Detecting Building Damage in Gaza Using Sentinel-1 Data with Machine Learning

Matteo Larrode

Abstract

This study evaluates machine learning models for detecting building damage in Gaza using Sentinel-1 Synthetic Aperture Radar (SAR) time series data. Through a pixel-wise classification framework that processes patterns of backscatter change, the random forest classifier achieves an F1-score of 66.3% and AUC of 72.3% when validated against unseen UNOSAT damage assessments, and demonstrates cross-location transferability between different urban environments. The method extracts statistical features from pre-conflict (October 2022 - October 2023) and post-onset of the conflict (November 2023 - February 2024) time windows to capture backscatter signatures resulting from structural damage. The proposed methodology tackles humanitarian actors' need for rapid information in active conflict zones by utilising freely available data unaffected by weather conditions.

1 Introduction

The ongoing conflict in Gaza has resulted in widespread destruction of infrastructure, with 60% of buildings estimated to be damaged or destroyed, and 90% of the population displaced (Graham-Harrison et al., 2025). Early and accurate detection of building damage is crucial for humanitarian response planning, resource allocation, and eventual reconstruction efforts in conflict zones. However, traditional damage assessment methods, which rely on field surveys and ground teams, are virtually impossible to deploy in active conflict zones due to security concerns and access restrictions. This limitation is particularly acute in Gaza, where humanitarian access has been severely constrained (BRC, 2025).

Satellite remote sensing has emerged as a prominent alternative for monitoring conflict-affected areas. While very high-resolution (VHR) optical imagery offers detailed visual information, its acquisition is expensive, often impeded by cloud cover, and requires significant human resources for manual interpretation. Conversely, Synthetic Aperture Radar (SAR) imagery offers distinct advantages for conflict monitoring. As an active sensor system that emits pulses and captures backscattered signals, SAR can operate regardless of weather conditions or time of day (ESA, 2024). Freely available Sentinel-1 SAR data presents an opportunity for sustainable, large-scale monitoring that overcomes the limitations of optical imagery.

This study asks: How accurately can building damage be detected by applying machine learning to SAR data in the Gaza conflict? By adapting methodologies that have shown promise in other conflict settings, I assess their effectiveness in identifying structural changes in Gaza's densely built urban environment. Findings demonstrate that machine learning applied to Sentinel-1 SAR time series can detect building damage in Gaza with moderate success (F1: 66.3%, AUC: 72.3%), with cross-location transferability between different urban environments, and important implications for humanitarian response and damage assessment in ongoing and future conflicts.

2 Related works

2.1 Remote monitoring of building damage using satellite imagery

Advances in machine learning have enabled the automation of damage assessment using very high-resolution (VHR) optical imagery, particularly for natural disasters. The release of the xBD dataset, containing over 850,000 annotated building footprints, has facilitated significant progress in this domain (Gupta et al., 2019). Deep learning approaches have demonstrated high accuracy for damage induced by both natural and anthropogenic causes (Gholami et al., 2022; Gueguen & Hamid, 2015; Kaur et al., 2023; Shen et al., 2021). Holail et al. (2024) developed a Siamese convolutional neural network with semi-supervised learning, finding that about 58% of the total housing units in the Gaza Strip had been destroyed (as of March 2024).

However, conflict-induced damage presents distinct challenges compared to natural disasters. While natural disasters typically involve localized destruction from single events, war damage accumulates gradually across large areas over extended periods (Sticher et al., 2024). Mueller et al. (2021)'s study of the Syrian civil war achieved an AUC of 0.9 using convolutional neural networks (CNN) trained on VHR imagery and UNOSAT labels, but revealed significant generalisation challenges, with accuracy dropping by up to 29% when models trained on one city were applied to others. Additionally, the cost of VHR imagery makes large-scale, continuous monitoring prohibitively expensive for most humanitarian actors. Indeed, unlike in the case of natural disasters, commercial providers usually do not provide open access to their data in the case of conflicts (Dietrich et al., 2025).

2.2 SAR for damage assessment

Synthetic Aperture Radar (SAR) offers a promising alternative to optical imagery for damage assessment, with weather-independent imagery and regular revisit times (ESA, 2024). SAR-based damage detection methods, which have also mostly focused on response to natural disasters (Ge et al., 2019), broadly fall into two categories: amplitude-based and coherence-based approaches.

Amplitude-based methods compare backscatter intensity changes between pre- and post-event images (Matsuoka & Yamazaki, 2004; Uprety et al., 2013). These approaches detect significant changes in buildings' scattering mechanisms, such as the transition from double-bounce scattering (typical of intact structures) to diffuse scattering (characteristic of rubble). However, backscatter amplitude can be affected by environmental factors (e.g. snow), and by properties such as the roofing of the building, undermining the detection of damage when only using a pair of images (Arciniegas et al., 2007; Liu & Yamazaki, 2017).

Coherence-based methods utilise both amplitude and phase information through interferometry (InSAR). These techniques can detect minute changes due to Sentinel-1's C-band signals with 5.6-centimetre wavelength, making them generally more accurate than their counterparts (Gamba et al., 2007; Mastro et al., 2022; Stephenson et al., 2021), especially for earthquake damage assessment (Matsuoka & Yamazaki, 2000; Yonezawa & Takeuchi, 2001). However, this sensitivity to minor change often leads coherence methods to produce high false positive rates (Meyer, 2019). Another drawback of these methods is the necessary reduction in resolution, from 10 to 40 metres per pixel – a resolution at which most buildings are smaller than individual pixels (Ballinger, 2024).

2.3 SAR-based damaged assessment during armed conflict

Recent studies have applied SAR-based methods specifically to armed conflicts. For Ukraine, Aimaiti et al. (2022) achieved 58% accuracy on UNOSAT labels in the Kyiv region using features from Sentinel-1 and Sentinel-2, while Huang et al. (2023) reached approximately 60% accuracy for destruction in Mariupol using optical and coherence SAR time series. Both using coherence-based approaches, but at different scales, Tavakkoliestahbanati et al. (2024) monitored damage from the Kakhovka dam collapse, and Scher and Van Den Hoek (2023) generated nationwide coherence maps.

Finally, I turn to the two most comparable studies to mine. Ballinger (2024) developed the Pixel-Wise T-Test (PWTT), an unsupervised approach comparing pre- and post-conflict Sentinel-1 SAR time series. This lightweight method achieved F1 scores of 0.68 (AUC: 75.7) and 0.64 (AUC: 0.81) for Ukraine and Gaza respectively, demonstrating that even simple techniques can effectively detect battle damage when applied to multi-temporal data. Building on this approach, Dietrich et al. (2025) trained a random forest classifier using features extracted from Sentinel-1 time series. Their method, implemented on Google Earth Engine, achieved an F1 score of 0.75 (AUC: 81.3) for building-level damage detection in Ukraine.

These approaches and the present study tackle key challenges in conflict monitoring: the need for verifiable impartiality with reproducible, explainable results based on open-access data; the cost barriers of VHR imagery; and the requirement for continuous monitoring over extended periods. Moreover, by leveraging Sentinel-1 time series rather than single image pairs, the approach mitigates the susceptibility of amplitude-based methods to dielectric variations from environmental factors (Canty et al., 2019). Simultaneously, it avoids the resolution degradation and high false positive rates characteristic of coherence-based techniques (Meyer, 2019), maintaining the original 10×10m resolution useful for building-level analysis in Gaza's dense urban landscape.

3 Methods

As shown in Fig. 1, the specific area of interest (AOI) of this study is a section of Gaza City of about 36 km^2 that experienced significant destruction following the onset of hostilities on October 7, 2023, and particularly after the siege of Gaza City began on November 2, 2023 (BBC, 2025). The analysis concentrates on building damage that occurred between November 2023 and January 2024.

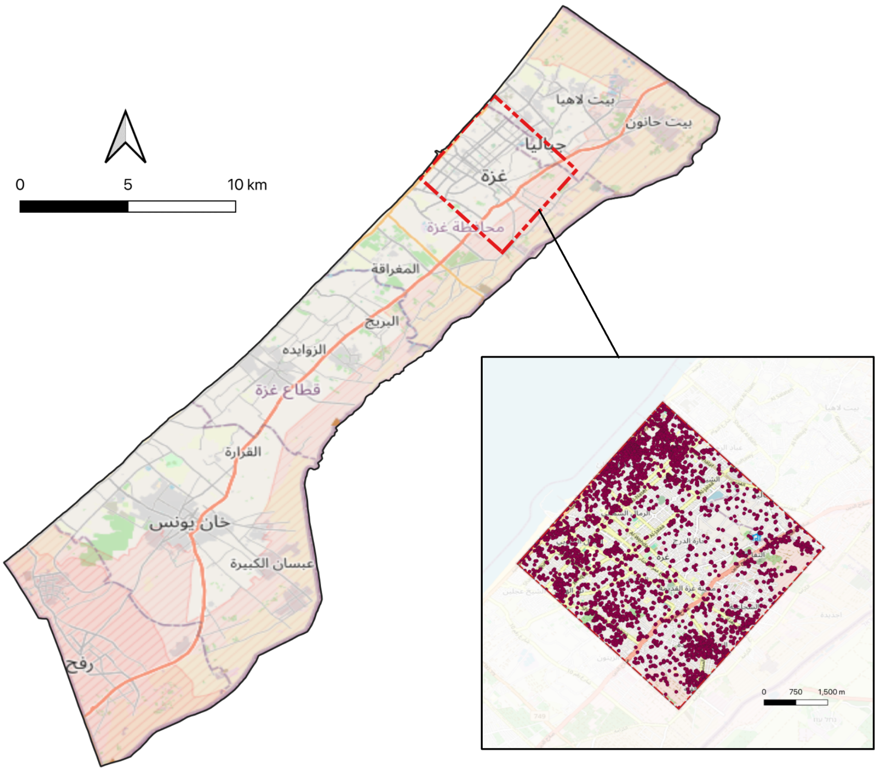

Figure 1. Area of interest for this study within the Gaza Strip. The inset highlights damaged buildings, as highlighted by UNOSAT on November 26th, 2023. Background map: OpenStreetMap.

3.1 Data

3.1.1 Damage labels and pseudo-negative labelling process

For positive damage labels, I used the public United Nations Satellite Centre (UNOSAT) damage assessment dataset from November 26, 2023 (UNOSAT, 2024). UNOSAT analysts produce these assessments by manually comparing pre- and post-event very high-resolution commercial satellite imagery using a standardized damage classification scale (Ballinger, 2024; Dietrich et al., 2025). Following Dietrich et al. (2025), I focused only on the two most severe damage categories—buildings classified as "destroyed" or "severely damaged" – as these are most likely to produce detectable changes in SAR imagery. These point-based labels only indicate locations of damaged buildings without information on the precise date of destruction or their spatial extent. For this study, I assumed that all observed damage occurred after October 7, 2023, starting date of the conflict.

For external validation purposes, a second independent UNOSAT damage assessment from January 6, 2024, covering Khan Yunis was also utilised. This dataset provided an opportunity to test the model's spatial and temporal transferability to a less densely built urban area approximately six weeks after the training data period.

Importantly, UNOSAT labels focus on damaged or destroyed buildings. To create a balanced training dataset, I developed a systematic approach for identifying *undamaged buildings*. Unlike Ballinger (2024), who assumed that all unlabelled building footprints were undamaged, I implemented a more conservative strategy to reduce potential mislabelling. First, I obtained building footprints for the AOI from a dataset curated by Ballinger (2024) sourced from he Microsoft Building Footprints dataset (Microsoft, 2025). I then created a 30-metre buffer around all UNOSAT damage points to account for potential proximity effects. Buildings falling outside this buffer were considered candidates for negative (undamaged) labels. From these candidates, I randomly sampled building centroids to create a dataset with a balanced ratio of damaged to undamaged points.

3.1.2 Satellite data

Finally, this study uses Synthetic Aperture Radar (SAR) data from the European Space Agency's Sentinel-1 mission (ESA, 2024), which provides weather-independent and day/night imaging capabilities critical for monitoring in conflict zones (Ballinger, 2024; Dietrich et al., 2025). The Sentinel-1 mission consists of two satellites moving in the same orbital plane, phased 180° apart, which each utilise C-band SAR with distinct polarisations: single co-polarisation (vertical transmit/vertical receive, or VV) and dual-band cross-polarisation (vertical transmit/horizontal receive, or VH). These polarisations capture different scattering mechanisms, such as double-bounce scattering from intact vertical structures, or diffuse scattering from rubble (Meyer, 2019). Both polarisations are included in the analysis, as their combination has been found to improve building damage detection accuracy (Karimzadeh & Mastuoka, 2017).

I accessed the Ground Range Detected (GRD) products from the Alaska Satellite Facility[1] in Interferometric Wide-Swath mode at 10m spatial resolution. I then pre-processed these products, consisting of log-amplitudes for the VV and VH polarisations, using the

[1] Available at https://search.asf.alaska.edu

SNAP software (SNAP, 2025) with orbit correction, thermal and border noise removal, radiometric calibration, and terrain correction, as commonly done for this type of analysis (Dietrich et al., 2025).

3.2 Machine learning framework

The damage detection approach employs a pixel-wise binary classification model based on temporal changes in SAR backscatter.

3.2.1 Time series extraction

For each labelled location (both damaged and undamaged), I extracted pixel-wise Sentinel-1 backscatter values in both VV and VH polarisations over the whole study period (October 2022 to February 2024). Following Dietrich et al. (2025), I processed each orbital direction separately to account for SAR's side-looking geometry. This resulted in one multivariate time series x_i per reference point, leading to a total of about 8,300 distinct time series. See Fig. 2 for examples of time-series at selected UNOSAT locations.

For the purpose of this analysis, I then defined specific temporal acquisition windows within the study period:

- **Reference period (T_{ref})**: A 12-month window from October 5, 2022, to October 6, 2023, representing pre-conflict conditions. It is assumed that building sustained no damage during this period.
- **Assessment period (T_{post})**: The three-month window following the conflict onset, and after the UNOSAT labelling, from November 26, 2023, to February 10, 2024.

For each time series x_i, I distinguished two segments, corresponding to the reference period and the assessment period, respectively $x_{i,\text{ref}}$ and $x_{i,\text{post}}$. Then, for each segment, polarisation, and orbit, I computed seven statistical features: minimum, maximum, mean, median, standard deviation, kurtosis, skewness. These features characterise the statistical properties of backscatter signals and their changes between periods. For each location and orbit, this process generated a feature vector of 56 dimensions (7 features $\times$ 2 polarisations $\times$ 2 orbits $\times$ 2 time periods), denoted as $\phi(x_i) \in \mathbb{R}^{56}$.

Notably, only relying on pixel time series ignores any spatial context, crucial when analysing optical satellite imagery with CNNs for example. However, as argued and demonstrated by (Dietrich et al., 2025), time series analysis contains the majority of relevant information for damage detection, likely because war-induced structural damage tends to be localized and relatively small in scale.

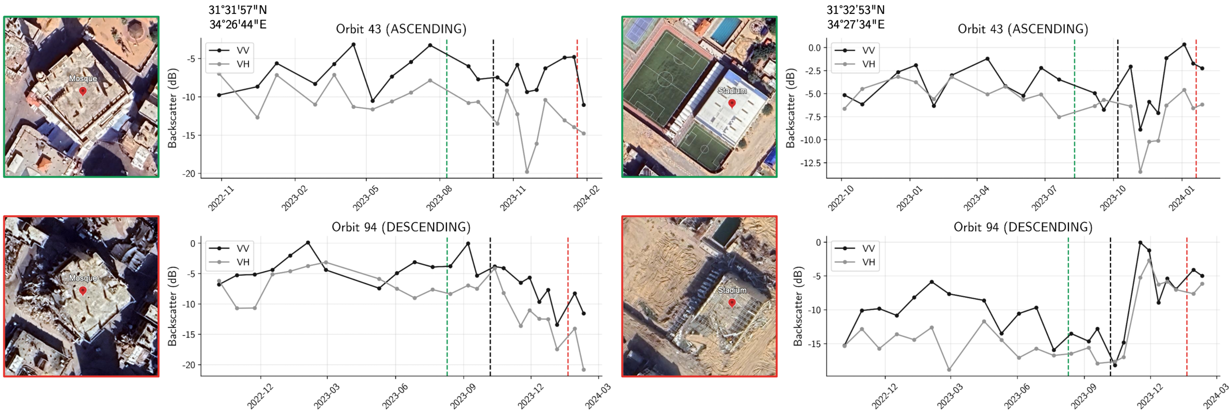

Figure 2. Example of Sentinel-1 time series. Sentinel-1 backscatter time series showing VV (black) and VH (gray) polarisations for selected locations in Gaza, with satellite imagery displaying pre-conflict (green border) and post-conflict (red border) conditions. Vertical dashed lines mark the onset of conflict on October 7, 2023 (black), the dates of the pre- and post-destruction images (respectively green and red). This figure indicates that building damage can produce varied SAR signatures depending on flight direction and incidence angle: some cases show clear backscatter drops, others show increases or more complex patterns. The multi-orbit, multi-polarisation approach captures these diverse signatures. Satellite imagery: Google Earth.

3.2.2 Classification model

Building damage detection was formulated as a supervised binary classification problem, where the model estimates the probability $\hat{y}_i \in [0, 1]$ that damaged occurred between the reference and assessment period for each feature vector.

The labelled dataset was split into training (70%) and test sets (30%), while maintaining class balance through stratification. After standardising features in the training data, a random search with 5-fold stratified cross-validation was conducted to optimise the hyperparameters of the random forest classifier. The optimal combination consisted of 200 estimators, a minimum of 10 samples required to split an internal node, a minimum of 4 samples required to be at a leaf node, and no maximum depth constraint. See Figure C.1 and C.2 for more details. Notably, other classification models were considered and optimised with random searches, including different neural network architectures. The random forest showed the best performance, confirming its already-demonstrated effectiveness for SAR-based damage detection (Dietrich et al., 2025).

To evaluate model performance, I calculated precision, recall, F1-score, and Area Under the Curve (AUC) metrics on the held-out test set. Additionally, I analysed the impact of different classification thresholds on the precision/recall trade-off.

4 Results

Table B.1 displays descriptive statistics of the SAR backscatter (dB) values used as features in our model, and an apparent absence of consistent patterns between damaged and undamaged buildings. As previously observed in the literature (Dietrich et al., 2025), the same damage event can produce significantly different backscatter responses depending on various factors. This variability in SAR signatures highlight the value of utilising a machine learning approaches, which can identify subtle, non-linear relationships across multiple dimensions that would otherwise remain undetectable through traditional analysis methods.

4.1 Building damage detection

Table 1 presents the performance metrics of the random forest model for detecting building damage in Gaza. The model achieves an overall accuracy of 67.3% with an AUC of 72.3%, demonstrating moderate discriminative ability. With the classification probability threshold set at 0.5, the precision of 68.5% and recall of 64.2% lead to an F1-score of 66.3%. These results indicate that while the model can detect building damage from SAR data with reasonable reliability, there remains substantial room for improvement.

Label	Precision	Recall	F_1	Support	Accuracy	AUC
Damaged	68.5 (80.9)	64.2 (32.1)	66.3 (46.0)	1267	67.3 (62.3)	72.3
Undamaged	66.3 (57.7)	70.5 (92.4)	68.3 (71.0)	1268		

Table 1. Evaluation metrics of the damage detection model (%). Performance based on a balanced test sample of previously unseen UNOSAT annotations and pseudo-negative points sampled from buildings 30m away from damaged buildings. The threshold decision value is 0.5, with values in parentheses indicate performance with a confidence threshold of 0.65, reaching 80% precision.

As visualised in Fig. 3, increasing the threshold from 0.5 to 0.65 allows to reach a target precision (proportion of positive predictions that are correct) of 80%, at the cost of halving recall (proportion of positive labels that are correctly classified) from 64% to 32%. While 0.5 was set as the baseline threshold, for the balance it achieves in this trade-off plotted in Fig. D.1, different applications of this model might warrant choosing other thresholds. For instance, a rather conservative threshold, such as 0.65, could be more appropriate when using this model across the whole Gaza strip to map aggregate patterns of destruction (as Scher and Van Den Hoek (2023) did with Ukraine). Conversely, a lower threshold may prove more useful to detect overlooked or isolated damage at smaller scales.

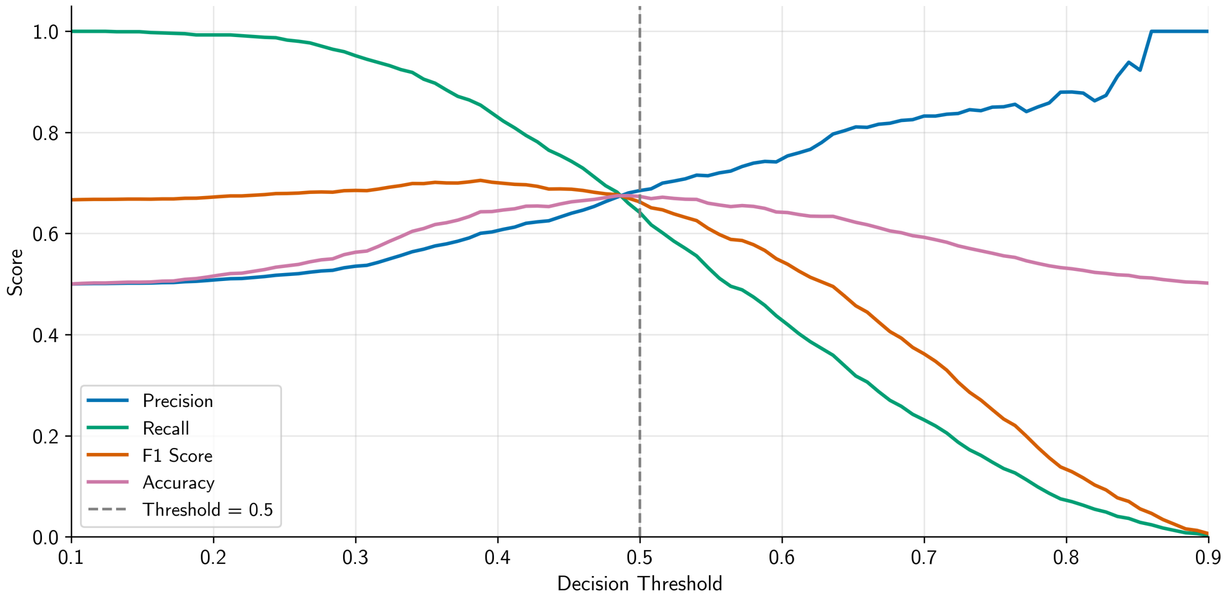

Figure 3. Performance of the model for different classification thresholds.

4.2 External validation in Khan Yunis

To further evaluate the model's generalisability, I conducted an external validation by applying the random forest classifier trained on Gaza City to a different geographic area and time period. Specifically, I tested the model on Khan Yunis using UNOSAT damage assessments from January 6, 2024.

This cross-location validation serves two important purposes: testing spatial transferability within Gaza to a less densely built urban area, and assessing temporal robustness as the conflict evolved. As shown in Figure 4, the model maintained relatively consistent performance despite these shifts in context, achieving an F1-score of 63.6% and AUC of 71.1%. These metrics are only marginally lower than those observed in the original test set (F1: 66.3%, AUC: 72.3%), suggesting good generalisability of the approach.

These results are particularly encouraging given the differences between the two urban environments. Khan Yunis features a less compact urban layout than Gaza City, with more dispersed buildings and different structural patterns. Additionally, the nature of destruction had evolved by January 2024, with different weapons and tactics potentially producing varied damage signatures in the SAR data.

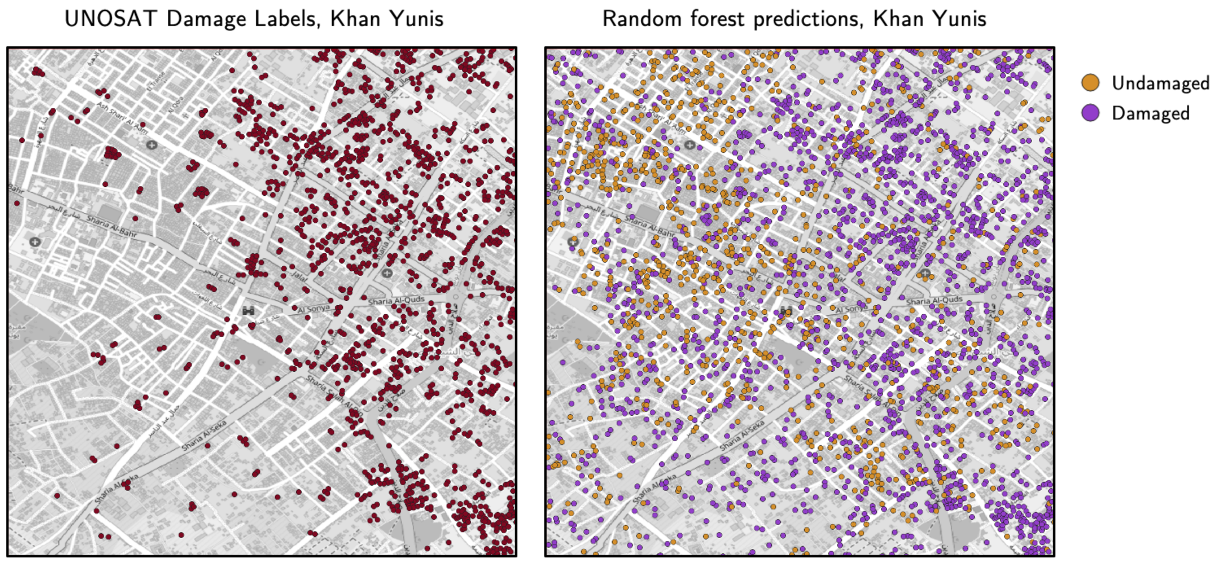

Figure 4. Comparison of UNOSAT damage labels and random forest predictions for Khan Yunis. Left: UNOSAT damage assessment points from January 6, 2024, showing buildings classified as destroyed or severely damaged (red). Right: Results of the random forest model trained on Gaza City data and applied to Khan Yunis, with predicted damaged (purple) and undamaged (orange) buildings. Threshold at 0.5. Background map: OpenStreetMap

5 Discussion

The results demonstrate that machine learning applied to Sentinel-1 SAR time series data can detect building damage in Gaza with moderate success (F1-score: 66.3%, AUC: 72.3%), providing a potentially valuable tool for humanitarian organizations operating in regions with limited ground access. Aligning with Dietrich et al. (2025), the random forest classifier outperformed other tested models, including neural networks, suggesting that ensemble methods are well suited to handle SAR data. Performance metrics of the model trained in this study are somewhat lower than those reported by Ballinger (2024) for Gaza (F1: 64%, AUC: 81%), and Dietrich et al. (2025) for Ukraine (F1: 74.9%, AUC: 81.3%). This difference may be attributed to several factors. First, this paper's model was trained on a single damage assessment instance rather than using dynamic labelling across multiple assessment windows as in Dietrich et al. (2025). As a consequence, damage happening during the assessment period (between November 26th, 2023 and February 10th, 2024) to buildings that had been labelled as undamaged might have hindered performance of the model. Additionally, the extreme density of Gaza's urban environment creates challenges for 10m resolution SAR data, with buildings frequently smaller than individual pixels.

Encouragingly, the external validation in Khan Yunis demonstrated the model's generalisability across both space and time. Despite differences in urban density and building patterns between Gaza City and Khan Yunis, and the six-week temporal gap between assessments, the model maintained consistent performance (F1: 63.6%, AUC: 71.1%). This suggests that the statistical features extracted from SAR time series capture fundamental signatures of building damage that transcend specific urban contexts and conflict phases, supporting the approach's potential applicability across the entire Gaza Strip.

For humanitarian applications, utilising freely available Sentinel-1 data eliminates the financial barriers associated with commercial VHR imagery, making continuous monitoring financially sustainable even for resource-constrained organisations. The model's performance could position it as a complementary tool in a broader damage assessment workflow. With a default threshold of 0.5 balancing precision and recall, it could serve as an initial screening mechanism to identify areas requiring closer examination with higher-resolution imagery. Adjusting the threshold to 0.65 increases precision to 80.9%, making it more suitable for generating reliable aggregate damage statistics, though at the cost of missing over two-thirds of damaged buildings.

However, significant limitations must be acknowledged. At 10m resolution, Sentinel-1 cannot reliably detect damage to small structures, which constitute a substantial portion of Gaza's building stock. The lack of precise destruction dates in the UNOSAT dataset complicates both training and evaluation, potentially introducing noise into the model. Moreover, the merging of multiple data sources – each with its own geographic accuracy limitations – introduces compound spatial errors that likely impact model performance. UNOSAT damage annotations, while manually verified by analysts, rely on georeferenced very high-resolution imagery that itself contains inherent positional uncertainties of 1-3 meters (UNOSAT, 2024). When these points are used to train models operating on Sentinel-1 data with 10-meter pixel resolution, even minor misalignments can assign damage labels to incorrect pixels. This problem is exacerbated by SAR's side-looking geometry, which creates distortions that vary with incidence angle and orbit direction, potentially displacing features by several metres, especially in areas with rapid elevation changes or tall structures (Ballinger, 2024).

This study highlights several promising avenues for future research. Integrating coherence based features with amplitude-based methods could potentially enhance detection accuracy, though at the cost of reduced spatial resolution. This trade-off might be acceptable for certain humanitarian applications focusing on neighbourhood-level damage patterns rather than individual buildings. Moreover, developing transfer learning approaches to adapt models trained on data-rich conflicts (like Ukraine or Gaza) to emerging crises with limited reference data would be valuable for rapid deployment in new emergency settings. Perhaps most critically, this research underscores the need for open standardised datasets for conflict damage assessment. While natural disaster response benefits from resources like the xBD dataset (Gupta et al., 2019), equivalent benchmark datasets for conflict zones, like the one developed by Ballinger (2024) would accelerate methodological advancements in this area.

6 Conclusion

This study has demonstrated that machine learning applied to Sentinel-1 SAR time series can detect building damage in Gaza with moderate success (F1: 66.3%, AUC: 72.3%). While performance is constrained by the complex urban environment and limited annotated data, the approach offers significant advantages: weather-independent monitoring, free data access, and regular coverage without requiring ground access. As conflicts continue to impact civilian infrastructure worldwide, the ability to detect building damage remotely has significant implications for humanitarian response, damage assessment, and reconstruction planning in conflict zones where ground access is limited. Future improvements could focus on developing transfer learning approaches, and establishing standardised benchmark datasets specifically for conflict damage assessment.

References

Aimaiti, Y., Sanon, C., Koch, M., Baise, L. G., & Moaveni, B. (2022). War Related Building Damage Assessment in Kyiv, Ukraine, Using Sentinel-1 Radar and Sentinel-2 Optical Images. *Remote Sensing*, *14*(24), 6239.

Arciniegas, G. A., Bijker, W., Kerle, N., & Tolpekin, V. A. (2007). 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.

Ballinger, O. (2024). Open Access Battle Damage Detection via Pixel-Wise T-Test on Sentinel-1 Imagery. *arXiv:2405.06323*.

BBC. (2025). Gaza Strip in maps: How 15 months of war have drastically changed life in the territory. *BBC News*. Retrieved April 26, 2025, from https://www.bbc.com/news/world-middle-east-20415675

BRC. (2025). How is aid getting into Gaza? Retrieved May 1, 2025, from https://www.redcross.org.uk/stories/disasters-and-emergencies/world/how-is-aid-getting-into-gaza

Canty, M. J., Nielsen, A. A., Conradsen, K., & Skriver, H. (2019). Statistical analysis of changes in Sentinel-1 time series on the Google Earth Engine. *Remote Sensing*, *12*(1), 46.

Dietrich, O., Peters, T., Sainte Fare Garnot, V., Sticher, V., Ton-That Whelan, T., Schindler, K., & Wegner, J. D. (2025). An open-source tool for mapping war destruction at scale in Ukraine using Sentinel-1 time series. *Communications Earth & Environment*, *6*(1), 1–10.

ESA. (2024). Sentinel-1 - Missions - Sentinel Online, 2024. Retrieved April 30, 2025, from https://sentiwiki.copernicus.eu/web/s1-mission

Gamba, P., Dell'Acqua, F., & Trianni, G. (2007). 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.

Ge, P., Gokon, H., Meguro, K., Graduate School of Engineering, The University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan, & Institute of Industrial Science, The University of Tokyo, Tokyo, Japan. (2019). Building Damage Assessment Using Intensity SAR Data with Different Incidence Angles and Longtime Interval. *Journal of Disaster Research*, *14*(3), 456–465.

Gholami, S., Robinson, C., Ortiz, A., Yang, S., Margutti, J., Birge, C., Dodhia, R., & Ferres, J. L. (2022). On the deployment of post-disaster building damage assessment tools using satellite imagery: A deep learning approach. *IEEE International Conference on Data Mining Workshops (ICDMW)*, 1029–1036.

Graham-Harrison, E., Scruton, P., Swan, L., Ahmedzade, T., Sheehy, F., & Boulinier, L. (2025). A visual guide to the destruction of Gaza. *The Guardian*. Retrieved May 1, 2025, from https://www.theguardian.com/world/2025/jan/18/a-visual-guide-to-the-destruction-of-gaza

Gueguen, L., & Hamid, R. (2015). Large-scale damage detection using satellite imagery. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1321–1328.

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., & Gaston, M. (2019). xBD: A Dataset for Assessing Building Damage from Satellite Imagery. *arXiv:1911.09296*.

Holail, S., Saleh, T., Xiao, X., Xiao, J., Xia, G.-S., Shao, Z., Wang, M., Gong, J., & Li, D. (2024). Time-series satellite remote sensing reveals gradually increasing war damage in the Gaza Strip. *National Science Review*, *11*(9), nwae304.

Huang, Q., Jin, G., Xiong, X., Ye, H., & Xie, Y. (2023). Monitoring urban change in conflict from the perspective of optical and SAR satellites: The case of Mariupol, a city in the conflict between RUS and UKR. *Remote Sensing*, *15*(12), 3096.

Karimzadeh, S., & Mastuoka, M. (2017). Building damage assessment using multisensor dual-polarized synthetic aperture radar data for the 2016 M 6.2 Amatrice Earthquake, Italy. *Remote Sensing*, *9*(4), 330.

Kaur, N., Lee, C.-C., Mostafavi, A., & Mahdavi-Amiri, A. (2023). Large-scale building damage assessment using a novel hierarchical transformer architecture on satellite images. *Computer-Aided Civil and Infrastructure Engineering*, *38*(15), 2072–2091.

Liu, W., & Yamazaki, F. (2017). Extraction of collapsed buildings in the 2016 Kumamoto earthquake using multi-temporal PALSAR-2 data. *Journal of Disaster Research*, *12*(2), 241–250.

Mastro, P., Masiello, G., Serio, C., & Pepe, A. (2022). Change Detection Techniques with Synthetic Aperture Radar Images: Experiments with Random Forests and Sentinel-1 Observations. *Remote Sensing*, *14*(14), 3323.

Matsuoka, M., & Yamazaki, F. (2000). Use of interferometric satellite SAR for earthquake damage detection. *Sat*, *2*, z1.

Matsuoka, M., & Yamazaki, F. (2004). Use of satellite SAR intensity imagery for detecting building areas damaged due to earthquakes. *Earthquake Spectra*, *20*(3), 975–994.

Meyer, F. (2019). Spaceborne Synthetic Aperture Radar: Principles, data access, and basic processing techniques. *Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*, 21–64.

Microsoft. (2025, April). Microsoft/GlobalMLBuildingFootprints. Retrieved April 27, 2025, from https://github.com/microsoft/GlobalMLBuildingFootprints

Mueller, H., Groeger, A., Hersh, J., Matranga, A., & Serrat, J. (2021). Monitoring war destruction from space using machine learning. *Proceedings of the National Academy of Sciences*, *118*(23), e2025400118.

Scher, C., & Van Den Hoek, J. (2023). Nationwide mapping of damage to human settlements across Ukraine using Sentinel-1 InSAR coherence change detection. *AGU Fall Meeting Abstracts*, *2023*, GC23A–01.

Shen, Y., Zhu, S., Yang, T., Chen, C., Pan, D., Chen, J., Xiao, L., & Du, Q. (2021). 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.

SNAP. (2025). SNAP - ESA's SentiNel Application Platform. Retrieved May 1, 2025, from https://github.com/senbox-org

Stephenson, O. L., Köhne, T., Zhan, E., Cahill, B. E., Yun, S.-H., Ross, Z. E., & Simons, M. (2021). Deep learning-based damage mapping with InSAR coherence time series. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–17.

Sticher, V., Dietrich, O., Pfeifle, B., & Wegner, J. D. (2024). Watching Armed Conflicts from Space. *CSS Analyses in Security Policy*, *336*.

Tavakkoliestahbanati, A., Milillo, P., Kuai, H., & Giardina, G. (2024). Pre-collapse spaceborne deformation monitoring of the Kakhovka dam, Ukraine, from 2017 to 2023. *Communications Earth & Environment*, *5*(1), 145.

UNOSAT. (2024, December). UNOSAT Gaza Strip Comprehensive Damage Assessment. Retrieved April 27, 2025, from https://unosat.org/products/4047

Uprety, P., Yamazaki, F., & Dell'Acqua, F. (2013). Damage detection using high-resolution SAR imagery in the 2009 L'Aquila, Italy, Earthquake. *Earthquake Spectra*, *29*(4), 1521–1535.

Yonezawa, C., & Takeuchi, S. (2001). Decorrelation of SAR data by urban damages caused by the 1995 Hyogoken-nanbu earthquake. *International Journal of Remote Sensing*, *22*(8), 1585–1600.

Appendix A. Data collection details

Data Type	Source	Description
SAR Imagery	Sentinel-1 GRD (Alaska Satellite Facility)	C-band SAR imagery, 10m resolution, VV and VH polarizations. Acquisitions from Oct 2022 to Feb 2024.
Damage Labels	UNOSAT Damage Assessment (UNOSAT, 2024)	Point-based damage annotations from November 26, 2023, focusing on buildings classified as "destroyed" or "severely damaged."
Building Footprints	Microsoft (2025)	Building polygon geometries used for generating negative labels and performing building-level aggregation.

Table A.1. Data sources used in this study.

Period	Date Range	Role
T_{ref}	Oct 5, 2022 - Oct 6, 2023	Reference period (pre-conflict)
-	Oct 7, 2023 - Nov 25, 2023	Discarded post-conflict onset period before UNOSAT labelling
T_{post}	26 Nov, 2023 - Feb 10, 2024	Assessment period

Table A.2. Temporal windows used for the analysis.

Appendix B. Descriptive statistics of the model features dataset

			Undamaged		Damaged	
Orbit	**Pol.**	**Statistic**	$T_{\mathbf{ref}}$	$T_{\mathbf{post}}$	$T_{\mathbf{ref}}$	$T_{\mathbf{post}}$
Sample Size			4224	4224	4224	4224
ASC Orbit						
	VH	Mean	-11.24 ± 2.36	-11.54 ± 2.10	-11.89 ± 2.52	-12.46 ± 2.37
		Median	-11.12 ± 2.36	-11.36 ± 2.23	-11.73 ± 2.48	-12.29 ± 2.38
		Min	-15.43 ± 3.31	-14.96 ± 2.81	-16.26 ± 3.77	-15.85 ± 3.59
		Max	-7.76 ± 2.32	-8.59 ± 2.25	-8.32 ± 2.45	-9.56 ± 2.31
		Std	2.21 ± 0.62	2.43 ± 0.97	2.28 ± 0.69	2.40 ± 1.11
	VV	Mean	-5.99 ± 2.23	-5.96 ± 1.93	-6.64 ± 2.34	-6.11 ± 1.85
		Median	-5.89 ± 2.25	-5.80 ± 2.11	-6.54 ± 2.35	-5.98 ± 1.99
		Min	-9.87 ± 2.71	-9.26 ± 2.32	-10.57 ± 2.89	-9.24 ± 2.49
		Max	-2.59 ± 2.33	-3.11 ± 2.19	-3.20 ± 2.41	-3.33 ± 1.99
		Std	2.09 ± 0.52	2.33 ± 0.88	2.11 ± 0.51	2.25 ± 0.89
DESC Orbit						
	VH	Mean	-12.26 ± 2.79	-11.94 ± 2.41	-12.74 ± 3.17	-12.88 ± 2.76
		Median	-12.07 ± 2.72	-11.86 ± 2.50	-12.51 ± 3.03	-12.73 ± 2.77
		Min	-16.99 ± 4.71	-15.58 ± 3.27	-17.79 ± 5.65	-16.66 ± 4.44
		Max	-8.52 ± 2.71	-8.50 ± 2.64	-8.97 ± 3.02	-9.60 ± 2.52
		Std	2.41 ± 0.92	2.61 ± 1.06	2.50 ± 1.09	2.60 ± 1.33
	VV	Mean	-6.70 ± 2.54	-6.31 ± 2.23	-7.13 ± 2.93	-6.39 ± 2.26
		Median	-6.63 ± 2.56	-6.30 ± 2.38	-7.04 ± 2.94	-6.34 ± 2.45
		Min	-10.69 ± 3.04	-9.86 ± 2.73	-11.22 ± 3.59	-9.67 ± 2.86
		Max	-3.06 ± 2.78	-2.74 ± 2.67	-3.48 ± 3.09	-3.24 ± 2.28
		Std	2.18 ± 0.59	2.62 ± 0.98	2.20 ± 0.62	2.37 ± 0.94

Table B.1. Descriptive statistics of SAR backscatter (dB) for undamaged and damaged buildings, by orbit and polarization. Values are reported as mean ± standard deviation.

Appendix C. Results of the random searches

Random forests

n_estimators	min_samples_split	min_samples_leaf	max_depth	Mean F_1	Std F_1
200	10	4	None	0.663	0.005
200	2	1	20	0.658	0.010
200	10	1	None	0.658	0.012
100	5	2	20	0.656	0.010
100	10	4	20	0.656	0.007

Figure C.1. Hyperparameter combinations of the five best performing random forests after random searches. Mean and standard deviation of F_1 score are based on 5-fold cross-validation.

Neural networks

Learning Rate	Alpha	Activation	Hidden Layers	Mean F_1	Std F_1
0.01	0.0001	ReLU	(30, 10, 120)	0.622	0.007
0.01	0.001	ReLU	(60, 60)	0.610	0.017
0.01	0.0001	ReLU	(120, 60, 10)	0.608	0.018
0.01	0.001	ReLU	(120, 30)	0.606	0.013
0.001	0.0001	ReLU	(60, 120, 30)	0.606	0.019

Figure C.2. Top five neural network configurations from random search. Mean and standard deviation of F_1 score are based on 5-fold cross-validation.

Appendix D. Additional model performance metrics

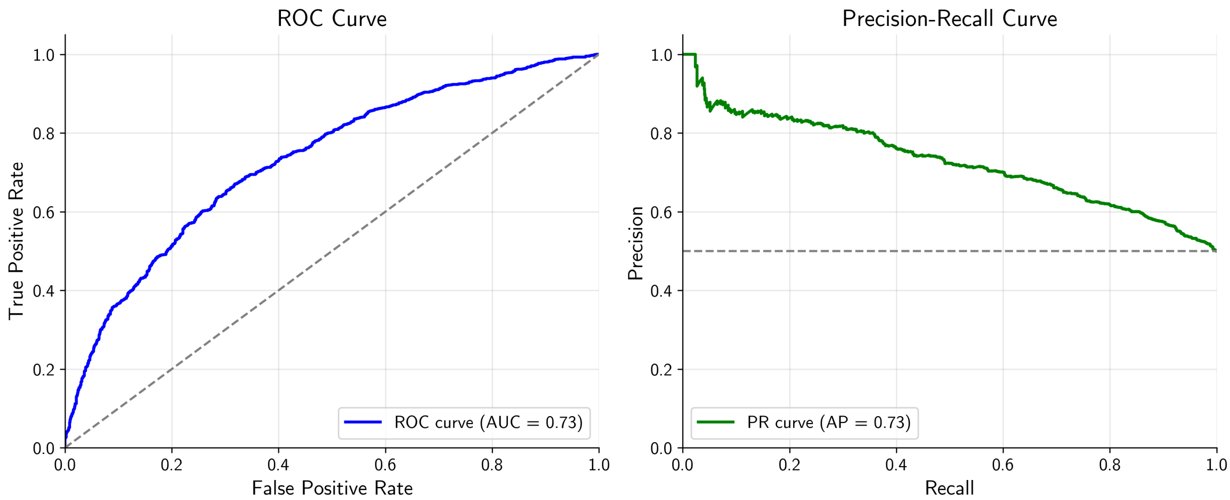

Figure D.1. ROC and Precision-Recall curves. The Receiver Operating Characteristic (ROC) curve illustrates the relationship between true positive rate (specificity) and false positive rate (1-specificity) across various classification thresholds. The Precision-Recall curve displays the trade-off between precision (positive predictive value) and recall (true positive rate) at different thresholds.

Appendix E. Code

For transparency and replicability, all code used for data pre-processing, visualisation, and analysis has been joined to this submission as zip file. It is structured as in a Github repository, with a `README.md` giving all necessary explanations to reproduce the findings. The repository can be found at: https://github.com/MatteoLarrode/BuildingDamageViaSAR