517] Zongzhe Xu, Ritvik Gupta, Wenduo Cheng, Alexander Shen, Junhong Shen, Ameet Talwalkar, and Mikhail Khodak. Specialized foundation models struggle to beat supervised baselines. In *The Thirteenth International Conference on Learning Representations*, 2025.

[518] Zhuoqun Xue, Xiaojian Zhang, David O Prevatt, Jennifer Bridge, Susu Xu, and Xilei Zhao. Post-hurricane building damage assessment using street-view imagery and structured data: A multimodal deep learning approach. *arXiv preprint arXiv:2404.07399*, 2024.

[519] Ritu Yadav, Andrea Nascetti, Hossein Azizpour, and Yifang Ban. Self-supervised contrastive model for flood mapping and monitoring on sar time-series. In *EGU General Assembly Conference Abstracts*, pages EGU–14375, 2023.

[520] Ritu Yadav, Andrea Nascetti, Hossein Azizpour, and Yifang Ban. Unsupervised flood detection on sar time series using variational autoencoder. *International Journal of Applied Earth Observation and Geoinformation*, 126:103635, 2024.

[521] Ritu Yadav, Andrea Nascetti, and Yifang Ban. Attentive dual stream siamese u-net for flood detection on multi-temporal sentinel-1 data. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 5222–5225. IEEE, 2022.

[522] Yoshio Yamaguchi. Disaster monitoring by fully polarimetric sar data acquired with alos-palsar. *Proceedings of the IEEE*, 100(10):2851–2860, 2012.

[523] Fumio Yamazaki, T. Thuy Vu, and Masashi Matsuoka. Context-based detection of post-disaster damaged buildings in urban areas from satellite images. In *2007 Urban Remote Sensing Joint Event*, pages 1–5, 2007.

[524] Heng Yan, Ailong Ma, Hong Shu, Yuting Wan, Liangpei Zhang, and Yanfei Zhong. Ultra-high-resolution sar and optical image registration: From global benchmark dataset to frequency-guided registration method. *ISPRS Journal of Photogrammetry and Remote Sensing*, 235:190–210, 2026.

[525] Chuan Yang, Yueqin Zhu, Jiantong Zhang, Xiaoqiang Wei, Haomeng Zhu, and Zhehui Zhu. A feature fusion method on landslide identification in remote sensing with segment anything model. *Landslides*, 22(2):471–483, 2025.

[526] Songxi Yang, Bo Peng, Tang Sui, Meiliu Wu, and Qunying Huang. Advancing self-supervised learning for building change detection and damage assessment: Unified denoising autoencoder and contrastive learning framework. *Remote Sensing*, 17(15), 2025.

[527] Wanting Yang, Xianfeng Zhang, and Peng Luo. Transferability of convolutional neural network models for identifying damaged buildings due to earthquake. *Remote Sensing*, 13(3), 2021.

[528] Wentao Yang, Ming Wang, and Peijun Shi. Using modis ndvi time series to identify geographic patterns of landslides in vegetated regions. *IEEE Geoscience and Remote Sensing Letters*, 10(4):707–710, 2012.

[529] Yanchen Yang, Chou Xie, Bangsen Tian, Yihong Guo, Yu Zhu, Shuaichen Bian, Ying Yang, Ming Zhang, and Yimin Ruan. Advanced Post-earthquake Building Damage Assessment: SAR Coherence Time Matrix with Vision Transformer. *International Journal of Applied Earth Observation and Geoinformation*, 133:104133, 2024.

[530] Yanchen Yang, Chou Xie, Bangsen Tian, Yihong Guo, Yu Zhu, Ying Yang, Haoran Fang, Shuaichen Bian, and Ming Zhang. Large-scale building damage assessment based on recurrent neural networks using SAR coherence time series: A case study of 2023 Turkey–Syria earthquake. *Earthquake Spectra*, 40(4):2285–2305, July 2024.

[531] Yanchen Yang, Chou Xie, Bangsen Tian, Yihong Guo, Yu Zhu, Ying Yang, Haoran Fang, Shuaichen Bian, and Ming Zhang. Large-scale building damage assessment based on recurrent neural networks using sar coherence time series: A case study of 2023 turkey–syria earthquake. *Earthquake Spectra*, 40(4):2285–2305, 2024.

[532] Yifan Yang, Lei Zou, Bing Zhou, Daoyang Li, Binbin Lin, Joynal Abedin, and Mingzheng Yang. Hyperlocal disaster damage assessment using bi-temporal street-view imagery and pre-trained vision models. *Computers, Environment and Urban Systems*, 121:102335, 2025.

[533] Ting Yao, Yiheng Zhang, Zhaofan Qiu, Yingwei Pan, and Tao Mei. Seco: Exploring sequence supervision for unsupervised representation learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12):10656–10664, May 2021.

[534] Xin Ye, Mingchao Liu, Jun Wang, Qiming Qin, Huazhong Ren, Jianhua Wang, and Jian Hui. Building-based damage detection from postquake image using multiple-feature analysis. *IEEE Geoscience and Remote Sensing Letters*, 14(4):499–503, 2017.

[535] Naoto Yokoya, Kazuki Yamanoi, Wei He, Gerald Baier, Bruno Adriano, Hiroyuki Miura, and Satoru Oishi. Breaking the limits of remote sensing by simulation and deep learning for flood and debris flow mapping. *arXiv preprint arXiv:2006.05180*, 2020.

[536] Chen Yu, Bin Hu, Xiuchuan Cheng, Guangqiang Yin, and Zhiguo Wang. Remote sensing building damage assessment with a multihead neighbourhood attention transformer. *International Journal of Remote Sensing*, 44(16):5069–5100, 2023.

[537] Zhengbo Yu, Zhe Chen, Zhongchang Sun, Huadong Guo, Bo Leng, Ziqiong He, Jinpei Yang, and Shuwen Xing. Segdetector: A deep learning model for detecting small and overlapping damaged buildings in satellite images. *Remote Sensing*, 14(23), 2022.

[538] Qian Yuan and Zhijuan Du. Dcibcd: A dual-branch cooperative interaction method for interference-resistant binary change detection. *Expert Systems with Applications*, 301:130210, 2026.

[539] Sang-Ho Yun, Kenneth Hudnut, Susan Owen, Frank Webb, Mark Simons, Patrizia Sacco, Eric Gurrola, Gerald Manipon, Cunren Liang, Eric Fielding, Pietro Milillo, Hook Hua, and Alessandro Coletta. Rapid damage mapping for the 2015 M w 7.8 Gorkha earthquake using synthetic aperture radar data from COSMO–SkyMed and ALOS-2 satellites. *Seismological Research Letters*, 86(6):1549–1556, October 2015.

[540] Vivien Zahs, Katharina Anders, Julia Kohns, Alexander Stark, and Bernhard Höfle. Classification of structural building damage grades from multi-temporal photogrammetric point clouds using a machine learning model trained on virtual laser scanning data. *International Journal of Applied Earth Observation and Geoinformation*, 122:103406, 2023.

[541] Rufai Yusuf Zakari, Owais Ahmed Malik, and Ong Wee-Hong. An enhanced wildfire spread prediction using multimodal satellite imagery and deep learning models. *Remote Sensing Applications: Society and Environment*, page 101632, 2025.

[542] Xiangqiang Zeng and Ying Qu. Building damage mapping through heterogeneous feature consistency and knowledge integration. In *IGARSS 2025 - 2025 IEEE International Geoscience and Remote Sensing Symposium*, pages 233–236, 2025.

[543] Wei Zhai and Chunlin Huang. Fast building damage mapping using a single post-earthquake PolSAR image: a case study of the 2010 yushu earthquake. *Earth, Planets and Space*, 68(1):86, 2016.

[544] Wei Zhai and Zhong-Ren Peng. Damage assessment using google street view: Evidence from hurricane michael in mexico beach, florida. *Applied Geography*, 123:102252, 2020.

[545] Wei Zhai, Jingfa Zhang, Xiulai Xiao, Jinghan Wang, Haoran Zhang, Xinxin Yin, and Zhen Wu. Damaged building extraction from post-earthquake polsar data based on the fourier transform. *Remote Sensing Letters*, 12(6):594–603, 2021.

[546] Yihao Zhan, Wen Liu, and Yoshihisa Maruyama. Damaged building extraction using modified mask r-cnn model using post-event aerial images of the 2016 kumamoto earthquake. *Remote Sensing*, 14(4), 2022.

[547] Chenxiao Zhang, Peng Yue, Deodato Tapete, Liangcun Jiang, Boyi Shangguan, Li Huang, and Guangchao Liu. A deeply supervised image fusion network for change detection in high resolution bi-temporal remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:183–200, 2020.

[548] Haiming Zhang, Guorui Ma, Di Wang, and Yongxian Zhang. M3icnet: A cross-modal resolution preserving building damage detection method with optical and sar remote sensing imagery and two heterogeneous image disaster datasets. *ISPRS Journal of Photogrammetry and Remote Sensing*, 221:224–250, 2025.

[549] Haiming Zhang, Mingchang Wang, Yongxian Zhang, and Guorui Ma. Tda-net: A novel transfer deep attention network for rapid response to building damage discovery. *Remote Sensing*, 14(15), 2022.

[550] Haiming Zhang, Yongxian Zhang, Di Wang, and Guorui Ma. Damaged building object detection from bitemporal remote sensing imagery: A cross-task integration network and five datasets. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–27, 2024.

[551] Hao Zhang, Jin-Jian Xu, Hong-Wei Cui, Lin Li, Yaowen Yang, Chao-Sheng Tang, and Niklas Boers. When geoscience meets

foundation models: Toward a general geoscience artificial intelligence system. *IEEE geoscience and remote sensing magazine*, 2024.

[552] Jianli Zhang, Liwei Tao, Wenbo Wei, Pengfei Ma, and Mengdi Shi. Cmsnet: A sam-enhanced cnn–mamba framework for damaged building change detection in remote sensing imagery. *Remote Sensing*, 17(23), 2025.

[553] Jing-Fa Zhang, Li li Xie, and Xia xin Tao. Change detection of earthquake-damaged buildings on remote sensing image and its application in seismic disaster assessment. In *2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings*, volume 4, pages 2436–2438 vol.4, 2003.

[554] Limeng Zhang, Zenghui Zhang, Tao Zhang, Gui Gao, and Wenxian Yu. Cdprompt: Multimodal change detection with in-domain prompt in missing modality scenarios. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–18, 2025.

[555] Ming Zhang, Chou Xie, Bangsen Tian, Yanchen Yang, Yihong Guo, Yu Zhu, and Shuaichen Bian. Application of gated recurrent unit neural network for flood extraction from synthetic aperture radar time series. *Water*, 15(21):3779, 2023.

[556] Puzhao Zhang, Yifang Ban, and Andrea Nascetti. Learning u-net without forgetting for near real-time wildfire monitoring by the fusion of sar and optical time series. *Remote Sensing of Environment*, 261:112467, 2021.

[557] Puzhao Zhang, Xikun Hu, Yifang Ban, Andrea Nascetti, and Maoguo Gong. Assessing sentinel-2, sentinel-1, and alos-2 palsar-2 data for large-scale wildfire-burned area mapping: Insights from the 2017–2019 canada wildfires. *Remote Sensing*, 16(3):556, 2024.

[558] Xiaokang Zhang, Weikang Yu, Man-On Pun, and Wenzhong Shi. Cross-domain landslide mapping from large-scale remote sensing images using prototype-guided domain-aware progressive representation learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 197:1–17, 2023.

[559] Yuchen Zhang, Mingsheng Long, Kaiyuan Chen, Lanxiang Xing, Ronghua Jin, Michael I. Jordan, and Jianmin Wang. Skilful nowcasting of extreme precipitation with NowcastNet. *Nature*, 619(7970):526–532, 2023.

[560] Yuwei Zhang, Peng Liu, Lajiao Chen, Mengzhen Xu, Xingyan Guo, and Lingjun Zhao. A new multi-source remote sensing image sample dataset with high resolution for flood area extraction: Gf-floodnet. *International Journal of Digital Earth*, 16(1):2522–2554, 2023.

[561] Zhijie Zhang, Jonathan Giezendanner, Rohit Mukherjee, Beth Tellman, Alexander Melancon, Matt Purri, Iksha Gurung, Upmanu Lall, Kobus Barnard, and Andrew Molthan. Assessing inundation semantic segmentation models trained on high-versus low-resolution labels using floodplanet, a manually labeled multi-sourced high-resolution flood dataset. *Journal of Remote Sensing*, 5:0575, 2025.

[562] Binru Zhao, Qiang Dai, Lu Zhuo, Shaonan Zhu, Qi Shen, and Dawei Han. Assessing the potential of different satellite soil moisture products in landslide hazard assessment. *Remote Sensing of Environment*, 264:112583, 2021.

[563] Bofei Zhao, Haigang Sui, and Junyi Liu. Siam-dwenet: Flood inundation detection for sar imagery using a cross-task transfer siamese network. *International Journal of Applied Earth Observation and Geoinformation*, 116:103132, 2023.

[564] Fei Zhao, Chengcui Zhang, Runlin Zhang, and Tianyang Wang. Visual prompt learning of foundation models for post-disaster damage evaluation. *Remote Sensing*, 17(10), 2025.

[565] Jie Zhao, Ming Li, Yu Li, Patrick Matgen, and Marco Chini. Urban flood mapping using satellite synthetic aperture radar data: A review of characteristics, approaches, and datasets. *IEEE Geoscience and Remote Sensing Magazine*, 2024.

[566] Jie Zhao, Yu Li, Patrick Matgen, Ramona Pelich, Renaud Hostache, Wolfgang Wagner, and Marco Chini. Urban-aware u-net for large-scale urban flood mapping using multitemporal sentinel-1 intensity and interferometric coherence. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–21, 2022.

[567] Jie Zhao, Zhitong Xiong, and Xiao Xiang Zhu. Urbansarfloods: Sentinel-1 slc-based benchmark dataset for urban and open-area flood mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 419–429, 2024.

[568] Lingli Zhao, Jie Yang, Pingxiang Li, Liangpei Zhang, Lei Shi, and Fengkai Lang. Damage assessment in urban areas using post-earthquake airborne polsar imagery. *International Journal of Remote Sensing*, 34(24):8952–8966, 2013.

[569] Zhuo Zheng, Yanfei Zhong, Zijing Wan, Liangpei Zhang, and Stefano Ermon. Neural disaster simulation for transferable building damage assessment. *Remote Sensing of Environment*, 331:114979, 2025.

[570] Zhuo Zheng, Yanfei Zhong, Junjue Wang, and Ailong Ma. Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4095–4104, 2020.

[571] Zhuo Zheng, Yanfei Zhong, Junjue Wang, Ailong Ma, and Liangpei Zhang. Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters. *Remote Sensing of Environment*, 265:112636, 2021.

[572] Zhuo Zheng, Yanfei Zhong, Junjue Wang, Ailong Ma, and Liangpei Zhang. Farseg++: Foreground-aware relation network for geospatial object segmentation in high spatial resolution remote sensing imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11):13715–13729, 2023.

[573] Zhuo Zheng, Yanfei Zhong, Liangpei Zhang, Marshall Burke, David B. Lobell, and Stefano Ermon. Towards transferable building damage assessment via unsupervised single-temporal change adaptation. *Remote Sensing of Environment*, 315:114416, 2024.

[574] Zhuo Zheng, Yanfei Zhong, Ji Zhao, Ailong Ma, and Liangpei Zhang. Unifying remote sensing change detection via deep probabilistic change models: From principles, models to applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 215:239–255, 2024.

[575] Zixiang Zhou, Jie Gong, and Xuan Hu. Community-scale multi-level post-hurricane damage assessment of residential buildings using multi-temporal airborne lidar data. *Automation in Construction*, 98:30–45, 2019.

[576] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867, 2020.

[577] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017.

[578] Xiaoyu Zhu, Junwei Liang, and Alexander Hauptmann. MSNet: A Multilevel Instance Segmentation Network for Natural Disaster Damage Assessment in Aerial Videos. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2022–2031, 2021.

[579] Zhe Zhu, Shixiong Wang, and Curtis E. Woodcock. Improvement and expansion of the fmask algorithm: cloud, cloud shadow, and snow detection for landsats 4–7, 8, and sentinel 2 images. *Remote Sensing of Environment*, 159:269–277, 2015.

[580] Zhe Zhu and Curtis E. Woodcock. Object-based cloud and cloud shadow detection in landsat imagery. *Remote Sensing of Environment*, 118:83–94, 2012.

[581] Zhe Zhu, Junxue Zhang, Zhiqiang Yang, Amal H. Aljaddani, Warren B. Cohen, Shi Qiu, and Congliang Zhou. Continuous monitoring of land disturbance based on landsat time series. *Remote Sensing of Environment*, 238:111116, 2020. Time Series Analysis with High Spatial Resolution Imagery.

[582] Zhifu Zhu, Xiping Yuan, Shu Gan, Raobo Li, Weidong Luo, Rui Bi, and Cheng Chen. Texture and structure interaction guided generative adversarial network for multimodal remote sensing image change detection. *GIScience & Remote Sensing*, 62(1):2565866, 2025.