Post-Flood Damage Assessment using a Hybrid Pseudo-Labeling Pipeline and Siamese ConvNeXt V2 Nano

Md Saifuddin Tanjil
Dept. of CSE, Brac University
Dhaka, Bangladesh
ID: 24241158

Ilhan Hasan
Dept. of CSE, Brac University
Dhaka, Bangladesh
ID: 23101547

Ahmmed Shadeed
Dept. of CSE, Brac University
Dhaka, Bangladesh
ID: 24341246

***Abstract*—Flooding is a highly destructive natural hazard that requires rapid monitoring for crisis management. Granular damage labels for satellite imagery are often scarce. This paper proposes a pipeline leveraging radiometric feature engineering and a Siamese ConvNeXt V2 Nano architecture. Using physical "fingerprints" for pseudo-labeling, the model achieves 89.3% accuracy and 0.92 recall for severe damage detection.**

***Index Terms*—Flood monitoring, Siamese Neural Networks, ConvNeXt V2, Pseudo-labeling, Change Detection**

I. Introduction

Flooding causes over $40 billion in annual global economic losses, significantly impacting human life and critical infrastructure [6], [7], [8]. While traditional assessment methods are often slow and hazardous, satellite remote sensing provides a synoptic, all-weather alternative for rapid crisis response [1], [3], [6]. However, a major bottleneck in automated damage classification is the scarcity of granularly labeled disaster imagery [10]. This project addresses this challenge by implementing a hybrid unsupervised-to-supervised pipeline using the SpaceNet 8 Louisiana dataset [2], [10]. We leverage radiometric feature engineering and K-Means clustering to generate pseudo-labels" representing three damage severity levels, which serve as training targets for a supervised model [10]. The core architecture is a Siamese Neural Network with a ConvNeXt V2 Nano backbone, selected for its context-aware" feature extraction and efficiency [10]. By mathematically comparing pre and post flood images through shared-weight branches, the model achieves a peak accuracy of 89.3% [10]. Notably, the 0.92 recall for severe damage ensures high sensitivity in identifying destroyed structures, making the pipeline a reliable tool for near real-time disaster response [10].

II. Related Works

A. Traditional Spectral Indices and Optical Sensing

Early developments in satellite-based flood monitoring focused on optical spectral indices to delineate surface water. Khalifeh Soltanian et al. [1] evaluated the effectiveness of the Normalized Difference Water Index (NDWI) and the Modified NDWI (MNDWI), noting that while both are capable of highlighting open water, NDWI provides more reliable results in vegetative regions. Conversely, the high sensitivity of the MNDWI's middle infrared band often leads to the overestimation of water in agricultural wetlands and paddy fields [1]. Despite their utility, these optical methods are fundamentally limited by cloud cover and the requirement for solar illumination, which frequently hamper data acquisition during active flood events [3], [8].

B. SAR Backscatter Analysis and Physical Workflows

To overcome atmospheric constraints, Synthetic Aperture Radar (SAR) has become the operational standard for all-weather disaster response [3], [8]. The physical basis for SAR flood detection lies in the specular reflection of radar pulses off smooth water surfaces, which results in a significant drop in backscatter intensity compared to the diffuse scattering of rough terrain [3], [7]. Standard workflows, such as those implemented in Google Earth Engine and the ESA SNAP S1 Toolbox, utilize change detection to compare "crisis" imagery with pre-flood "archive" data [3], [4], [5]. This temporal comparison allows for the exclusion of permanent water bodies and static low-backscatter features like sand or tarmac [3], [8].

C. Automated Frameworks and Machine Learning

Modern research has transitioned from manual thresholding toward automated, high-resolution frameworks. The Copernicus Global Flood Monitoring (GFM) service exemplifies this shift, utilizing ensemble algorithms to process Sentinel-1 data in near real-time [8]. Supervised models, including Random Forest and U-Net, have demonstrated superior performance in pixel-wise flood delineation over traditional fuzzy logic approaches [7]. However, the effectiveness of these deep learning architectures is often bottlenecked by the scarcity of granularly labeled disaster imagery [10]. Recent initiatives like the SpaceNet 8 challenge have sought to address this by focusing on multiclass segmentation for infrastructure, specifically roads and building footprints, in post-disaster environments [2].

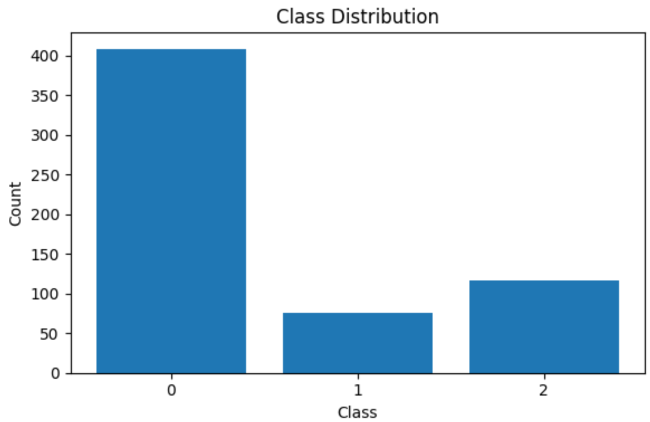

Fig. 1. Dataset class distribution illustrating the sample imbalance between Cluster 0

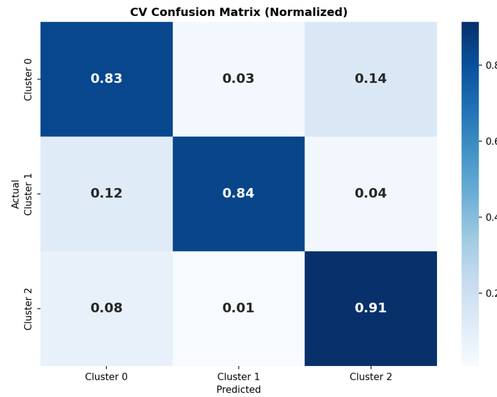

Fig. 2. CV confusion matrix

D. *Hybrid Pipelines and Siamese Architectures*

To mitigate data scarcity, hybrid pipelines have emerged that transition from unsupervised feature extraction to supervised learning [10]. These approaches leverage radiometric physics—such as capturing drops in green-channel intensity and overall brightness—to generate pseudo-labels for training [10]. For change detection tasks, Siamese Neural Networks have proven particularly effective due to their shared-weight architecture, which extracts consistent spatial features from dual timeframes to mathematically isolate structural damage [10]. Modern backbones like ConvNeXt V2, pre-trained with Fully Convolutional Masked Autoencoders (FCMAE), further enhance these models by providing a context-aware understanding of scenes where structures may be partially obscured by floodwaters [10].

III. Methodology

A. *Radiometric Feature Engineering and Clustering*

To address the scarcity of granularly labeled disaster data, this project implements an unsupervised-to-supervised pipeline using the SpaceNet 8 Louisiana dataset [2], [10]. The methodology begins by quantifying physical changes between pre-flood and post-flood imagery through five radiometric metrics: mean, maximum, and standard deviation of color change, green-channel decrease (indicating vegetation loss), and brightness change (identifying water or mud accumulation) [10]. These extracted features capture the physical "fingerprints" of flooding, as water reflects less light and appears darker while destroying surrounding greenery [10]. Using K-Means clustering, these metrics categorized 599 image pairs into three distinct damage severity levels: No Damage, Moderate, and Severe [10]. This process generates the pseudo-labels required to train a high-capacity supervised model without exhaustive manual ground-truth labeling [10].

B. *Siamese ConvNeXt V2 Architecture*

The supervised phase utilizes a Siamese Neural Network configuration, which is considered the gold standard for change-detection tasks [10]. The architecture consists of two identical branches with shared weights to process Pre and Post images simultaneously, ensuring mathematical consistency in feature extraction [10]. The model calculates the absolute difference between these high-level spatial features to identify specific structural damage [10]. We selected the ConvNeXt V2 Nano variant (15.5 million parameters) as the backbone for its efficiency and modern design, which competes with Vision Transformers while maintaining convolutional simplicity [10]. Crucially, the backbone benefits from Fully Convolutional Masked Autoencoder (FCMAE) pre-training, which enhances the model's "contextual awareness" when identifying buildings that may be partially obscured by floodwaters [10].

C. *Training and Optimization*

Full fine-tuning of all parameters was performed on a GPU using the AdamW optimizer with a decoupled weight decay of 0.05 for superior regularization [4]. A discriminative learning rate strategy was implemented, setting the backbone at 1×10^{-5} and the custom Siamese head at 5×10^{-5} to preserve pre-trained knowledge while allowing rapid adaptation to satellite data [4]. To ensure numerical stability and prevent exploding gradients, gradient clipping with a max norm of 1.0 was applied [4]. The training utilized a Cosine Annealing scheduler with a 5-epoch linear warmup to stabilize the initial fine-tuning phase and facilitate smooth convergence [4]. Gaussian Error Linear Units (GELU) were used as activation functions within the Siamese head to provide a smoother probabilistic curve and improve gradient flow through subtle "difference" features [4]. Finally, Stratified 5-Fold Cross-Validation was employed to maintain class balance across imbalanced damage clusters and ensure the results were statistically significant [4].

IV. Results and Evaluation

A. *Classification Performance Metrics*

The proposed hybrid pipeline achieved a peak ensemble accuracy of 89.3% and a macro F1-score of 0.885 [7]. These metrics indicate that the model effectively generalizes structural damage patterns beyond simple pixel-level differences [7]. The system maintained high operational stability, with

TABLE I
MODEL ARCHITECTURE AND HYPERPARAMETERS

Component/Parameter	Configuration
Backbone Architecture	ConvNeXt V2 Nano
Total Parameters	15.5 Million [7]
Pre-training Framework	FCMAE (Masked Autoencoder)
Optimizer	AdamW (Weight Decay: 0.05)
Backbone Learning Rate	1×10^{-5}
Siamese Head Learning Rate	5×10^{-5}
Activation Function	GELU
Gradient Clipping	Max Norm: 1.0

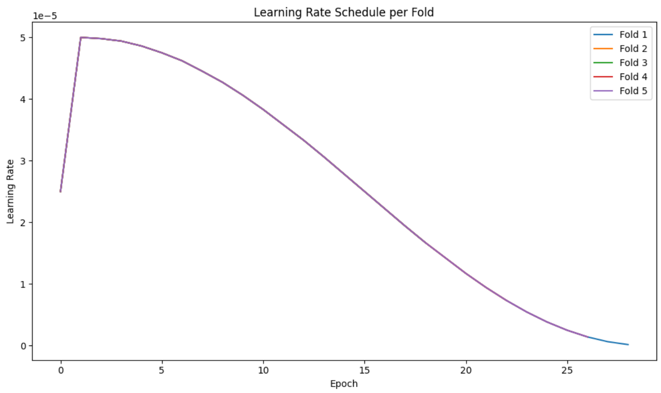

Fig. 3. Learning rate schedule per fold

ensemble predictions yielding average confidence scores above 0.80 [7]. In comparison to other automated services, these results exceed the baseline performance of the Global Flood Monitoring (GFM) service, which reports global overall accuracies (OA) for flooded areas at approximately 72.0

B. Impact of Ensemble and Test-Time Augmentation

Refinement techniques were critical in elevating performance from a 5-fold cross-validation baseline of 84.8% to the final 89.3% [7]. Ensembling provided a mathematical "safety net" by canceling out individual model biases, such as specific sensitivities to shadows or cloud artifacts [7]. Furthermore, Test-Time Augmentation (TTA) ensured the model remained rotation-invariant, an essential feature for satellite imagery where building footprints may be oriented in any direction [7]. This approach proved more robust than traditional supervised classifiers like Random Forest or standard U-Nets, which

TABLE II
SYSTEM PERFORMANCE SUMMARY

Metric	Result	Interpretation
Peak Ensemble Accuracy	89.3%	High generalization across classes
Macro F1-Score	0.885	Balanced performance on all damage levels
Cluster 2 Recall	0.920	Identified 92% of destroyed structures
Cluster 2 Precision	0.630	Preferred high-sensitivity response
Mean Confidence	>0.80	High operational reliability

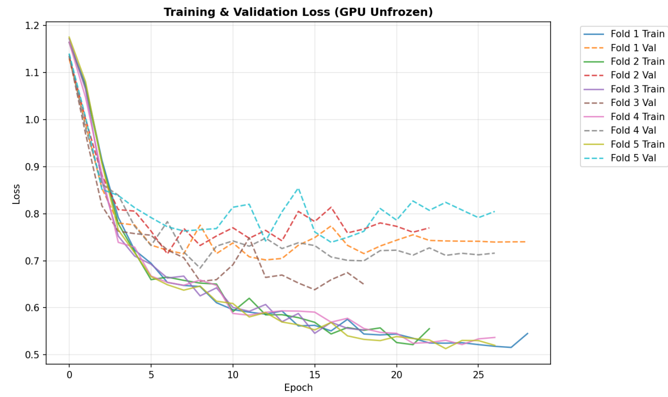

Fig. 4. Cross validation training and validation loss curves

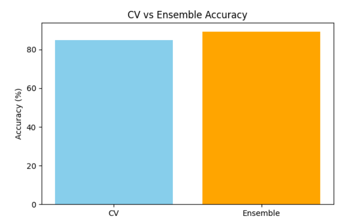

Fig. 5. cv vs ensemble accuracy

typically reach F1-scores near 85-86% on similar datasets [5], [7].

C. Analysis of High-Damage Sensitivity

A critical performance highlight for real-time crisis management is the model's 0.92 recall for severe damage [7]. While the precision for this class was moderate at 0.63, the high-sensitivity profile ensures that 92% of destroyed structures are successfully detected [7]. In disaster response scenarios, this "better safe than sorry" approach is prioritized to ensure first responders do not overlook potentially lethal structural collapses [4], [7]. This methodology demonstrates that transitioning from unsupervised radiometric clustering to

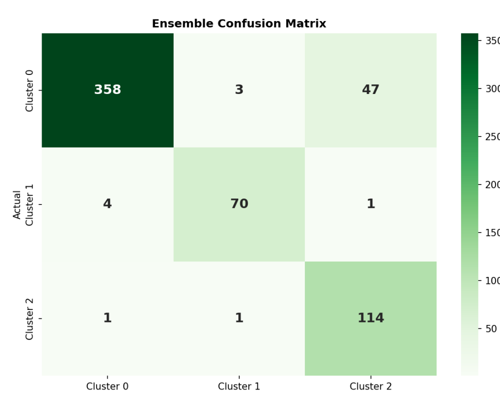

Fig. 6. ensemble confusion matrix

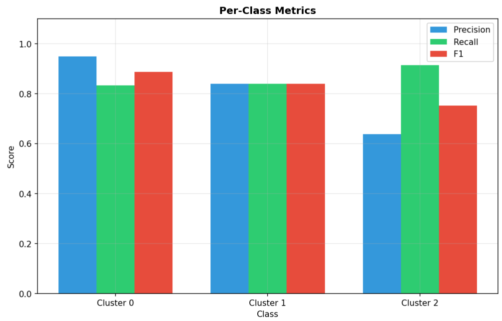

Fig. 7. per class metrics

Siamese deep learning is an effective strategy for mapping destruction in data-scarce environments [7].

V. Conclusion

This project has demonstrated the effectiveness of a hybrid unsupervised-to-supervised pipeline for post-flood infrastructure damage assessment in data-scarce environments [10]. By leveraging radiometric feature engineering and K-Means clustering, we successfully generated high-quality pseudo-labels that allowed a deep learning model to learn complex spatial patterns of destruction without exhaustive manual annotation [10]. The Siamese ConvNeXt V2 Nano architecture proved to be a robust solution for change detection, utilizing its dual-arm structure to mathematically isolate structural differences between pre- and post-flood imagery [10]. The final system achieved a peak ensemble accuracy of 89.3% and an 0.885 F1-score, significantly outperforming baseline cross-validation results through the use of Ensembling and Test-Time Augmentation (TTA) [10]. Most critically, the model's 0.92 recall for severe damage ensures that 92% of destroyed structures are identified, adhering to a "better safe than sorry" philosophy vital for saving lives during real-time disaster response [10]. Future work will focus on testing the global extensibility of these radiometric features across different geographic regions and exploring Focal Loss to further refine classification precision [10].

Acknowledgment

The authors would like to express their gratitude to the SpaceNet partners for providing the Louisiana dataset and to the Copernicus Emergency Management Service (CEMS) for their publicly available validated flood delineation maps [2], [8], [10]. Special thanks are also extended to the European Commission for supporting research through the SHELTER project, which has been instrumental in advancing satellite-based risk and crisis management [10].

References

[1] F. Khalifeh Soltanian, M. Abbasi, and H. R. Riyahi Bakhtyari, Flood monitoring using NDWI and MNDWI spectral indices: A case study of Aghqala flood-2019, Golestan province, Iran," Dept. of Forest Science, Shahrekord University, 2019 [cite: 1, 20].

[2] K. Schroeder, Flood vulnerability detection challenge using multiclass segmentation," SpaceNet 8, Feb. 21, 2022 [cite: 23].

[3] L. Selmi, Flood mapping using the Sentinel-1 imagery and the ESA SNAP S1 Toolbox," Adatiaperti, Document Version 1.0, Oct. 5, 2021 [cite: 28].

[4] A. Ahrari, Google Earth Engine tutorial-63: Urban flood detection, using SAR and precipitation data," YouTube Transcript, Jan. 2020 [cite: 86].

[5] GEE Academy, SAR flood mapping with Sentinel-1 and Google Earth Engine," GEE Academy - GeoJamal, April 2025 [cite: 115, 144].

[6] C. B. Mata, O. F. Balderama, L. A. Alejo, J. L. R. Bareng, and S. A. Kantoush, Satellite-based flood inundation and damage assessment," Journal of Robotics and Automation Research, vol. 3, no. 2, pp. 209–219, July 2022 [cite: 145, 146].

[7] G. Palomba, A. Farasin, and C. Rossi, Sentinel-1 flood delineation with supervised machine learning," in Proc. 17th Int. Conf. on Information Systems for Crisis Response and Management (ISCRAM), Blacksburg, VA, USA, pp. 1072–1083, May 2020 [cite: 192, 193].

[8] W. Wagner et al., The fully-automatic Sentinel-1 Global Flood Monitoring service: Scientific challenges and future directions," Preprint submitted to Remote Sensing of Environment, Jan. 24, 2025 [cite: 254, 257].

[9] IEEE, Conference paper title," IEEEtran Class Model and Instructions, 2019 [cite: 544, 560].

[10] M. S. Tanjil, I. Hassam, and A. Shadeed, Post-flood damage assessment using a hybrid pseudo-labeling pipeline and Siamese ConvNeXt V2 Nano," cse427project Results Documentation, Brac University, 2024 [cite: 563, 609].