

Enhancing Post-Disaster Damage Detection and Recovery Monitoring by Addressing Class Imbalance in Satellite Imagery Using Enhanced Super-Resolution GANs (ESRGAN)

Umut Lagap¹, Saman Ghaffarian¹

¹ Department of Risk and Disaster Reduction, University College London, Gower Street, London, WC1E 6BT UK –
umutlagap@gmail.com, s.ghaffarian@ucl.ac.uk

Keywords: Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN), Post-disaster damage detection and recovery monitoring, Remote sensing, Class imbalance mitigation.

Abstract

Access to very high-resolution (HR) satellite imagery is often limited, delayed, or cost-prohibitive, restricting accurate and timely post-disaster damage detection and recovery monitoring (PDDRM). Additionally, class imbalance in disaster classification datasets further complicates deep learning (DL)-based assessments. This study addresses these challenges by leveraging ESRGAN to enhance low-resolution (LR) satellite imagery, thereby improving damage classification accuracy and the ability to monitor post-disaster recovery over time with three state-of-the-art DL models: Vision Transformer (ViT), ConvNeXt, and MaxViT for PDDRM classification across four key recovery states: Not Damaged, Not Recovered, Recovered, and New Buildings. To generate super-resolution (SR) images, LR images were first paired with HR images to train ESRGAN. Numerical evaluations using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) between SR and HR images confirm that ESRGAN effectively reconstructs high-resolution features, with Not Damaged (PSNR: 29.2, SSIM: 0.78) and New Buildings (PSNR: 30.3, SSIM: 0.81) exhibiting the highest reconstruction quality. ESRGAN-generated SR images were then compared against LR images in terms of classification accuracy and reliability. The results demonstrate that SR improves classification accuracy and precision, particularly for ViT and ConvNeXt, with ViT achieving an accuracy of 84% and ConvNeXt 82% on SR images, compared to 79% and 78% on LR images. We also employed Grad-CAM++ visualizations to interpret model predictions, which highlighted reliability improvements in certain classes. This study demonstrates that SR is a scalable and cost-effective alternative to very high-resolution satellite imagery, reducing dependency on expensive data sources while improving classification accuracy for PDDRM.

1. Introduction

The growing frequency and intensity of disasters highlight the urgent need for accurate and timely assessment of both post-disaster damage and recovery (Lallemant et al., 2017). This comprehensive assessment is critical to minimizing the adverse effects of the disasters and improving community resilience by ensuring effective emergency response, resource allocation, and desired long-term recovery outcomes (Sheykhoumousa et al., 2019). Remote sensing data, particularly satellite imagery, has emerged as a powerful tool for post-disaster damage detection and recovery monitoring (PDDRM) by providing large-scale, high-frequency observations that enable systematic assessments across vast geographic areas (Ghaffarian and Emtehani, 2021). However, conventional damage detection studies focus solely on immediate post-disaster conditions, often neglecting the subsequent recovery phase. This narrow approach limits the ability to evaluate long-term resilience and prevents a full understanding of recovery processes, as it overlooks the gradual restoration of urban areas. By extending post-disaster damage detection to include recovery monitoring, it becomes possible to track progress, identify areas lagging in reconstruction, and support evidence-based strategies for resilient recovery (Ghaffarian and Emtehani, 2021).

Accurate and timely PDDRM present several challenges, particularly when relying on satellite imagery. One of the primary obstacles is the scarcity of very-high resolution satellite imagery immediately after disasters. Very-high resolution satellite data is often expensive, difficult to acquire in real-time, or subject to cloud cover and operational delays, restricting the ability to perform precise and timely assessments (Hodgson et al., 2010). Another critical challenge is class imbalance, as certain recovery

states, such as newly constructed buildings, are underrepresented in datasets (Ren et al., 2020). This imbalance might skew model predictions, leading to biases where less frequent recovery patterns may be misclassified or overlooked entirely. The temporal complexity of recovery monitoring further complicates the task as it requires tracking structural and environmental changes across at least three key time frames: before the disaster, at the event time, and after the disaster (Ghaffarian and Emtehani, 2021). Identifying and distinguishing between damage, ongoing recovery, and new construction over these time frames demands more sophisticated feature extraction techniques. Moreover, the role of image resolution in detecting fine-grained damage and recovery states remains unexplored. While very-high resolution images provide greater detail, their availability is inconsistent, and their actual impact on model performance has not been thoroughly investigated.

PDDRM requires robust, scalable, and efficient methodologies to process vast amounts of satellite imagery. Conventional approaches, such as manual interpretation and classical machine learning techniques, often struggle with the high volume of data and the complexity of identifying structural changes across multiple time frames (Sublime and Kalinicheva, 2019). To address these limitations, DL models have emerged as a powerful solution for automating post-disaster damage detection efforts. These models can learn hierarchical representations of satellite imagery, enabling the detection of subtle structural damage and reconstruction patterns that may not be easily identifiable through conventional methods (Chaudhuri and Bose, 2020). However, despite their accuracy and efficiency, DL models are often regarded as black boxes, providing little transparency into how decisions are made (Rai, 2020). This lack of interpretability poses challenges, particularly in post-disaster response and recovery

scenarios where trust and accountability are critical (Ghaffarian and Taghikhah, 2023). Explainable AI (XAI) methods aim to address interpretability and reliability issues of DL models by offering insights into model predictions, allowing decision-makers to understand which image features influence classification outcomes (Ghaffarian and Taghikhah, 2023; Kakogeorgiou and Karantzas, 2021).

DL techniques have been widely used in post-disaster damage detection due to their ability to handle complex spatial patterns and automate large-scale classification tasks (Wang et al., 2024). For example, Rahnemoonfar et al. (2023) introduced RescueNet, a high-resolution UAV dataset for post-disaster semantic segmentation, featuring images collected after Hurricane Michael. The study evaluated PSPNet, DeepLabv3+, Segformer, and Attention U-Net, with Attention U-Net achieving the highest mean Intersection over Union (mIoU) of 98.47%. Safavi and Rahnemoonfar (2022) evaluated real-time semantic segmentation models on FloodNet, a high-resolution UAV dataset collected after Hurricane Harvey (2017), containing images with pixel-level annotations for flood-related damage. The study compared encoder–decoder models (e.g., UNet, HarDNet, SegFormer) and two-pathway models (e.g., BiSeNet, DDRNet, PIDNet) for flood damage segmentation, and SegFormerB0 achieved the highest mIoU of 61.6% and pixel accuracy of 89.5%. Gupta et al. (2019) introduced xBD with severe class imbalance, the largest publicly available dataset for building damage assessment from pre- and post-disaster satellite imagery, covering 19 disasters, including hurricanes, wildfires, floods, tsunamis, earthquakes, volcanic eruptions, and tornadoes. The study implemented a U-Net-based localization model, achieving an IoU of 0.66 for buildings, and a ResNet50-based classifier obtaining an overall weighted F1 score of 0.2654. However, most previous DL research focuses on damage detection, which is crucial for immediate response, and long-term recovery tracking with DL remains underexplored.

XAI techniques have gained attention in remote sensing applications, offering greater interpretability for AI-driven models. Cheng et al. (2022) developed an uncertainty-aware convolutional neural network (CNN) for AI-assisted post-disaster damage assessment, trained on UAV-based DoriaNET and satellite-based xBD datasets. To enhance explainability, Grad-CAM was applied, revealing that the models primarily focus on roof damage and structural failures in high-wind disasters. Kakogeorgiou and Karantzas (2021) evaluated ten Explainable AI (XAI) techniques for multi-label classification in remote sensing, using DenseNet and ResNet trained on BigEarthNet and SEN12MS datasets. Their study found that Occlusion, Grad-CAM, and LIME provided the most reliable and interpretable explanations, with Grad-CAM achieving the lowest Max-Sensitivity score (0.14) and AUC-MoRF (23.62) for SEN12MS. Even though there have been advancements in applying XAI to post-disaster damage detection, its application to post-disaster recovery states, where structural conditions evolve over time, remains an underexplored area with significant potential.

SR techniques have been extensively researched in remote sensing to improve image quality, particularly when high-resolution data is unavailable (Wang et al., 2022). SRGAN, a class of DL models based on Generative Adversarial Networks (GAN), have demonstrated remarkable success in enhancing satellite imagery by reconstructing fine details lost in LR images. Xiong et al. (2020) developed an Improved SR GAN (ISRGAN) for remote sensing image SR to enhance the generalization capability across different locations and satellite sensors. The

model demonstrated strong generalization across geographic and sensor variations, achieving PSNR of 35.82 and SSIM of 0.99 in cross-location tests, and PSNR of 38.09 and SSIM of 0.99 in cross-sensor tests. Dou et al. (2020) developed a 3D Attention-based SR Generative Adversarial Network (3DASRGAN) for hyperspectral image SR, addressing the spectral distortion problem by integrating 3D convolution layers and an attention mechanism. Compared to traditional SR methods and CNN-based approaches, 3DASRGAN achieved the highest PSNR (32.27) and SSIM (0.97) for the Washington DC Mall dataset and PSNR (33.27) and SSIM (0.91) for the Urban dataset. Moreover, Fu et al. (2022) proposed an SRGAN-based end-to-end framework for enhancing post-disaster damage detection performance using the xBD dataset, specifically using the 2018 Sunda Strait tsunami and 2018 Sulawesi earthquake and tsunami events. The SR module enhanced image quality, with PSNR up to 36.22 and SSIM reaching 0.939, demonstrating that SRGAN can generate high-quality damage maps when high-resolution imagery is unavailable. Efforts to implement SRGAN techniques for post-disaster damage detection have been limited, and research on their application to post-disaster recovery monitoring remains entirely absent.

This study advances the field of post-disaster damage detection and recovery monitoring by addressing key challenges in remote sensing, DL, Explainable AI, Generative AI.

- This research extends the post-disaster damage detection analysis to include post-disaster recovery monitoring. By tracking structural changes over time, the study offers a more comprehensive understanding of recovery dynamics, shedding light on how affected areas transition from damaged to recovered.
- ESRRGAN-generated SR images significantly improve classification performance compared to LR images, particularly for challenging and underrepresented recovery classes, such as new construction and recovery stages. These enhancements help overcome the limitations of LR imagery in PDDRM.
- By training and evaluating models on both LR and HR images, this study systematically examines how image resolution affects classification accuracy. This analysis helps determine whether SR techniques truly enhance model performance or if high resolution is only beneficial in specific cases.
- The study integrates Explainable AI (XAI) techniques, such as Grad-CAM++, to analyze how DL models interpret damage and recovery states across different resolutions. This approach enhances model transparency and reliability, ensuring that AI-driven PDDRM are based on interpretable and meaningful visual features rather than unintended biases.

2. Methods and Materials

2.1 Study Area and Dataset

Typhoon Haiyan (locally known as Yolanda) was one of the most powerful tropical cyclones ever recorded, making landfall in the Philippines on November 8, 2013. With storm surges exceeding 5 meters, it caused widespread devastation, particularly in Tacloban, Leyte and resulted in over 7,000 fatalities and massive destruction of infrastructure, displacing millions and severely impacting coastal settlements (van Loenhout et al., 2018). Figure 1 shows the effects of Typhoon Haiyan on Tacloban, Leyte, and the subsequent recovery process. The left panel presents a regional map of the Philippines, showing Leyte Island and Tacloban City, a heavily affected area. The storm's trajectory and

impact zone are overlaid, with a red line marking the typhoon's path and varying shades indicating the intensity of damage across affected regions. The right panel comprises three very-high resolution satellite images, illustrating Tacloban's landscape at three key time points: pre-disaster (October 18, 2013), event time (November 27, 2013), and post-disaster (September 18, 2015).

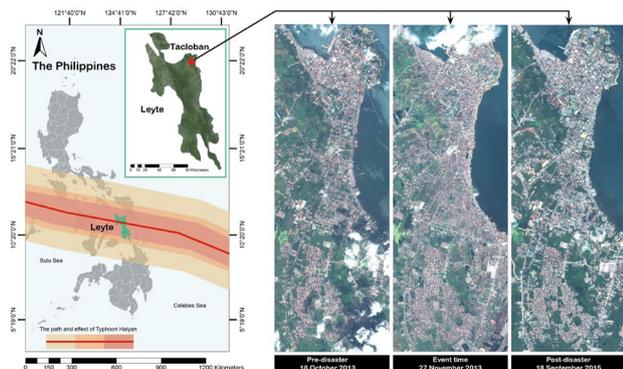


Figure 1: Map showing Typhoon Haiyan and pre-disaster, event time, and post-disaster satellite imagery.

These satellite images provide crucial insights into the damage extent and recovery patterns following Typhoon Haiyan. To systematically classify post-disaster conditions, areas are categorized into four classes based on the visually determined damage ratio at the event time (T1) and the level of recovery at a later period (T2) compared to pre-disaster conditions (T0):

- **Not damaged:** areas where the visually determined damage ratio is below 50% at T1 (event time), indicating minimal or no structural impact.
- **Not recovered:** areas where the damage ratio exceeded 50% at T1 but remained below 50% recovery at T2, indicating slow or stalled reconstruction efforts.
- **Recovered:** areas where the damage ratio exceeded 50% at T1 but experienced over 50% recovery by T2 (post-disaster recovery period), showing significant reconstruction efforts.
- **New buildings:** areas where completely new structures have been constructed after the disaster, representing land-use changes and new developments.

Figure 2 presents example images from the dataset. These samples illustrate how satellite imagery captures varying levels of structural impact and reconstruction across different regions. The visual distinctions between the classes highlight the progression of disaster damage and recovery patterns, emphasizing the necessity of accurate classification methods for PDDRM.

Table 1 shows that the dataset exhibits a noticeable class imbalance, with the "Not damaged" class containing the highest number of training images (625), while the "New buildings" class is the least represented, with only 180 training samples. This imbalance poses a challenge for DL models, as underrepresented classes may lead to biased predictions, where the model favors more frequent classes.

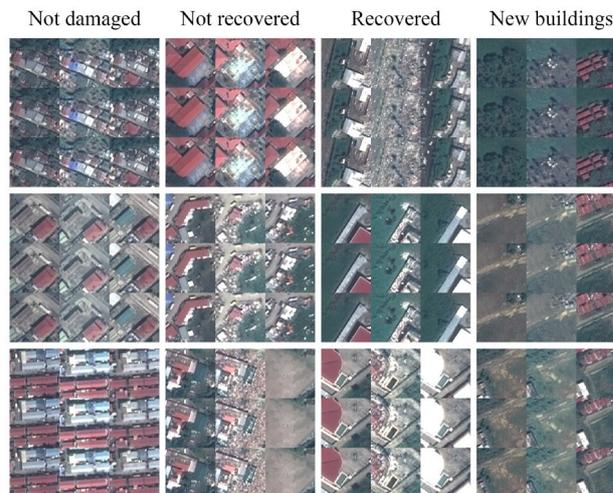


Figure 2: Example images from the dataset representing four classes: "not damaged", "not recovered", "recovered", and "new buildings".

Table 1: Distribution of training and test images across PDDRM classes.

Class	Training images	Test images
Not damaged	625	100
Not recovered	275	100
Recovered	450	100
New buildings	180	100

2.2 Methodology

The proposed framework integrates ESRGAN and DL models to improve PDDRM. The first stage is data preparation, where multi-temporal satellite images, with three bands: Red, Green, and Blue, are collected and categorized into three timeframes: T0 (Pre-disaster), T1 (Event time), and T2 (Post-disaster). These images are then processed into low-resolution (LR) ($75 \times 75 \times 3$) from high-resolution (HR) ($300 \times 300 \times 3$) patches for model training and evaluation. In the second stage, SR enhancement is performed using ESRGAN. LR images are input into the ESRGAN model, which consists of convolutional layers, residual blocks, and an upsampling layer to reconstruct HR images. The accuracy of the generated super-resolved images is evaluated using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). The third stage involves PDDRM using DL models. Both LR and generated HR images are used to train separate DL models, allowing for direct comparison of classification accuracy across different resolutions. The final stage is accuracy and reliability assessment. The classification performance of the DL models is assessed using Precision and Accuracy. To ensure model reliability, XAI technique Grad-CAM++ is applied, highlighting the critical image regions used in model decision-making.

ESRGAN is an improved version of SRGAN, designed to generate high-resolution images from LR inputs by enhancing fine textures and reducing artifacts (Wang et al., 2018). The ESRGAN framework consist of two components: generator G , that takes a LR image I_{LR} as input and produces a super-resolved image I_{SR} , and discriminator D , distinguishes between generated SR images I_{SR} and ground truth high-resolution images I_{HR} . The generator function can be expressed as:

$$I_{SR} = G(I_{LR}; \theta_G), \quad (1)$$

where θ_G represents the learnable parameters of the generator. The discriminator function follows the relativistic average GAN approach by comparing real and fake images instead of simply classifying them as real or fake:

$$D(I_{HR}, I_{SR}) = \sigma(C(I_{HR}) - \mathbb{E}[C(I_{SR})]), \quad (2)$$

where σ is the sigmoid function and C represents the discriminator's feature extraction function. To improve perceptual quality, ESRGAN optimizes a multi-component loss function comprising adversarial loss, perceptual loss, and pixel loss as explained in Wang et al. (2018).

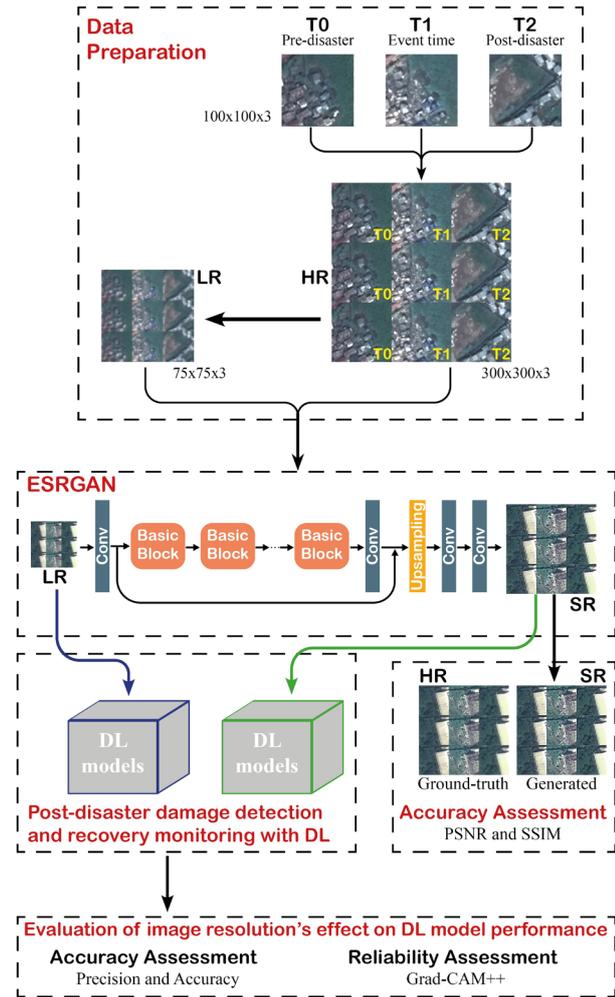


Figure 3: Overview of the Methodology.

For PDDRM classification, this study employs three state-of-the-art DL architectures: Vision Transformer (ViT) (Dosovitskiy, 2020), MaxViT (Tu et al., 2022), and ConvNeXt (Liu et al., 2022). These models leverage different architectural designs to extract spatial and structural features from satellite imagery, enhancing classification performance.

2.3 Accuracy Assessment

For ESRGAN model, two standard image quality metrics are applied to assess the quality of the super-resolved images: PSNR and SSIM. PSNR evaluates the pixel level accuracy of a generated high-resolution image I_{SR} by measuring the reconstruction error compared to the ground truth image I_{HR} . A higher PSNR value indicates better image quality, meaning the

super-resolved image closely resembles the original high-resolution reference.

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX^2}{MSE} \right), \quad (3)$$

where MAX represents the maximum possible pixel value (255 for 8-bit images) and MSE (Mean Squared Error) is given by:

$$MSE = \frac{1}{N} \sum_{i=1}^N (I_{HR} - I_{SR})^2, \quad (4)$$

SSIM evaluates the perceptual similarity between the super-resolved image and the ground truth by considering luminance, contrast, and structure. SSIM values range from 0 to 1, with 1 indicating perfect similarity.

$$SSIM(I_{HR}, I_{SR}) = \frac{(2\mu_{HR}\mu_{SR} + C_1)(2\sigma_{HR,SR} + C_2)}{(\mu_{HR}^2 + \mu_{SR}^2 + C_1)(\sigma_{HR}^2 + \sigma_{SR}^2 + C_2)}, \quad (5)$$

where:

- μ_{HR}, μ_{SR} are the mean intensities of the images.
- $\sigma_{HR}^2, \sigma_{SR}^2$ are the variances of the images.
- $\sigma_{HR,SR}$ is the covariance between the images.
- C_1 and C_2 are small constants to avoid instability.

The DL models used for PDDRM are assessed using two widely used classification metrics: Accuracy and Precision.

$$Precision = \frac{TP}{TP + FP}, \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (7)$$

where TP is true positives, TN is true negatives, FP means false positives, and FN represents false negatives.

2.4 Reliability Assessment

To ensure the reliability of DL models for PDDRM this study employs Gradient-weighted Class Activation Mapping++ (Grad-CAM++). Grad-CAM++ is an improved version of Grad-CAM, which enhances visualization precision by considering pixel-wise importance weights rather than relying solely on coarse feature maps and particularly useful for multi-class classification tasks (Chattopadhyay et al., 2018).

3. Results and Discussion

3.1 ESRGAN for Satellite Image Resolution Enhancement

The ESRGAN-generated SR images are evaluated in terms of PSNR and SSIM (Table 2). These metrics quantify the quality of the super-resolved images by comparing them to high-resolution ground truth images. Higher PSNR values indicate lower reconstruction error, while higher SSIM values reflect more significant structural similarity to the original images.

The Not Damaged class achieved an average PSNR of 29.2 and an average SSIM of 0.78, suggesting that ESRGAN effectively restores fine details in areas with minimal structural changes. The

New Buildings class exhibited the highest PSNR (30.3) and SSIM (0.81), indicating that newly constructed structures were well-reconstructed with a high degree of similarity to the ground truth. These results suggest that the ESRGAN model performs particularly well in regions with clear, well-defined structures.

Table 2: Quantitative evaluation of ESRGAN-generated SR images.

Class	PSNR	SSIM
Not damaged	29.2	0.78
Not recovered	24.7	0.72
Recovered	26.6	0.74
New buildings	30.3	0.81
Average	27.7	0.76

In contrast, the Recovered and Not recovered classes yielded lower PSNR and SSIM values, with PSNR values of 26.6 and 24.7, respectively, and SSIM values of 0.74 and 0.72. The lower scores in these classes suggest that ESRGAN faces greater difficulty in reconstructing areas with visible structural damage or complex textures, such as debris, partially collapsed buildings, or areas undergoing repair. The Not recovered category, in particular, had the lowest PSNR and SSIM values, which may be attributed to the presence of mixed textures resulting from disaster damage and ongoing reconstruction, making the SR process less effective in generating highly accurate reconstructions.

Overall, the average PSNR and SSIM values across all categories were 27.7 dB and 0.76, respectively. Despite effectively enhancing LR images, the performance of the ESRGAN model varies across different damage and recovery categories. The relatively high PSNR and SSIM values for Not Damaged and New Buildings suggest that ESRGAN performs well in stable environments with clearly defined structures. However, the lower values for Not Recovered and Recovered indicate that complex post-disaster environments, characterized by debris and incomplete structures, present challenges for SR reconstruction.

Figure 4 presents a visual comparison of satellite images at three different resolutions: Low-Resolution (LR), Super-Resolution (SR), and High-Resolution (HR). Each row (labeled a–h) represents a different example, highlighting how SR reconstruction compares to original LR and HR imagery. Row a and b shows that the LR images in Not damaged category exhibit pixelation and blurriness, making it difficult to distinguish fine structural details such as rooftops and road boundaries. The SR images, generated by ESRGAN, significantly improve clarity, restoring building edges and preserving color consistency, demonstrating the model's ability to reconstruct fine textures with high fidelity. Row g and h demonstrate that the SR images effectively restore the shapes and colors of new buildings, providing clearer outlines of rooftops and construction patterns. A comparison with HR images reveals that ESRGAN-generated SR outputs closely match the true HR data, indicating that SR can be a valuable tool for detecting new developments in post-disaster recovery analysis.

Row c, d, e, and f shows that the SR images provide a better view of damaged structures; however, the model face challenges in generating finer details such as debris of the buildings. Despite these challenges, the model enhances damage visibility and improves clarity in SR images, which is particularly useful for PDDRM.

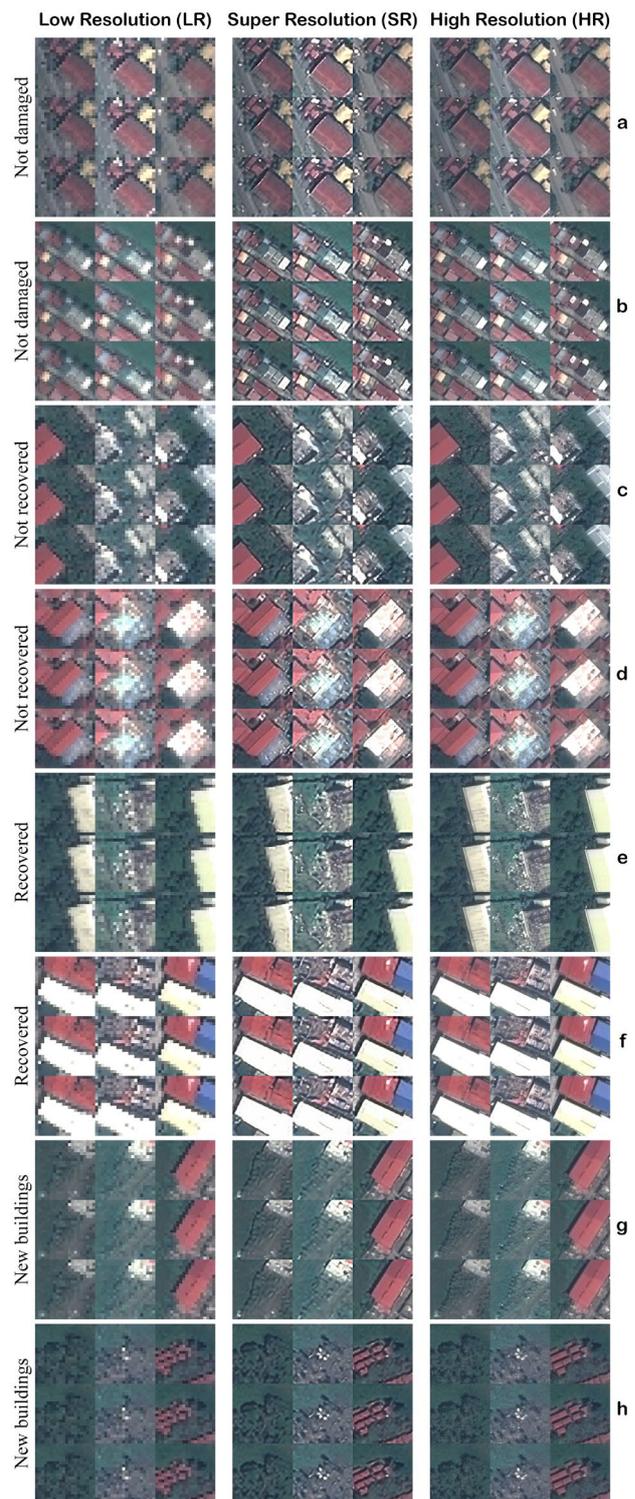


Figure 4: Comparison of example LR, SR, and HR images.

3.2 PDDRM Classification with DL Models

Table 3 presents the classification accuracy and precision of three DL models when trained on LR images and ESRGAN-generated SR images. The results demonstrate the effectiveness of SR enhancement in improving classification performance, with notable variations across models. For ViT, the classification accuracy improved from 0.79 (LR) to 0.84 (SR), while precision increased from 0.80 to 0.84, suggesting that ESRGAN model successfully enhances the model's ability to distinguish between PDDRM classes. Similarly, ConvNeXt exhibited an increase in

accuracy from 0.78 to 0.82 and in precision from 0.79 to 0.83. In contrast, MaxViT showed no significant improvement in classification accuracy, remaining at 0.74 for both LR and SR, with only a marginal increase in precision from 0.75 to 0.76. This suggests that MaxViT, despite its hybrid architecture integrating convolutional and transformer-based mechanisms, does not leverage SR information as effectively as ViT or ConvNeXt for this particular task.

Table 3: Classification performance of DL models on LR and SR images.

Model	Accuracy		Precision	
	LR	SR	LR	SR
ViT	0.79	0.84	0.80	0.84
ConvNeXt	0.78	0.82	0.79	0.83
MaxViT	0.74	0.74	0.75	0.76

Figure 5 presents confusion matrices comparing the classification performance of ViT, ConvNeXt, and MaxViT on LR and SR images. Each confusion matrix shows the number of correctly classified and misclassified instances across four PDDRM classes, enabling an assessment of how SR influences classification accuracy.

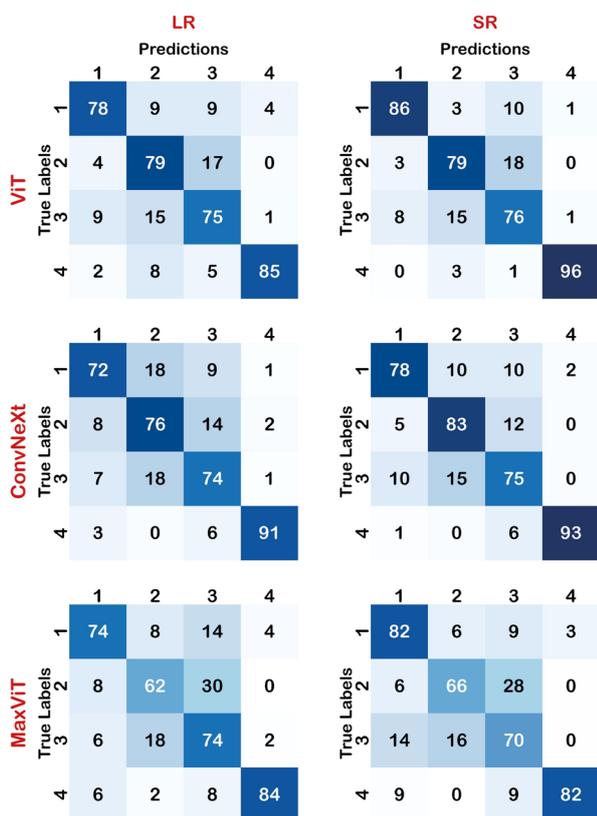


Figure 5: Confusion matrices for LR and SR images across DL models (1: Not damaged, 2: Not recovered, 3: Recovered, and 4: New buildings).

For ViT, the SR-enhanced model demonstrates an increase in correct classifications across some classes. The most notable improvement occurs in class 4 (New Buildings), where correct predictions increase from 85 (LR) to 96 (SR), reducing misclassifications in other classes. Additionally, class 1 (Not Damaged) benefits significantly, with correct classifications improving from 78 (LR) to 86 (SR), indicating that ESRGAN-generated SR images improve ViT's ability to differentiate new

buildings and undamaged structures from other classes. However, no significant improvement is observed in class 2 (Not Recovered) and class 3 (Recovered), suggesting that SR does not substantially aid ViT in distinguishing between different levels of damage and recovery.

For ConvNeXt, SR images lead to consistent improvements across all classes. The correct classification of class 2 increases from 76 (LR) to 83 (SR), reducing misclassifications into other categories. Class 1 (Not Damaged) improves from 72 (LR) to 78 (SR). The relatively small misclassification rates suggest that ConvNeXt benefits from spatial detail enhancement, particularly in distinguishing damaged and undamaged structures. Unlike ViT and ConvNeXt, MaxViT shows minimal improvements with SR. The model still struggles with distinguishing between damaged and recovered structures.

Figure 6 presents Grad-CAM++ visualizations for three DL models applied to LR and ESRGAN-enhanced SR (SR) images across different PDDRM classes. Grad-CAM++ highlights the image regions that contribute most to the model's decision-making process, with warmer colors (red) indicating high attention areas and cooler colors (blue) representing less significant regions. These visualizations help assess how SR influences model reliability and focus.

For ViT, the Grad-CAM++ visualizations reveal a clear improvement in attention when using SR images. For Not damaged category, the model exhibits weak and scattered activations, failing to effectively capture structural integrity in the LR case. With SR images, the model learns to systematically compare all three temporal frames (pre-disaster, event time, and post-disaster) to assess whether a structure has remained intact. In the New Buildings category, ViT's attention in LR images is dispersed and unfocused, with activations failing to highlight defining structural features. When SR is applied, the model learns to directly focus on newly constructed structures in the post-disaster frame, effectively identifying areas where reconstruction has occurred. This shift in attention suggests that ViT leverages SR to enhance its ability to detect buildings that have not suffered damage across different time periods, and new urban developments in post-disaster scenarios, leading to improved classification performance. However, for Not Recovered and Recovered categories, the Grad-CAM++ visualizations show no significant improvement with SR images, aligning with the confusion matrix results. The model continues to struggle with distinguishing between partially damaged, unrepaired, and recovered structures, as activations remain diffuse and inconsistent across both LR and SR cases.

For ConvNeXt, the Grad-CAM++ visualizations reveal that the model learns to focus more effectively on damaged areas when using SR images, particularly in Row b, which corresponds to the Not Recovered category. In the LR case, the model's attention appears scattered and inconsistent, with activations failing to highlight key structural damages such as collapsed roofs, debris, and missing infrastructure in event time. With SR images, the model shifts its attention towards damaged regions, focusing more precisely on structural deformations that indicate unrepaired buildings in event time and post-disaster period. This improved focus explains ConvNeXt's superior performance in distinguishing Class 2 (Not Recovered) when using SR images, as confirmed by the increase in classification accuracy and precision observed in the confusion matrix. However, despite these improvements, ConvNeXt still exhibits some limitations when using SR images. In some cases, misclassifications persist between Not Recovered and Recovered categories, indicating

that the model struggles to distinguish between ongoing reconstruction and persistent damage. Its activations remain overly concentrated in certain areas, sometimes neglecting overall structural context as shown in Row a.

For MaxViT, the Grad-CAM++ visualizations reveal minimal improvement in attention distribution when SR images. Unlike ViT and ConvNeXt, which show clear benefits from SR,

MaxViT's activations remain scattered and inconsistent across both LR and SR inputs. These findings suggest that MaxViT's multi-scale spatial attention mechanism does not fully utilize the additional details provided by ESRGAN-generated images, leading to limited performance gains compared to ViT and ConvNeXt.

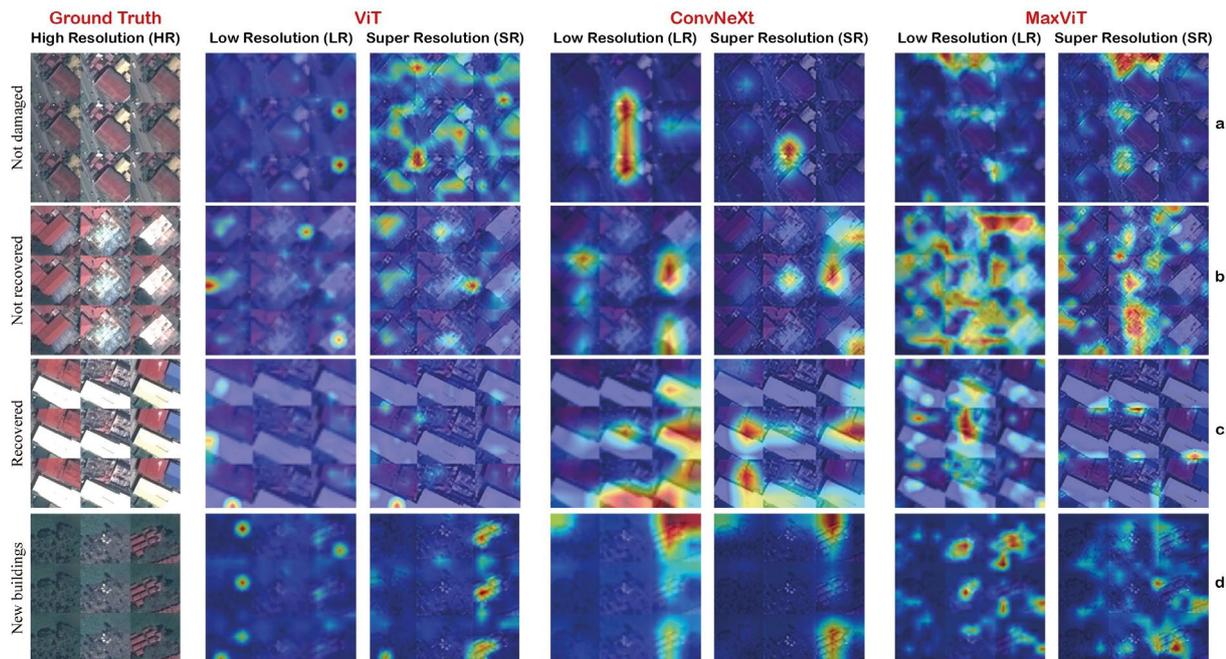


Figure 6: Grad-CAM++ visualizations for ViT, ConvNeXt, and MaxViT on LR and SR images across different PDDRM classes.

4. Conclusions

This study demonstrates that ESRGAN technique provide a cost-effective and scalable solution for PDDRM, particularly when very high-resolution satellite imagery is unavailable, delayed, or expensive. By leveraging ESRGAN-generated SR images, DL models can enhance classification accuracy without requiring additional high-resolution data sources, making this approach valuable for real-time disaster response and long-term recovery assessments. The ability to reconstruct fine structural details from LR images ensures that critical decisions related to damage assessment, resource allocation, and rebuilding efforts can be made more efficiently, even in resource-constrained settings.

The findings confirm that SR imagery improves DL-based classification models, particularly ViT and ConvNeXt, by enhancing feature extraction and structural clarity. Numerical evaluations using PSNR and SSIM indicate that ESRGAN effectively restores high-resolution features, with the highest reconstruction quality observed in Not Damaged and New Buildings categories. The confusion matrix analysis shows that SR improves classification accuracy and precision, for ViT and ConvNeXt.

The Grad-CAM++ visualizations further reveal how SR influences model decision-making. ViT learns to analyze all three temporal frames (pre-disaster, event time, and post-disaster), enhancing its ability to detect intact structures and newly built areas, suggesting that the model leverages SR to improve its temporal reasoning when assessing structural changes. ConvNeXt, in contrast, focuses more effectively on damaged regions, improving its ability to distinguish unrepaired

buildings, which aligns with its improved classification performance for Not Recovered structures in the confusion matrix results. However, these improvements were not consistent across all classes, raising reliability concerns about these models despite their performance improvements.

A key limitation of this study is its exclusive reliance on optical satellite imagery, which may not fully capture structural damage in areas with dense urban environments, cloud cover, or debris accumulation. Despite better image clarity and classification performance, the improvements are not uniform across all damage and recovery categories, particularly in distinguishing partially damaged from recovered structures. Additionally, the study focuses on single-timeframe classification, limiting the ability to assess longitudinal recovery trends. Future research should focus on enhancing model reliability through explainability-driven training strategies, such as attention refinement techniques or uncertainty-aware learning, to ensure that classification decisions align more closely with meaningful disaster-related features. Furthermore, incorporating temporal modeling techniques, such as spatiotemporal transformers or recurrent neural networks, could improve the ability to track recovery trajectories over time, making the approach more applicable for long-term disaster monitoring and policy planning.

Acknowledgements

The satellite images were provided by the Digital Globe Foundation and the 7 European Space Agency (Airbus Defence and Space), granted for a project at ITC, University 8 of Twente, titled "Post-disaster recovery assessment using remote sensing data and spatial 9 economic modeling".

References

- Chattopadhyay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. *2018 IEEE winter conference on applications of computer vision (WACV) (IEEE)*, 839-47.
- Chaudhuri, N., & Bose, I. (2020). Exploring the role of deep neural networks for post-disaster decision support. *Decis. Support Syst.*, 130, 113234.
- Cheng, C.-S., Behzadan, A. H., & Noshadravan, A. (2022). Uncertainty-aware convolutional neural network for explainable artificial intelligence-assisted disaster damage assessment. *Structural Control and Health Monitoring*, 29(10), e3019.
- Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Dou, X., Li, C., Shi, Q., & Liu, M. (2020). Super-Resolution for Hyperspectral Remote Sensing Images Based on the 3D Attention-SRGAN Network. *Remote Sensing*, 12(7), 1204.
- Fu, X., Kouyama, T., Yang, H., Nakamura, R., & Yoshikawa, I. (2022, 17-22 July 2022). Toward Faster and Accurate Post-Disaster Damage Assessment: Development of End-to-End Building Damage Detection Framework with Super-Resolution Architecture. *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 1588-91.
- Ghaffarian, S., & Emtehani, S. (2021). Monitoring urban deprived areas with remote sensing and machine learning in case of disaster recovery [Article]. *Climate*, 9(4), Article 58.
- Ghaffarian, S., Taghikhah, F. R., & Maier, H. R. (2023). Explainable artificial intelligence in disaster risk management: Achievements and prospective futures. *International Journal of Disaster Risk Reduction*, 98, 104123.
- 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 preprint arXiv:1911.09296*, 1(2), 5.
- Hodgson, M. E., Davis, B. A., Cheng, Y., & Miller, J. R. (2010). Modeling Remote Sensing Satellite Collection Opportunity Likelihood for Hurricane Disaster Response. *Cartography and Geographic Information Science*, 37, 15 - 17.
- Kakogeorgiou, I., & Karantzalos, K. (2021). Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 103, 102520.
- Lallemant, D., Soden, R., Rubinyi, S., Loos, S., Barns, K., & Bhattacharjee, G. (2017). Post-Disaster Damage Assessments as Catalysts for Recovery: A Look at Assessments Conducted in the Wake of the 2015 Gorkha, Nepal, Earthquake. *Earthquake Spectra*, 33, 435 - 451.
- Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11976-86.
- Rahneemofar, M., Chowdhury, T., & Murphy, R. (2023). RescueNet: a high resolution UAV semantic segmentation dataset for natural disaster damage assessment. *Scientific Data*, 10(1), 913.
- Rai, A. (2020). Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science*, 48(1), 137-141.
- Ren, Y., Zhang, X., Ma, Y., Yang, Q., Wang, C., Liu, H., & Qi, Q. (2020). Full Convolutional Neural Network Based on Multi-Scale Feature Fusion for the Class Imbalance Remote Sensing Image Classification. *Remote. Sens.*, 12, 3547.
- Safavi, F., & Rahneemofar, M. (2022). Comparative study of real-time semantic segmentation networks in aerial images during flooding events. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16, 4-20.
- Sheykhoumousa, M., Kerle, N., Kuffer, M., & Ghaffarian, S. (2019). Post-Disaster Recovery Assessment with Machine Learning-Derived Land Cover and Land Use Information. *Remote Sensing*, 11(10), 1174.
- Sublime, J., & Kalinicheva, E. (2019). Automatic Post-Disaster Damage Mapping Using Deep-Learning Techniques for Change Detection: Case Study of the Tohoku Tsunami. *Remote. Sens.*, 11, 1123.
- Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., & Li, Y. (2022). Maxvit: Multi-axis vision transformer. *European conference on computer vision (Springer)*, 459-79.
- van Loenhout, J. A. F., Gil Cuesta, J., Abello, J. E., Isiderio, J. M., de Lara-Banquesio, M. L., & Guha-Sapir, D. (2018). The impact of Typhoon Haiyan on admissions in two hospitals in Eastern Visayas, Philippines. *PLoS ONE*, 13(1), e0191516.
- 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.
- Wang, P., Bayram, B., & Sertel, E. (2022). A comprehensive review on deep learning based remote sensing image super-resolution methods. *Earth-Science Reviews*, 232, 104110.
- Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. *Proceedings of the European conference on computer vision (ECCV) workshops*.
- Xiong, Y., Guo, S., Chen, J., Deng, X., Sun, L., Zheng, X., & Xu, W. (2020). Improved SRGAN for Remote Sensing Image Super-Resolution Across Locations and Sensors. *Remote Sensing*, 12(8), 1263.