A Deep Learning Framework for Rapid Building Damage Detection through Multimodal Data Fusion: Application to the 2025 Myanmar Earthquake

Luigi Russo[1], Deodato Tapete[2], Silvia Liberata Ullo[3], Paolo Gamba[1]

[1]Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy
luigi.russo02@universitadipavia.it; paolo.gamba@unipv.it
[2]Italian Space Agency (ASI), Rome, Italy – deodato.tapete@asi.it
[3]Department of Engineering, University of Sannio, Benevento, Italy – ullo@unisannio.it

Keywords: SAR, optical imagery, building damage, earthquake, deep learning.

Abstract

Rapid and reliable assessment of building damage after major earthquakes is essential for effective emergency response and recovery planning. This study formulates post-disaster building damage detection (BDD) as a binary image classification task (damaged vs. undamaged buildings) using multimodal satellite data and a unified ResNet-18 backbone to enable a controlled comparison of fusion strategies.
The analysis focuses on the M_w 7.7 Myanmar earthquake of 28 March 2025 and integrates post-event COSMO-SkyMed Second Generation (CSG) dual-polarization (HH, HV) SAR imagery, Maxar optical data, OpenStreetMap (OSM) building footprints, and UNOSAT damage annotations. Three fusion paradigms are evaluated: Early Fusion (EF), Late Fusion (LF), and a novel Middle Fusion (MF) approach.
The proposed MF framework introduces a Footprint-Guided Cross-Attention (FGCA) mechanism that uses building geometry as a spatial prior to guide feature-level interaction between SAR and optical representations. Five-fold cross-validation results show that MF consistently outperforms EF and LF, achieving higher precision, F1-score, and robustness across modality configurations.
By jointly exploiting SAR structural sensitivity, optical detail, and footprint-based spatial context, the proposed Footprint-Guided Middle Fusion (FGMF) framework enables accurate and scalable building damage mapping from heterogeneous Earth Observation (EO) data.

1. Introduction

Rapid and reliable assessment of building damage after major earthquakes is critical for emergency response and recovery planning. Traditional ground-based inspections, although accurate, are too slow and resource-intensive to support timely decision-making at this scale.
Remote sensing (RS) data have therefore become central to post-disaster mapping, offering wide-area coverage, timely acquisitions, and the ability to operate even in inaccessible regions (Dell'Acqua and Gamba, 2012).
Among the most widely used modalities, Synthetic Aperture Radar (SAR) and optical imagery offer complementary information. SAR is unaffected by lighting or weather conditions and captures changes in structural scattering, while optical imagery provides very-high-resolution (VHR) spectral and textural details that are more intuitive to interpret. In parallel, building footprints from sources like OpenStreetMap (OSM) provide useful geometric priors that help to focus the analysis on relevant structures and reduce background noise.
In this work, post-disaster building damage detection (BDD) is formulated as a binary image classification task at the building level (damaged vs. undamaged structures), implemented using a ResNet-18 convolutional neural network (CNN) backbone. Within this framework, the choice of input modalities and fusion strategies plays a crucial role in overall performance.
SAR-based damage detection has shown promising results, but also limitations when applied in isolation. Coherence-based approaches (ElGharbawi and Zarzoura, 2021; Ge et al., 2020) are widely used to detect damage, yet they face challenges in dense urban environments due to variations in scattering and viewing geometry (Rao et al., 2023). Optical imagery is also widely adopted for damage mapping, but remains sensitive to cloud cover and acquisition conditions. To overcome these limitations, multimodal fusion approaches have gained increasing attention. Brunner et al. (2010) provided one of the earliest demonstrations of fusing VHR SAR and optical imagery for earthquake damage assessment, showing clear improvements over single-sensor approaches. More recently, Serifoglu Yilmaz et al. (2023) combined Sentinel-1 SAR and Sentinel-2 optical imagery to detect building damage from the 2023 Kahramanmaraş earthquakes, reporting overall accuracies exceeding 75%. In parallel, large-scale multimodal benchmarks have been introduced to support reproducible research. Notably, Sun et al. (2024) released the *QuickQuakeBuildings* dataset, which adopts a deep learning (DL) approach that combines SAR and optical data in a late fusion (LF) setting for rapid BDD. The dataset highlights both the potential and the limitations of decision-level fusion, where modality-specific predictions are integrated only at the final stage.
Fusion strategies are typically categorized as early (EF), middle (MF), and late (LF) (Jiao et al., 2024). EF concatenates raw inputs, potentially diluting modality-specific information, while LF integrates modality-specific predictions only at the decision stage (Sun et al., 2024). MF enables feature-level interaction and has shown stronger potential. Recent works have explored homogeneous feature fusion (Jiang et al., 2020) and building-guided cross-modal learning (Li et al., 2025), highlighting the importance of object-level priors in multimodal integration.
A key operational challenge in damage mapping is the reliance on both pre- and post-event data, often requiring cloud-free and temporally aligned imagery that may be unavailable in emer-

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-927-2026

gency scenarios (Dong and Shan, 2013). Such dependency limits rapid deployment. In contrast, approaches based solely on post-event data offer a more practical alternative (Macchiarulo et al., 2025; Li et al., 2024), especially when combined with spatial priors such as building footprints.

In light of these developments, the growing use of deep learning (DL) has significantly advanced large-scale building damage mapping from remote sensing (RS) data. A recent review by Wang et al. (2024) surveyed 242 studies and highlighted persistent challenges, including class imbalance, limited generalization across events, and poor interpretability. Several DL architectures specifically tailored for building damage assessment (BDA) have been proposed, often exploiting pre- and post-disaster imagery to capture structural changes. Notably, Shen et al. (2022) introduced BDANet, a multiscale convolutional neural network (CNN) with cross-directional attention designed to fuse paired satellite images, while Xia et al. (2023) applied a BDANet-based framework to ultra-high-resolution optical data for the 2023 Türkiye earthquake, demonstrating its practical applicability in real emergency-response settings. However, generalization across diverse disaster scenarios remains difficult. Domain adaptation methods (Hertel et al., 2025) are emerging as promising strategies, while studies using VHR optical data (Ainscoe et al., 2025) report improved performance but also underline the dependence on cloud-free acquisitions. Similar limitations have been observed across multiple test sites, where deep models trained on specific events tend to lose accuracy when applied to different regions (Wiguna et al., 2024).

From this perspective, several research gaps remain: (i) many existing methods underexploit cross-modal relationships by fusing either too early or too late, (ii) building footprints are rarely employed as active guidance mechanisms, being used mostly as spatial masks rather than drivers of the fusion process, and (iii) generalization across different events remains a major challenge. In this study, these gaps are addressed by introducing a Footprint-Guided Middle Fusion (FGMF) framework for post-earthquake BDD that synergistically combines high-resolution (HR) SAR, optical, and building footprint data. The core innovation lies in the Footprint-Guided Cross-Attention (FGCA) mechanism, which explicitly leverages building geometry to guide feature-level interactions between SAR and optical modalities. By doing so, the model enhances the contribution of complementary information while suppressing background interference, enabling more accurate and interpretable damage mapping.

Unlike conventional approaches that depend on pre- and post-event optical pairs, the proposed framework relies solely on post-event acquisitions complemented by pre-existing building footprint data, allowing for rapid deployment immediately after an earthquake. This design leverages the distinct advantages of each data source: SAR provides structural sensitivity, optical imagery contributes visual interpretability, and footprint data offers spatial context, resulting in a robust and operationally scalable post-disaster assessment.

The remainder of the paper is structured as follows: Section 2 introduces the study area and datasets; Section 3 describes the proposed methodology and fusion strategies; Section 4 presents the experimental results and evaluation; and Section 5 discusses the final remarks and potential extensions of the framework for operational earthquake damage mapping.

2. Study Area and Data

This study focuses on the Mandalay Region in central Myanmar, traversed by the right-lateral *Sagaing Fault*, the country's principal plate-boundary structure accommodating India–Sunda relative motion at rates of approximately $\sim$ 18-20 $\mathrm{mm\,yr^{-1}}$ (Socquet et al., 2006).

On 28 March 2025 at 12:50 MMT, a M_w 7.7 earthquake struck near Mandalay, exhibiting right-lateral strike-slip faulting consistent with rupture along the central segment of the Sagaing Fault. A second M_w 6.7 event occurred shortly after. As of 25 April 2025, at least 3,800 fatalities, 5,100 injuries, and 116 missing persons were reported nationwide, with total damages and losses estimated at approximately USD 11 billion and more than 17 million people affected (United Nations, 2025; World Bank, 2025). Emergency mapping was activated through the Copernicus Emergency Management Service (EMSR798)[1], with complementary UNOSAT[2] analyses supporting rapid damage assessment. Within the AOI defined in this study, Mandalay and Sagaing experienced particularly high concentrations of damage and are therefore selected as representative urban case studies. Figure 1 delineates the AOI and highlights these two heavily affected sub-regions.

2.1 Satellite Data

The **COSMO-SkyMed Second Generation (CSG)** constellation, developed and operated by the **Italian Space Agency (ASI)**, represents an advanced VHR SAR system building upon the heritage of the first-generation COSMO-SkyMed (CSK) mission (Italian Space Agency, 2021). CSG provides enhanced radiometric stability, multi-polarization capabilities, and flexible acquisition modes optimised for rapid response to emergencies. Within ASI's operational framework, a dedicated disaster acquisition plan ensures priority tasking and accelerated data delivery following major events, supporting international response initiatives such as the *Copernicus Emergency Management Services (CEMS)* and the *Committee on Earth Observation Satellite (CEOS) Working Group on Disasters*. In addition, systematic background acquisitions over hazard-prone regions enable immediate access to pre-event imagery for post-disaster assessments. This strategy has been successfully applied during recent crises, including the 2023 Türkiye earthquake sequence.

For the 28 March 2025 M_w 7.7 Myanmar earthquake, a dual-polarization (HH, HV) SAR acquisition from CSG was collected on 30 March 2025 by the *CSG-SSAR1* sensor in *StripMap HIMAGE* mode during a descending pass. The native ground resolution is approximately 3 m $\times$ 3 m, providing X-band backscatter with consistent co- and cross-polarization geometry. Level-1A products were processed to GTC (Geocoded Terrain Corrected) Level-1D.

Preprocessing was conducted in ESA SNAP and included radiometric calibration of HH and HV intensity layers and Range–Doppler Terrain Correction using the Copernicus 30 m Digital Elevation Model (DEM). Pixels lacking valid DEM elevation values were masked to avoid geometric artifacts.

Complementary optical data were obtained from the **Maxar Open Data Program**[3], which provides VHR satellite imagery for humanitarian response via the *MGP Xpress*[4] platform. Post-event imagery acquired on 31 March 2025 was used, providing orthorectified multispectral data with effective ground sampling distances of approximately 0.5 m (panchromatic) and 1.2 m (multispectral). For computational efficiency, only the RGB bands were retained.

[1] https://emergency.copernicus.eu/mapping/list-of-components/EMSR798
[2] https://unosat.org/products/4093
[3] https://registry.opendata.aws/maxar-open-data/
[4] https://xpress.maxar.com/

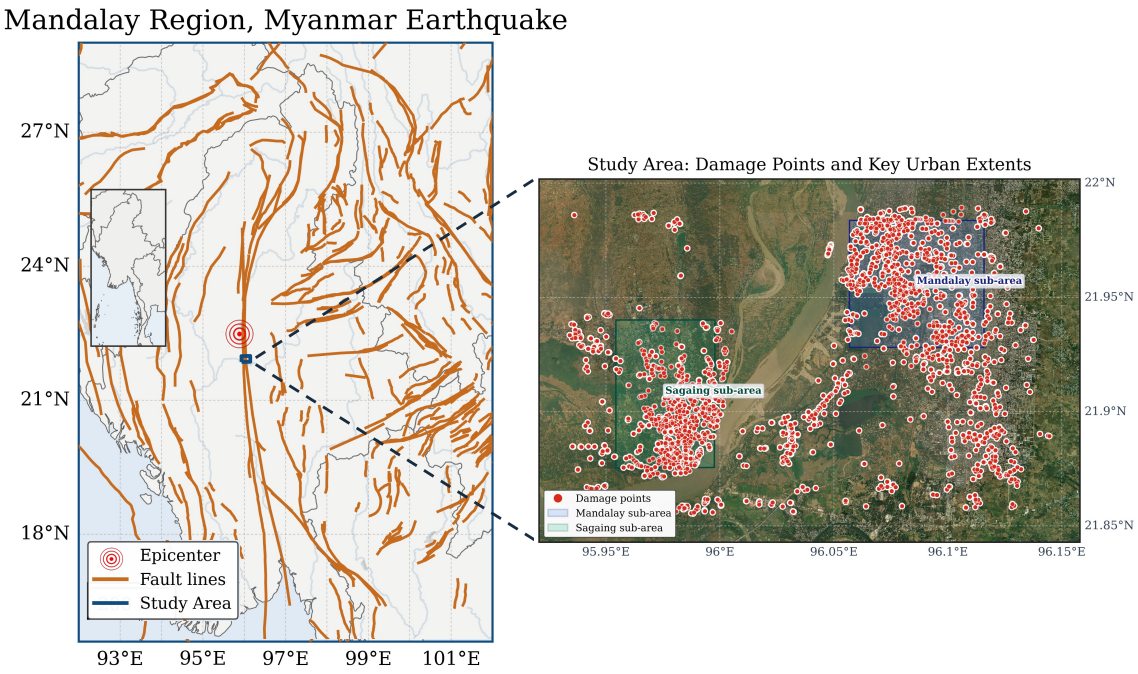

Figure 1. Regional seismic context and detailed view of the study area affected by the 28 March 2025 Myanmar earthquake. The epicenter and active faults define the tectonic setting of central Myanmar. The zoomed view on the right shows a VHR post-event optical basemap with damage points (magenta) and highlights the AOI, including the heavily affected urban areas of Mandalay and Sagaing.

Provider	Date	Spatial Resolution	Processing Level	Acquisition Mode	Spectral / Polarimetric Content	Coverage Area
CSG	30 Mar 2025	∼2 m (after resampling)	L1A (cal.) → L1D	StripMap HIMAGE	X–band (HH, HV)	∼318.47 km^2
Maxar ODP	31 Mar 2025	∼2 m (after resampling)	ARD (orthorectified mosaic)	–	RGB (3 bands)	∼318.47 km^2

Table 1. Summary of the characteristics of the satellite datasets employed in this study after preprocessing.

SAR and optical mosaics were co-registered in QGIS and resampled to a common 2 m grid to ensure spatial alignment between modalities. A summary of the satellite datasets after preprocessing is reported in Table 1.

2.2 Reference and Ancillary Data

Building damage annotations were obtained from the **UNITAR/UNOSAT** Rapid Mapping service, which provides geospatial damage assessments derived from VHR satellite imagery. As part of the international response to the March 2025 Myanmar earthquake, *UNOSAT* analysts manually interpreted post-event imagery to identify damaged structures across affected areas.

To ensure high label reliability and focus on cases of confirmed structural damage, only annotations corresponding to the highest severity level in the dataset were retained. These annotations, provided as point features representing the centroid of affected structures, were used as positive samples for model training and evaluation.

To spatially link the damage points with the corresponding building structures, polygonal building footprints were extracted from **OSM** data using the Myanmar country extract provided by **Geofabrik**[5]. These vector footprints delineate individual buildings and enable object-level localization of the mapped damage evidence.

All geospatial layers were reprojected to a common coordinate system to ensure geometric consistency with satellite imagery. A spatial intersection was then performed between the damage points and the building polygons: only those footprints intersecting at least one high-confidence damage annotation were retained as positive samples. Finally, all geometries were filtered and clipped to the spatial extent covered by the available post-event satellite imagery, ensuring alignment between reference data and the input features used in our experiments.

2.3 Dataset Preparation

Building upon the preprocessed and co-registered SAR and optical mosaics described in Section 2.1, a georeferenced patch-level dataset was constructed to support building-level damage detection analysis. The preparation pipeline integrates multimodal imagery with vector building footprints, ensuring spatially aligned samples with consistent geographic context across all modalities.

Post-event acquisitions from CSG (HH/HV) and Maxar ODP (RGB), both resampled to a common GSD of 2 m, were combined with OSM building polygons. The **UNOSAT** damage points served as reference indicators of affected structures. Each building footprint retained through the above-described intersection process was labeled as *damaged*, while all remaining buildings were assigned to the *intact* class. This spatial association yielded a total of 1,729 damaged buildings within the study area.

For each labeled building, a square patch was extracted and centered on the footprint centroid, covering a physical extent of 80×80 m (i.e., 40×40 pixels at 2 m GSD). The patch size captures both the target building and its immediate spatial context. The SAR–optical component of each sample was derived from the mosaics described in Section 2.1. Specifically, a two-channel

[5] `https://download.geofabrik.de/asia/myanmar.html`

Ancillary datasets		Class distribution (AOI)		
Source	**Data summary**	**Class**	**Building Count**	**Sampling ratio**
UNOSAT	Damage points from VHR image interpretation	Damaged	1,729	1
OSM (Geofabrik, Myanmar)	Building footprints (polygon geometries)	Intact (sampled)	34,580	20
		Total	**36,309**	–

Table 2. Summary of ancillary geospatial datasets and corresponding class distribution after dataset construction within the study area.

SAR stack (HH/HV) from the CSG acquisition and the corresponding RGB optical patch from the Maxar composite were extracted over each building footprint. In parallel, a one-channel footprint mask was generated by rasterizing each building polygon onto the SAR grid, providing a spatial reference precisely aligned with the satellite inputs. This three-modal configuration (SAR, optical, footprint) constituted the foundation of the dataset used for multimodal fusion and subsequent classification of damaged structures.

To mitigate the strong class imbalance typical of post-disaster contexts, a fixed-ratio sampling strategy was applied, whereby intact buildings were randomly selected in a $20:1$ ratio relative to damaged instances. This preserves the natural post-disaster damage class asymmetry.

The resulting dataset is geometrically consistent and suitable for reproducible multimodal learning under the fusion strategies detailed in Section 3. Table 2 summarizes the dataset composition and ancillary data sources.

3. Methodology

This study investigates three multimodal fusion strategies for BDD, leveraging multimodal Earth Observation data and auxiliary geospatial layers. The input to each model includes:

- SAR dual-polarization backscatter: $\mathbf{X}_{SAR} \in \mathbb{R}^{P\times P\times 3}$, including HH, HV, and the ratio HH/HV;
- Optical imagery: $\mathbf{X}_{OPT} \in \mathbb{R}^{P\times P\times 3}$ (RGB channels from Maxar ODP);
- Building footprint mask: $\mathbf{X}_{FTP} \in \{0,1\}^{P\times P}$, binarized and replicated across 3 channels for compatibility with the *ResNet-18* backbone.

Each input patch has spatial dimensions $P \times P$, where $P = 40$, corresponding to an 80 m $\times$ 80 m area at 2 m ground sampling distance. Prior to being fed into the CNN, patches are bilinearly resampled to 224×224 pixels to ensure compatibility with the ResNet-18 backbone architecture and its standard receptive field configuration.

3.1 Early Fusion

In the EF strategy, all the above described input modalities are concatenated along the channel dimension at the input level, forming a unified tensor that captures raw cross-modal information before any feature extraction process.

The resulting 9-channel input tensor is as follows:

$$\mathbf{f}'_{EF} = \mathbf{X}_{SAR} \,\|\, \mathbf{X}_{OPT} \,\|\, \mathbf{X}_{FTP} \in \mathbb{R}^{P\times P\times 9}, \tag{1}$$

where $\|$ denotes channel-wise concatenation.

This fused tensor is processed by a modified *ResNet-18* backbone (noted as E). The extracted features are then passed through a two-layer multilayer perceptron (MLP) classifier with *ReLU* activation function and dropout:

$$\hat{z}_{EF} = \mathrm{MLP}\big(\mathrm{GAP}(E(\mathbf{f}_{EF}))\big), \qquad \hat{p}_{EF} = \sigma(\hat{z}_{EF}), \tag{2}$$

where GAP denotes the global average pooling operator and σ is the sigmoid activation function. The workflow of this configuration is shown in Figure 2.

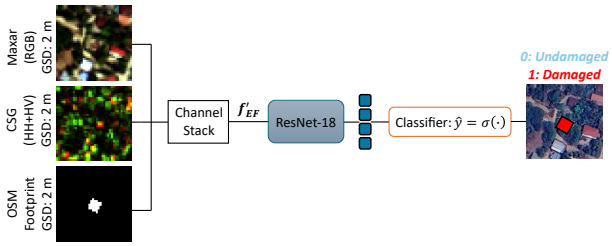

Figure 2. EF architecture: concatenation of SAR (HH, HV, HH/HV), RGB, and footprint mask (replicated over 3 channels) processed by a single *ResNet-18* backbone.

3.2 Late Fusion

In the LF strategy, each modality is processed independently through a dedicated *ResNet-18* encoder, and fusion is performed at the representation level prior to final classification.

Given the modality-specific inputs $\mathbf{X}_{SAR}$, $\mathbf{X}_{OPT}$, and $\mathbf{X}_{FTP}$, each branch applies a full *ResNet-18* backbone followed by GAP and a linear projection (W) to obtain a modality-specific embedding vector of dimension d:

$$\mathbf{f}'_{SAR} = W_{SAR}\,\mathrm{GAP}(E_{SAR}(\mathbf{X}_{SAR}))\,, \tag{3}$$
$$\mathbf{f}'_{OPT} = W_{OPT}\,\mathrm{GAP}(E_{OPT}(\mathbf{X}_{OPT}))\,, \tag{4}$$
$$\mathbf{f}'_{FTP} = W_{FTP}\,\mathrm{GAP}(E_{FTP}(\mathbf{X}_{FTP}))\,, \tag{5}$$

where $\mathbf{f}'_m \in \mathbb{R}^d$ and $d = 128$ in our implementation.

The modality-specific embeddings are concatenated along the feature dimension:

$$\mathbf{f}_{LF} = \mathbf{f}'_{SAR} \,\|\, \mathbf{f}'_{OPT} \,\|\, \mathbf{f}'_{FTP} \in \mathbb{R}^{nd}, \tag{6}$$

where n is the number of modalities.

The fused representation is then processed by a two-layer multilayer perceptron (MLP):

$$\hat{z}_{LF} = \mathrm{MLP}(\mathbf{f}_{LF}), \qquad \hat{p}_{LF} = \sigma(\hat{z}_{LF}). \tag{7}$$

This design preserves modality-specific feature learning while enabling interaction at the embedding level prior to final classification. The overall architecture is illustrated in Figure 3.

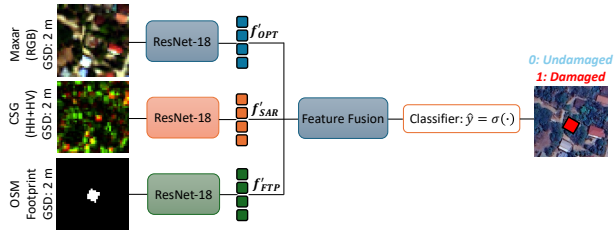

Figure 3. LF: each modality is processed by a dedicated *ResNet-18* encoder; logits are fused at the decision level.

3.3 Middle Fusion with Footprint-Guided Cross Attention

The MF strategy proposed in this work introduces a cross-modal interaction mechanism guided by building footprints. As illustrated in Figure 4, the core idea is to allow SAR and optical features to interact at the feature level, with the binary building mask acting as a spatial prior that focuses attention on the region of interest. This design aims to enhance discriminative capability by suppressing irrelevant background and encouraging object-aware fusion.

Each modality input (i.e. $\mathbf{X}_{SAR}$, $\mathbf{X}_{OPT}$, and $\mathbf{X}_{FTP}$), is processed through the first three convolutional blocks (*Layers 1-3*) of a modality-specific *ResNet-18* encoder:

$$\begin{aligned} \mathbf{f}'_{SAR} &= E^{1\text{-}3}_{SAR}(\mathbf{X}_{SAR}), \\ \mathbf{f}'_{OPT} &= E^{1\text{-}3}_{OPT}(\mathbf{X}_{OPT}), \\ \mathbf{f}'_{FTP} &= E^{1\text{-}3}_{FTP}(\mathbf{X}_{FTP}). \end{aligned} \tag{8}$$

The resulting feature maps are reshaped into token sequences:

$$\begin{aligned} \mathbf{Z}_m = \mathrm{Flatten}(\mathbf{f}'_m) \in \mathbb{R}^{N\times d}, m \in \{SAR, OPT, FTP\}, \\ H' = W' = 14, \quad N = H'W' = 196. \end{aligned} \tag{9}$$

Subsequently, the footprint embedding is used to generate the query representation, while the SAR and optical embeddings provide the keys and values:

$$\begin{aligned} \mathbf{Q}_{FTP} &= \mathbf{Z}_{FTP}\mathbf{W}_Q, \\ \mathbf{K}_m &= \mathbf{Z}_m\mathbf{W}_K, \\ \mathbf{V}_m &= \mathbf{Z}_m\mathbf{W}_V, \quad m \in \{SAR, OPT\}. \end{aligned} \tag{10}$$

Keys and values from the SAR and optical branches are concatenated:

$$\mathbf{K} = [\mathbf{K}_{SAR} \,\|\, \mathbf{K}_{OPT}], \qquad \mathbf{V} = [\mathbf{V}_{SAR} \,\|\, \mathbf{V}_{OPT}], \tag{11}$$

where $\|$ denotes token-wise concatenation along the sequence dimension.

The footprint-guided cross attention (FGCA) is computed as:

$$\mathbf{Z}_{FGCA} = \mathrm{softmax}\left(\frac{\mathbf{Q}_{FTP}\mathbf{K}^\top}{\sqrt{d}}\right)\mathbf{V}, \tag{12}$$

where $\mathbf{Q}_{FTP} \in \mathbb{R}^{N\times d}$, $\mathbf{K}, \mathbf{V} \in \mathbb{R}^{(2N)\times d}$, and d is the current feature embedding dimension ($d = 256$), following the standard attention formulation Vaswani et al. (2017) to ensure numerical stability of the dot-product.

The footprint queries $\mathbf{Q}_{FTP}$ act as spatial selectors, attending to relevant building regions in the SAR and optical representations. Intuitively, the cross-attention mechanism operates at the spatial token level, where each token corresponds to a local region in the original image (i.e., a spatial location in the 14×14 feature grid). For each footprint token acting as a query, the attention module computes similarity scores with all SAR and optical tokens (keys), thereby determining how much information should be aggregated from each modality at that spatial position. This mechanism allows the model to adaptively combine cross-modal features depending on whether a location belongs to a building (foreground) or to the surrounding area (background). In practice, tokens corresponding to building regions highlighted by the footprint prior tend to assign higher attention weights to structurally relevant SAR and optical responses, while suppressing irrelevant background signals. Therefore, the proposed FGCA does not simply merge modalities globally, but performs spatially-aware, object-guided feature integration.

The attended tokens $\mathbf{Z}_{FGCA}$ are reshaped into spatial feature maps $\mathbf{f}_{FGCA} \in \mathbb{R}^{d\times H'\times W'}$, which are then passed through the final convolutional block (*Layer 4*) of the footprint encoder branch:

$$\mathbf{f}'_{MF} = E^{4}_{MF}(\mathbf{f}_{FGCA}). \tag{13}$$

The resulting representation is adaptively pooled and classified using the MLP layer:

$$\hat{z}_{MF} = \mathrm{MLP}\left(\mathrm{GAP}(\mathbf{f}'_{MF})\right), \qquad \hat{p}_{MF} = \sigma(\hat{z}_{MF}). \tag{14}$$

This architecture effectively combines the advantages of both early and late fusion. It preserves modality-specific feature learning, as in LF, while introducing explicit cross-modal interaction at the intermediate feature level, guided by building footprints that provide spatial context. This intermediate integration stage, absent in conventional EF and LF schemes, enables more targeted information exchange between modalities. As shown in Section 4, the proposed MF approach achieves consistently superior performance across all modality configurations, demonstrating enhanced robustness and predictive reliability.

3.4 Training Strategy

Given the strong class imbalance between positive and negative examples, the training process incorporated a weighted binary cross-entropy loss. For a binary label $y \in \{0, 1\}$ and predicted logit $\hat{z}$, the loss is defined as:

$$\mathcal{L}_{\mathrm{BCE}} = -\left[w_{pos}\, y \log \sigma(\hat{z}) + (1-y)\log\left(1-\sigma(\hat{z})\right)\right], \tag{15}$$

where $\sigma(\cdot)$ denotes the sigmoid activation function. The positive-class weight is computed as:

$$w_{pos} = \frac{N_{neg}}{N_{pos}}, \tag{16}$$

with N_{neg} and N_{pos} representing the number of intact and damaged samples within the training split, respectively. This weight-

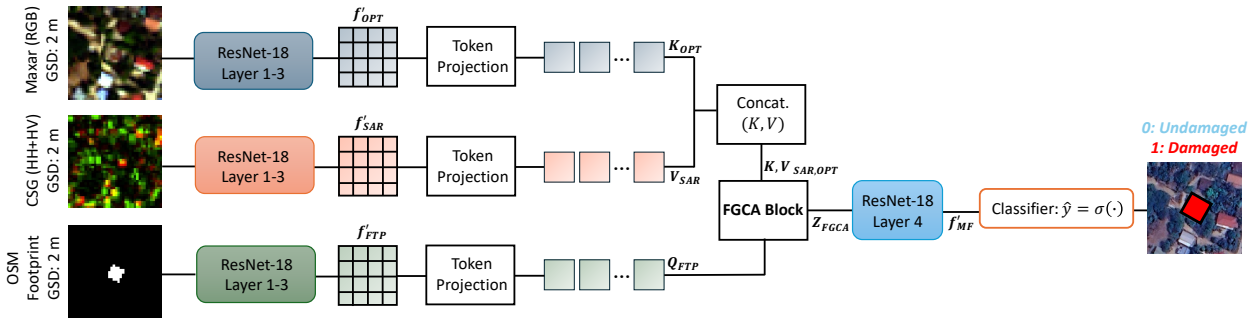

Figure 4. MF with FGCA: footprint-guided attention fuses SAR and optical features at the middle stage, before the final classification.

Metric	*SAR + FTP*			*OPT + FTP*			*SAR + OPT + FTP*		
	EF	LF (Sun et al., 2024)	MF (this study)	EF	LF (Sun et al., 2024)	MF (this study)	EF	LF (Sun et al., 2024)	MF (this study)
Precision	0.247 ± 0.073	0.274 ± 0.084	**0.567 ± 0.197**	0.341 ± 0.044	0.333 ± 0.038	**0.442 ± 0.050**	0.362 ± 0.112	0.361 ± 0.029	**0.420 ± 0.040**
Recall	0.281 ± 0.116	0.261 ± 0.080	0.185 ± 0.033	0.362 ± 0.052	**0.366 ± 0.035**	0.320 ± 0.046	0.339 ± 0.045	0.340 ± 0.070	**0.353 ± 0.023**
F1–Score	0.231 ± 0.030	0.248 ± 0.028	**0.266 ± 0.023**	0.346 ± 0.022	0.346 ± 0.014	**0.367 ± 0.027**	0.336 ± 0.025	0.345 ± 0.036	**0.384 ± 0.030**
κ	0.191 ± 0.038	0.210 ± 0.036	**0.248 ± 0.018**	0.312 ± 0.024	0.311 ± 0.016	**0.340 ± 0.021**	0.303 ± 0.032	0.314 ± 0.034	**0.355 ± 0.032**
AUROC	0.729 ± 0.011	0.738 ± 0.005	**0.739 ± 0.012**	0.824 ± 0.017	0.818 ± 0.017	**0.826 ± 0.021**	0.819 ± 0.010	0.810 ± 0.012	**0.829 ± 0.014**

Table 3. Results comparison of different fusion strategies across different modality configurations. Best results per metric are in bold.

ing scheme increases the penalty assigned to misclassified damaged buildings and mitigates bias toward the majority class, while preserving the original data distribution.

4. Results and Discussion

The experimental evaluation was designed to assess the performance of the proposed FGMF framework in comparison with the EF and LF strategies under consistent training conditions. To ensure a robust and reliable evaluation, all models were trained and validated using a five-fold cross-validation scheme. The complete data set comprises 36,309 building-centered samples (Table 2), including 1,729 damaged and 34,580 intact buildings. Under the five-fold cross-validation scheme, approximately 80% of the samples (about 29,000 buildings) were used for training at each fold, while the remaining 20% were reserved for validation. The dataset was partitioned into five non-overlapping subsets through a stratified split, preserving the proportion of damaged and undamaged buildings in each fold. At each iteration, four folds (approximately 80% of the data) were used for training, while the remaining one was held out to monitor performance and compute the final metrics. The checkpoint achieving the highest AUROC on the excluded fold was retained as the best model for that iteration. This process was repeated until each sample had served once as test data, and the reported results correspond to the mean and standard deviation across the five folds, with all metrics referring to the positive *damaged* class. The quantitative results are summarized in Table 3. For each test fold, precision, recall, F1-score, and Cohen’s Kappa (κ) were computed from binary predictions obtained by thresholding the model outputs at the value that maximized the F1-score on the corresponding precision–recall curve. This adaptive threshold selection ensures that all metrics are evaluated at the model’s optimal operating point with respect to the *damaged* class, which represents the primary focus in BDD. Furthermore, an ablation study was conducted within each fusion strategy to evaluate the impact of different modality combinations. Specifically, three configurations were tested: SAR with footprint data (*SAR + FTP*), optical with footprint data (*OPT + FTP*), and the full combination of SAR, optical, and footprint inputs (*SAR + OPT + FTP*). The results in Table 3 reveal a clear and consistent trend across all modality configurations. The proposed MF strategy systematically outperforms both EF and LF baselines in nearly all metrics, confirming that introducing cross-modal interactions at the feature level leads to more discriminative and stable representations. The improvement is particularly evident in precision, F1-score, and κ, indicating that MF yields more reliable predictions with fewer false positives and stronger agreement between predicted and reference labels. When applied to the configuration combining SAR and footprint data, MF achieves the largest relative gain, highlighting the benefit of explicitly modeling structural information in radar backscatter when guided by spatial priors such as building footprints. Unlike LF, where modality integration occurs only at the decision level, MF performs fusion at the intermediate feature stage, enabling the network to learn richer spatial relationships between geometric structure and backscatter intensity. A comparable benefit is observed when using optical and footprint data alone, where the attention mechanism enhances focus on building regions while effectively suppressing background interference. Overall, the integration of all three modalities yields the strongest performance. In this full multimodal setup, MF attains the highest F1-score, κ, and AUROC across all possible combinations, confirming that the joint exploitation of SAR, optical, and footprint information provides the most comprehensive representation of damage-related features. This result highlights the complementary strengths of optical imagery and SAR: the former captures fine-scale textural and radiometric variations indicative of surface damage, whereas the latter provides structural and scattering information that remains largely unaffected by illumination or atmospheric conditions. The joint exploitation of these two domains enhances sensitivity to different manifestations of building damage, thereby improving the overall effectiveness of damage detection. Taken together, these findings demonstrate that the proposed FGMF framework achieves an effective balance between modality-specific learning and cross-modal interaction, resulting in improved generalization and reliability for BDD from heterogeneous EO data.

4.1 Evaluation on the unseen Sagaing area

To further evaluate the generalization capability of the proposed framework under realistic deployment conditions, the FGMF

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-927-2026 | © Author(s) 2026. CC BY 4.0 License.

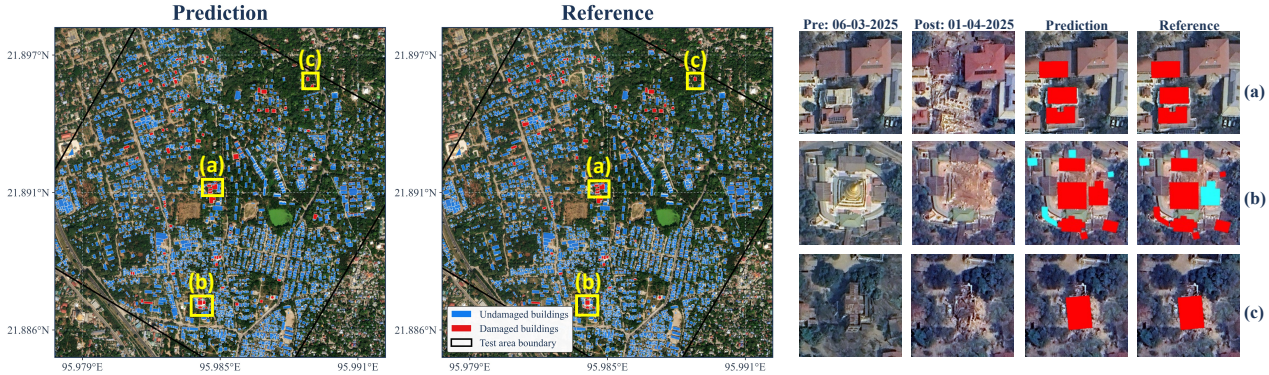

Figure 5. Comparison between model predictions and manual annotations for a subset of the Sagaing area. The maps on the left show the spatial distribution of predicted and reference building damage from the proposed FGMF framework. Damaged and undamaged buildings are marked in red and blue, respectively, within the test area boundary. Zoomed examples (a–c) on the right illustrate a comparison between model outputs and manual annotations, supported by VHR pre- and post-event imagery from Google Earth.

model was tested on an independent portion of the Sagaing subarea. As described in Section 2, Sagaing together with Mandalay was among the most severely affected urban centres within the considered AOI, providing a meaningful and challenging benchmark for assessing model transferability to unseen spatial domains.

For this experiment, all available data outside the selected test area were used for model training, while a subset of the Sagaing region was completely excluded from the learning process and reserved for inference only. Model predictions over this area were subsequently compared with the reference building damage annotations to evaluate the spatial consistency and reliability of the inferred damage map. The resulting comparison is presented in Figure 5.

As illustrated in the figure, the predicted damage distribution produced by the FGMF framework closely matches the reference annotations, correctly identifying the majority of damaged buildings across the test zone. A limited number of false positives can be observed, which is a common behavior in post-disaster scenarios where slight spectral or structural alterations may resemble actual damage. Nonetheless, these over-detections are generally preferable to missed cases, as they help ensure that potentially affected structures are not overlooked during emergency response and damage assessment efforts.

The right-hand panels of Figure 5 present detailed visual comparisons for three representative examples (a–c), supported by VHR pre- and post-event optical imagery from Google Earth (acquired on *6 March 2025* and *1 April 2025*, respectively). Example (a) shows a densely built residential block where multiple collapsed houses are correctly identified by the model. Example (b) depicts a religious structure that collapsed completely, consistent with reports indicating that historical buildings in Sagaing were among the most severely affected types of structures. Finally, example (c) presents an isolated mid-rise residential building of older construction, which also exhibits evident structural failure captured by both the model and the reference annotations. These findings confirm the strong generalization of the proposed FGMF framework, which maintains accurate and spatially consistent predictions in real deployment scenarios. The model reliably captures major damage patterns while minimizing missed detections, supporting its use in large-scale post-disaster mapping.

4.2 Limitations and generalization considerations

Despite the encouraging results, some limitations should be acknowledged. First, the reliance on OSM building footprints may introduce spatial inaccuracies or incompleteness, particularly in rapidly evolving urban areas, potentially affecting both training and inference. Second, although the independent evaluation on the Sagaing subarea demonstrates promising spatial transferability, the model was trained and tested within the same earthquake event, and its robustness across different seismic contexts remains to be systematically assessed. Finally, the intrinsic class imbalance typical of post-disaster scenarios may still influence decision thresholds and precision–recall trade-offs, despite the adopted sampling strategy.

5. Conclusions

This study presented a multimodal framework for post-earthquake BDD that integrates HR SAR, optical, and building footprint data through a Footprint-Guided Middle Fusion (FGMF) strategy. The proposed approach enables effective feature-level interaction between SAR and optical representations under explicit footprint guidance, enhancing cross-modal complementarity and focusing the analysis on building-related information. Experiments on the 2025 M_w 7.7 Myanmar earthquake demonstrated that FGMF consistently outperforms conventional early and late fusion schemes, achieving higher precision, F1-score, and Cohen’s Kappa metrics. The independent evaluation on an unseen portion of the Sagaing subarea further confirmed the robustness and generalization capability of the framework under realistic deployment conditions. Beyond its methodological contribution, the FGMF framework highlights the value of integrating multimodal and geospatial information for operational post-disaster mapping. By jointly exploiting the structural sensitivity of SAR, the visual detail of optical imagery, and the spatial context of building footprints, the proposed approach supports accurate, interpretable, and scalable BDD. These findings highlight the potential of multimodal fusion to transform heterogeneous satellite observations into actionable information for rapid and reliable earthquake response. Future work will explore extending the FGMF framework to additional hazard types, as well as investigating transfer learning and domain adaptation strategies to improve cross-event generalization and operational deployment.

6. Acknowledgements

This research was performed in the framework of a PhD funded by NextGenerationEU, Action 4, DM n. 118, 02/03/2023.

COSMO-SkyMed Product of the Italian Space Agency (ASI) was accessed under a license to use granted by ASI within the Project Card MyanmarEQ2025.

References

Ainscoe, E., Swaminathan, R., Way, L., Modugno, S., Chin, S. T., Panta, N., Crevoisier, T., Yun, S.-H., 2025. Earthquake damage mapped more comprehensively and accurately by radar satellites than optical imagery. *Communications Earth & Environment*, 6, 631. doi.org/10.1038/s43247-025-02623-4.

Brunner, D., Lemoine, G., Bruzzone, L., 2010. Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 2403–2420. doi.org/10.1109/TGRS.2009.2038274.

Dell’Acqua, F., Gamba, P., 2012. Remote Sensing and Earthquake Damage Assessment: Experiences, Limits, and Perspectives. *Proceedings of the IEEE*, 100(10), 2876–2890. doi.org/10.1109/JPROC.2012.2196404.

Dong, L., Shan, J., 2013. A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, 84, 85–99. doi.org/10.1016/j.isprsjprs.2013.06.011.

ElGharbawi, T., Zarzoura, F., 2021. Damage detection using SAR coherence statistical analysis, application to Beirut, Lebanon. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 1–9. doi.org/10.1016/j.isprsjprs.2021.01.001.

Ge, P., Gokon, H., Meguro, K., 2020. A review on synthetic aperture radar-based building damage assessment in disasters. *Remote Sensing of Environment*, 240, 111693. doi.org/10.1016/j.rse.2020.111693.

Hertel, V., Geiß, C., Wieland, M., Taubenböck, H., 2025. Rapid domain adaptation for disaster impact assessment: Remote sensing of building damage after the 2021 Germany floods. *Science of Remote Sensing*, 12, 100287. doi.org/10.1016/j.srs.2025.100287.

Italian Space Agency, 2021. COSMO-SkyMed Second Generation (CSG): Mission and Products Description. `https://www.asi.it/wp-content/uploads/2021/03/CSG-Mission-and-Products-Description-defpdf-1.pdf`. Technical report, February 2021.

Jiang, X., Li, G., Liu, Y., Zhang, X.-P., He, Y., 2020. Change detection in Heterogeneous Optical and SAR Remote Sensing Images via Deep Homogeneous Feature Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 1551–1566. doi.org/10.1109/JSTARS.2020.2983993.

Jiao, T., Guo, C., Feng, X., Chen, Y., Song, J., 2024. A Comprehensive Survey on Deep Learning Multi-Modal Fusion: Methods, Technologies and Applications. *Computers, Materials and Continua*, 80, 1–35. doi.org/10.32604/cmc.2024.053204.

Li, J., Huang, H., Sheng, Y., Guo, Y., He, W., 2025. Building-Guided Pseudo-Label Learning for Cross-Modal Building Damage Mapping. *arXiv preprint arXiv:2505.04941*. arxiv.org/abs/2505.04941.

Li, Q., Zhang, J., Jiang, H., 2024. Incorporating multi-source remote sensing in the detection of earthquake-damaged buildings based on logistic regression modelling. *Heliyon*, 10(12), e32851. doi.org/10.1016/j.heliyon.2024.e32851.

Macchiarulo, V., Giardina, G., Milillo, P., Aktas, Y. D., Whitworth, M. R. Z., 2025. Integrating post-event very high resolution SAR imagery and machine learning for building-level earthquake damage assessment. *Bulletin of Earthquake Engineering*, 23, 5021–5047. doi.org/10.1007/s10518-024-01877-1.

Rao, A., Jung, J., Silva, V., Molinario, G., Yun, S.-H., 2023. Earthquake building damage detection based on synthetic-aperture-radar imagery and machine learning. *Natural Hazards and Earth System Sciences*, 23(2), 789–807. doi.org/10.5194/nhess-23-789-2023.

Serifoglu Yilmaz, C., Yilmaz, V., Tansey, K., Aljehani, N. S. O., 2023. Automated detection of damaged buildings in post-disaster scenarios: a case study of Kahramanmaraş (Türkiye) earthquakes on February 6, 2023. *Natural Hazards*, 119, 1247–1271. doi.org/10.1007/s11069-023-06154-z.

Shen, Y. et al., 2022. 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. doi.org/10.1109/TGRS.2021.3080580.

Socquet, A., Vigny, C., Chamot-Rooke, N., Simons, W., Rangin, C., Ambrosius, B., 2006. India and Sunda plates motion and deformation along their boundary in Myanmar determined by GPS. *Journal of Geophysical Research: Solid Earth*, 111(B5), B05406. doi.org/10.1029/2005JB003877.

Sun, Y., Wang, Y., Eineder, M., 2024. QuickQuakeBuildings: Post-Earthquake SAR-Optical Dataset for Quick Damaged-Building Detection. *IEEE Geoscience and Remote Sensing Letters*, 21, 1–5. doi.org/10.1109/LGRS.2024.3406966.

United Nations, 2025. OCHA Myanmar: Earthquake Response Situation Report No. 4. Available at: `https://myanmar.un.org/en/293275-ocha-myanmar-earthquake-response-situation-report-no-4`.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wang, L., Wu, J., Yang, Y., Tang, R., Ya, R., 2024. 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. doi.org/10.1109/JSTARS.2024.3449097.

Wiguna, S., Adriano, B., Mas, E., Koshimura, S., 2024. 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. doi.org/10.1109/JSTARS.2024.3367853.

World Bank, 2025. Earthquake compounds Myanmar’s economic challenges . Available at: `https://www.worldbank.org/en/news/press-release/2025/06/12/earthquake-compounds-myanmar-s-economic-challenges`.

Xia, H., Wu, J., Yao, J., Zhu, H., Gong, A., Yang, J., Hu, L., Mo, F., 2023. 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. doi.org/10.1007/s13753-023-00526-6.