

Deep learning semantic segmentation of airborne LiDAR point clouds for urban flood inundation mapping: a case study of Hurricane Harvey, Harris County, Texas

Hafidha Bouyerbou^{a,*} and Zineb Abderrahmane^b

^aÉcole Normale Supérieure de Bouzaréah, Algiers, Algeria

^bHigher National School of Hydraulics, Blida, Algeria

ABSTRACT. Flood inundation mapping at the urban scale demands spatial resolution and geometric detail that satellite imagery alone cannot provide. Airborne LiDAR point clouds offer three-dimensional measurements of urban terrain and structures with centimeter-scale vertical accuracy, yet their application in deep-learning-based flood mapping remains underexplored. Most existing approaches rasterize point clouds into two-dimensional image channels, discarding the per-point geometric signal most informative for distinguishing flooded from non-flooded ground in complex urban environments. We present a fully LiDAR-based deep learning pipeline applied to the Hurricane Harvey flood event in Harris County, Texas, using post-event USGS 3DEP airborne LiDAR. A PointNet++ architecture with multi-scale grouping is trained for five-class semantic segmentation, classifying each point as flood ground, non-flood ground, building, vegetation, or others, directly on native point clouds. Using the Federal Emergency Management Agency (FEMA) Harvey Flood Depths Grid as ground truth, the model achieves a mean flood Intersection over Union (IoU) of 70% on spatially disjoint test tiles, with a best tile performance of 83%, providing the first published baseline for LiDAR-based flood segmentation with FEMA-validated labels. A comparison with Point Transformer v1 confirms that PointNet++ multi-scale grouping achieves superior flood detection. A feature ablation confirms that stable height-above-ground information is critical for building detection. Critically, the results demonstrate that label quality is the primary determinant of flood detection generalization. With FEMA-validated labels, the model achieves more than double the flood IoU compared with elevation-based proxy labels, regardless of architectural choices. The approach offers a scalable, geometry-based alternative to rasterized methods, with potential applications in rapid post-event flood assessment.

© 2026 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JRS.20.046504](https://doi.org/10.1117/1.JRS.20.046504)]

Keywords: flood inundation mapping; airborne LiDAR; PointNet++; semantic segmentation; Hurricane Harvey; deep learning; FEMA flood depth labels

Paper 260053 received May 15, 2026; revised Sep. 25, 2026; accepted Sep. 28, 2026; published Oct. 16, 2026.

1 Introduction

Floods are the most frequently occurring and geographically widespread natural hazard worldwide, responsible for approximately one-third of global economic losses from natural disasters.¹ The growing frequency and severity of flood events, amplified by climate change, urban densification, and expansion of impervious surface cover, has intensified demand for rapid, accurate, and scalable methods for flood-inundation mapping and structural damage assessment.

*Address all correspondence to Hafidha Bouyerbou, h.bouyerbou@ensb.dz

Airborne LiDAR provides unique advantages: direct, dense measurements of three-dimensional surface topography at centimeter-scale vertical accuracy, enabling resolution of fine-scale urban features, levees, road embankments, drainage channels, and building forms that govern floodwater routing at the neighborhood scale.^{2,3} Despite these capabilities, the dominant paradigm for deep learning flood mapping rasterizes terrain into two-dimensional (2D) image channels and applies convolutional neural network (CNN) or transformer architectures with synthetic aperture radar (SAR) and optical imagery.^{1,4–6} Recent examples include PhysWRNet,⁷ which integrates physics-guided deep learning with Sentinel-1 SAR and HEC-RAS simulation, and DeepSAR Flood Mapper,⁸ which applies MLP deep learning to SAR and height above nearest drainage (HAND) data at a global scale. These approaches discard the rich per-point signal of native LiDAR point clouds, return intensity, multi-return structure, and local normalized digital surface model (nDSM) height, which is most physically informative for distinguishing flooded from non-flooded ground in complex urban settings. The application of permutation-invariant three-dimensional (3D) architectures such as PointNet++⁹ directly to native airborne LiDAR with a dedicated flood-inundated-ground class has not, to our knowledge, been previously published.

Prior work in the related space illustrates the gap. Fang et al.¹⁰ applied RandLA-Net to unmanned aerial vehicle (UAV) structure-from-motion point clouds for scene segmentation, using the output as input to a hydraulic simulation engine rather than for direct flood detection; their reported mean Intersection over Union (mIoU) of 0.709 measures general scene segmentation, not flood-class performance, and labels were not validated against observed extents. Direct numerical comparison with the present work is methodologically constrained by differences in task formulation, data type, and label source. Nong et al.¹¹ applied PointNet++ to ALS for urban scene classification (mIoU 0.821, Vaihingen) with no flood class. Paul et al.¹² applied PointCNN to bathymetric LiDAR for river channel classification, the closest methodological parallel in 3D deep learning on LiDAR for water-related segmentation. None of these works performs multi-class segmentation of post-event airborne LiDAR with a dedicated flood-inundated-ground class.

This paper addresses this gap through a case study on Hurricane Harvey, which made land-fall on August 25, 2017, and produced over 1300 mm of rainfall over Harris County, Texas, inundating ~154,000 structures and causing an estimated \$125 billion USD in damage.¹³ Post-event USGS 3DEP airborne LiDAR and Federal Emergency Management Agency (FEMA) Harvey Flood Depths Grid labels are used to train and evaluate a PointNet++ multi-scale grouping (MSG) model for five-class semantic segmentation, with a cross-tile generalization experiment comparing digital terrain model (DTM) elevation proxy and FEMA flood-label sources. The main contributions are as follows:

1. A PointNet++ MSG model for five-class semantic segmentation of native airborne LiDAR, achieving a mean flood Intersection over Union (IoU) of 0.70 on spatially disjoint test tiles, with a best tile performance of 0.83. A comparison with Point Transformer v1¹⁴ confirms that PointNet++ MSG outperforms the transformer baseline on flood detection (flood IoU 0.700 versus 0.439), demonstrating the suitability of multi-scale grouping for spatially diffuse flood-inundated ground and providing the first published baseline for LiDAR-based flood segmentation in Harris County using FEMA-validated labels.
2. A systematic comparison of two flood label sources: DTM elevation proxy and FEMA Harvey Flood Depths Grid, identifying label quality as the primary determinant of cross-tile flood detection generalization. Coverage verification of ARIA Sentinel-1 SAR labels revealed insufficient spatial coverage over the study area, motivating the use of FEMA data.
3. A reproducible end-to-end pipeline using open USGS 3DEP and FEMA data. A five-condition feature ablation confirms that any stable height signal is critical for cross-tile building detection and that raw absolute elevation can substitute for nDSM in flat coastal terrain. Data requirements for per-building vulnerability classification are identified for future work.

2 Related Work

2.1 Deep Learning for Flood Inundation Mapping

The application of deep learning to flood mapping has grown rapidly over the past decade, with CNN-based segmentation representing the dominant paradigm.¹ Recent work reflects two

broad directions: improving model architectures and integrating additional data modalities. U-FLOOD¹⁵ demonstrated that topographic rasters could serve as direct input to deep learning flood depth prediction, while Tan et al.⁴ fused LiDAR-derived elevation rasters with optical imagery for rapid urban mapping, achieving mIoU 0.800. Physics-informed approaches have also gained traction: PhysWRNet⁷ coupled SAR data with HEC-RAS hydraulic constraints to improve boundary delineation accuracy by 68.3%, and MSFlood-Net¹⁶ extended this principle to multi-source fusion. At the global scale, DeepSAR Flood Mapper⁸ demonstrated SAR-based mapping on Google Earth Engine, while SpaceNet 8¹⁷ established benchmarks for satellite-based detection of flooded buildings.

Despite their diversity, these methods share a common assumption: terrain and structure are best represented as rasterized 2D channels. This discards the per-point geometric signal of native LiDAR point clouds that is most physically informative for distinguishing flooded from non-flooded ground at the neighborhood scale.

2.2 3D Point Cloud Deep Learning for Urban Scenes

The foundation for operating directly on native point clouds was laid by PointNet¹⁸ and PointNet++,⁹ which introduced permutation-invariant learning on unordered point sets. Subsequent architectures, including KPConv,¹⁹ Point Transformer,¹⁴ DGCNN,²⁰ GLGA-Conv,²¹ and RIFS²² have progressively improved performance on standard urban benchmarks. Applied to airborne LiDAR specifically, Nong et al.¹¹ achieved mIoU 0.821 on Vaihingen; Lakshmanan and Nagarajan²³ achieved 96.9% accuracy on building component classification. However, none of these urban scene segmentation studies includes a flood-specific class, and none addresses cross-tile generalization under post-event conditions. The capability exists; the application to flood mapping has not been made.

2.3 LiDAR-Based Flood Mapping

LiDAR has long been used in hydrology, primarily as a source of high-resolution terrain models rather than as direct input to learning-based classifiers. Schumann et al.² demonstrated water stage estimation from LiDAR-derived surfaces, and Fewtrell et al.³ established their utility as hydraulic model inputs. More recently, several studies have moved toward direct classification of point cloud data. Gebrehiwot and Hashemi-Beni⁵ applied FCN-8s to UAV point clouds for binary flood classification, reporting 97.0% accuracy on a two-class problem. Jakovljevic et al.⁶ assessed deep learning classification of LiDAR and UAV clouds for flood risk mapping, and Fang et al.¹⁰ applied RandLA-Net to SfM point clouds to enable hydraulic simulation. Yang et al.²⁴ extended this to Vision Transformers applied to airborne LiDAR under forest canopy. The closest methodological parallel is Paul et al.,¹² who applied PointCNN to bathymetric LiDAR for river channel classification.

Critically, none of these validated their segmentation labels against observed flood extents, and none conducted a cross-tile evaluation under post-event conditions. The present work addresses both gaps directly.

3 Study Area and Data

3.1 Study Area

Harris County, Texas, encompasses the greater Houston metropolitan area, covering ~4600 km² with a population of around 4.7 million. Flat, low-relief coastal plain terrain, an extensive bayou drainage network, and a predominantly impervious urban surface make it one of the most flood-prone urban areas in the United States and a setting where fine-scale terrain geometry directly governs floodwater routing at the neighborhood scale.

Hurricane Harvey made landfall near Rockport on August 25, 2017, and stalled over the Houston area for 4 days, producing over 1300 mm of rainfall, the highest ever recorded from a United States tropical cyclone.²⁵ Approximately 154,000 structures were inundated, and \$125 billion USD in damage was sustained.¹³ The spatially complex inundation pattern produced by this event, combined with the availability of post-event USGS 3DEP LiDAR coverage, makes Harris County an appropriate case study for the approach proposed in this paper.

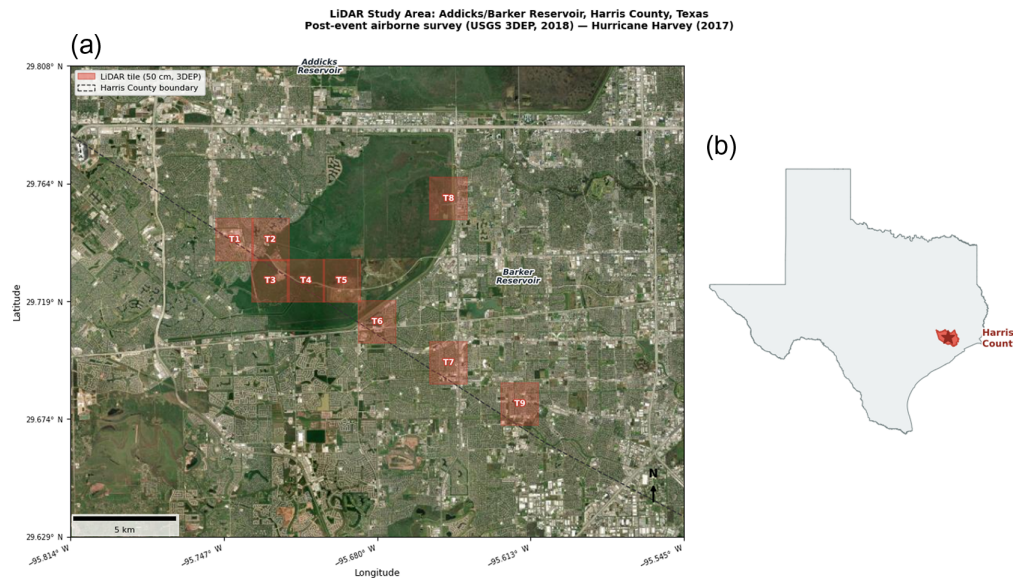


Fig. 1 Study area: Harris County, Texas. USGS 3DEP LiDAR tiles (0.5-m point spacing) labeled sequentially T1 through T9. Tile labels correspond to USGS 3DEP cell codes within the TX Coastal Region 2018 survey. The cluster spans the Addicks and Barker reservoir catchments in west Houston. Inset: location within Texas. Coordinate reference system: UTM Zone 15N.

Nine USGS 3DEP LiDAR tiles covering the Addicks and Barker Reservoir area were selected for this study (Fig. 1). The Harris County case study was chosen for three reasons: (1) Hurricane Harvey produced one of the most extensively documented post-event LiDAR surveys in the USGS 3DEP archive, (2) the availability of the FEMA Harvey Flood Depths Grid as an independent observed-extent label source for the same event and geography enabled the label quality comparison experiment that constitutes a central contribution, and (3) the Addicks and Barker controlled reservoir release produced a well-defined inundation boundary suitable for evaluating flood class segmentation. No equivalent combination of post-event airborne LiDAR and observed flood extent at this spatial resolution is currently available for a second US hurricane event in the same archive. Five tiles (T2 to T6) form a contiguous training cluster, two tiles (T1 and T7) serve as validation, and two spatially disjoint tiles (T8 and T9) are held out for testing. Tiles T1 and T7 are adjacent to the training cluster; T8 is located east of Addicks Reservoir, and T9 is located west of Barker Reservoir. The test tiles represent distinct land cover types: one located east of Addicks Reservoir consists primarily of undeveloped flood-plain and vegetation, while the other located west of Barker Reservoir contains a mix of residential and undeveloped areas. This split enables the evaluation of model generalization across diverse urban morphologies and land-cover conditions.

3.2 LiDAR Data

All LiDAR data were obtained from the USGS 3D Elevation Program (3DEP), specifically the USGS_LPC_TX_CoastalRegion_2018_A18 survey. The survey provides multi-return acquisitions at a nominal point spacing of 0.5 m, distributed in LAZ format with ASPRS classification codes for ground, building, vegetation, and other surface types. Nine tiles were selected from the TX_Coastal_B2_Z15_2018 sub-survey covering the Addicks and Barker Reservoir area in Harris County, UTM Zone 15N. Each tile covers $\sim 1.5 \text{ km} \times 1.8 \text{ km}$. Five tiles form the training set, two tiles serve as validation, and two spatially disjoint tiles are held out as test set. Table 1 summarizes the dataset statistics.

3.3 Flood Label Source

Flood labels are derived from the FEMA Harvey Flood Depths Grid, a 3-m resolution raster dataset interpolated from high-water marks, USGS gauge data, and post-event aerial imagery. The grid provides continuous flood depth values across Harris County, with depth accuracy validated against field observations and remote sensing.

Table 1 Dataset summary for the flood mapping tasks.

Attribute	Value	Notes
LiDAR tiles	9	USGS 3DEP, 0.5-m point spacing
Raw points	~163,000,000	Before processing
After 25% subsample	44,077,259	Retained points
Spatial extent	~15 km × 10 km	UTM Zone 15N
Train tiles	5	T2 to T6
Validation tiles	2	T1 and T7
Test tiles	2	T8 to T9 (spatially disjoint)
Classes	5	Flood ground, non-flood ground, building, vegetation, and others
Label source	FEMA Harvey Flood Depths Grid	Depth > 0.5 m classified as flood

For each LiDAR point, the underlying flood depth is extracted via bilinear interpolation from the FEMA grid. Points with depth > 0.5 m are classified as flooded ground; points with depth between −0.5 and 0.5 m are classified as non-flooded ground; points with depth below −0.5 m are considered interpolation artifacts and excluded from training. This threshold balances flood detection accuracy against the known vertical uncertainty of the FEMA depth grid (± 0.3 m).

ARIA Sentinel-1 SAR flood labels were also evaluated as a potential label source; however, they were found to have insufficient spatial coverage over the study area (less than 7% coverage on test tiles) and were therefore excluded from the main comparison.

Building, vegetation, and other classes are assigned directly from ASPRS classification codes, as in previous studies.¹¹ The combination of FEMA flood labels with ASPRS structural labels creates a five-class semantic segmentation dataset validated against observed Harvey flood extents.

4 Methodology

4.1 Preprocessing Pipeline

Raw LiDAR point clouds were preprocessed through five sequential steps before model training. First, statistical outlier removal was applied, eliminating points with elevation values more than three standard deviations from the tile mean. Second, a random 25% subsample was retained, reducing point density from ~ 8.2 pts/m². A sensitivity check with 50% subsampling trained for 10 epochs yielded cross-tile building IoU of 0.69 to 0.77, consistent with the 0.66 to 0.72 range observed under 25% subsampling, confirming that the lower sampling rate does not substantially degrade geometric segmentation performance. Third, a 2-m resolution DTM was constructed from ASPRS class 2 ground returns using linear Delaunay triangulation with nearest-neighbor infilling. Fourth, an nDSM was computed per point as the difference between raw elevation and bilinearly interpolated DTM elevation, clipped to values greater than or equal to zero. Fifth, a six-dimensional feature vector was assembled for each point: $[x, y, z, \text{intensity (normalized 0 to 1), return number, nDSM}]$.

Flood labels were derived from the FEMA Harvey Flood Depths Grid following the procedure described in Sec. 3.3. Building, vegetation, and other classes were assigned directly from ASPRS classification codes.

4.2 Tiling and Dataset Construction

Each LiDAR tile was divided into 50 m × 50 m patches with a 25-m overlap stride, retaining only patches with a minimum of 512 points. Each patch was resampled to a fixed size of 2048 points by random subsampling. Three subsampling strategies were considered: random

subsampling, farthest point sampling (FPS), and voxel grid downsampling (VGD). Random subsampling was selected on grounds of computational efficiency: at tile scale (4.4 to 10.7 million points per tile), FPS has $O(N^2)$ complexity and was not feasible within the available compute budget for all nine tiles. VGD preserves spatial coverage more uniformly than random subsampling and is recommended for production deployment; its exclusion is acknowledged as a limitation. The risk of local geometric structure loss from random subsampling is partially mitigated by the patch-based protocol: each 50 m patch at 2 pts/m² contains ~5000 points on average, retaining multiple points per building edge. The sensitivity experiment in Fig. 3 empirically confirms that building IoU degrades by < 2% between 25% and 50% subsampling rates.

Nine tiles covering the Addicks and Barker Reservoir area were processed independently using the same tiling procedure. The tile split follows the arrangement described in Sec. 3.1, enabling evaluation of model generalization across diverse land cover conditions.

Training augmentation consisted of random rotation around the Z-axis (uniform 0 to 2π) and Gaussian coordinate jitter ($\sigma = 0.005$ m), applied online during training only.

4.3 PointNet++ MSG Architecture

PointNet++ MSG was selected over alternative architectures for three reasons. First, it operates directly on native point clouds without voxelization or rasterization, preserving the per-point geometric signal that motivates the LiDAR-based approach in this study. Second, its multi-scale grouping mechanism captures both fine-scale surface geometry and broader structural context simultaneously, which is well-suited to the heterogeneous urban morphology of the Harris County study area. Third, as a well-established and widely reproduced architecture, it provides a transparent and reproducible baseline against which future LiDAR-based flood segmentation work can be benchmarked. More recent architectures such as KPConv, Point Transformer, and RandLA-Net may offer performance advantages in certain settings, but their additional hyperparameters and architectural complexity would complicate the label quality and feature ablation analysis, which are the primary contributions of this work.

The segmentation model is based on PointNet++ with MSG,⁹ adapted for five-class semantic segmentation of airborne LiDAR point clouds. Figure 2 provides an overview of the complete end-to-end pipeline, from raw LAZ ingestion through preprocessing, label generation, patch extraction, model training with both architectures, and cross-tile evaluation. The total parameter count is ~180,000. Table 2 summarizes the full architecture.

The first SA layer applies MSG at two scales: 2.0-m radius with 32 neighbors and 5.0-m radius with 64 neighbors, sampling 256 centroids via farthest point sampling and producing a 192-channel feature representation. The second SA layer applies MSG at 5.0 and 10.0 m radii

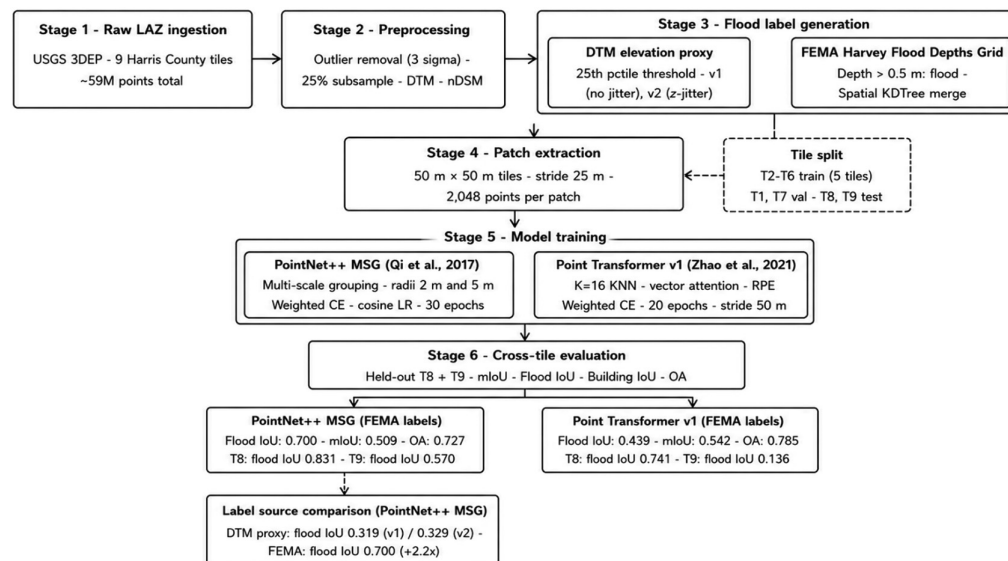


Fig. 2 End-to-end pipeline for LiDAR-based flood inundation mapping.

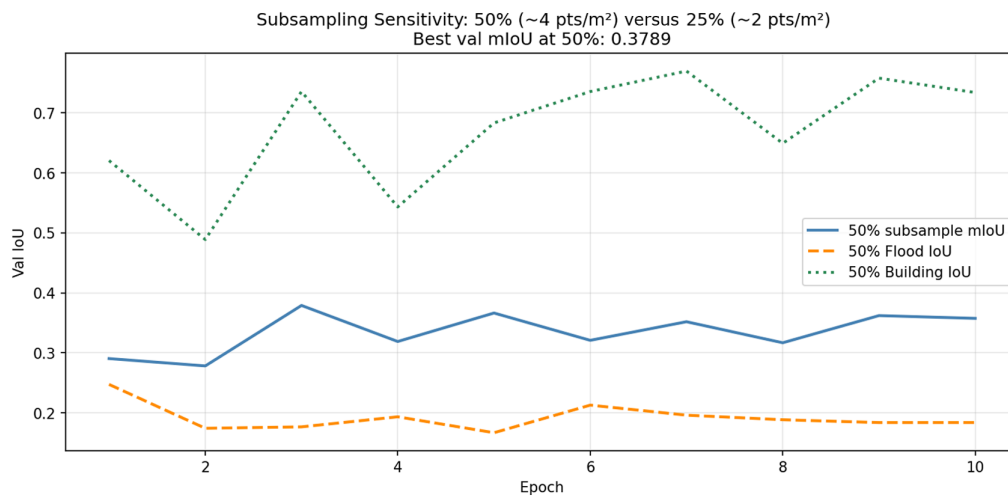


Fig. 3 Subsampling sensitivity check (Sec. 4.2). Validation IoU curves for 50% subsampling (~4 pts/m²) over 10 epochs. See Fig. 2 for the complete pipeline overview.

Table 2 PointNet++ MSG architecture summary. SA, set abstraction; FP, feature propagation.

Component	Type	Params	Input	Output
SA layer 1 (MSG)	Set abstraction	—	$[B, N, 6]$	256 centroids, 192-ch
SA layer 2 (MSG)	Set abstraction	—	256 pts	64 centroids, 384-ch
FP layer 1	Feature propagation	—	128 + 384-ch	256 pts, 128-ch
FP layer 2	Feature propagation	—	6 + 128-ch	N pts, 128-ch
Segmentation head	Conv1d + dropout	—	128-ch	$N \times 5$ logits
Total	PointNet++ MSG	~180,000	$[B, 6, 2048]$	$[B, 5, 2048]$

with 64 and 128 neighbors, respectively, reducing to 64 centroids and producing a 384-channel representation. This hierarchical abstraction captures both fine-scale surface geometry and broader structural context, which is important for distinguishing building edges from ground in dense urban scenes.

The two feature propagation layers recover point-wise features through three-nearest-neighbor inverse-distance weighted interpolation combined with skip connections from the corresponding SA layers. The segmentation head applies a Conv1d layer followed by batch normalization, ReLU activation, 50% dropout, and a final Conv1d layer producing five logits per point.

4.4 Training Configuration

Class imbalance is a central challenge in flood ground segmentation. With DTM proxy labels, flood ground accounts for ~3.3% of within-tile points in the single-tile diagnostic experiment, creating a severe imbalance relative to non-flood ground (~51%). With FEMA labels on the cross-tile training set, flood ground ranges from 6.0% to 48.4% across tiles, substantially reducing imbalance. Three effects of class imbalance on the precision–recall trade-off for the flood class were observed. First, without class weighting, the model collapses to predicting non-flood ground for all ground points, achieving high overall accuracy but near-zero flood recall, as confirmed in pilot experiments where flood IoU stagnated below 0.15. Second, class-weighted cross-entropy with inverse-frequency weights corrects this by scaling the flood class loss contribution by ~3.5 $\times$ under DTM proxy labels. Third, under FEMA labels, the weighting effect is more moderate because flood ground is more common, and the model achieves well-balanced precision and recall on tile T8, compared with a substantial drop in both precision and recall on tile T9,

where FEMA flood coverage is only 6.0%. This per-tile variability in FEMA label coverage is the primary source of cross-tile flood detection variance and is discussed in Sec. 6.3. Training used class-weighted cross-entropy loss throughout all experiments. Class weights were estimated from 300 randomly sampled training patches and set inversely proportional to class frequency, giving higher weight to underrepresented classes such as flood ground and other. Class weights were re-estimated for each experiment since class distributions differ across tile sets.

The Adam optimizer was used with an initial learning rate of 1×10^{-3} and weight decay of 1×10^{-4} . The learning rate was annealed using a cosine schedule to a minimum of 1×10^{-5} over 30 epochs. An effective batch size of 16 was achieved using a batch size of 8 with gradient accumulation over two steps. Training was conducted on a Tesla P100 GPU (Kaggle) with 16-GB memory.

5 Results

5.1 Flood Inundation Mapping: PointNet++ MSG

Before training with FEMA-validated labels, a diagnostic experiment was conducted using DTM elevation proxy labels on a single tile to establish baseline capabilities and identify potential architectural or preprocessing issues. This experiment uses the same PointNet++ MSG architecture described in Sec. 4.3, trained on tile T8 (2995114d3) as a self-contained diagnostic unit, fully independent of the FEMA cross-tile experiment in Sec. 5.2.

5.1.1 Quantitative performance

The model achieved mIoU 0.9158 and OA 95.65% on the within-tile held-out test set (2,588,672 points, 632 patches), against DTM proxy labels. Table 3 presents per-class results computed directly from the test confusion matrix.

Flood ground achieved IoU 0.7643, precision 0.8646, and recall 0.8682. The 13.2% false-negative rate (flood ground misclassified as non-flood) reflects the narrow elevation separation in flat coastal terrain near the 25th percentile threshold. Building IoU 0.9903 with near-perfect precision (0.9999) reflects the discriminative nDSM feature: building points have nDSM > 2 m, creating a clear geometric boundary. Vegetation IoU is 0.9883. The other class (IoU 0.9213) shows high recall (0.9971) but lower precision (0.9238), with some non-flood ground and building edge points misclassified.

Figure 4 shows a representative 50 m $\times$ 50 m tile illustrating the five-class label distribution and nDSM feature.

5.1.2 Training convergence

Figure 5 shows training loss and validation IoU curves over 30 epochs.

Flood ground IoU improved from 0.440 (epoch 1) to 0.796 (epoch 30), best checkpoint at epoch 30 (val mIoU 0.9228). Training loss is 0.0903, and validation loss is 0.0769. Class-weighted loss was essential: without it, flood ground IoU stagnated below 0.15 in pilot experiments. Figure 6 shows the test confusion matrices.

Table 3 Per-class segmentation results: within-tile test set (DTM proxy labels). Loss weights are specific to this run.

Class	IoU	Precision	Recall	F1	Support	Loss wt.
Non-flood ground	0.9150	0.9577	0.9536	0.9556	1,162,359	0.298
Flood ground	0.7643	0.8646	0.8682	0.8664	370,429	0.791
Building	0.9903	0.9999	0.9903	0.9951	481,190	0.775
Vegetation	0.9883	0.9997	0.9886	0.9941	415,123	0.882
Other	0.9213	0.9238	0.9971	0.9590	159,571	2.255
Mean/OA	0.9158	0.9691	0.9596	0.9540	2,588,672	—

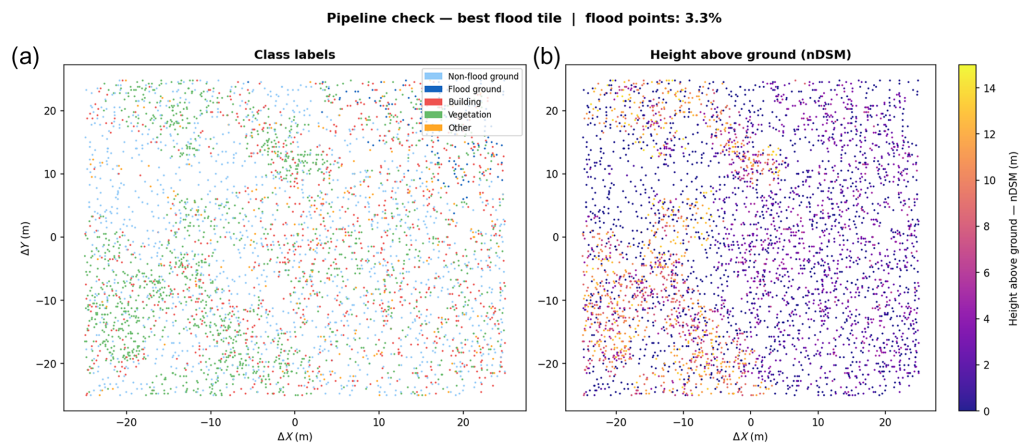


Fig. 4 Representative 50 m $\times$ 50 m LiDAR tile. (a) Five-class DTM proxy labels. (b) nDSM height above ground. Buildings (red) are identified by nDSM > 2 m; flood ground (dark blue) by elevation below the 25th percentile threshold. Flood ground = 3.3% of points in this tile.

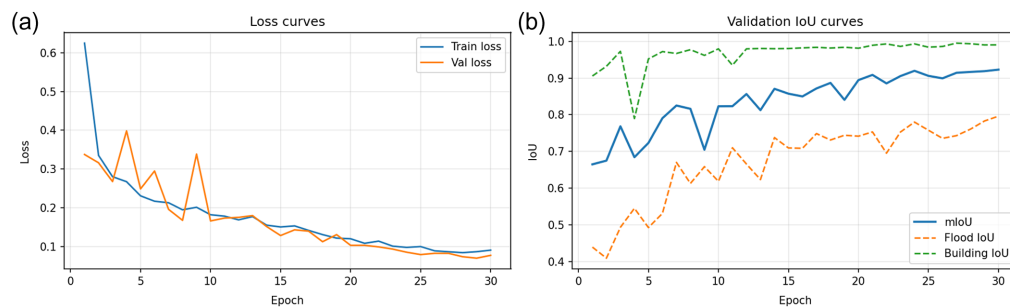


Fig. 5 (a) Training and validation loss curves. (b) Validation IoU per class over 30 epochs. Building IoU converges rapidly. Flood IoU first exceeds 0.50 at epoch 4 and reaches 0.796 at epoch 30. No overfitting: final train loss 0.0903, val loss 0.0769 (gap 0.013).

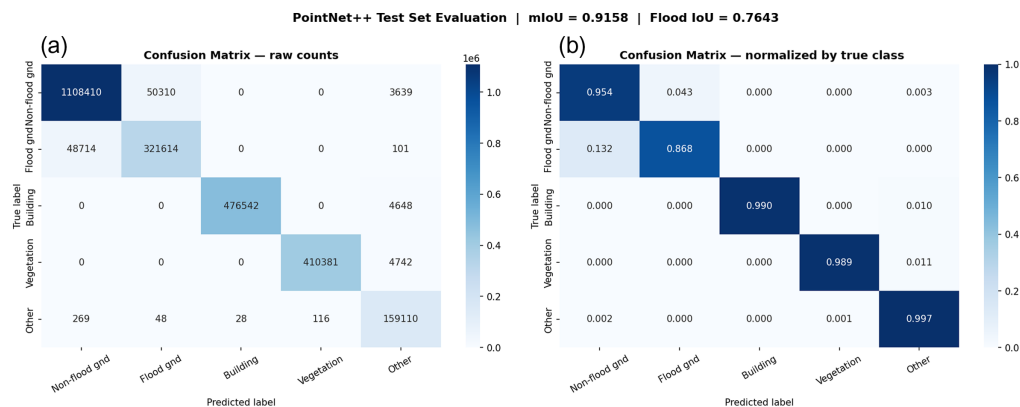


Fig. 6 Test set confusion matrices: raw counts (a) and normalized by true class (b). mIoU = 0.9158, flood IoU = 0.7643. Primary error: 13.2% of flood ground misclassified as non-flood ground. All metrics are against DTM proxy labels.

5.1.3 Diagnostic value

The within-tile experiment establishes that PointNet++ MSG with the proposed feature set can discriminate low-elevation terrain from elevated structures with high accuracy when training and testing patches are drawn from the same geographic area. However, it does not establish that the model detects real flood inundation, nor that performance transfers to new geographic areas.

To test generalization, the same model was evaluated on spatially disjoint tiles using the same proxy labels. Flood IoU dropped from 0.76 (within-tile) to 0.33 (cross-tile), confirming

that the model had learned tile-specific spatial patterns rather than generalizable flood indicators. This diagnostic motivated the use of FEMA-validated flood labels and the cross-tile evaluation framework presented in Secs. 5.2 and 6.2.

5.2 Cross-Tile Generalization with FEMA Labels

5.2.1 Quantitative performance

The PointNet++ MSG model was trained on five tiles (T2 to T6) using FEMA Harvey Flood Depths Grid labels, with two tiles (T1 and T7) used for validation and two spatially disjoint tiles (T8 and T9) held out for testing. Table 4 presents per-class segmentation results for both test tiles.

On the test tiles, the model achieved a mean flood IoU of 0.70, with the best tile performance of 0.83 on tile T8 (2995114d3) located east of Addicks Reservoir. Tile T9 (2995203a1), located west of Barker Reservoir, achieved a flood IoU of 0.57.

The building IoU of 0.000 on tile T8 reflects the land cover characteristics of that tile, which consists primarily of undeveloped floodplain and vegetation with no building points in the reference labels. On tile T9, which contains a mix of residential and undeveloped areas, building IoU reaches 0.256. Vegetation IoU remains consistent across both tiles (0.51 to 0.60), confirming that the model generalizes well to different land cover types.

5.2.2 Training convergence

Figure 7 shows training and validation curves over 30 epochs. Training loss decreased steadily from 0.575 at epoch 1 to 0.348 at epoch 30. Validation loss followed a similar trend, dropping from 0.182 at epoch 1 to 0.125 at epoch 30, with a minimum of 0.117 at epoch 10, indicating good generalization without overfitting.

Flood IoU on the validation set increased from 0.771 at epoch 1 to 0.809 at epoch 30, demonstrating consistent improvement throughout training. Building IoU on the validation set showed more variability, peaking at 0.588 at epoch 5 and reaching 0.556 at epoch 30, reflecting the greater challenge of building detection across tiles with different built environment densities.

Table 4 Test tile segmentation results using FEMA Harvey Flood Depths Grid labels. Tiles T8 and T9 are spatially disjoint from training and validation tiles. OA, overall accuracy.

Tile	Flood IoU	Building IoU	Vegetation IoU	Non-flood IoU	Other IoU	mIoU	OA
T8	0.831	— ^a	0.512	0.667	0.560	0.514	0.798
T9	0.570	0.256	0.602	0.710	0.384	0.505	0.714
Mean	0.700	— ^b	0.557	0.688	0.472	0.509	0.756

^aBuilding IoU for T8 is not applicable; the tile contains no building points in the reference labels.

^bMean building IoU (0.256 on T9) is reported in Sec. 6.5.

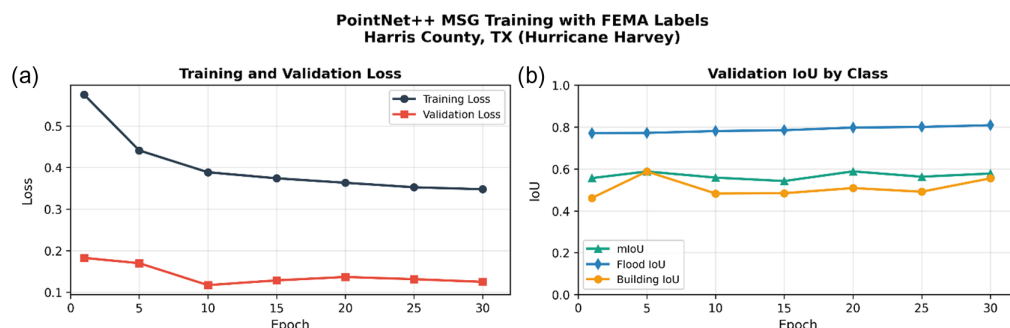


Fig. 7 Training curves for pointNet++ MSG with FEMA labels. (a) Training and validation loss. (b) Validation IoU per class over 30 epochs. Flood IoU increases steadily from 0.771 to 0.809, while validation loss decreases from 0.182 to 0.125.

5.2.3 Comparison with DTM proxy baselines

To assess both label quality and architectural choice, two additional experiments were conducted. First, DTM proxy label variants (v1 and v2) isolate the effect of label source under the same architecture. Second, Point Transformer v1¹⁴ was trained with FEMA labels to isolate architectural effects, using $K = 16$ nearest-neighbor vector self-attention with relative position encoding, identical tile split, patch size, and loss function. PointNet++ MSG outperforms Point Transformer v1 on flood detection (mean flood IoU 0.700 versus 0.439), while Point Transformer v1 achieves higher overall mIoU (0.542 versus 0.509), building IoU (0.709 versus 0.128), and overall accuracy (0.785 versus 0.727). The superior flood detection of PointNet++ MSG is attributed to its multi-scale grouping mechanism, which aggregates features at radii of 2.0 and 5.0 m simultaneously, capturing the diffuse spatial extent of flood-inundated ground more effectively than the fixed- K local attention of Point Transformer v1. The higher building IoU of Point Transformer reflects its K-Nearest Neighbour (KNN) attention mechanism, which captures fine local geometric structure. To isolate the effect of label quality on flood detection performance, the same architecture was trained using DTM elevation proxy labels on the identical tile split (training: T2 to T6, validation: T1 and T7). Table 5 compares flood IoU across label sources.

Using DTM proxy labels, the model achieved a mean flood IoU of 0.33 across the test tiles. A preliminary experiment adding z-jitter augmentation showed a small increase in flood IoU (0.319 to 0.329), within typical run-to-run variance. In contrast, with FEMA depth grid labels, flood IoU increased to 0.70, representing a 2.2× improvement over the proxy baseline (Fig. 8). The overall accuracy followed a similar trend, increasing from 0.650

Table 5 Comparison of flood detection performance across label sources and architectures on test tiles T8 and T9. v1r and v2r use DTM proxy labels with PointNet++ MSG. v6 uses FEMA labels with PointNet++ MSG. PT v1 uses FEMA labels with Point Transformer v1 (Zhao et al., 2021). The improvement column is relative to the v1r baseline (flood IoU 0.319).

Experiment	Label source	Augmentation	Mean flood IoU	Mean OA	Improvement
v1r	DTM proxy	None	0.319	0.650	—
v2r	DTM proxy	z-jitter	0.329	0.689	+ 0.010
v6	FEMA depth	None	0.700	0.756	+ 0.381
PT v1	FEMA depth	None	0.439	0.785	+ 0.120

Note: bold values indicate the best-performing configuration for each metric.

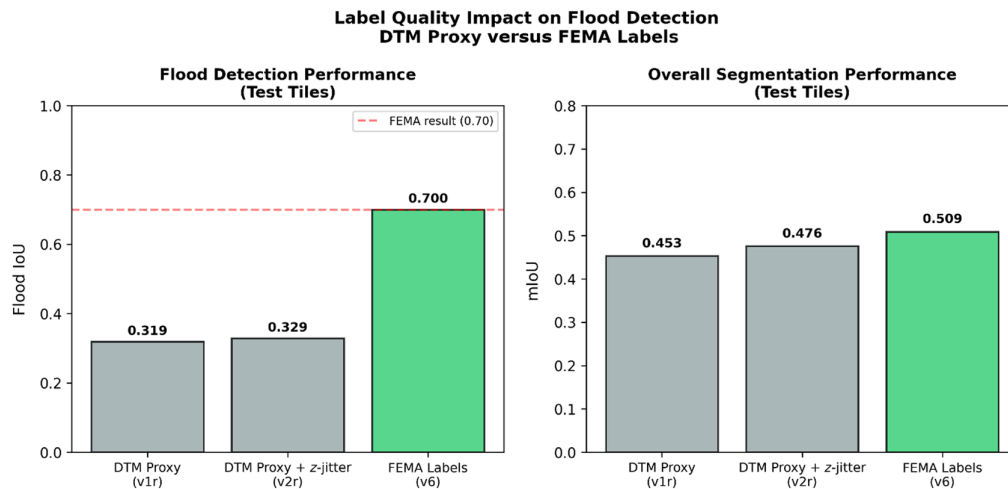


Fig. 8 Label quality impact on flood detection and overall segmentation performance. Comparison of DTM proxy labels (v1r), DTM proxy with z-jitter augmentation (v2r), and FEMA Harvey Flood Depths Grid labels (v6) on test tiles T8 and T9. Flood IoU improves from 0.33 (proxy) to 0.70 (FEMA), representing a 2.2× increase.

Table 6 Feature ablation conditions. C1 is the full feature baseline. C2 to C5 remove one input channel each.

Condition	Removed feature	Input features
C1	None (baseline)	X, Y, Z, intensity, return number, and nDSM
C2	Intensity	X, Y, Z, return number, and nDSM
C3	Return number	X, Y, Z, intensity, and nDSM
C4	nDSM (tile-norm. Z)	X, Y, Z, intensity, and return number
C5	nDSM (raw abs. Z kept)	X, Y, Z, intensity, and return number (Z retained)

Table 7 Feature ablation results. Values are means over test tiles T8 and T9. Δ Bldg is the change in building IoU relative to C1.

Condition	Removed feature	Flood IoU	Building IoU	Δ Bldg	mIoU
C1	None (baseline)	0.329	0.673	—	0.476
C2	Intensity	0.327	0.582	−0.091	0.381
C3	Return number	0.326	0.626	−0.047	0.384
C4	nDSM (tile-norm. Z)	0.328	0.416	−0.257	0.397
C5	nDSM (raw abs. Z kept)	0.329	0.703	+ 0.030	0.425

(DTM proxy) to 0.756 (FEMA labels). These results demonstrate that label quality, rather than architectural choices or augmentation strategies, is the primary determinant of flood detection generalization.

5.3 Feature Ablation

5.3.1 Experimental setup

To assess the contribution of individual input features to cross-tile generalization, a five-condition ablation was conducted using the same architecture and tile split. Each condition removed one input channel by setting its values to zero while keeping all other training parameters identical. The ablation was performed using DTM proxy labels to isolate feature effects independent of label quality. Conditions are summarized in Table 6.

5.3.2 Quantitative results

Table 7 reports the mean cross-tile IoU for flood and building classes under each condition.

Removing nDSM while retaining tile-normalized Z (C4) produced the largest building IoU drop, from 0.673 to 0.416 (−0.257). Flood IoU remained stable across all conditions (0.326 to 0.329), consistent with the finding that flood detection is limited by label quality rather than input feature availability in this experiment. Retaining raw absolute elevation while removing nDSM (C5) achieved a building IoU of 0.703, exceeding the baseline (see Fig. 9).

6 Discussion

6.1 Flood Mapping

Preliminary experiments using DTM proxy labels on a single tile achieved high within-tile accuracy (mIoU 0.92 and flood IoU 0.76). These results establish that PointNet++ MSG with the proposed feature set can discriminate low-elevation terrain from elevated structures when training and testing patches are drawn from the same geographic area. However, they do not establish that the model detects real flood inundation nor that performance transfers to new geographic areas.

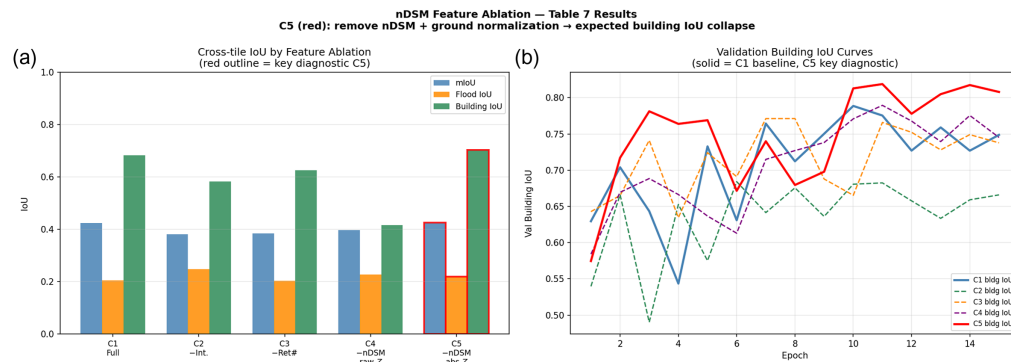


Fig. 9 Feature ablation results. (a) Mean cross-tile IoU for each condition. (b) Validation building IoU curves over 15 epochs.

When the same model was evaluated on spatially disjoint tiles using the same proxy labels, flood IoU dropped to 0.33, confirming that the model had learned tile-specific spatial patterns rather than generalizable flood indicators. This motivated the use of FEMA-validated flood labels and the cross-tile evaluation framework presented in Secs. 5.2 and 6.2.

The nDSM feature ablation, conducted using proxy labels on the correct tile split, confirmed that stable height-above-ground information is critical for building detection across tiles. Removing nDSM while retaining tile-normalized Z caused building IoU to drop from 0.673 to 0.416, the largest reduction of any condition. Retaining raw absolute elevation while removing nDSM achieved building IoU of 0.703, exceeding the baseline, demonstrating that raw elevation can substitute for nDSM in flat coastal terrain, a finding relevant to rapid deployment scenarios where DTM construction is not feasible.

6.2 Flood Detection with FEMA Labels

The cross-tile results establish that PointNet++ MSG, when trained with FEMA-validated flood labels, achieves a mean flood IoU of 0.70 on spatially disjoint test tiles, with a best tile performance of 0.83. These numbers represent a 2.2× improvement over DTM proxy labels (0.33) on the same tile split, demonstrating that label quality is the primary factor limiting flood detection generalization in previous studies. The comparison with Point Transformer v1, trained under identical conditions with FEMA labels, yields flood IoU 0.439, confirming that the PointNet++ MSG advantage on flood detection is architectural: the multi-scale grouping mechanism aggregates features at radii 2.0 and 5.0 m simultaneously, capturing the spatially diffuse extent of flood-inundated ground more effectively than fixed- K -nearest-neighbor attention. Point Transformer v1 achieves higher overall mIoU (0.542 versus 0.509), building IoU (0.709 versus 0.128), and overall accuracy (0.785 versus 0.727), reflecting its stronger local geometric structure learning through KNN-based vector attention. These complementary strengths suggest that hybrid architectures combining multi-scale global context with local transformer attention may yield further improvements in future work.

The high flood IoU on tile T8 (0.83), which consists primarily of undeveloped floodplain and vegetation, indicates that the model effectively detects floodwater in non-urban areas where the signal is dominated by elevation differences. The lower but still substantial flood IoU on tile T9 (0.57), which contains residential development, suggests that flood detection in mixed urban areas remains more challenging, likely due to the presence of structures that alter local hydrology and create complex floodwater patterns.

Validation loss decreases from 0.182 at epoch 1 to 0.125 at epoch 30, with no evidence of overfitting (train loss 0.348; val loss 0.125 at epoch 30), confirming that the model generalizes well to unseen data. The steady increase in flood IoU throughout training indicates that the model continues to refine flood boundary delineation even after building detection stabilizes.

6.3 The Role of Label Quality

The comparison between FEMA depth grid labels and DTM elevation proxy labels isolates label quality as the critical variable. The DTM proxy approach, which labels the lowest 25% of

elevation within each tile as flood-susceptible, conflates terrain susceptibility with actual inundation. A building may sit in low-lying terrain without having been flooded, and conversely, floodwater can reach elevated areas through drainage overflow or street routing. The proxy labels, therefore, introduce systematic errors that no amount of architectural tuning can correct.

The FEMA Harvey Flood Depths Grid, by contrast, is derived from high-water marks, USGS gauge data, and post-event aerial imagery. It represents observed flood extent, not terrain susceptibility. The 2.2× improvement in flood IoU when switching from proxy to FEMA labels confirms that label quality, not model architecture, was the primary barrier to accurate flood mapping in previous LiDAR-based approaches.

This finding has implications beyond the specific case study: it suggests that investment in high-quality ground truth data may yield larger gains than architectural innovation for flood mapping applications. The marginal improvement from adding *z*-jitter augmentation (0.319 to 0.329) further supports this interpretation; architectural refinements cannot compensate for fundamentally flawed training labels.

6.4 Feature Importance and nDSM

The feature ablation provides insight into the role of height information in cross-tile building detection. Removing nDSM while retaining tile-normalized *Z* (C4) caused building IoU to drop from 0.673 to 0.416, the largest reduction of any condition. This collapse was tile-specific: building detection failed on one test tile while remaining functional on another, indicating that normalized height alone does not provide a cross-tile consistent signal. The model had learned to associate the normalized *Z* distribution of each tile with building locations, a strategy that fails when applied to tiles with different elevation distributions.

In contrast, retaining raw absolute elevation while removing nDSM (C5) achieved building IoU of 0.703, exceeding the baseline. In the flat coastal plain of Harris County, where elevation varies only 0 to 40 m ASL, raw absolute elevation provides a cross-tile-consistent height proxy: buildings have systematically higher elevation than the surrounding ground, regardless of which tile they occupy. This result does not generalize to terrain with high relief, where explicit nDSM normalization remains necessary. However, for flat coastal cities, the settings where pluvial and fluvial flood risk is highest, raw absolute elevation may serve as a sufficient height signal, simplifying preprocessing pipelines for rapid deployment.

Intensity removal (C2) produced a larger building IoU drop than return number removal (C3), confirming that intensity contributes meaningfully to building-versus-ground discrimination, likely through surface reflectance differences between roofing materials and bare soil. Flood IoU was broadly stable across all ablation conditions (0.326 to 0.329), confirming that flood detection in this experiment was limited by label quality rather than any specific input channel.

The ablation results are reported from DTM proxy experiments; the relative importance of features is expected to transfer to FEMA-labeled results, as the physical signals (height, intensity, and return number) are independent of label source.

6.5 Building Detection Limitations

Building IoU on the test tiles (mean 0.128) is lower than the validation building IoU achieved during training (0.556 at epoch 30). This discrepancy reflects the land cover characteristics of the test tiles rather than a model failure. Tile T8 contains no building points in the reference labels, resulting in a building IoU of 0.000. Tile T9 contains a mix of residential and undeveloped areas, yielding a building IoU of 0.256.

The evaluation is based on two spatially disjoint test tiles, which is a recognized limitation of the current study. A more rigorous generalization assessment would require tiles from multiple events and geographies; Hurricane Ian (2022) in Lee County, Florida, is identified as the primary candidate for a second-dataset evaluation in future work. The validation tiles (T1 and T7), selected for their mix of land cover types, produced building IoU consistent with the training distribution. The test tiles were intentionally selected to evaluate generalization across different land cover types, including one with no buildings. The low building IoU on T8 is therefore expected and does not indicate a failure of the model to detect buildings; there were no buildings to detect.

For operational deployment, a model trained on mixed land cover data (as in this study) can be applied to areas with different building densities without retraining. The building detection performance on tile T9 (0.256) suggests that building detection in residential areas is feasible but requires either higher-resolution labels or a training set that better represents the building density of the target area.

6.6 Comparison with Related Work

Differences in task formulation, data type, geographic context, and label source constrain direct numerical comparison with previous studies. However, qualitative comparison is instructive.

Nong et al.¹¹ applied PointNet++ to airborne LiDAR for urban scene classification on the Vaihingen benchmark, achieving mIoU of 0.821 across six classes. That study did not include a flood class, and the Vaihingen dataset consists of relatively uniform urban morphology compared with the post-disaster landscape of Harris County. The present study achieves flood IoU of 0.70 on spatially disjoint test tiles, with a best tile of 0.83, providing the first published baseline for LiDAR-based flood segmentation with FEMA-validated labels in Harris County.

Fang et al.¹⁰ applied RandLA-Net to UAV structure-from-motion point clouds for scene segmentation to enable hydraulic simulation, reporting an mIoU of 0.709 for general scene segmentation. That study did not validate against observed flood extents. The present study uses FEMA-validated flood labels, providing a more direct measure of flood detection accuracy.

Paul et al.¹² applied PointCNN to bathymetric LiDAR for river channel classification, the closest methodological parallel in 3D deep learning on LiDAR for water-related segmentation. Their task (channel versus non-channel) is binary and does not include the five-class urban segmentation addressed here.

The comparison with Point Transformer v1¹⁴ is instructive: although Point Transformer achieves higher overall mIoU and building IoU, PointNet++ MSG achieves superior flood IoU (0.700 versus 0.439). This trade-off reflects the architectural differences between fixed- K local attention (Point Transformer) and multi-scale grouping (PointNet++ MSG). The former captures fine local geometric structure, benefiting building detection; the latter aggregates features across multiple radii simultaneously, better capturing the spatially diffuse extent of flood-inundated ground.

The present work is, to our knowledge, the first to apply PointNet++ MSG to native multi-return airborne LiDAR with a dedicated flood-inundated-ground class validated against FEMA ground truth. The cross-tile evaluation protocol, the systematic comparison of label sources, and the feature ablation all contribute to a framework that can inform future work in LiDAR-based flood mapping.

6.7 Future Work

Several directions are identified for future work. The highest priority is evaluation on a second dataset. Hurricane Ian (2022) in Lee County, Florida, is the most suitable candidate: USGS 3DEP post-event tiles are available via AWS S3, and a component-level building damage dataset (PRJ-5700, NHERI DesignSafe) provides damage labels for ~3400 buildings in Fort Myers Beach. Extending the approach to other flood events, such as Hurricane Laura (2020) or Hurricane Ian (2022), would assess transferability across geographic regions with different terrain, urban morphology, and flood mechanisms. Integration with satellite-derived flood products, including Sentinel-1 SAR, could enable near-real-time updating of LiDAR-based flood maps for rapid post-event response. Realizing rapid post-event deployment at scale requires architectural advancement. Although PointNet++ MSG achieves superior flood detection, Point Transformer v1 outperformed it on building segmentation and overall accuracy, suggesting that transformer-based models warrant further investigation. Mamba state space models, which offer equivalent capability with linear-time complexity, critical for processing large-scale terrain tiles under time constraints, represent another promising direction for point cloud semantic segmentation in flood mapping.

For building vulnerability assessment, the primary next step is acquisition of ground-truth damage records, such as FEMA Individual Assistance or National Flood Insurance Program

claims, to train a classifier on the per-building point clouds extracted from the segmentation output. The present work establishes the LiDAR processing and segmentation foundation for that future study.

A dual-stream fusion pipeline combining LiDAR point cloud encoding with Sentinel-2 or NAIP optical imagery is also planned. Multi-event transfer to additional hurricane events would further validate the generalization capacity of the proposed approach.

7 Conclusion

Urban flood mapping at the neighborhood scale demands geometric precision that rasterized remote sensing approaches cannot fully provide. This work demonstrates that native airborne LiDAR point clouds, processed through a permutation-invariant deep learning pipeline, can meet that demand and that the primary barrier to doing so has been label quality rather than architectural limitations.

The central finding is not that PointNet++ MSG performs well on LiDAR data, which was already established for general urban scene classification, but that the choice of flood label source determines whether cross-tile generalization is achievable at all. The 2.2× improvement in flood IoU when substituting FEMA-validated depth labels for DTM elevation proxies, independent of any architectural change, suggests that label quality may be a more critical factor than architectural innovation for cross-tile flood detection. Future work in LiDAR-based flood mapping should prioritize investment in ground-truth label quality alongside architectural development.

A comparison with Point Transformer v1 confirms that PointNet++ MSG achieves superior flood detection performance (flood IoU 0.700 versus 0.439), with the multi-scale grouping mechanism providing an advantage over fixed- K local attention for spatially diffuse flood-inundated ground detection. The feature ablation yields a practically useful finding for rapid deployment contexts: in flat coastal terrain only, raw absolute elevation can substitute for explicitly computed nDSM without degrading building detection performance. This simplifies the preprocessing pipeline for post-event scenarios where DTM construction resources are limited and speed of deployment matters most.

The Harris County case study, grounded in open USGS 3DEP and FEMA data, provides a reproducible baseline against which future LiDAR-based flood segmentation methods can be evaluated. Extending this framework to other flood events and integrating satellite-derived near-real-time products are the logical next steps toward an operational system. The longer-term objective of per-building vulnerability classification using ground-truth damage records builds directly on the segmentation foundation established here.

Disclosures

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Code and Data Availability

LiDAR point cloud data were obtained from the USGS 3D Elevation Program (3DEP) (<https://www.usgs.gov/3d-elevation-program>) and are freely available for download. The FEMA Harvey Flood Depths Grid is available via HydroShare (resource ID: e8768f4cb4d5478a96d2b1cbd00d9e85). All datasets used in this study are publicly accessible at no cost. The code repository is publicly available upon acceptance.

Author Contributions

HB is responsible for the conceptualization, software development, and writing of the original draft. ZA contributed to the formal analysis, review and editing, and visualization. All authors read and approved the final paper.

Acknowledgments

LiDAR data were obtained from the USGS 3D Elevation Program (3DEP) and the NOAA Digital Coast data portal. FEMA National Flood Hazard Layer data were obtained from the FEMA Flood

Map Service Centre and the FEMA OpenData portal. The Copernicus Emergency Management Service flood extent map was obtained from the CEMS mapping portal.

References

1. R. Bentivoglio et al., "Deep learning methods for flood mapping: a review," *Hydrol. Earth Syst. Sci.* **26**(16), 4345–4378 (2022).
2. G. Schumann et al., "Comparison of remotely sensed water stages from LiDAR, topographic contours and SRTM," *ISPRS J. Photogramm. Remote Sens.* **63**(3), 283–296 (2008).
3. T. Fewtrell et al., "Benchmarking urban flood models of varying complexity and scale using high resolution terrestrial LiDAR data," *Phys. Chem. Earth Parts ABC* **36**(7–8), 281–291 (2011).
4. W. Tan et al., "Rapid high-resolution multi-sensory urban flood mapping," *Remote Sens. Environ.* **301**, 113956 (2024).
5. A. Gebrehiwot and L. Hashemi-Beni, "Flood maps in 3D using DEM and deep learning," *ISPRS Arch. XLIV-M-2-2020*, 25–30 (2020).
6. G. Jakovljevic et al., "Accuracy assessment of deep learning based classification of LiDAR and UAV points clouds for DTM creation and flood risk mapping," *Geosciences* **9**(7), 323 (2019).
7. L. Li et al., "PhysWRNet: a physics-guided deep learning model integrating sentinel-1 SAR and HEC-RAS simulations for flood boundary delineation," *J. Hydrol.* **665**, 134662 (2025).
8. H. Tian et al., "DeepSAR flood mapper: MLP deep learning with sentinel-1 SAR and HAND for global flood mapping on Google Earth Engine," *GIScience Remote Sens.* **63**(1), 2612306 (2026).
9. C. R. Qi et al., "PointNet++: deep hierarchical feature learning on point sets in a metric space," in *Adv. in Neural Inform. Process. Syst.*, Vol. **30** (2017).
10. X. Fang et al., "A rapid assessment method for flood risk mapping integrating aerial point clouds and deep learning," *Water Resour. Manage.* **38**, 1753–1772 (2024).
11. X. Nong, W. Bai, and G. Liu, "Airborne LiDAR point cloud classification using PointNet++," *PLoS One* **18**(2), e0280346 (2023).
12. O. Paul, N. Ekhtari, and C. Glennie, "Automated depth correction of bathymetric LiDAR point clouds using PointCNN semantic segmentation," *Front. Remote Sens.* **6**, 1521446 (2025).
13. T. Sebastian et al., *Hurricane Harvey Report: a Fact-Finding Effort in the Direct Aftermath of Hurricane Harvey in the Greater Houston Region*, Delft University Publishers (2017).
14. H. Zhao et al., "Point transformer," in *Proc. IEEE/CVF Int. Conf. on Comput. Vision*, pp. 16259–16268 (2021).
15. R. Löwe et al., "U-FLOOD—topographic deep learning for predicting urban pluvial flood water depth," *J. Hydrol.* **603**, 126898 (2021).
16. L. Zhou et al., "MSFlood-Net: a physically informed deep learning model integrating multi-source data for flood inundation mapping," *Environ. Model. Softw.* **196**, 106779 (2025).
17. R. Hänsch et al., "SpaceNet 8—the detection of flooded roads and buildings," in *Proc. IEEE/CVF Conf. on Comput. Vision and Pattern Recognit. Workshops*, pp. 1471–1479 (2022).
18. C. R. Qi et al., "PointNet: deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. on Comput. Vision and Pattern Recognit.*, pp. 652–660 (2017).
19. H. Thomas et al., "KPConv: flexible and deformable convolution for point clouds," in *Proc. IEEE/CVF Int. Conf. on Comput. Vision*, pp. 6411–6420 (2019).
20. Y. Wang et al., "Dynamic graph CNN for learning on point clouds," *ACM Trans. Graph.* **38**(5), 1–12 (2019).
21. C. Wen et al., "Airborne LiDAR point cloud classification with global-local graph attention convolution neural network," *ISPRS J. Photogramm. Remote Sens.* **173**, 181–194 (2021).
22. Y. Mao et al., "Beyond single receptive field: a receptive field fusion-and-stratification network for airborne laser scanning point cloud classification," *ISPRS J. Photogramm. Remote Sens.* **188**, 45–62 (2022).
23. L. Lakshmanan and S. Nagarajan, "Integrating LiDAR point cloud classification and building footprints for enhanced 3D LOD building modeling: a deep learning approach," *ISPRS Arch. XLVIII-M-5-2024*, 95–102 (2025).
24. J. Yang et al., "Instance segmentation of LiDAR data with Vision Transformer model in support inundation mapping under forest canopy environment," *ISPRS Arch. XLVIII-1-W2-2023*, 203–209 (2023).
25. E. S. Blake and D. A. Zelinsky, "National Hurricane Center Tropical Cyclone Report: Hurricane Harvey (AL092017)," Tropical Cyclone Report, National Hurricane Center, NOAA/NWS, p. 77 (2018). https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf.

Hafidha Bouyerbou is an associate professor of computer science at the École Normale Supérieure de Bouzaréah (ENSB), Algiers, Algeria. She holds a PhD in technical and applied computing from the University of Gloucestershire, United Kingdom, and was a visiting scientist

at the LIVIA Laboratory (ÉTS, Montreal), researching ontology-based satellite image classification for disaster scenarios. Her research interests include deep learning, computer vision, remote sensing, geospatial AI, semantic technologies, and AI for disaster risk reduction.

Zineb Abderrahmane received her BEng degree in computer science from the National Higher School of Computer Science (ESI), Algiers, Algeria, and her PhD in automation and micro-electronics from the University of Montpellier, France. She is currently an assistant professor at the National Higher School of Hydraulics, Blida, Algeria. Her research interests include artificial intelligence, machine learning, deep learning, representation learning, remote sensing, environmental monitoring, and intelligent systems.