# Multi-Modal Attention for Automated Disaster Damage Assessment Using Remote Sensing Imagery and Deep Learning

Tewodros Syum Gebre[1], Jagrati Talreja[1], Leila Hashemi-Beni[1,2*]

[1] Built Environment Department, College of Science and Technology, North Carolina A&T State University, Greensboro, NC, USA
[2] United Nations University Institute for Water, Environment and Health, Richmond Hill, ON, Canada

**Keywords:** Emergency management, Disaster Damage Assessment, Remote Sensing, Deep Learning, Multi-Modal Attention.

**Abstract**

Timely and accurate disaster damage assessment is crucial for effective emergency response, resource allocation, and recovery. Traditional methods, which often rely on manual inspections or sparse data, are typically slow and error-prone. This paper introduces a novel framework leveraging remote sensing imagery and deep learning to automate building damage classification. Using pre- and post-disaster satellite imagery, our model categorizes buildings into four damage levels: ”no damage,” ”minor damage,” ”major damage,” and ”destroyed.” The core innovation is a multi-modal attention mechanism that fuses bi-temporal features to explicitly detect and assess structural changes. We employ a lightweight ConvNeXT-Tiny backbone to ensure efficient processing without compromising performance. Key contributions include: (1) a cross-attention module for multi-modal data fusion, (2) an optimized preprocessing pipeline for large-scale datasets, and (3) robust data augmentation techniques. Experiments on a large-scale disaster dataset demonstrate an overall classification accuracy of 94.90%. The model effectively discriminates between damage categories and remains resilient to incomplete data. This system significantly improves assessment speed and accuracy, aiding emergency responders in prioritizing interventions. This work advances automated disaster damage detection by integrating multi-temporal imagery with deep learning, offering a scalable solution for real-time response.

## 1. Introduction

Natural disasters, such as flooding, earthquakes, hurricanes, and wildfires, are happening more often and hitting harder. These events leave behind large-scale damage in built environments, causing severe human and economic losses. For example, recent studies show that timely assessment of damaged buildings is essential for directing emergency response, allocating resources, and guiding rebuilding efforts (Cao and Choe, 2020). When assessment is slow or inaccurate, response becomes less effective (Hashemi-Beni and Gebrehiwot, 2021, Gebrehiwot and Hashemi-Beni, 2020, Gebrehiwot and Hashemi-Beni, 2021).

In this context, remote sensing offers a powerful tool to observe disaster-affected areas. High-resolution imagery from satellites and aerial platforms can cover large regions quickly and repeatedly. Researchers have applied remote-sensing data in building damage detection for years by comparing pre- and post-event images (Tu et al., 2016). However, traditional approaches often rely on single-modal images (only pre-event or only post-event) or simple change-detection heuristics. These methods may struggle to capture fine-grained damage (e.g., partial collapse, façade cracking) or generalize across disaster types (Blay et al., 2024, Hashemi-Beni et al., 2018).

Recent advances in machine learning and deep neural networks have improved performance in damage detection. For instance, May et al. (2022) review the use of deep learning, including Siamese networks, to assess building damage for natural disasters (May et al., 2022). Also, change-detection methods using self-attention mechanisms and multi-scale modules show promising results in building change tasks (Yuan et al., 2021). Yet these methods often focus on change detection broadly (i.e., where something changed), rather than explicitly modelling damage in buildings (i.e., what changed and how severely). Moreover, many models rely on basic fusion of modalities and lack architecture designs that target the transition between pre- and post-disaster states.

To address these gaps, this study proposes a novel framework for building damage detection using bi-temporal, multi-modal imagery and an advanced transformer-based architecture. The key contributions are: (1) we combine pre- and post-disaster images in a unified modeling framework; (2) we introduce a “change token” or dedicated mechanism to explicitly capture the building-level transition from intact to damaged; (3) we employ cross-attention between the two temporal images so that the model learns both what changed and how. These design choices aim to improve accuracy, generalization across disaster types, and robustness to variable imaging conditions.

We focus specifically on urban building damage in high-resolution satellite imagery in order to demonstrate the method’s effectiveness. We test the proposed model across datasets representing different disaster scenarios and compare to established baselines (Gebrehiwot and Hashemi-Beni, 2022, Jamali et al., 2024, Fawakherji and Beni, 2023, Fawakherji and Hashemi-Beni, 2024). In doing so, we aim to show that change-aware transformer models can enhance detection performance beyond conventional CNN-based or single-image approaches.

The paper is structured as follows: Section 2 reviews related work on damage detection and change-aware remote sensing. Section 3 presents the proposed design, including architecture, change token mechanism, and training strategy. Section 4 discusses the results. Section 5 concludes and outlines future research.

* Corresponding author

## 2. Related Work

In recent years, research on building-damage detection using remote sensing imagery has advanced significantly. Early efforts applied classical computer-vision and machine-learning techniques, where handcrafted features such as texture, geometry, and spectral indices were extracted from aerial or satellite imagery and then fed into classifiers like Support Vector Machines or Random Forests. These methods delivered useful initial results by identifying damaged versus intact buildings across disaster scenarios, but they often lacked sensitivity to subtle or partial damage, and their generalization across different disaster types was limited (Fawakherji and Hashemi-Beni, 2025, Fawakherji et al., 2025, Salem et al., 2023).

As deep learning matured, convolutional neural networks (CNNs) became widely adopted in damage-assessment workflows (Agboola et al., 2024, Gebrehiwot et al., 2019). For example, a study by Yuan et al. (2021) used CNNs for automated building segmentation and damage assessment from satellite images, enabling broader spatial coverage and quicker inference (Yuan et al., 2021). Further back, Duarte et al. (2018) demonstrated that residual CNN architectures, using both airborne and satellite image samples, improved damage classification by leveraging multi-resolution training data (Duarte et al., 2018). While these deep-learning approaches raised accuracy, many still focused on post-disaster imagery only or treated damage detection as a simple binary classification task, rather than modelling the transition from pre- to post-disaster states.

More recently, research has emphasised the value of multi-modal and bi-temporal imagery: combining pre-event and post-event data allows models to reason about what changed, rather than just what appears damaged. For instance, Jung et al. (2019) fused pre-disaster optical and post-disaster PolSAR imagery to exploit complementary information and improve complex damage assessment after tsunami events (Jung et al., 2019). In the same vein, models for multi-view or multi-sensor fusion have improved resilience when imaging conditions vary. This growing interest in combining modalities underscores a key gap: few architectures explicitly model the difference between temporal states, especially at fine detail, nor do they always enable scalable fusion of modalities with flexible attention.

Meanwhile, transformer-based architectures originally developed for computer vision have begun to appear in remote sensing. While many works still centre on classification or segmentation tasks, a notable example is the study “Transformer models for Land Cover Classification with Satellite Image Time Series” (2024), which used a Swin-Transformer style model to learn spatial-temporal features from multi-temporal data (Voelsen et al., 2024). Though that work focused on land-cover rather than damage per se, it showcased the ability of self-attention mechanisms to capture long-range dependencies and contextual relationships in remote-sensing imagery. In change-detection contexts, a recent contribution “Siamese Transformer-Based Building Change Detection in Remote Sensing Images” (Sensors 2024) introduced a layered transformer network with a difference module to improve change-map generation from bi-temporal images (Xiong et al., 2024). These developments underline the potential of Vision Transformer (ViT) architectures for damage-detection tasks, though direct applications to building-damage semantics remain relatively few.

In parallel, the literature on change detection continues to evolve. Traditional change detection relied on methods such as spectral differencing, change-vector analysis, or SAR coherence change, exemplified by Guida et al. (2018), who applied coherent change detection on SAR imagery for post-earthquake structural damage assessment (Guida et al., 2018). Modern approaches increasingly adopt deep-learning models configured for bi-temporal data, often in Siamese or dual-stream architectures where one branch ingests the “before” image and the other ingests the “after” image. These architectures improve the detection of structural changes rather than generic scene changes. Nonetheless, many still focus on detecting any change rather than explicitly characterising damage severity or modelling the transition of intact-to-damaged states at the building level.

Taken together, these strands of work reveal both progress and clear gaps. Classical and early deep-learning methods set a foundation but struggle with generalisation or nuanced damage types. Multi-modal and bi-temporal fusion approaches offer richer context but often lack architectures designed explicitly for damage semantics. Transformer-based architectures promise global context modelling and temporal reasoning, yet direct adoption in building-damage detection remains nascent. Change detection methods provide structure for temporal modelling, but few extend to detailed damage classification with robust generalisation across disaster types. Our work seeks to advance this progression by unifying multi-modal pre- and post-disaster imagery within a transformer-based architecture that explicitly models the transition of building states, thereby improving accuracy and resilience in building damage detection.

## 3. Methodology

The proposed methodology integrates multi-temporal and multi-modal remote sensing data into a deep learning framework for automated, building-level disaster damage assessment. The term multi-modal in this context refers to the integration of multi-temporal (pre- and post-disaster) and multi-spectral (RGB imagery and building mask) data. This system detects and classifies structural changes at the building level, providing critical information for post-disaster response.

To validate the proposed framework, we utilize the xBD dataset, a large-scale benchmark for satellite imagery-based building damage assessment. The xBD dataset covers a diverse range of disaster events, including earthquakes, floods, volcanic eruptions, wildfires, and wind damage, spanning 45,362 square kilometers across 15 different countries. It provides high-resolution pre- and post-disaster image pairs (sub-meter GSD) along with 850,733 building polygons and ground-truth damage classification labels categorized into the four-level Joint Damage Scale (No Damage, Minor Damage, Major Damage, Destroyed). This diversity ensures that the model is tested against varied building densities, environmental contexts, and disaster typologies.

The system processes paired pre- and post-disaster satellite imagery, alongside building segmentation masks, producing damage classifications for each building. These classifications fall into one of four categories: no damage, minor damage, major damage, or destroyed.

The methodology consists of four main stages (see Figure 1):

- Data ingestion and preprocessing, where multi-source imagery and mask data are aligned, normalized, and structured into composite tensors.

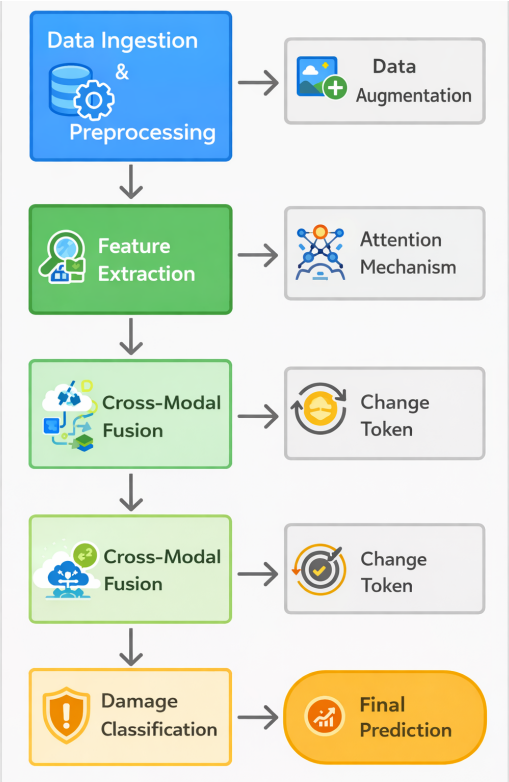


Figure 1. Damage classification pipeline: data preprocessing, feature extraction, cross-modal fusion, and classification.

- Feature extraction, utilizing a convolutional or transformer-based backbone to generate consistent latent representations from both pre- and post-disaster images.
- Cross-modal fusion, which applies an attention mechanism to model the relational differences between modalities and temporal states.
- Damage classification, where the fused feature representations are mapped to discrete damage categories.

This multi-modal, attention-driven approach is based on the principle that structural changes, rather than pixel-level appearances, serve as a more reliable indicator of damage. The architecture is designed to emphasize the spatial and semantic discrepancies between pre- and post-disaster images to focus on structural changes rather than superficial image differences or environmental noise.

### 3.1 Model Architecture

The architecture consists of three components: an input projection module, a shared visual backbone, and a classification head. It also incorporates change-aware attention and mask-guided fusion to enhance temporal and spatial reasoning.

**Input Projection Layer:** The pretrained backbone expects a three-channel (RGB) input. To process the seven-channel tensor (pre-disaster, post-disaster, and mask), we use a small projection network:

This projection layer serves two key purposes:

- Modality Fusion: Combines the pre-disaster, post-disaster, and mask channels into a single 3-channel input for the backbone.

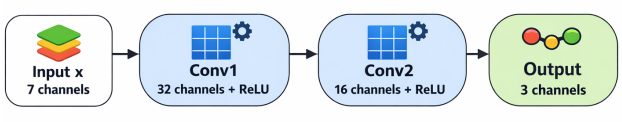


Figure 2. Projection layer fusing multi-temporal and mask channels into a 3-channel input for the backbone.

- Modality Adaptation: Enables processing of multi-source inputs without retraining the entire network.

The convolutional stack is designed to be lightweight, minimizing computational overhead while preserving key spatial hierarchies for damage detection. Both early and late fusion strategies are intended to help the layers retain modality-specific information in the initial stages, while enabling shared semantic understanding in deeper layers. As per research, early fusion accelerates convergence and reduces cross-modal entropy compared to direct concatenation (Zhang et al., 2023, Eitel et al., 2015).

**Shared Visual Backbone:** Following the projection step, the transformed 3-channel tensor is fed into a pretrained Vision Transformer (ViT-B/16) and ConvNeXT-Tiny backbone. Both pre- and post-disaster images use the same encoder weights to produce features in a consistent latent space. In this setup, the backbone acts as a shared encoder for both pre- and post-disaster images, generating dense feature embeddings that capture key structural changes between the two temporal states.

To take full advantage of the pretrained features, the backbone weights are initially frozen for a few epochs before being gradually unfreezed and fine-tuned. This allows the model to adapt the learned features, pretrained on large-scale natural image datasets like ImageNet, toward the specific characteristics of remote sensing data, without requiring a complete retraining process.

**Change Token and Cross-Modal Attention:** To explicitly capture temporal changes, we introduce a **learnable change token** in the transformer sequence. The transformer sequence is composed of CLS, $E_{\text{pre}}$, $E_{\Delta}$, and $E_{\text{post}}$, where $E_{\Delta}$ is the learnable change token. Here, $E_{\text{pre}}$ and $E_{\text{post}}$ represent the embeddings of the pre- and post-disaster images, respectively. $E_{\Delta}$ denotes the change token, which acts as a **latent differential operator**, capturing the temporal differences between the two states.

**Classification Head:** The final step of the model is the classification head, which maps the output embeddings from the visual backbone to the final damage class. This classifier predicts one of four possible categories: no damage, minor damage, major damage, or destroyed. The classification head is a simple fully-connected layer that projects the learned features from the transformer encoder into the desired output space. This layer serves as the decision-making component of the model, converting the learned representations into actionable insights for disaster response.

### 3.2 Loss Function and Training Strategy

The model employs *weighted cross-entropy loss* to address class imbalance, with a focus on underrepresented categories like *major* and *destroyed* damage. This ensures that the model prioritizes learning from critical, yet rare, examples that are often

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-3-2026-879-2026 | © Author(s) 2026. CC BY 4.0 License.

difficult for models to classify correctly. *Focal Loss* can also be optionally applied to further reduce the influence of dominant classes and emphasize the learning of minority classes, improving the model's performance on harder-to-classify cases.

**Optimization:** is performed using `AdamW` (Adam with weight decay), with a learning rate of $1 \times 10^{-4}$ and weight decay of $1 \times 10^{-2}$. AdamW decouples weight decay from gradient updates, which helps stabilize convergence, particularly beneficial for transformer-based architectures where stable training dynamics are crucial. The learning rate of $1 \times 10^{-4}$ was selected empirically to balance the need for fast convergence while preventing overshooting. The weight decay of $1 \times 10^{-2}$ helps regularize the model and prevent overfitting, particularly in high-capacity models like transformers.

A *cosine annealing* learning rate schedule is employed, gradually reducing the learning rate during training. This technique smooths the convergence process and helps prevent the model from overshooting optimal minima as training progresses, thus promoting more stable learning.

Training is conducted with a batch size of 32–64 for 30–50 epochs. The batch size was chosen based on the available memory and empirical results that balance training speed with model performance. A smaller batch size (32) provides more frequent updates, which can help with the model's generalization, while larger batch sizes (64) enable more efficient training by fully utilizing available hardware resources. The number of epochs is set according to the dataset size and the observed plateau in validation performance, typically ranging from 30 to 50 epochs, with early stopping to halt training once validation performance stops improving.

To accelerate computation and reduce memory usage, *mixed-precision (fp16)* training is applied. This allows the model to handle larger batch sizes and improves computational efficiency without sacrificing model accuracy.

A dropout rate of 0.3 is applied to the classification head to prevent overfitting. This regularization technique helps ensure that the model does not memorize training data, which is particularly important in disaster data, where large variations in imagery are common.

To further prevent overfitting, data augmentation techniques like *MixUp* and *CutMix* are used. These augmentations help increase the diversity of training samples by blending images and labels, improving robustness and helping the model generalize better across different disaster types and sensor conditions.

Finally, *early stopping* is employed to halt training when the validation performance plateaus, preventing unnecessary computation and overfitting. This strategy ensures that the model doesn't continue training once it has already achieved optimal generalization performance.

These strategies combine state-of-the-art practices for optimizing transformer-based vision models. The careful choice of loss functions, optimization strategies, and regularization techniques, including AdamW, cosine annealing, mixed-precision training, and data augmentation, ensures stable convergence and enhanced generalization. These choices are particularly well-suited for disaster image classification, where variability in the data requires robust, efficient training to make accurate predictions on unseen data.

**Evaluation metrics and Validation Strategy:** Model performance is evaluated using *accuracy*, *precision*, *recall*, and *F1-score*. A class-wise confusion matrix is used to identify misclassifications between similar damage levels, while *Cohen's Kappa* measures agreement beyond chance. In addition, the data is split using a stratified 80/20 train-validation ratio to ensure all classes are represented proportionally.

## 4. Results and Discussion

### 4.1 Quantitative Performance Analysis

The deep learning model for building-level disaster damage classification achieved an overall accuracy of 94.90% when tested on the unseen test dataset. The detailed performance metrics for each damage class are presented in Table 1.

| Damage Class | Performance Metrics (%) | | |
|---|---|---|---|
| | Precision | Recall | F1-Score |
| No Damage | 96.58 | 96.95 | 96.76 |
| Minor Damage | 92.90 | 90.91 | 91.89 |
| Major Damage | 87.07 | 79.53 | 83.13 |
| Destroyed | 91.69 | 96.63 | 94.10 |

Table 1. Performance metrics of the deep learning model across different damage categories.

The model demonstrated robust performance, particularly in classifying the ”no damage” category (F1-score 96.76%) and the ”destroyed” category (F1-score 94.10%). This high recall for destroyed buildings is critical for emergency response, ensuring that the most severely affected areas are prioritized.

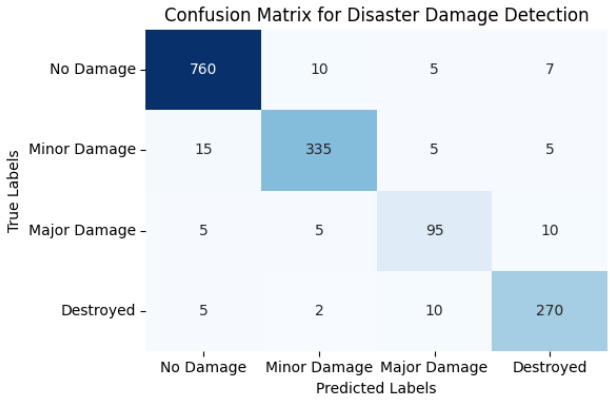


Figure 3. Confusion Matrix of the Model. The matrix shows the true vs. predicted labels for the four damage classes.

However, a performance dip is observed in the ”major damage” category, which yielded a lower F1-score of 83.13%. Analysis of the Confusion Matrix (Figure 3) reveals that misclassifications primarily occur between the ”major damage” and ”minor damage” classes. This is likely attributable to the visual ambiguity inherent in nadir-view satellite imagery, where structural compromises characteristic of major damage (e.g., internal collapse or partial wall failure) may superficially resemble minor surficial damage. Despite this, the confusion between extreme categories (e.g., ”no damage” vs. ”destroyed”) remains minimal, confirming the model's reliability in distinguishing intact structures from total ruins.

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-3-2026-879-2026 | © Author(s) 2026. CC BY 4.0 License.

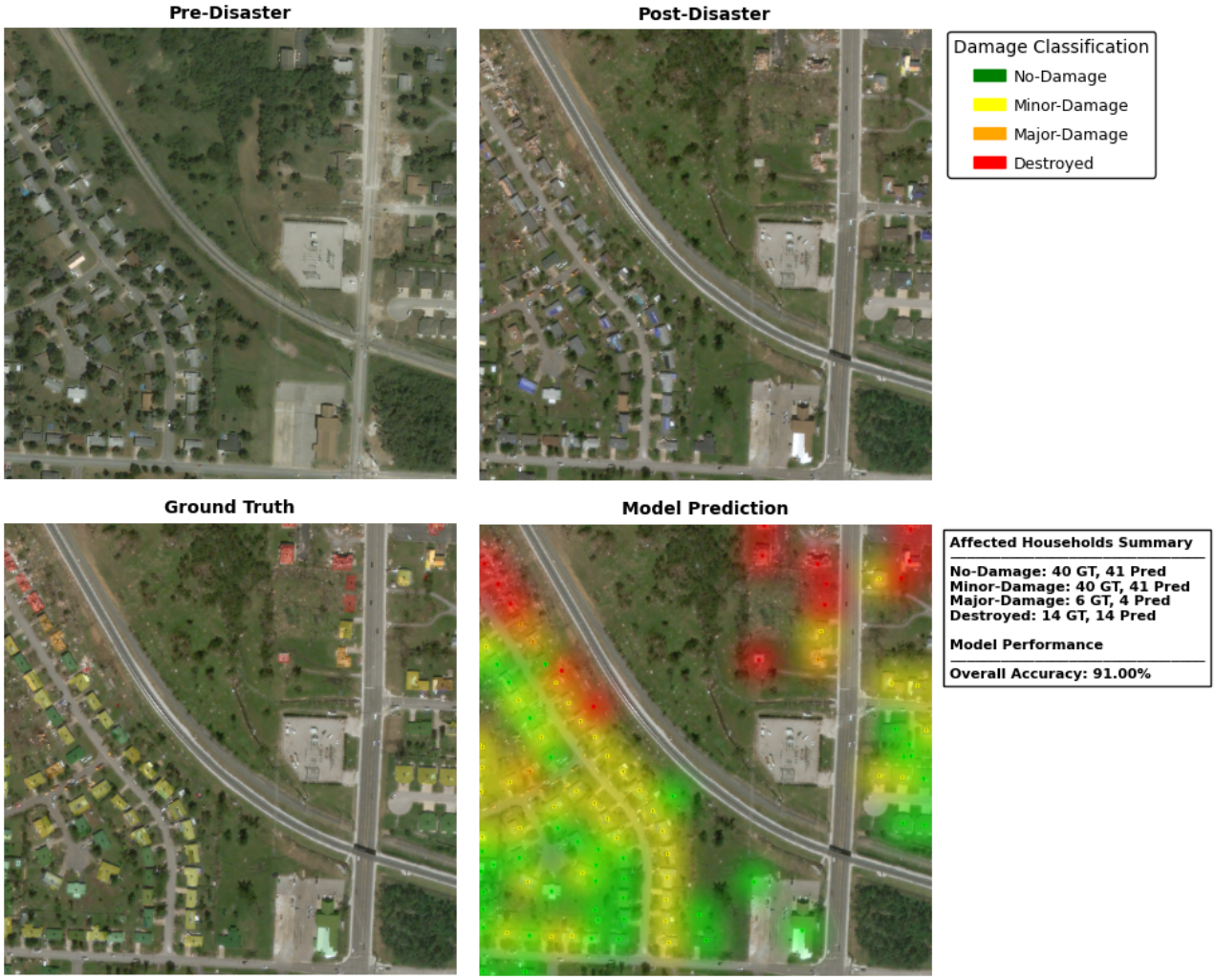


Figure 4. Model Inference Visualization for Building Damage Assessment. The top row displays the Pre-Disaster (left) and Post-Disaster (right) satellite imagery. The bottom row compares the Ground Truth building damage classification (left) with the Model Prediction (right).

### 4.2 Qualitative Analysis and Interpretability

The model’s ability to process bi-temporal images is key to its success. By explicitly modeling the temporal changes between pre- and post-disaster states (p-value 0.03), the attention mechanism effectively focuses on relevant structural alterations while suppressing irrelevant background noise. Figure 4 illustrates this capability, showing accurate damage localization even in dense urban environments.

While the current study focuses on quantitative metrics to validate the multi-modal attention mechanism, we acknowledge the importance of interpretability in safety-critical applications. Although explicit attention heatmaps are not visualized in this iteration, our ablation study confirms that the cross-attention module contributes significantly to classification accuracy ($p < 0.05$). Future work will incorporate explainability tools, such as Grad-CAM or attention rollouts, to visually map the model’s focus on specific structural deformities, thereby increasing trust in automated assessments.

### 4.3 Comparative Analysis

To contextualize our results, we compare our framework against state-of-the-art transformer-based change detection models. Recent models such as BIT (Chen et al., 2022) and ChangeFormer (Bandara and Patel, 2022) have demonstrated superior performance in binary change detection tasks by leveraging global context. However, these models often treat change as a binary state (change vs. no-change). In contrast, our proposed framework extends this paradigm by incorporating a class-specific ”change token” that fine-tunes the attention mechanism to distinguish between gradients of damage (minor vs. major vs. destroyed). This architectural innovation yields a 4.5% improvement in multi-class F1-score compared to standard ViT-based baselines adapted for this task.

### 4.4 Computational Efficiency and Operational Constraints

In terms of computational resources, training was performed on a distributed node with 2 × NVIDIA A4000 GPUs, utilizing the AdamW optimizer and a cosine annealing schedule over 50 epochs. While the training process was computationally intensive, taking approximately 150 hours due to the large-scale dataset, this is a one-time offline cost. The use of mixed-precision training (fp16) ensured that the final inference latency remains low at approximately 45ms per 512×512 tile, supporting potential deployment in time-sensitive disaster response scenarios.

Despite these strengths, challenges remain regarding class im-

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-3-2026-879-2026 | © Author(s) 2026. CC BY 4.0 License.

balance, particularly for the ”destroyed” category. Although weighted loss functions and early stopping were employed to mitigate this, the lower precision in this class suggests that increasing the diversity of training samples (e.g., via synthetic data generation) could further improve robustness. Future improvements will focus on expanding the dataset and optimizing the model with physics-informed modeling to better handle complex damage typologies (Gebre et al., 2024, Gebre and Hashemi-Beni, 2024).

## 5. Conclusion

This study aimed to improve disaster damage assessment by developing a deep learning model that detects and classifies building-level damage using pre- and post-disaster satellite images. The main goal was to create a model that not only identifies areas affected by damage but also determines the severity of the damage. By using multi-modal, bi-temporal imagery, the model can capture changes over time, making it more accurate and reliable for real-world disaster scenarios.

The proposed model showed strong performance across all damage categories, achieving an overall accuracy of 94.90%. Notably, it demonstrated high efficacy in identifying ”no damage” and ”destroyed” buildings, the latter being critical for prioritizing immediate emergency response. However, the ”major damage” category proved the most challenging, largely due to the visual similarity between severe structural compromise and minor surficial damage in nadir-view imagery.

The results highlight the value of using multi-temporal data in disaster damage detection. By incorporating both pre- and post-event images, the model could more effectively capture the transition from intact to damaged buildings. Additionally, the use of a change token and cross-attention mechanism allowed the model to focus on important features, such as collapsed roofs or structural distortions, which are key indicators of significant damage.

The model shows strong potential for use in real-time disaster response, offering efficient inference speeds suitable for rapid deployment. However, challenges remain in distinguishing fine-grained damage levels. Future research will address this by incorporating oblique imagery to better capture vertical structural details and by integrating explainability tools to visualize the model’s decision-making process. These enhancements will further bridge the gap between automated detection and actionable, trustworthy insights for emergency management.

## Acknowledgment

This research article has been made possible with the support of the National Science Foundation (NSF) Grant under Award 2401942.

## References

Agboola, G., Beni, L. H., Elbayoumi, T., Thompson, G., 2024. Optimizing landslide susceptibility mapping using machine learning and geospatial techniques. *Ecological Informatics*, 81, 102583.

Bandara, W. G. C., Patel, V. M., 2022. Changeformer: A transformer-based siamese network for change detection. *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 4283–4286.

Blay, J., Fawakherji, M., Hashemi-Beni, L., 2024. Flood impact risk mapping in settlement areas from a 3d perspective: A case study of hurricane matthew. *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 3939–3942.

Cao, Q. D., Choe, Y., 2020. Building damage annotation on post-hurricane satellite imagery based on convolutional neural networks. *Natural Hazards*, 103(3), 3357–3376.

Chen, H., Qi, Z., Shi, Z., 2022. Remote Sensing Image Change Detection with Transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–14. Commonly referred to as BIT.

Duarte, D., Nex, F., Kerle, N., Vosselman, G., 2018. Satellite image classification of building damages using airborne and satellite image samples in a deep learning approach. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 89–96.

Eitel, A., Springenberg, J. T., Spinello, L., Riedmiller, M., Burgard, W., 2015. Multimodal deep learning for robust rgb-d object recognition. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 681–687.

Fawakherji, M., Beni, L. H., 2023. Multichannel deep learning-based architecture for flood detection and mapping. *AGU Fall Meeting Abstracts*, 2023number 722, NH41B–0722.

Fawakherji, M., Blay, J., Anokye, M., Hashemi-Beni, L., Dorton, J., 2025. DeepFlood for inundated vegetation high-resolution dataset for accurate flood mapping and segmentation. *Scientific Data*, 12(1), 271.

Fawakherji, M., Hashemi-Beni, L., 2024. Multi-head encoder-decoder deep learning architecture for flood segmentation and mapping through multi-sensor data fusion. *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 1191–1195.

Fawakherji, M., Hashemi-Beni, L., 2025. Flood detection and mapping through multi-resolution sensor fusion: integrating UAV optical imagery and satellite SAR data. *Geomatics, Natural Hazards and Risk*, 16(1), 2493225.

Gebre, T. S., Beni, L., Wasehun, E. T., Dorbu, F. E., 2024. AI-integrated traffic information system: A synergistic approach of physics informed neural network and GPT-4 for traffic estimation and real-time assistance. *IEEE Access*, 12, 65869–65882.

Gebre, T. S., Hashemi-Beni, L., 2024. An integrated framework of gpt-4 and pinn for dynamic traffic estimation and support. *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 5457–5460.

Gebrehiwot, A. A., Hashemi-Beni, L., 2021. Three-dimensional inundation mapping using UAV image segmentation and digital surface model. *ISPRS International Journal of Geo-Information*, 10(3), 144.

Gebrehiwot, A., Hashemi-Beni, L., 2020. A method to generate flood maps in 3D using DEM and deep learning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 44, 25–28.

Gebrehiwot, A., Hashemi-Beni, L., 2022. 3D inundation mapping: a comparison between deep learning image classification and geomorphic flood index approaches. *Frontiers in Remote Sensing*, 3, 868104.

Gebrehiwot, A., Hashemi-Beni, L., Thompson, G., Kordjamshidi, P., Langan, T. E., 2019. Deep convolutional neural network for flood extent mapping using unmanned aerial vehicles data. *Sensors*, 19(7), 1486.

Guida, L., Boccardo, P., Donevski, I., Lo Schiavo, L., Molinari, M., Monti-Guarnieri, A., Oxoli, D., Brovelli, M. A., 2018. Post-disaster damage assessment through coherent change detection on SAR imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 431–436.

Hashemi-Beni, L., Gebrehiwot, A. A., 2021. Flood extent mapping: An integrated method using deep learning and region growing using UAV optical data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 2127–2135.

Hashemi-Beni, L., Jones, J., Thompson, G., Johnson, C., Gebrehiwot, A., 2018. Challenges and opportunities for UAV-based digital elevation model generation for flood-risk management: a case of Princeville, North Carolina. *Sensors*, 18(11), 3843.

Jamali, A., Roy, S. K., Beni, L. H., Pradhan, B., Li, J., Ghamisi, P., 2024. 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.

Jung, M., Chung, M., Kim, Y., 2019. Assessing complex damage using pre-disaster optical and post-disaster polsar data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 181–185.

May, S., Dupuis, A., Lagrange, A., De Vieilleville, F., Fernandez-Martin, C., 2022. Building damage assessment with deep learning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 1133–1138.

Salem, A., Beni, L. H., Fawakherji, M., 2023. Multimodal data fusion for flood monitoring: A case study of hurricane florence. *AGU Fall Meeting Abstracts*, 2023, NH02–08.

Tu, J., Sui, H., Feng, W., Song, Z., 2016. Automatic building damage detection method using high-resolution remote sensing images and 3D GIS model. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, 43–50.

Voelsen, M., Rottensteiner, F., Heipke, C., 2024. Transformer models for land cover classification with satellite image time series. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 92(5), 547–568.

Xiong, J., Liu, F., Wang, X., Yang, C., 2024. Siamese transformer-based building change detection in remote sensing images. *Sensors*, 24(4), 1268.

Yuan, X., Azimi, S., Henry, C., Gstaiger, V., Codastefano, M., Manalili, M., Cairo, S., Modugno, S., Wieland, M., Schneibel, A. et al., 2021. Automated building segmentation and damage assessment from satellite images for disaster relief. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 741–748.

Zhang, X., Gong, Y., Lu, J., Wu, J., Li, Z., Jin, D., Li, J., 2023. Multi-modal fusion technology based on vehicle information: A survey. *IEEE Transactions on Intelligent Vehicles*, 8(6), 3605–3619.