UCL DEPARTMENT OF GEOGRAPHY

YEAR 2024-25

EXAM CANDIDATE ID:	**LTKB8**
MODULE CODE:	**GEOG0125**
MODULE NAME:	**Advanced Topics in Social & Geographic Data Sciences**
COURSE PAPER TITLE:	**Modified Pyramid Pooling Module-Based Semi-Siamese Network (PPM-SSNet) for Classifying Post Disaster Building Damage Label**
WORD COUNT:	**1498 (excluding tables & figures contents and references)**
CODE REPOSITORY LINK:	**https://drive.google.com/drive/folders/1JkR_sNVziBVmumNVn9_YDLSTKhrlF2z0?usp=sharing**

I acknowledge the use of Artificial Intelligence (AI) of ChatGPT (https://chatgpt.com/) to assist in improving my writing and troubleshooting code. I used ChatGPT to enhance the clarity and conciseness of my writing after drafting initial content, while ensuring that the core ideas and intended meaning remained intact. Additionally, I used ChatGPT to assist with troubleshooting errors and improving my code. For example, I encountered challenges while building components of my modified PPM-SSNet model from scratch in Jupyter Notebook. To address this, I used the prompt: *"I want to use PPM-SSNet but only post-disaster images as input."* While ChatGPT provided useful guidance, the initial output was not fully functional. I cross-referenced its suggestions with the logic from the original model author and several trusted GitHub repositories, particularly those covering the construction of residual blocks, SE blocks, and PPM modules. These sources are cited in this report as Engelberts, (2018); Chen, (2020a); Zhao, (2020); Tan and Le, (2019); Hug, (2023); Bai et al., (2020) in this document. To ensure transparency, I documented and credited all ChatGPT-assisted code sections both in the Jupyter Notebook and in this report. My use of ChatGPT focused on improving code readability and resolving specific implementation issues, with all outputs carefully verified against reliable sources.

1. Introduction

In early 2025, the world was devastated by catastrophic wildfires in Los Angeles and major earthquakes in Myanmar caused widespread destruction, displacing thousands and damaging numerous homes (O'Hara-Glaspie, 2025; Horton et al., 2025; Burgess & Hagan, 2025). In these large-scale disasters, accurate and timely damage identification and assessment are essential to enable optimal resource allocation and reduce additional casualties (Benedict et al., 2024; Weber & Kan, 2020; Chen, 2020b; Wang, Chew, & Zhang, 2022).

With the increasing availability of high-resolution satellite imagery, convolutional neural networks (CNNs) have become essential tools for automating post-disaster damage assessment, due to its efficiency in identifying critical areas in need of assistance (Nex et al., 2019; Bai et al., 2020; Benedict et al., 2024). Recent studies have explored various CNN architectures for post-disaster building damage classification using satellite imagery, including DenseNet121 with dilated convolutions (Nex et al., 2019), Mask R-CNN with a Feature Pyramid Network (Weber & Kan, 2020), and the Pyramid Pooling Module Semi-Siamese Network (PPM-SSNet) by Bai et al. (2020). Building on the work of Bai et al. (2020), this study conducts an exploratory research on a modified PPM-SSNet using the xBD dataset to further assess the model effectiveness in post-disaster damage classification.

2. Data and Methods

The xBD dataset comprises pre- and post-disaster satellite imagery with over 850,000 annotated building polygons affected by six categories of natural disasters: earthquakes, tsunamis, volcanic eruptions, floods, wildfires, and hurricanes. Each building is labeled with a corresponding damage level (Gupta et al., 2019; Defense Innovation Unit, n.d.). A detailed breakdown of the damage classification is provided in Table 1.

Label	**Description**
No damage	Structure is undisturbed with no visible signs of water exposure, structural issues, burn marks, or any other form of damage.
Minor damage	Buildings show slight impacts like partial burning, nearby volcanic flow, surrounding water, missing roof elements, or visible cracks.
Major damage	Significant damage is present, including partial wall or roof collapse, direct impact from volcanic flow, or the structure being surrounded by water or mud.
Destroyed	Structure is severely impacted like completely collapsed, scorched, partially or fully submerged in water or mud, or entirely missing.
Unclassified	Unclassified

Table 1: xBD Dataset Damage Description

Regarding the model, Bai et al. (2020) introduced PPM-SSNet by integrating dilated convolutions, a squeeze-and-excitation (SE) attention mechanism, and a Pyramid Pooling Module (PPM) to enhance feature extraction for pixel-level prediction tasks. Dilated convolutions expand the receptive field to effectively capture both global context and local details (Bai et al., 2020; Li et al., 2021). The SE mechanism further improves performance by emphasizing informative features and suppressing less relevant ones. Additionally, the PPM aggregates contextual information at multiple spatial scales through hierarchical pooling (Bai et al., 2020; Zhao et al., 2017). These design choices make PPM-SSNet well-suited for building damage classification tasks, which is why the author selected it for further experiments to evaluate its effectiveness in post-disaster damage assessment

Specifically, in this study, a modified version of the PPM-SSNet model is utilized for damage classification. The modification is different from the original implementation, which uses a Semi Siamese network architecture to compare pre- and post-disaster images. As the xBD dataset includes annotated building damage labels, this study utilizes these annotations to extract individual buildings and classify their damage levels. The approach intentionally omits the localization stage and the comparison between pre- and post-disaster images, aiming to assess the effectiveness of the model on a simplified classification pipeline. Specifically, the study evaluates model performance using only post-disaster imagery to determine whether acceptable accuracy can still be achieved without the localisation step.

The second key difference lies in the network architecture. To accommodate limited computational resources, a shallower version of the PPM-SSNet model is implemented, using input images of 64×64 instead of the 1024×1024 images used by Bai et al. (2020). The model adopts a progressive channel-doubling strategy, starting with a small number of channels (input = 3 for RGB images, output = 16) and doubling at key stages up to a maximum of 128 channels. This design efficiently increases the network's representational capacity while maintaining low memory usage (Tan and Le, 2019). Model training was conducted using the Adam optimizer, learning rate of 0.0002, and Focal Loss to address class imbalance (Wu et al., 2021; Chen, 2020b; Bai et al., 2020; Benedict et al., 2024; Nex et al., 2019; Zhao et al., 2024). The detailed architecture is presented in Figure 1.

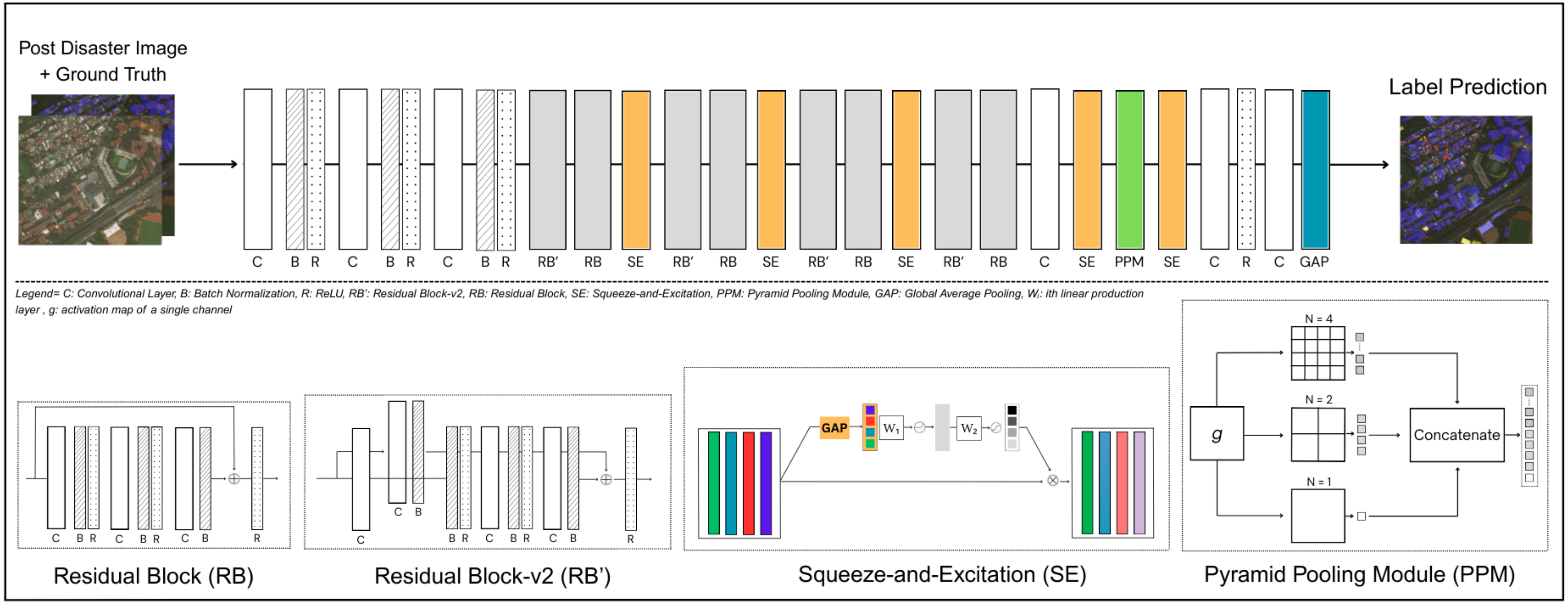
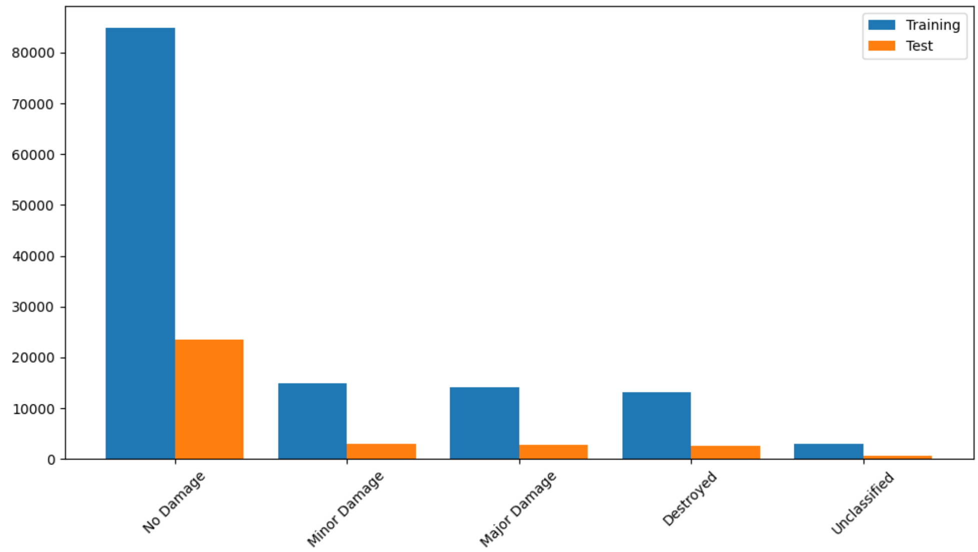
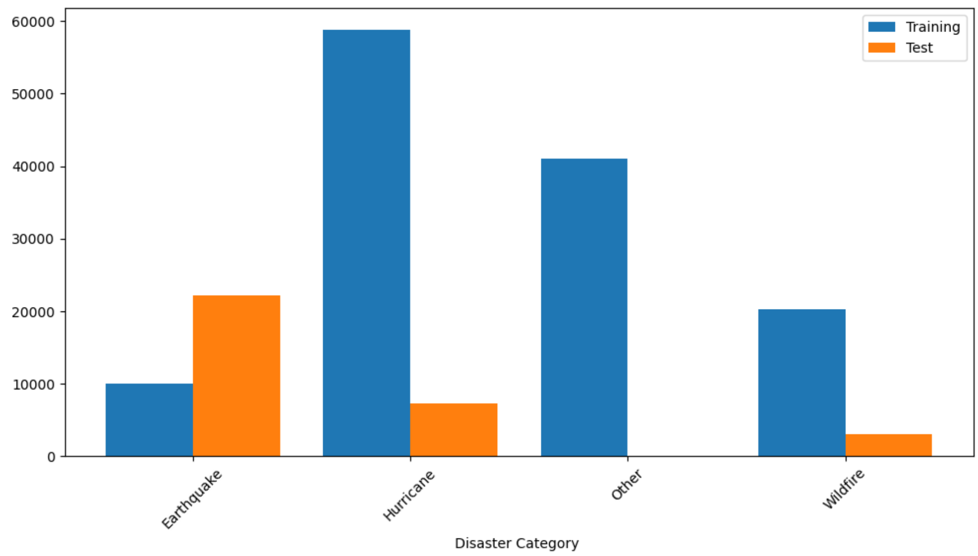


Figure 1: Author's architecture of the modified PPM-SSNet, based on Bai et al. (2020). Sources: Engelberts (2018); Chen (2020a); Zhao (2020); Tan & Le (2019); Hug (2023); Bai et al. (2020); OpenAI (2025a, 2025b).

The dataset consists of 1,799 images at a resolution of 1024×1024 pixels. Each image was cropped into small 64×64 images, which centered on each individual building, and subsequently resized to 224×224 pixels to match the standard input dimensions for CNN (Perez, Tah, & Mosavi, 2019). A stratified split was applied, allocating 80% of the data for training and 20% for testing, resulting in 130,229 building images for training and 32,558 for testing. Figure 2 illustrates the distribution of the training and test datasets across damage labels and disaster categories. Due to stratification, the test set includes only three disaster categories.



(a) Damage label distribution (Train vs Test)

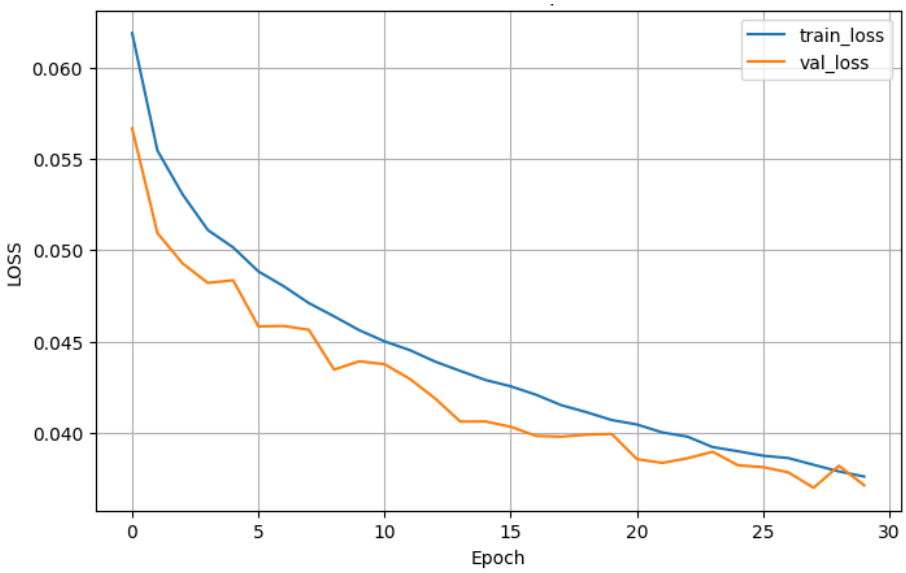


(b) Disaster category distribution (Train vs Test)

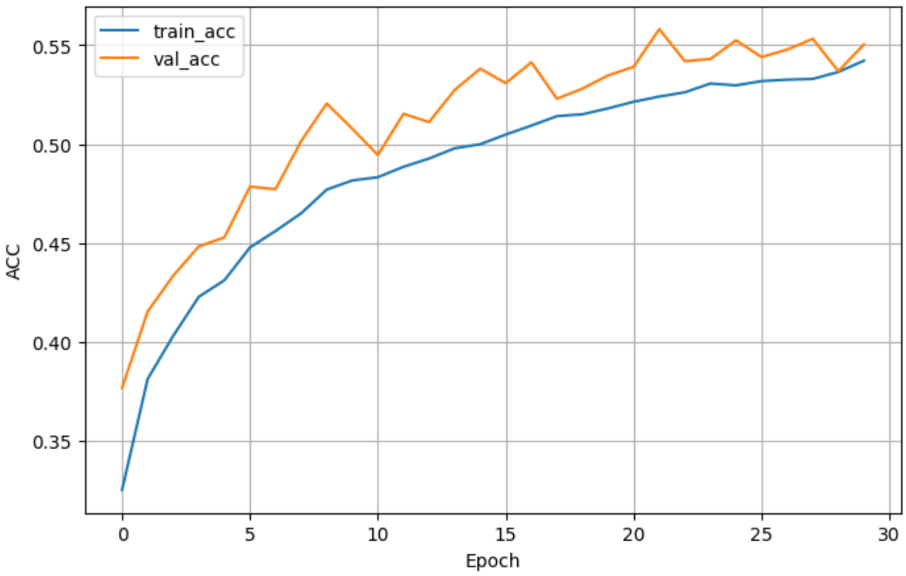
Figure 2: Dataset Distribution.

3. Results

After 30 epochs, the loss values for both the training and test sets gradually declined, indicating a consistent learning process. The model achieved an overall precision of 0.47 and recall of 0.57, suggesting a tendency toward higher false positives while performing relatively better in minimizing false negatives, thus resulting in an F1-score of 0.51 (Buhl, 2023).



(a) Training and Test loss



(b) Training and Test Accuracy

Figure 3: Training and Test Performance.

The overall performance in this study is understandably lower than that reported by Bai et al. (2020). However, the model achieved comparable results and even outperformed the original model in the minor damage class (see Table 2). The F1 scores for the undamaged and destroyed classes were relatively high at 0.75 and 0.63, respectively. This is likely due to the distinct structural characteristics of these classes, allowing the model to more accurately differentiate them. In contrast, the model performed poorly in detecting minor and major damage, with F1 scores around 0.4. This may be attributed to the less distinct visual features of these damage levels compared to completely destroyed structures, which leads to greater variability and model confusion. Furthermore, the appearance of structural damage can differ significantly by disaster type, and even within the same disaster type, variations in building materials may result in diverse damage patterns and thus complicating classification results even further (Nex et al., 2019; Zhao et al., 2024).

Model	**Loss**	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Modified Model	0.04	0.56	0.47	0.57	0.51

Label	**Precision**	**Recall**	**F1-Score**	
			Modified	Original
No Damage	0.96	0.62	0.75	0.90
Minor Damage	0.40	0.55	0.47	0.41
Major Damage	0.31	0.63	0.42	0.65
Destroyed	0.58	0.68	0.63	0.70
Unclassified	0.10	0.71	0.17	–

Table 2: Performance Metrics.

The confusion matrix in Figure 4 summarizes the model's overall classification performance across the three disaster categories in the test dataset. The results suggest relatively consistent performance across the disaster categories. However, the model predominantly classifies buildings as having no damage. This is likely due to class imbalance, with the no damage category comprising over 72% of the training data. The model frequently misclassified minor and major damage, often predicting them interchangeably or labeling both as no damage. Similarly, unclassified are frequently misclassified as no damage, resulting in a low precision of 0.1 for this class. This outcome is somewhat expected, as unclassified buildings may visually resemble undamaged structures. These limitations highlight the importance of further investigation, particularly due to the limited range of disaster categories represented in the test dataset.

To evaluate the model's generalization capability, three images separate from the original dataset were selected for additional testing. These images were introduced to the trained model to assess its performance on entirely unseen data. For the wildfire and hurricane images, where ground truth labels included a high proportion of destroyed buildings, the model demonstrated relatively accurate damage predictions. In contrast, performance was less accurate for the earthquake dataset, which only included labels for major, minor, and no damage. The model showed difficulty distinguishing between these categories, highlighting an area for potential improvement in future studies (See Figure 5).

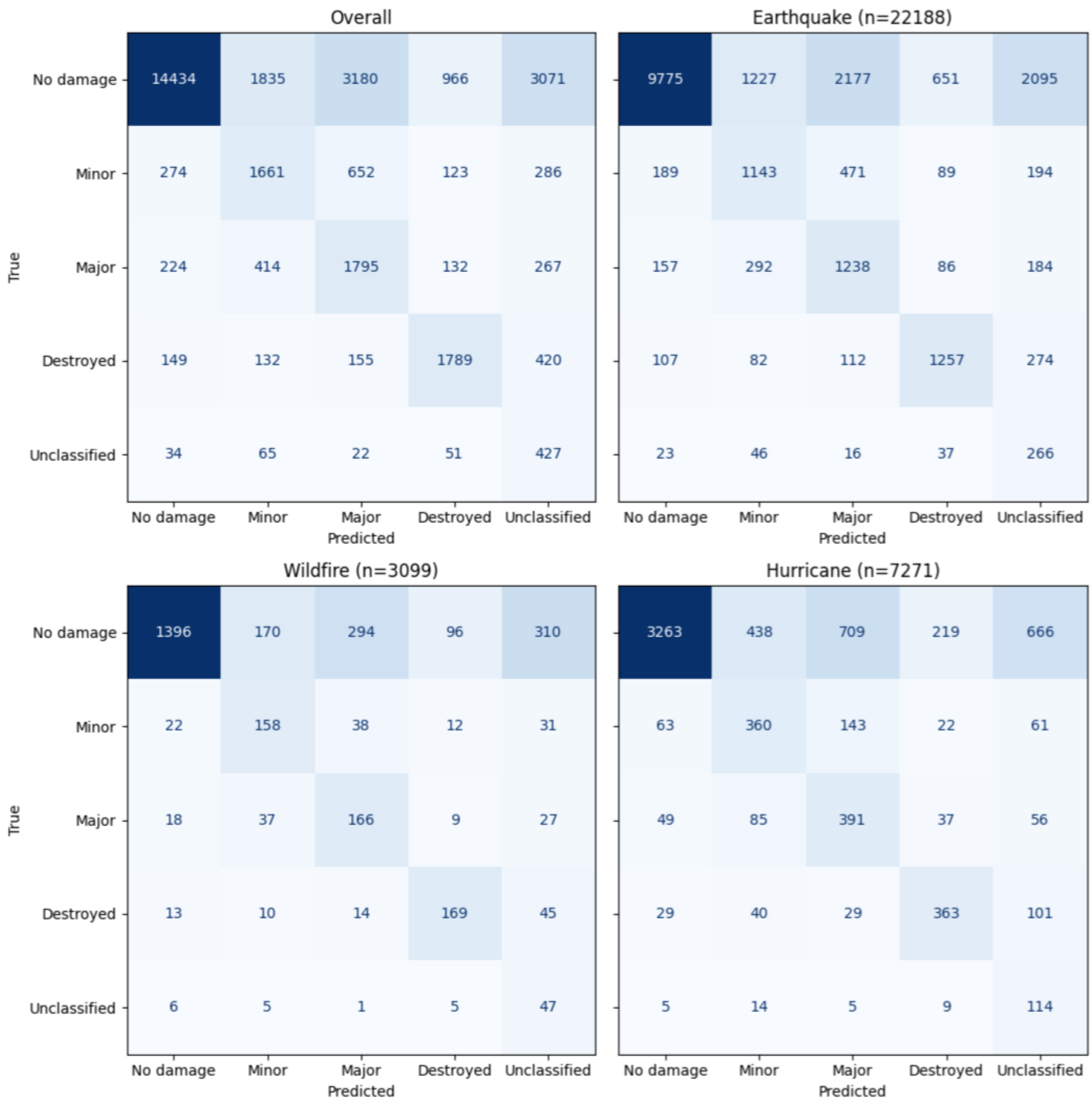


Figure 4: Confusion matrix.

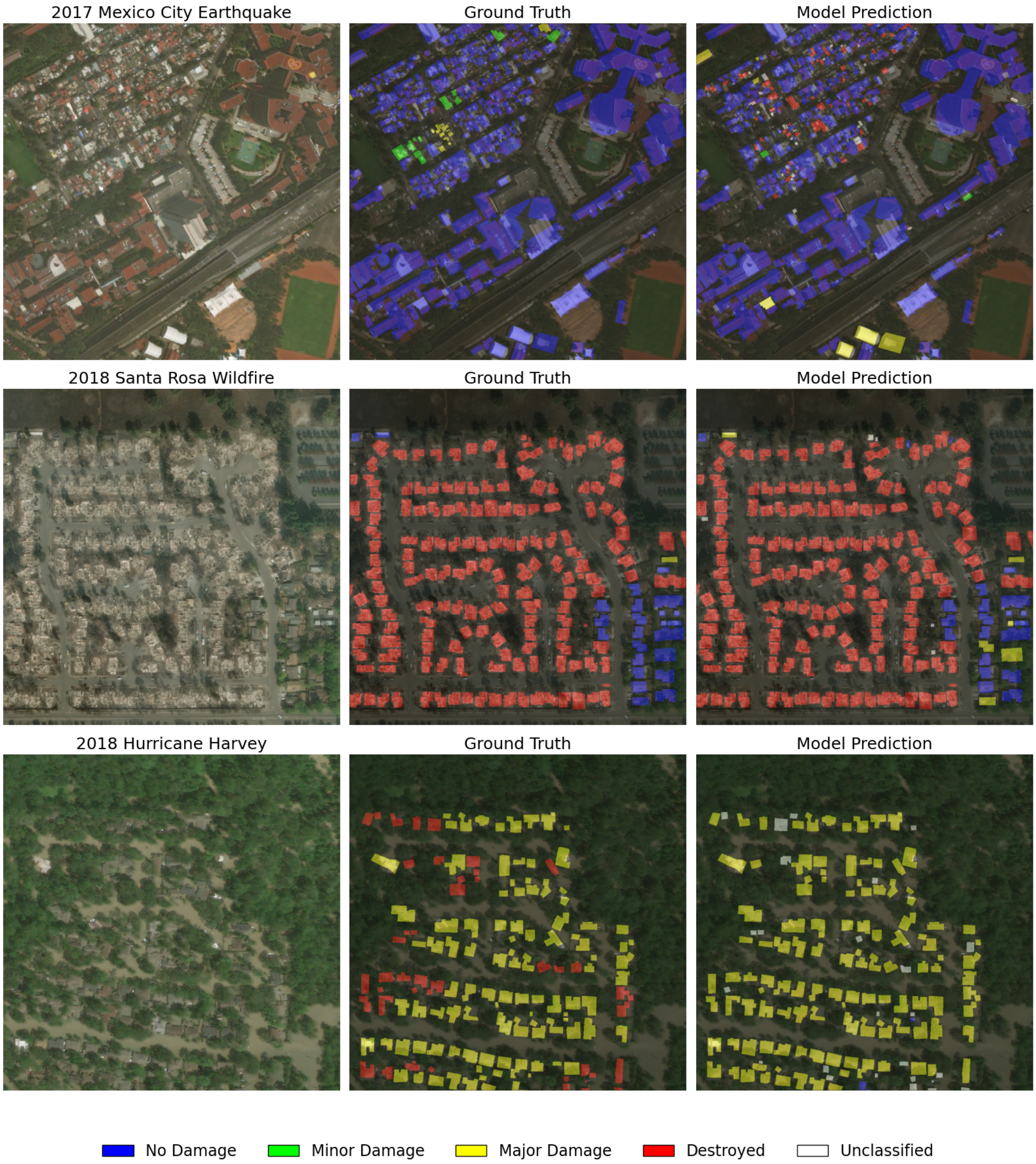


Figure 5: Model performance on unseen post-disaster images.

4. Conclusion

In this study, the xBD dataset is utilized with a simplified approach that differs from the original PPM-SSNet model proposed by Bai et al. (2020). After 30 training epochs, the modified model achieved F1 scores of 0.75 (no damage), 0.47 (minor damage), 0.42 (major damage), 0.63 (destroyed), and 0.17 (unclassified). The model demonstrated competitive results and, in some cases, outperformed the original implementation by Bai et al. (2020).

However, this study has several limitations, particularly in two areas. First, the test set in this study is limited to three disaster categories. Incorporating a broader range of disaster scenarios is essential, as different events can produce distinct structural impacts and damage patterns. (Zhao et al., 2024). Second, due to constraints in computational resources and time, the study did not include a building segmentation task, which has been shown to improve classification performance. Previous research (Bai et al., 2020; Chen, 2020b) highlights that incorporating segmentation features enhances the model's ability to capture structural cues for a better damage classification accuracy. As such, integrating segmentation may further improve the current model's performance, which remains limited with an F1-score of 0.5.

This exploratory study further contributes to the literature in examining how CNN can support post-disaster damage assessment. The findings highlight the potential of deep learning models to assist governmental agencies and first responders in making timely, informed, and accurate decisions during disaster events to prevent additional casualties.

References

Bai, Y., Hu, J., Su, J., Liu, X., Liu, H., He, X., Meng, S., Mas, E. and Koshimura, S. (2020). *Pyramid Pooling Module-Based Semi-Siamese Network: A Benchmark Model for Assessing Building Damage from xBD Satellite Imagery Datasets*. Remote Sensing, 12(24). https://doi.org/10.3390/rs12244055

Benedict, R., Winartio, R.B., Adinata, M.F. and Irwansyah, E. (2024). *Comparison on Difference Deep Learning Models for Building Damage Assessment using xBD Dataset*. In: 2024 Arab ICT Conference (AICTC). IEEE, pp.181–186.

Buhl, N. (2023). *F1 Score in Machine Learning*. [online] encord.com. Available at: https://encord.com/blog/f1-score-in-machine-learning/

Burgess, J. and Hagan, R. (2025). *Myanmar earthquake: What we know*. BBC News. [online] 28 Mar. Available at: https://www.bbc.co.uk/news/articles/crlxlxd7882o

Chen, J. (2020a). *pytorch-lars*. [online] GitHub. Available at: https://github.com/JosephChenHub/pytorch-lars/blob/master/resnet.py

Chen, X. (2020b). *Using Satellite Imagery to Automate Building Damage Assessment: A case study of the xBD dataset*. In: Institute of Industrial and Systems Engineers Annual Conference.

Defense Innovation Unit (n.d.). *xView2*. [online] xview2.org. Available at: https://xview2.org/dataset

Engelberts, J. (2018). *Squeeze-Excitation-PyTorch*. [online] GitHub. Available at: github.com/jonnedtc/Squeeze-Excitation-PyTorch/blob/master/networks.py

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

Horton, M., Sedgwick, S.M., Adams, J., Wei, D. and Skyberg, M. (2025). *IMPACT OF 2025 LOS ANGELES WILDFIRES AND COMPARATIVE STUDY*. [online] Southern California Leadership Council (SCLC). Available at: https://laedc.org/wpcms/wp-content/uploads/2025/02/LAEDC-2025-LA-Wildfires-Study.pdf

Hug, N. (2023). *resnet.py*. [online] GitHub. Available at: https://github.com/pytorch/vision/blob/main/torchvision/models/resnet.py

Li, B., Hua, Y., Liu, Y. and Lu, M. (2021). *Dilated Fully Convolutional Neural Network for Depth Estimation from a Single Image*. Advances in Science, Technology and Engineering Systems Journal, 5(5). https://doi.org/10.25046/aj060292

Nex, F., Duarte, D., Tonolo, F.G. and Kerle, N. (2019). *Structural Building Damage Detection with Deep Learning: Assessment of a State-of-the-Art CNN in Operational Conditions*. Remote Sensing, 11(23). https://doi.org/10.3390/rs11232765

O'Hara-Glaspie, K.J. (2025). *California fires: Facts, FAQs, and how to help*. [online] World Vision. Available at: www.worldvision.org/disaster-relief-news-stories/california-fires-facts-faqs-how-to-help

OpenAI (2025a). *can you explain in detail about this code "self.layer2 = ResidualBlockV2(16, 32, stride=2, dilation=1)..."*. [online] ChatGPT (April 18, 2025 Version). Available at: https://chatgpt.com/c/67f21151-18ec-800a-81e7-5feff8a55e94

OpenAI (2025b). *I want to use PPM SSNET but only post disaster images as an input*. [online] ChatGPT (April 18, 2025 Version). Available at: https://chatgpt.com/c/67f10d2d-ba30-800a-b3c9-4789ea5e3806

Perez, H., Tah, J.H.M. and Mosavi, A. (2019). *Deep Learning for Detecting Building Defects Using Convolutional Neural Networks*. Sensors, 19(16). https://doi.org/10.3390/s19163556

Tan, M. and Le, Q.V. (2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. In: International Conference on Machine Learning. [online] PMLR, pp.6105–6114. Available at: https://arxiv.org/pdf/1905.11946

Wang, Y., Chew, A.W.Z. and Zhang, L. (2022). *Building damage detection from satellite images after natural disasters on extremely imbalanced datasets*. Automation in Construction, 140. https://doi.org/10.1016/j.autcon.2022.104328

Weber, E. and Kan, H. (2020). *Building disaster damage assessment in satellite imagery with multi-temporal fusion*. arXiv preprint arXiv:2004.05525.

Wu, C., Zhang, F., Xia, J., Xu, Y., Li, G., Xie, J., Du, Z. and Liu, R. (2021). *Building Damage Detection Using U-Net with Attention Mechanism from Pre- and Post-Disaster Remote Sensing Datasets*. Remote Sensing, 13(5). https://doi.org/10.3390/rs13050905

Zhao, H. (2020). *pspnet.py*. [online] GitHub. Available at: https://github.com/hszhao/semseg/blob/master/model/pspnet.py

Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J. (2017). *Pyramid Scene Parsing Network*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.2881–2890.

Zhao, Z., Wang, F., Chen, S., Wang, H. and Cheng, G. (2024). *Deep object segmentation and classification networks for building damage detection using the xBD dataset*. International Journal of Digital Earth, 17(1). https://doi.org/10.1080/17538947.2024.2302577