Deep Learning-Based Downscaling of Global Digital Elevation Models for Enhanced Urban Flood Modeling

Zanko Zandsalimi[a], **Sergio A. Barbosa**[a], **Negin Alemazkoor**[a], **Jonathan L. Goodall**[a], **Majid Shafiee-Jood**[a,*]

[a]Department of Civil and Environmental Engineering, University of Virginia, Charlottesville, VA, USA.
Email addresses: Z. Zandsalimi (`mfx2uq@virginia.edu`), S. A. Barbosa (`rqt5kh@virginia.edu`), N. Alemazkoor (`na7fp@virginia.edu`), J. L. Goodall (`goodall@virginia.edu`), M. Shafiee-Jood (`ms2dm@virginia.edu`).
[*]Corresponding author, `ms2dm@virginia.edu`

This is an Accepted Manuscript of an article published in the Journal of Hydrology, Vol 653
available online: https://www.sciencedirect.com/science/article/pii/S0022169425000253

Abstract

Urban flood modeling depends heavily on the quality of Digital Elevation Models (DEMs). However, accurate, high-resolution DEMs are often expensive and not widely available, particularly in data-limited regions. Consequently, researchers frequently rely on Global Digital Elevation Models (GDEMs), which suffer from vertical biases and limited spatial resolution. This limitation is especially critical in urban settings, where detailed terrain features are essential for accurate flood prediction. In this study, we introduce a novel methodology that leverages Convolutional Neural Network (CNN) architecture (U-Net) and utilizes GDEMs and other publicly available datasets (e.g., Landsat-8, Sentinel-1, and Sentinel-2) to produce an enhanced DEM with a 5-meter spatial resolution. Using USGS high-resolution DEMs as a reference, our results demonstrate that our method is able to generate DEMs with significantly lower vertical biases (82.1% lower RMSE and 87.8% lower MAE) compared to GDEMs. Additionally, the model produces a more detailed representation of urban features that are essential for flood pattern analysis. By applying this improved DEM within a flood simulation model, we show that the Probability of Detection increases by 12% increase and the False Alarm Ratio decreases by 13% compared to GDEMs. These findings underscore the potential of using deep learning and multi-source data to improve DEM quality for more accurate urban flood modeling and management in data-limited regions.

Keywords: Digital Elevation Models; Deep Learning; Convolutional Neural Network; Urban Flood Modeling; Remote Imagery Data; Downscaling.

1 Introduction

Flood events pose a significant global challenge and can have major, far-reaching social, economic, and environmental impacts in both developed and developing countries (F. Li et al., 2019). A recent study estimates that approximately 1.81 billion people, or 23% of the global population, are directly exposed to significant flood risks, with approximately $9.8 trillion of economic activity located in areas at significant flood risk, equivalent to about 12% of global GDP in 2020 (Rentschler et al., 2022). The increasing frequency and severity of floods are attributed to a combination of factors, including more intense storms, urban development, dense population, reduced soil infiltration capacity, and the accelerating pace of climate change (Fereshtehpour & Karamouz, 2018; Pasquier et al., 2019; O'Donnell & Thorne, 2020). With an estimated 68% of the global population expected to reside in cities by 2050 (UN, 2019), many of which are located in low-lying, flood-prone regions, the growing complexity and risk of flood events underscore the critical need for more accurate flood mapping and impact estimation in urban areas (Ford et al., 2019; Y. Liu et al., 2021).

Digital elevation models (DEMs) are essential tools for flood modeling, providing the critical topographic data needed to accurately simulate water flow and flood pattern. Therefore, the quality of DEMs directly influences the accuracy of flood model outputs, particularly in delineating inundation extents (Courty et al., 2019; McClean et al., 2020; Shastry & Durand, 2020; Zhang et al., 2019; Zandsalimi et al., 2024). DEMs are produced from a variety of sources, varying in resolution and cost, from highly accurate and detailed but expensive options to those with coarser resolution and lower cost. Among these, accurate DEMs can be obtained from airborne LiDAR surveys or from topographic surveys. Indeed, airborne LiDAR DEMs provide superior vertical accuracy at high spatial resolutions. However, it comes with high costs and limited availability; open-access LiDAR data covers merely 0.005% of the Earth's land surface. This scarcity is attributed to the advanced technology, resources, and significant acquisition expenses involved (Hawker et al., 2018; Jarihani et al., 2015; Schumann & Bates, 2018). Open-access spaceborne DEMs, developed based on techniques such as radar interferometry, optical stereo matching, or photogrammetry, offer alternative options (Hawker et al., 2018; Nandam & Patel, 2024; Zhang et al., 2019).

In the last two decades, a variety of global or nearly Global digital elevation models (GDEM) have been made freely accessible, including the Shuttle Radar Topography Mission (SRTM) DEM (Farr et al., 2007), the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) DEM (Hirt et al., 2010), the ALOS World 3D-30m (AW3D30) DEM (Tadono et al., 2016), the Copernicus DEM (Fahrland et al., 2020), and NASADEM (Crippen et al., 2016). Although spaceborne DEMs are valuable alternatives to airborne DEMs, particularly in data-scarce regions, they have several drawbacks, including significant observational errors, a lack of fine spatial resolution, and the presence of noise, all of which contribute to uncertainty in the flood modeling and risk assessment (Zandsalimi et al., 2024; Biswal et al., 2023; Hinkel et al., 2021; Nandam & Patel, 2024; Sampson et al., 2016; Yan et al., 2015; Fernández et al., 2016). Given these limitations, it becomes imperative to innovate and develop methodologies aimed at enhancing the quality of topographical data derived from these models. In this research, we introduce a deep learning-based methodology that combines different GDEMs with remotely sensed imagery, road networks, and building footprints to enhance the accuracy and spatial resolution of GDEMs tailored for urban flood modeling applications.

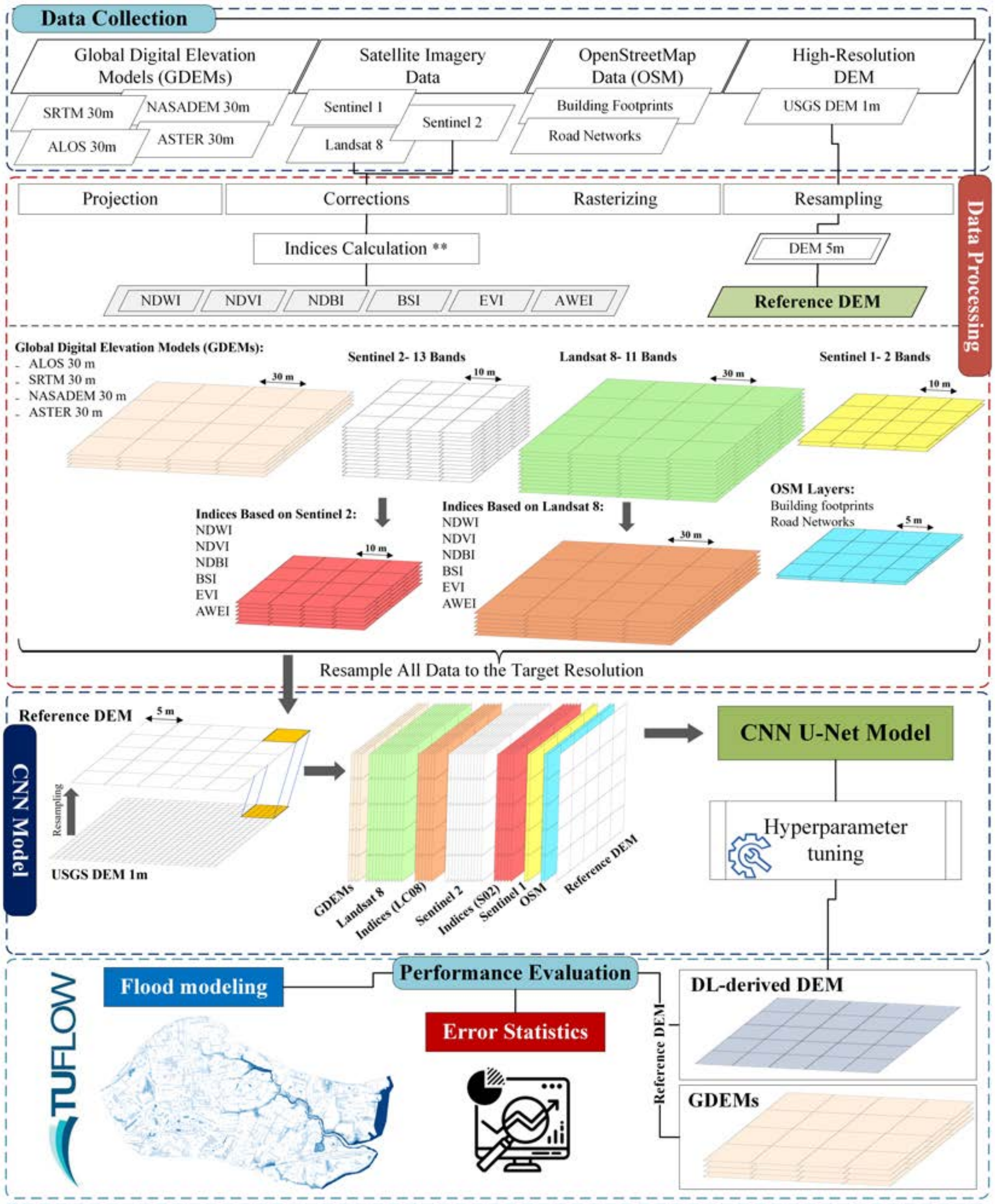
A wide range of algorithms has been developed to correct errors in DEMs. These algorithms can be broadly classified into two categories: parametric regression and non-parametric regression (Ouyang et al., 2023). Parametric regression models focus on identifying and correcting trends in DEM errors by applying models that fit the deviations found in DEM data with variables associated with errors, which are obtained from high-quality supplementary elevation measurements. These variables include factors such as canopy cover, slope, aspect, elevation, vegetation height, and density (Hawker et al., 2018; Y. Li et al., 2022; Magruder et al., 2021; Preety et al., 2022; Su et al., 2015). For instance, Su et al. (2015) employed a multiple linear regression model to correct errors in SRTM data across vegetated mountainous terrains, utilizing terrain slope and vegetation parameters as predictive variables. Zhou et al. (2020) also introduced a method focused on terrain to address both local and global inaccuracies in SRTM DEM data across mountainous regions. They adopted an adaptive approach for error modeling in the SRTM DEM, leveraging the Bayesian information criterion for initial modeling, and then applying an M-estimator for parameter determination. This method improves the elevation accuracy by 20%. Nonetheless, the implementation of this correction technique requires the determination of 15 model parameters.

The second group of models involve non-parametric machine learning algorithms. Recent advancements in machine learning have enabled the development of sophisticated, high-dimensional models that capture the complex, nonlinear, and stochastic relationships between dependent and independent variables. Indeed, the broad application of machine learning algorithms has proven effective in tackling difficult challenges (Shen, 2018), notably in refining DEMs to correct biases. A few studies in the literature have also demonstrated the efficacy of these machine learning techniques (Biswal et al., 2023; Kasi et al., 2020; Kim et al., 2020, 2021; Kulp & Strauss, 2018; Meadows & Wilson, 2021; Ouyang et al., 2023; Robinson et al., 2014; Wendi et al., 2016; Yue et al., 2017). For instance, Kulp and Strauss (2018) leveraged Artificial Neural Networks (ANNs) to enhance SRTM data in coastal regions, incorporating variables such as slope, canopy height, vegetation density, data from the ICESat, and population density as inputs to the ANN model, resulting in approximately a 50% improvement in the RMSE of the derived SRTM DEM. However, their method was constrained to forested areas and is less effective in densely populated urban regions. In another study, conducted by Kim et al. (2020), an ANN algorithm was employed in conjunction with SRTM data and Sentinel-2 multi-spectral imagery to enhance the precision of SRTM DEMs across urban landscapes, aiming for a target spatial resolution of 10 meters. A high-quality DEM served as the target data for training the ANN, which subsequently facilitated the generation of a high-quality DEM with an improvement in RMSE by approximately 38%. Furthermore, Meadows and Wilson (2021) applied three machine learning techniques, including Random Forest, Densely Connected Neural Networks, and Fully Convolutional Neural Networks, to adjust for biases in riverine bathymetry. This was achieved by analyzing the correlations between SRTM and LiDAR DEMs, alongside cloud-free composite imagery from Landsat and Sentinel satellites, complemented by 11 spectral indices. It is important to highlight that the regions examined by Meadows and Wilson (2021) predominantly consisted of open spaces. Additionally, Biswal et al. (2023) developed a method employing hybrid machine learning-based Multi-DEM Ensemble techniques with GDEMs like SRTM, ASTER, and ALOS-AW3D30 to correct river cross-sections. This approach improves hydrodynamic modeling in data-sparse regions, demonstrating the effectiveness of using multiple DEMs in accurately simulating river dynamics with publicly accessible DEMs.

Our literature review reveals three significant gaps in existing research on improving the accuracy ans spatial resolution of GDEMs. First, most studies have focused on enhancing GDEM accuracy in open spaces, where terrain is relatively simple compared to the complexities of urban environments. Second, when urban areas have been examined, efforts typically refined GDEMs to a spatial resolution (e.g., 10 meters or coarser) that is still insufficient for detailed hydraulic modeling and flood mapping. Third, many studies relied on a single DEM source and limited remote imagery data, significantly restricting the comprehensiveness and accuracy of the resulting models. In this study, we address these gaps by developing a deep learning-based methodology using the U-Net Convolutional Neural Network (CNN) architecture that is flexible to utilize and integrate a combination of multiple datasets, including GDEMs (i.e., ALOS, SRTM, NASADEM, and ASTER), satellite imagery from Landsat-8, Sentinel-2, Sentinel-1, and OpenStreetMap data for road networks and building footprints, and produces improved DEMs (referred to as deep learning-derived DEM or simply DE-derived DEM) with a spatial resolution of 5 meters. We use error indices to compare the performance of the DL-derived DEM with GDEMs. We then demonstrate the value of the proposed methodology by analyzing flood simulation results and comparing the accuracy of flood extent mapping.

2 Materials and Methods

Figure 1 illustrates the methodological framework proposed in this research. First, we collect data from various sources and process them (section 2.1). Then, we develop a CNN U-Net architecture tailored to integrate these various datasets (section 2.2). The data is then split into training, validation, and testing sets to ensure robust model performance (section 2.4). We then conduct hyperparameter tuning to optimize the accuracy of the model. The output of the model is DL-derived DEMs. Finally, we develop a flood simulation model to compare the performance of the DL-derived DEMs compared to GDEMs. In the following, we provide more detailed explanations of each step.



** NDWI: Normalized Difference Water Index, NDVI: Normalized Difference Vegetation Index, NDBI: Normalized Difference Built-up Index, BSI: Bare Soil Index, EVI: Enhanced Vegetation Index, and AWEI: Automated Water Extraction Index.

Figure 1. Schematic of the methodology proposed for enhancing GDEMs using CNN U-Net.

2.1 Data

2.1.1 Topographic Data and Remote Sensing Imagery Data

In this study, we utilized four different publicly available GDEMs: ALOS PALSAR (Tadono et al., 2016), SRTM (Farr et al., 2007), NASADEM (Bettiol et al., 2021), and ASTER (Hirt et al., 2010) (see Table 1). These GDEMs have a spatial resolution of 30 meters and are derived using different remote-sensing technologies, notably Interferometric Synthetic Aperture Radar (InSAR) and Stereography. InSAR is particularly effective at penetrating vegetation and canopy cover, offering more accurate representations of the surface elevation. Stereography, used in models like ALOS, relies on optical imagery which although effective in clear conditions, tends to overestimate elevation in areas with dense vegetation or cloud cover by mistaking canopy tops for ground level (Muench et al., 2022). This variance in technology not only affects the vertical accuracy of the data but also its applicability across different regions and environments. Incorporating these different datasets in our model enhances the deep learning algorithm by providing a richer training set that includes multiple perspectives of terrain representation. This ensures that the model is well-prepared to analyze and interpret diverse geographical features and elevation data more accurately, enhancing both its robustness and generalizability. Additionally, we collected the U.S. Geological Survey (USGS) high-resolution DEM as our reference or target DEM. USGS standard one-meter DEMs (with vertical accuracy of $RMSE = 10cm$) are produced exclusively from high-resolution light detection and ranging (Lidar) source data of one-meter or higher resolution through the 3D Elevation Program (USGS, 2016).

Table 1. The Global Digital Elevation Models (GDEMs) considered in this study.

Dataset	Spatial Resolution (m)	Method	Spatial Coverage	Estimated Vertical Accuracy (m)	Released Year
ALOS	30	Stereography	82°N to 82°S	5	2016
ASTER	30	Stereography	82°N to 82°S	5	2016
SRTM	30	InSAR	60°N to 56°S	9	Since 2015
NASADEM	30	InSAR	60°N to 56°S	5	2020

In this study, we also utilized both optical and radar-based remote sensing imagery, including Landsat 8, Sentinel 2, and Sentinel 1, to leverage the distinct advantages of each sensor system, allowing for generation of more accurate and high-resolution DEMs and improving their utility in applications like flood modeling and terrain analysis. Landsat 8 offers high radiometric resolution and capturing data in both visible and infrared spectra and is widely used for monitoring changes in land use and land cover (Alam et al., 2020). Landsat 8 product helps identify surface features that can contribute to refining DEMs. Sentinel 2 product aims at monitoring variability in land surface conditions at a high resolution, facilitating detailed observation of surface changes and enhancing the accuracy of DEMs by providing precise information on land cover changes (ESA, 2024). Sentinel 1 is especially effective in monitoring sea ice, detecting oil spills, and observing marine environments. It is also valuable for land-based applications such as forestry, agriculture, and mapping changes in land surfaces. The radar imaging capability of Sentinel 1 is crucial for emergency response in scenarios like floods and earthquakes, where real-time, weather-independent data is essential for rapid assessment and decision-making (Islam & Meng, 2022).

Additionally, we incorporated widely used spectral indices in our methodology due to their ability to enhance the model's capacity to distinguish between different land cover types, which is crucial for accurately capturing variations in urban landscapes. These indices, derived from Landsat 8 and Sentinel 2, include the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974), Normalized Difference Water Index (NDWI) (McFeeters, 1996), Normalized Difference built-up Index (NDBI) (Zha et al., 2003), Bare Soil Index (BSI), Enhanced Vegetation Index (EVI) (H. Q. Liu & Huete, 1995), and Automated Water Extraction Index (AWEI) (Feyisa et al., 2014). These indices have been proven effective in various applications, such as water body identification, urban growth tracking, and vegetation monitoring ("Assessing LULC changes and LST through NDVI and NDBI spatial indicators: A case of Bengaluru, India", n.d.; Sima et al., 2023; Kebede et al., 2022), helping identify key features like green spaces, buildings, and water bodies. By integrating these indices, our approach aims to reduce bias in these areas and improve the model's precision in detecting landscape features.

2.1.2 Open Street Map Layers

Building footprints and road networks are key features in urban environments, and incorporating these elements in our input data is expected to enhance the model's accuracy. To achieve this, we extracted building footprints and road networks from OpenStreetMap (OSM) (OSM, 2024) using the OSM plugin of the open-source Quantum Geographic Information System software (QGIS, 2024). Since OSM layers are in shapefile format, we rasterized the extracted features to match the grid resolution and alignment of other inputs, ensuring consistency across all data layers used in the analysis.

2.1.3 Data Processing

A comprehensive data processing is essential to ensure consistency across all the different input datasets used in the model. We obtained the remote imagery data based on the date of the high-resolution USGS DEM (1-meter resolution) to ensure temporal consistency across all datasets. A common coordinate system, NAD 1983 UTM Zone 18N, was defined for all data, and all datasets were resampled to a 5-meter spatial resolution to maintain consistency in spatial analysis.

For Landsat-8, processing steps included radiometric calibration to convert digital numbers to reflectance, atmospheric correction to address atmospheric effects, cloud masking to remove clouds and shadows, and geometric correction to ensure spatial accuracy. We excluded certain bands that were not relevant for our analysis, such as Cirrus band (1.36 - 1.38 µm) and Thermal Infrared Sensor bands (10.6 - 11.19 µm and 11.5 - 12.51 µm), which are primarily used for atmospheric correction and thermal measurements. Similar processing was performed for Sentinel-2, including atmospheric correction using Sen2Cor and the exclusion of the Water Vapour band (0.945 µm) and SWIR Cirrus band (1.375 µm), which are also used for atmospheric correction and cloud detection. Both datasets were resampled to a 5-meter spatial resolution using the Nearest Neighbor method to preserve the original values without introducing new interpolated data, which is critical for maintaining the integrity of the data when downsampling. Specific indices such as NDWI, NDVI, NDBI, BSI, EVI, and AWEI were then calculated for both datasets. Sentinel-1 data processing involved applying orbit files for orbital corrections, thermal and border noise removal, calibration to convert SAR data to the backscatter coefficient, speckle filtering, range-Doppler terrain correction using a DEM, and conversion to dB. OSM layers, including building footprints and road networks, were rasterized into 5-meter spatial resolution binary raster layers.

GDEMs (ALOS, ASTER, SRTM, NASADEM) were resampled from their original 30-meter resolution to 5 meters using the Nearest Neighbor method. This method was chosen for downsampling to minimize data distortion and preserve the integrity of the original elevation values. A critical step was evaluating different resampling methods (e.g., Nearest Neighbor, bilinear, cubic) for the high-resolution USGS DEM (1-meter resolution). We determined that Nearest Neighbor was the most effective method for upsampling to a 5-meter resolution, ensuring that the resampled DEM accurately represented the original high-resolution data. This resampled 5-meter DEM was then used as the reference DEM for the CNN U-Net model. The final step involved stacking all processed bands and indices into a single raster layer in the following order: GDEMs, Landsat 8 bands, indices from Landsat 8, Sentinel-2 bands, indices from Sentinel-2, Sentinel-1 bands, OSM layers, and the target 5-meter DEM. These meticulously prepared datasets were essential for training the CNN U-Net model to extract the DL-derived DEM.

2.2 CNN U-Net Model

Convolutional Neural Networks (CNNs) are a specific type of neural network created for tasks related to image processing and computer vision. They excel at recognizing spatial patterns within images, irrespective of the patterns' positions, and can process these patterns at various scales (Fukushima, 1988; Krizhevsky et al., 2017). Unlike traditional methods that analyze individual pixels or grid cells, CNNs also consider the features of surrounding neighbors, effectively capturing spatial relationships within the area. This capability has led to their widespread success in tasks such as image classification and segmentation (Ronneberger et al., 2015; Long et al., 2015), and they have been effectively applied in analyzing satellite imagery as well (Ma et al., 2019; Wieland & Martinis, 2019; Meadows & Wilson, 2021; Chang & Chen, 2024; Chudasama et al., 2024; Masafu & Williams, 2024).

Given the effectiveness of CNNs in analyzing spatial patterns and relationships between features in imagery, this study employs the U-Net architecture, a specialized form of CNN that was originally developed for biomedical image segmentation (Ronneberger et al., 2015). U-Net is particularly well-suited for the purpose of our study due to its efficient data fusion capabilities that allows leveraging multiple satellite data sources and GDEMs to enhance the predictive accuracy and detail of the model. As shown in Figure 2, the structure of the U-Net model is symmetrical, representing a variant of the Fully Connected Network. It consists of two sub-networks: an encoder (contracting path) and a decoder (expansive path), with skip connections between these units. The encoder employs two consecutive convolutions followed by a max pooling operation for progressive downsampling, while the decoder uses a symmetrical expansive path for upsampling and reconstructing the output segmented image.

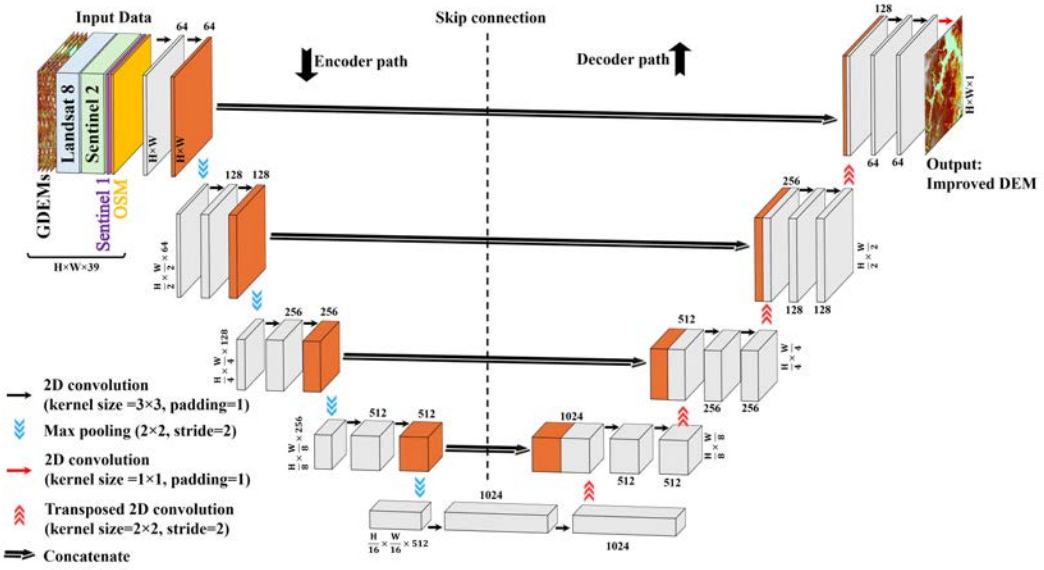


Figure 2. CNN U-Net Architecture for DEM Enhancement

The downsampling part consists of four convolution blocks, each containing two consecutive convolution operations using a 3 × 3 kernel size. These convolutions are crucial for extracting spatial features from the input images. After each convolution, a Rectified Linear Unit (ReLU) activation function is applied. This is followed by a 2 × 2 max pooling operation with a stride of 2 for downsampling, which reduces the spatial dimensions while simultaneously doubling the number of feature maps at each convolution block, thus progressively enhancing the feature depth.

Conversely, the decoder part of the U-Net architecture works to reconstruct the spatial dimensions of the input image, progressively restoring the detailed segmentation. Each of the four upsampling blocks in the decoder uses a 2 × 2 transposed convolution with a stride of 2 to incrementally increase the resolution of the feature maps. Following the upsampling, the feature maps from the corresponding encoder block are reintegrated via skip connections. This crucial step retrieves spatial details lost during the downsampling phase and helps preserve the context of the input image. After the reintegration, each block applies two consecutive 3 × 3 convolutions to refine the features further, followed by ReLU activation. The cycle of upsampling, feature integration, and convolution continues across all blocks, ensuring that the resolution and detail are progressively restored to match the original input dimensions. The decoder culminates in a final convolutional layer that applies a 1 × 1 convolution, producing the final output without additional activation scaling. These output values represent normalized elevation data, scaled according to the min-max normalization applied to the original DEM data. The final output reflects detailed and accurate elevation profiles, reinstating the DEM's original dimensions after undergoing transformations through the encoder and decoder pathways.

2.3 Study Area

To apply the proposed methodology, we selected the City of Portsmouth, in Virginia as our case study. Located in the coastal plain of southeastern Virginia, near the Hampton Roads Harbor where the James and Elizabeth Rivers meet the Chesapeake Bay (see Figure 3a), Portsmouth is highly vulnerable to coastal flooding caused by storm surges from nor'easters and tropical cyclones. The city's low elevation and limited drainage capacity also make it vulnerable to pluvial flooding during heavy rainfall events (U.S. Army Corps of Engineers, Norfolk District, 2015). The majority of the city encompasses developed areas with a broad spectrum of land uses, including various classes of residential development, commercial and industrial areas, as well as governmental installations. These conditions make Portsmouth an ideal case study for the purpose of our research.

The study area was divided into distinct datasets for training, validation, and testing, as illustrated in Figure 3. The training and validation datasets (areas) were used to train and fine-tune the the CNN U-Net model, and the test dataset was exclusively used to evaluate the model. Specifically we chose a watershed within the test area that is highly vulnerable to flooding and has an extensive flooding history. The study area excluding the test areas was randomly divided into training (green) and validation (blue) datasets, following an 80/20 ratio. This random partitioning was implemented to enhance the model's generalizability across varying geographical features. To further improve the generalizability of the model, the training dataset was selected to cover a sufficiently large area, incorporating a diverse range of land cover types, complex structures such as buildings and roads, as well as green spaces and water bodies. To ensure the continuity and smoothness of the elevation profiles across patch boundaries, an overlap of 10% was maintained between patches. The validation dataset was used to fine-tune and evaluate the model's performance during training.

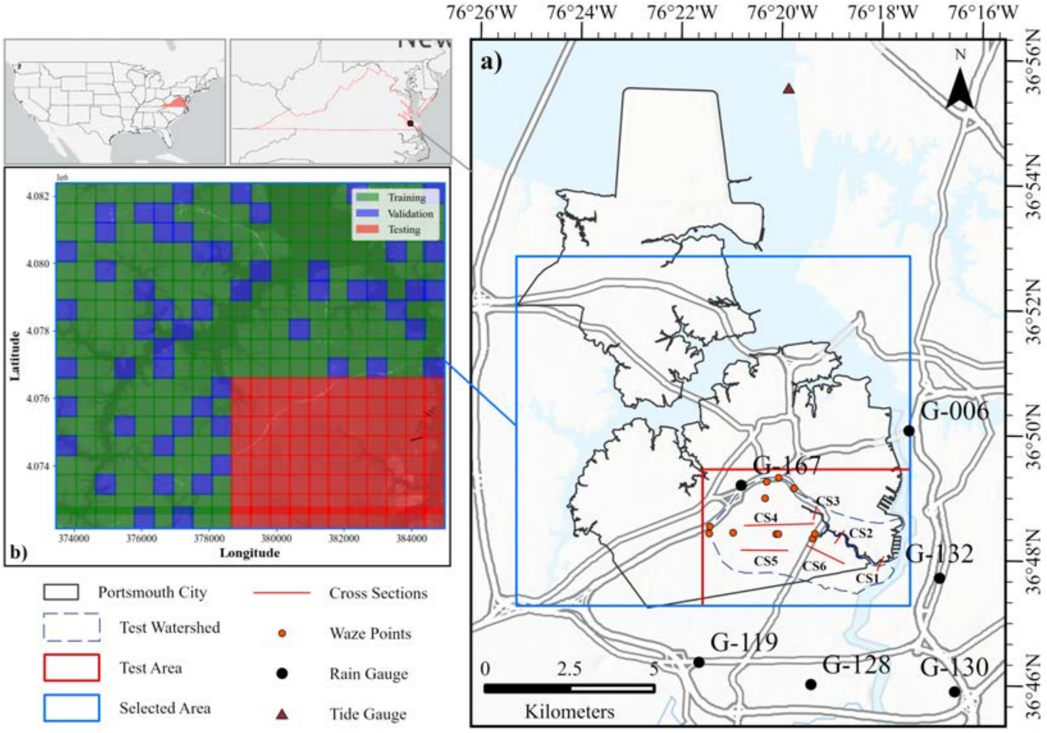


Figure 3. Study area in the City of Portsmouth in Coastal Virginia with different delineations used in the CNN U-Net mode. (a) shows area within the City of Portsmouth selected for our study, including the sub-area selected for testing the model (red box) and the watershed within the test area used for model evaluation (test watershed). CS1 to CS6 denote defined cross-sections for evaluating the DL-derived DEM performance. (b) shows the delineation of the study area into training (green), validation (blue), and testing (red) patches for the CNN U-Net model.

2.4 Hydrodynamic Model

Besides directly comparing our DL-derived DEM to GDEMs, it is important to quantify the extent to which improvements in the DEM translate into enhancements in flood modeling. Therefore, in this study, we use the Two-dimensional Unsteady FLOW (or TUFLOW) model (Syme, 2001) to develop a flood simulation model for a watershed within the test area that is located inside the Paradise Creek basin and covers about 12 square kilometers (Figure 3a). This area specifically selected due to its vulnerability to flooding, making it critical for evaluating our flood simulation strategies.

The model's domain and topography were defined using a high-resolution DEM, with the basin outlet connecting to the Southern Branch of the Elizabeth River. Data crucial for the simulations, including tide levels and rainfall, were sourced from the Sewells Point station and six rainfall gauges managed by the Hampton Roads Sanitation District, respectively. For this study, we specifically focused on an event characterized by high rainfall and normal tide levels, recorded on 11/12/2020. This event was chosen for its potential to cause significant flooding, with all gauges reporting an average precipitation of 132 mm. Details on these data sources, additional model parameters, and visualizations of the precipitation hydrograph and tide levels for this significant event are provided in the supplementary materials (Text S1).

For validating the hydrodynamic model, we utilized Waze-reported points from a significant flood event on 11/12/2020. This approach allowed us to compare and evaluate the model's predictions of flooded streets in Portsmouth with real-time, location-specific incident reports collected via the Waze mobile app. The locations of the Waze points used in this study are highlighted in Figure 3a. Details on the integration and analysis of these data points, as well as on the validation of the flood modeling, are provided in the supplementary materials (Text S1).

2.5 Error Analysis

This study performs an error analysis to evaluate the accuracy of both the DL-derived DEM and GDEMs compared to the reference DEM. The quantitative metrics used include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), and the Correlation Coefficient (R^2). RMSE measures overall error by accounting for both variance and bias, while MAE quantifies the systematic error, indicating how much predictions deviate on average from the ground truth. MPE provides insight into the relative error magnitude as a percentage, and R^2 assesses how well the model fits the data.

Additionally, we applied the flood extent evaluation metrics proposed by Wilks (2011) using a pixel-wise contingency table comparing predicted and observed flood extents. The contingency table includes Hits (Predicted and Observed), False Alarms (Predicted but Not Observed), Misses (Not Predicted but Observed), and Correct non-events (Not Predicted and Not Observed). From this, three key performance metrics were derived: Probability of Detection (POD), False Alarm Ratio (FAR), and Accuracy. POD measures the ratio of correctly predicted flood pixels to observed flood pixels, ranging from 0 to 1, where 1 indicates perfect detection. FAR quantifies the proportion of false flood predictions, with 0 being ideal. Accuracy reflects the percentage of correct flood and non-flood pixel predictions on a pixel-wise scale, with values ranging from zero to one, where one indicates perfect accuracy. Further details and formulations for all metrics are provided in the supplementary materials (Text S2).

3 Results

In this section, we present the results of our study, organized into three main sections. First, we evaluate the vertical accuracy of GDEMs in the test area, using the USGS high-resolution DEMs as our reference DEM (3.1). Next, we present details on the configuration of the CNN U-Net model (3.2). Finally, in section (3.3), we evaluate the performance of the proposed framework in DEM enhancement.

3.1 Evaluating the accuracy of GDEMs

Figure 4 and Table 2 compare the accuracy of the four GDEMs used in this study (i.e, ALOS, ASTER, SRTM, and NASADEM) with the USGS high-resolution DEM (at 5-m resolution) – the reference DEM – over the test watershed. Figure 4a-d show that the GDEMs generally cannot accurately represent the fine-scale features of the area (e.g., road networks and building footprints) due mainly to their low spatial resolution and low vertical accuracy. Because GDEMs are available at a 30-meter resolution, downscaling them to a 5-meter resolution using the Nearest Neighbor method leads to a smoothed topography when compared to a high-resolution DEM, where individual features would be more distinctly represented. Although some variations are visible in the elevation data of the GDEMs, these variations occur at a broader scale and are not as detailed or precise as those found in the high-resolution DEM. In terms of accuracy, GDEMs exhibited a higher range of elevation values compared to the reference DEM, ranging from a minimum of -17 to a maximum of 44 meters. Specifically, Figure 4e-h show significant biases and a wider elevation span in the GDEMs, indicating an overestimation which is further highlighted by error measures in Table 2. Among the four GDEMs, as shown in Table 2, NASADEM shows the closest alignment to the reference DEM with the lowest tendency for overestimation, indicating a more reliable representation of the terrain

($RMSE = 3.41(m)$, $MAE = 2.54(m)$, $MPE = 20\%$). After NASADEM, ALOS performs better compared to the other GDEMs, particularly in representing riverine and coastline features. On the other hand, ASTER and SRTM drastically overestimate the elevation with large $RMSE$, MAE, and R^2 values. These results underline the necessity of developing more accurate and reliable DEMs tailored for urban flooding applications, where minimizing both overestimation and underestimation is crucial for precise modeling.

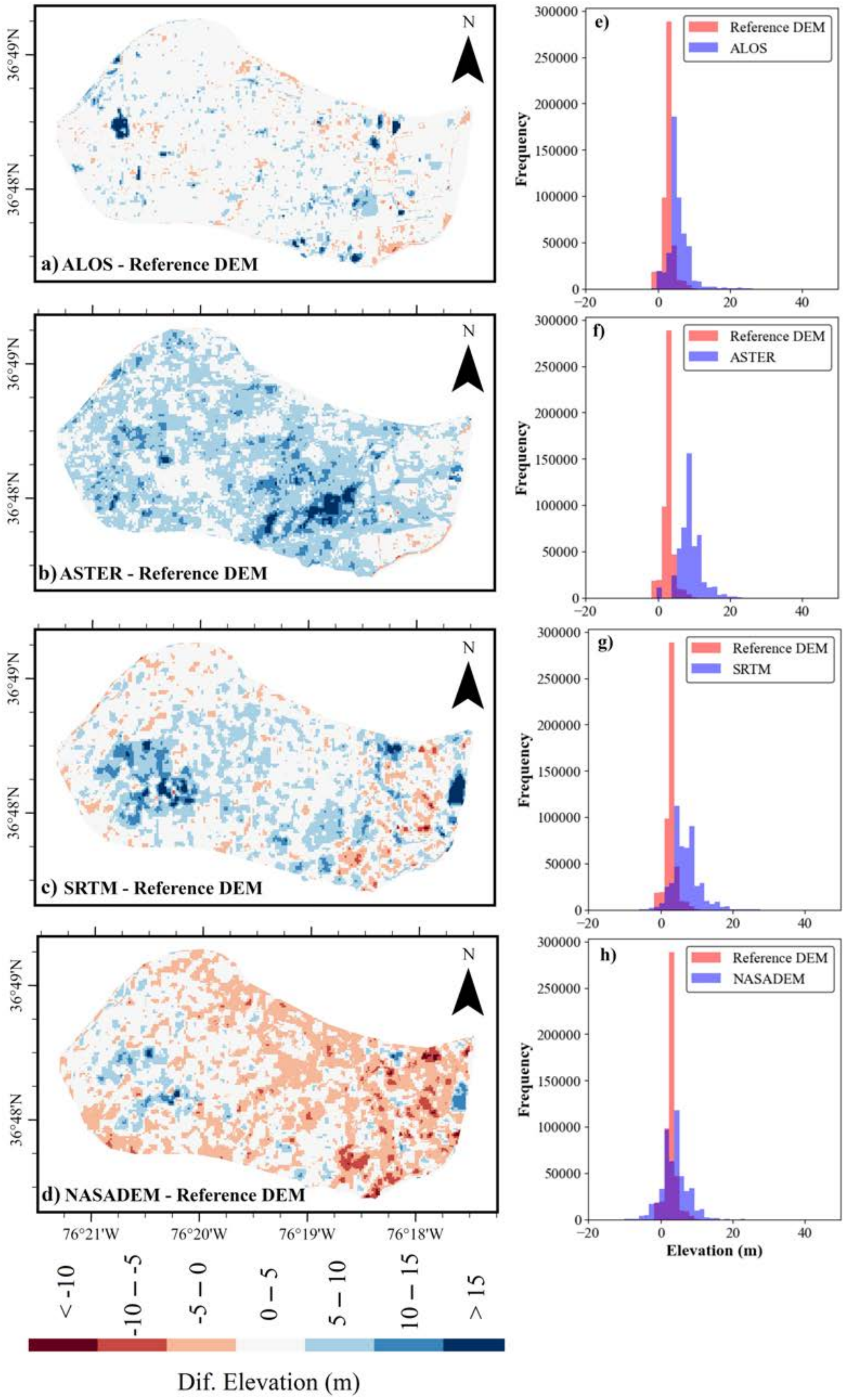


Figure 4. Comparison between the four GDEMs used in this study (ALOS, ASTER, NASADEM, SRTM) and the reference DEM. (a-d) show the elevation differences between the four GDEMs and the reference DEM. (e-i) compare the frequency distribution of elevation values between the GDEMs and the reference DEM.

Table 2. Evaluation metrics (RMSE, MAE, MPE, and R^2) comparing the performance of GDEMs against the reference DEM.

	$RMSE(m)$	$MAE(m)$	$MPE(\%)$	R^2
ALOS	3.79	2.76	171	0.19
ASTER	6.65	5.94	317	0.064
NASADEM	3.41	2.54	20	0.13
SRTM	5.32	4.20	178	0.056

3.2 Configuration of the CNN U-Net Model

The computational setup for model training involved using PyTorch on a system equipped with a 13th Gen Intel(R) Core(TM) i9-13900H processor, 64 GB of RAM, and an NVIDIA GeForce RTX 4070 GPU, leveraging CUDA for efficient processing. The optimal configuration of the CNN U-Net model was determined through an extensive hyperparameter tuning process, specifically optimized for handling the multi-dimensional input data. We chose ReLU as the activation function and RMSE as the loss function, as they showed superior performance in more accurately processing and analyzing elevation data (refer to Appendix A for more details about the CNN U-Net model configuration). The model's training dynamics revealed a sharp initial decrease in both training and validation losses, with stabilization occurring afterward (see Figure A1), indicating effective training and avoidance of overfitting. The close alignment between training and validation losses throughout the training further demonstrates the model's strong generalization capabilities. Following the completion of training and hyperparameter tuning, the best-performing model configuration was applied to the test area.

3.3 Evaluating the Performance of the CNN U-Net model in DEM enhancement

To assess the performance of the CNN U-Net model in DEM enhancement, we first compare the DL-derived DEM with reference DEM using error measures over the test watershed. Then, we provide a higher-resolution comparison between DL-derived DEM, reference DEM, and four GDEMs in six cross sections within the the test watershed (see Figure 3 for the cross sections). Finally, we demonstrate the extent to which the DL-derived DEM improves flood modeling compared to the four GDEMs.

3.3.1 Assessing the Accuracy of the DL-derived DEM

Figure 5(a-b) demonstrates that the DL-derived DEM closely replicates the spatial patterns of the reference DEM. Importantly, the model accurately captures details such as road networks and river paths, both of which are essential features in flood modeling in urban settings. Figure 5(c-f) further demonstrates that, although there are areas in the test watershed where the DL-derived DEM either underestimates or overestimates elevation, its elevation histogram closely aligns with that of the reference DEM. Additionally, the elevation differences between the DL-derived DEM and the reference DEM mostly fall within the range of -1 to +1 meter. The R^2 value of 0.78 shows a strong correlation between the DL-derived DEM and the reference DEM, meaning the model effectively captures elevation changes. The $RMSE$ of 0.86 meters indicates that the average elevation error is relatively small, reflecting the model's high accuracy. With an MAE of just 0.47 meters, the average error is less than half a meter, further confirming the model's precision. The error measures of $R^2 = 0.78$, $RMSE = 0.86(m)$, and $MAE = 0.47(m)$ further demonstrate the accuracy of the DL-derived DEM, particularly compared with GDEMs (refer to Table 2), highlighting the capability of the proposed model in generating DEMs that can capture detailed features with higher accuracy. It is worth noting that the reference DEM shows a higher frequency of elevations above 5 meters, likely due to the model's training limitations, e.g., when the training dataset lacks certain elevation ranges found in the test dataset, referred to as out-of-distribution data in deep learning (Yang et al., 2021).

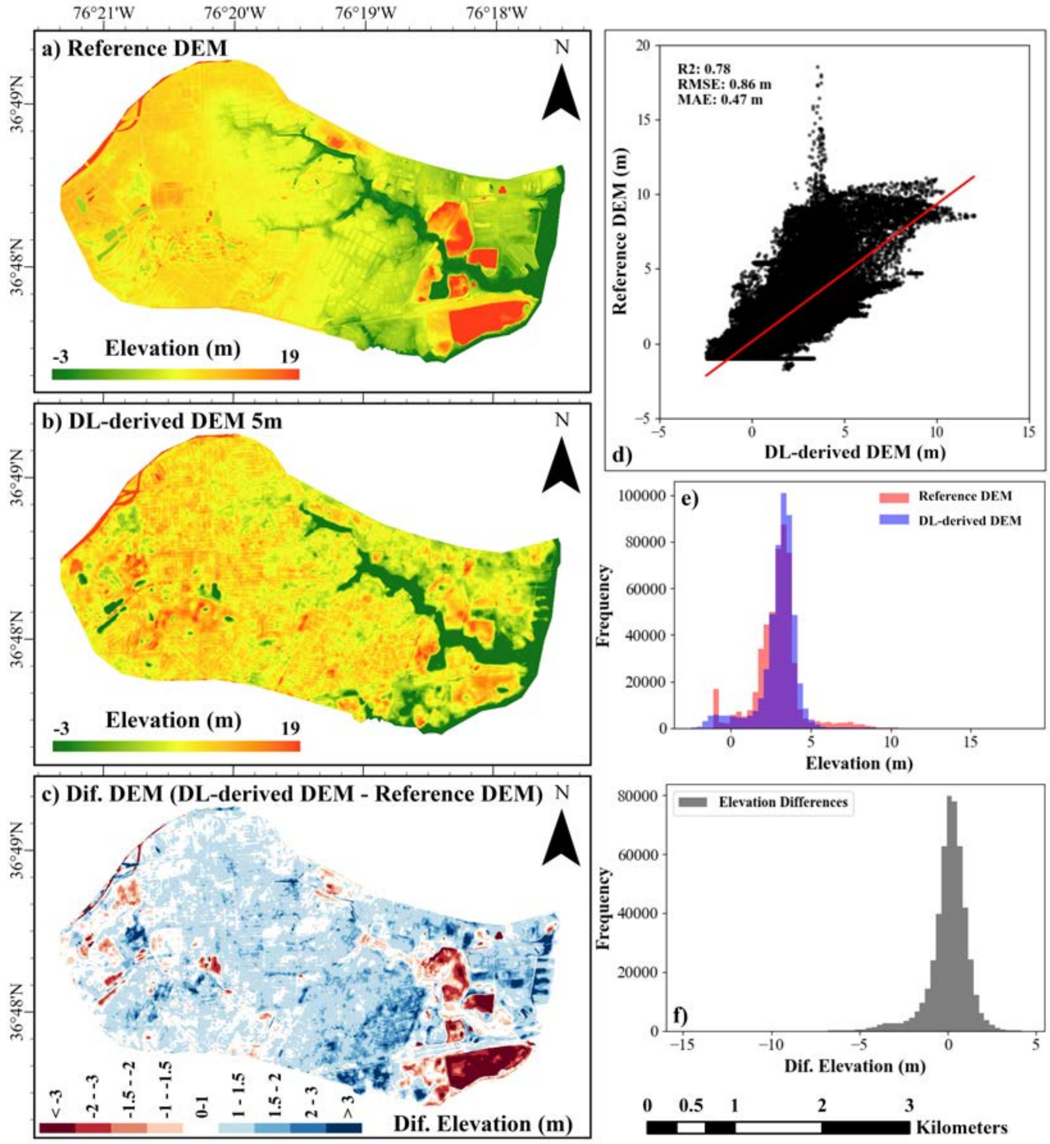


Figure 5. Comparison between DL-derived DEM and reference DEM. (a) shows the reference DEM over the test area; (b) shows the DL-derived DEM over the test area; (c) shows the elevation difference between the DL-derived and reference DEMs; (d-e) compare the DL-derived DEM with the reference DEM using a scatter plot and a histogram; and (f) shows the histogram of the error (i.e., the difference between DL-derived and reference DEMs).

3.3.2 Comparison of Elevation Profiles among Various DEMs

For a more detailed evaluation of the DL-derived DEM, we compared the DL-derived DEM with the reference DEM and the four GDEMs across six cross sections over the

the watershed. As shown in Figure 3, these cross sections were purposefully selected in different parts of the test area to represent various key features: three cross sections along the river path (i.e., CS1 at the beginning, CS2 in the middle, and CS3 at the end of the river), two cross sections in urban, developed areas (i.e., CS4 and CS6), and one cross section in green space areas (i.e., CS5). The elevation profiles over these cross-sections presented in Figure 6 reveal that the GDEMs fail to accurately represent the profiles across the cross-sections, whereas the DL-derived DEM exhibits significant improvements in accuracy (refer to Table S5 for $RMSE$, MAE, and R^2 values). For the river path profiles, Figure 6(a-c) shows that the DL-derived DEM closely aligns with the reference DEM (with average $RMSE$ of 0.78 (m)) while the GDMEs either overestimate or underestimate the elevation with average $RMSE$ of 3.24 (m) for ALOS, 7.14(m) for ASTER, 2.32 (m) for nasadem, and 4.16(m) for SRTM. The elevation profiles in urban areas and green space area (Figure 6(d-f)) show similar patterns. In CS4, as an example, where the GDEMs significantly overestimate the elevation, the DL-derived DEM improves both $RMSE$ and MAE by approximately 90%. These results further highlight the superior performance of the DL-derived DEM in accurately capturing elevation profiles across diverse terrain features compared with the GDEMs.

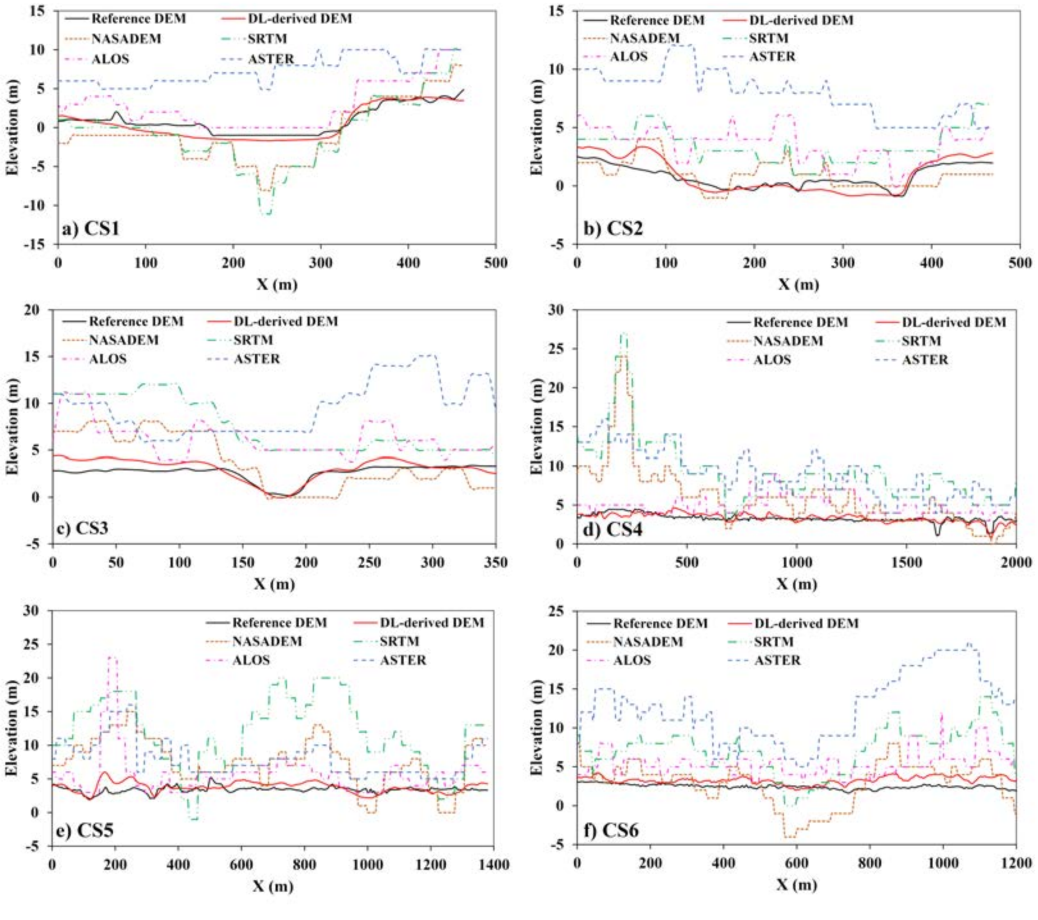


Figure 6. Elevation profiles in the six cross sections identified in the test watershed based on the four GDEMs, the DL-derived DEM, and the reference DEM.

3.3.3 Comparative Analysis of Flood Inundation Extents Using Various DEMs

The flood inundation extents generated from GDEMs (Figure 7a-d) display unrealistic patterns, notably lacking coherent flow patterns over road networks and incorrectly spreading from lower to higher elevation areas. This not only results in a spatial misrepresentation but also leads to an overestimation of flood extents, compromising the accuracy of flood mapping and risk assessments. In contrast, the reference DEM delineates a precise flood pattern that originates from coastal and riverine areas, subsequently spreading into floodplains (Figure 7e). This pattern accurately captures the dynamics of rainfall-driven flooding, highlighting inundation across street networks and low-lying areas first.

On the other hand, the DL-derived DEM (Figure 7f) significantly improves flood inundation mapping compared with GDEMs, closely matching the reference DEM in terms of coastal and riverine flood patterns and correctly modeling flood dispersion along roads, though it slightly overestimates flood extents in some areas. The flood extent derived from the DL-derived DEM also captures flooding in green spaces and artificial water bodies—features conspicuously absent in the GDEMs. Although the flood extents derived from GDEMs coincidentally align with some Waze points, suggesting potential flood occurrence at these locations, their overall accuracy is diminished by the lack of a logical flood progression and the tendency to overestimate. Conversely, the DL-derived DEM not only corresponds with Waze points but also aligns more closely with the reference DEM's flood patterns, offering greater reliability in predicting flood extents and supporting more accurate risk assessments.

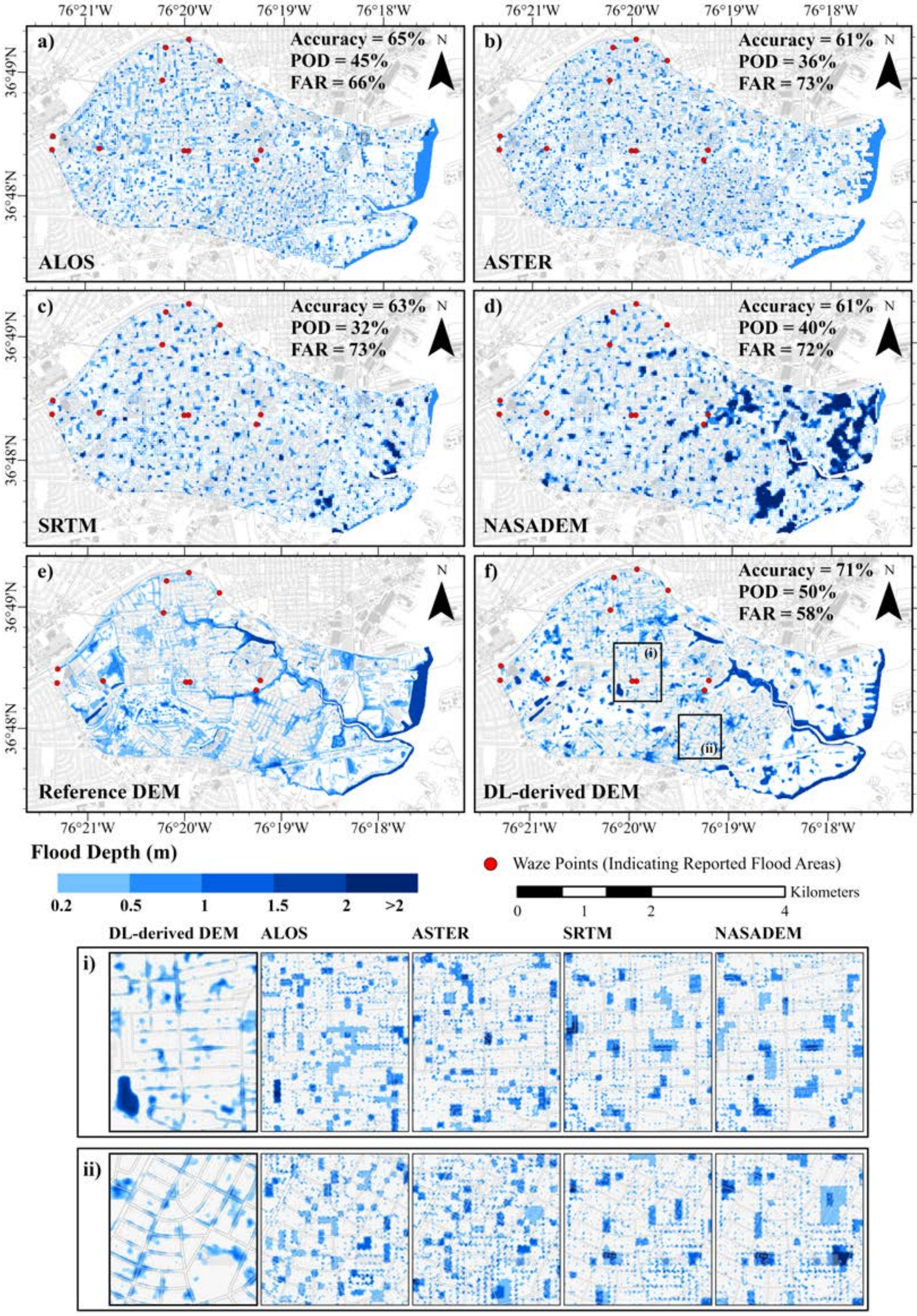


Figure 7. Comparison of flood inundation extent using different DEMs. (a-d) correspond to ALOS, ASTER, SRTM, and NASADEM, respectively. (e) corresponds to the reference DEM. (f) corresponds to the DL-derived DEM. (i) and (ii) show detailed flood mapping comparisons in road networks in the areas identified in (f).

Referring to the performance metrics displayed in Figure 7, our analysis shows that the DL-derived DEM, significantly enhances flood modeling compared with the GDEMs, highlighted by higher accuracy, higher Probability of Detection, and lower False Alaram Rate. This improvement can be attributed to several key factors inherent in our methodology. First, the CNN U-Net model's enhanced spatial pattern recognition capability allows the DL-derived DEM to closely follow the distribution of critical features in the reference DEM, such as road networks and water channels, which are essential for predicting flood flow paths in urban areas. Second, our model's ability to refine riverine and coastline features provides a more precise boundary condition for flood modeling, leading to better prediction of flood dynamics originating from these key areas. Additionally, the CNN U-Net model effectively reduces elevation errors, resulting in a more accurate representation of the landscape, which further minimizes false flood predictions.

4 Discussion

Previous studies highlighted the limitations of using GDEMs in urban flooding, including overestimation or underestimation of flood depth and extent, inaccuracies in spatial flood distribution, and flawed flood risk assessments (Zandsalimi et al. (2024); Biswal et al. (2023); Nandam and Patel (2024)). In this research, we develop a Deep Learning-based framework that utilizes CNN U-Net model and integrates various datasets to improve the resolution and the accuracy of GDEMs. By integrating multiple datasets, including various GDEMs and remote imagery products, our model leverages the unique strengths of each dataset. For instance, regarding GDEMs, NASADEM and SRTM are derived using the InSAR method, which can penetrate cloud cover and vegetation, offering a more precise representation of the ground surface compared with ALOS and ASTER (Muench et al., 2022). Additionally, incorporating remote imagery data, such as Landsat 8, which offers 30m resolution for multispectral bands and 15m for panchromatic bands (Roy et al., 2014), and Sentinel 2, which provides higher spatial resolution and more frequent revisit times (Drusch et al., 2012), enhances the model's capability to monitor environmental changes. Sentinel 1's SAR technology further strengthens the model by enabling accurate mapping in all weather conditions and during nighttime (Torres et al., 2012). This integration of multi-source data ensures that our CNN U-Net model is a practical solution for DEM enhancement, especially in data-limited areas.

Our results showed that given the robust capabilities of CNNs in detecting spatial patterns and uncovering relationships within imagery, the DL-derived DEM demonstrates significant improvements in capturing detailed features of urban environments compared with GDEMs (see Figure 5 and Figure 6). Specifically, the model excels in generating detailed road networks, which are crucial for flood modeling in urban settings. Additionally, other important features such as riverine and coastal areas, green spaces, and artificial water bodies are rendered with greater precision. Consequently, our results demonstrated substantial improvements in flood inundation mapping, (Figure 7f), closely mirroring the reference DEM in terms of coastal and riverine flood patterns and correctly depicting flood dispersion along road networks.

The reference DEM we used to train the CNN U-Net model was resampled from the USGS 1-meter DEM, which represents the topographic bare-earth surface and is hydrologically conditioned to flatten water bodies. To create bare-earth DEMs, LiDAR observations undergo a series of processes to filter out returns from vegetation, anthropogenic, and other features, followed by gridding the observations with resampling methods (Aristizabal et al., 2024; Passalacqua et al., 2015). Although the USGS DEM provides a high-quality representation of the bare-earth surface, because we use a wide range of products in our input dataset, the DL-derived DEM generated by the CNN U-Net model includes additional detailed features such as wharves or piers (Figure 8a), building footprints (Figure 8b-c), and roads over waterbodies (Figure 8d-e) that were missing in the USGS high-resolution DEM. While this highlights the capability of the proposed model in providing a detailed and accurate representation of various features, it is crucial to perform post-processing, including hydrologic conditioning (Pearson et al., 2023; Lidberg et al., 2017), to ensure that these generated features do not negatively impact water flow and other hydrological processes.

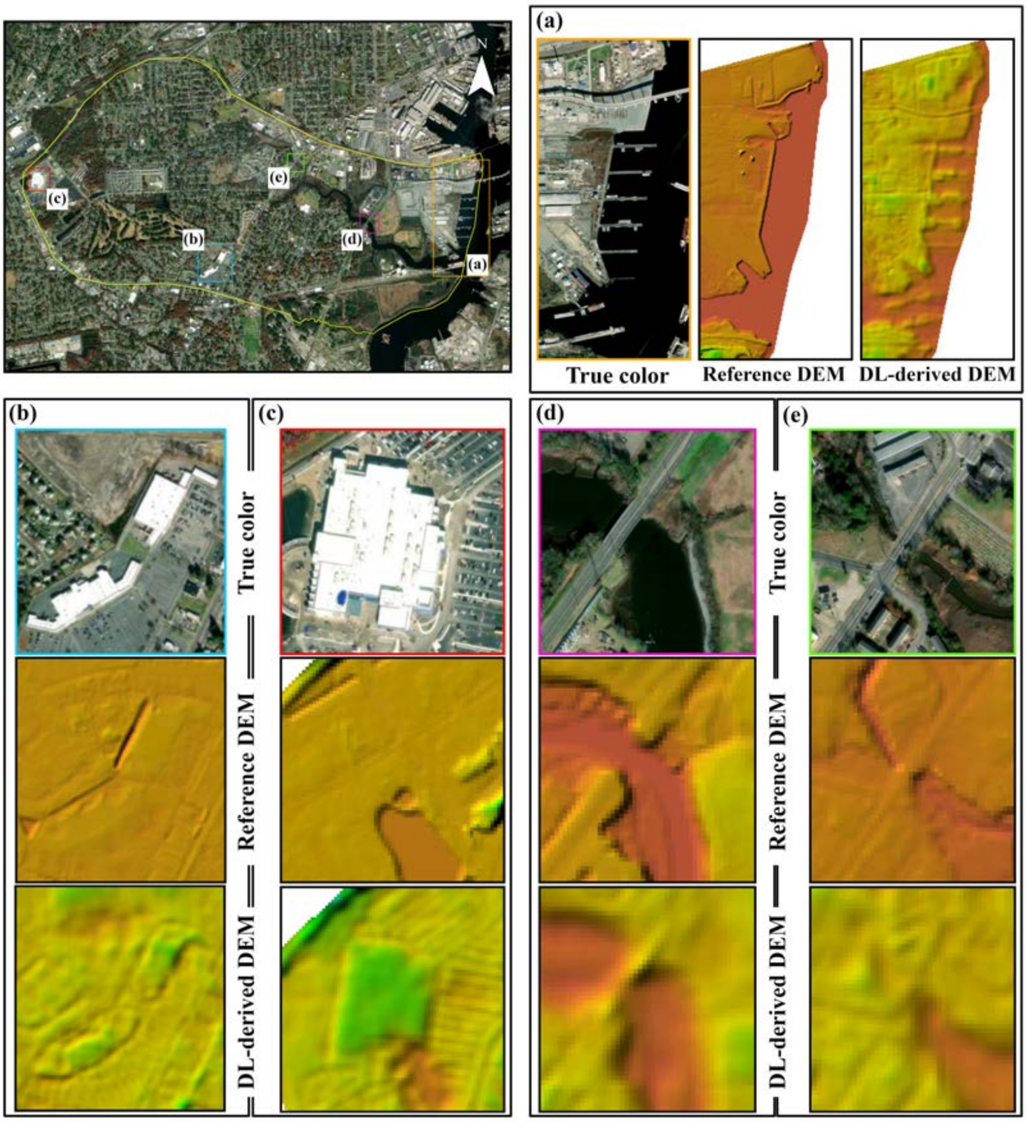


Figure 8. Visualization of enhanced topographic details Using CNN U-Net Model. Each subplot juxtaposes the true-color satellite image with corresponding sections of the reference and DL-derived DEMs, highlighting a terrain feature captured by the DL-derived DEM that is absent in the reference DEM.

One important assumption in our study was that the CNN U-Net model was specifically trained using the USGS high-resolution DEM for the case study area located in coastal Virginia, as presented in Figure 3. While the model was evaluated using a test dataset that was not utilized during training or validation (see Figure 3a), this assumption limits the generalizability of the model. To assess its generalizability, we applied the model to two new sites. The first site (Site 1) also lies in coastal Virginia but located in Norfolk, VA, entirely outside the boundary of the study area. The second site (Site 2) is in Wilmington, NC, another coastal location. These two sites were selected because they share similar characteristics with Portsmouth, VA, in terms of terrain, topography, development, and flood risk. Table 3 presents the performance of the model in terms of RMSE and MAE (refer to Figures S2 and S3 for detailed visual comparisons are provided). The results show that the DL-derived DEM performs better than the GDEMs at both sites. However, the improvement is smaller than in the case study area, as the model was not trained for these two new sites. Additionally, the results for Norfolk show a greater improvement compared to Wilmington, likely because Norfolk shares more similarities with the case study area.

Table 3. Comparison of the performance of the DL-derived DEM and GDEMs using evaluation metrics with respect to the reference DEM in two unseen study sites located within Norfolk, VA (Site 1) and Wilmington, NC (Site 2).

	Site 1 - Norfolk, VA		Site 2 - Wilmington, NC	
	$RMSE$ (m)	MAE (m)	$RMSE$ (m)	MAE (m)
DL-derived DEM	0.99	0.90	1.81	1.46
ALOS	4.16	3.04	6.23	4.80
ASTER	8.50	7.82	5.39	4.05
NASADEM	3.25	2.66	3.49	2.78
SRTM	4.47	3.92	2.78	2.12

5 Conclusion

Flood modeling is a crucial aspect of flood risk assessment, heavily dependent on on DEMs. However, accurate, high-resolution DEMs are not widely available. As a result, researchers and practitioners often rely on GDEMs, which are typically available at a 30-meter resolution and suffer from vertical biases. In this study, we demonstrated the significant potential of a deep learning-based approach, specifically the CNN U-Net model, combined with data fusion, in enhancing DEM resolution and accuracy by integrating multi-source data. The DL-derived DEM not only provides a high-resolution representation of urban features but also enhances flood simulation accuracy, making it a valuable tool for urban flood management. As such, our study effectively addresses the common limitations associated with GDEMs such as lower spatial resolution, the inability to capture detailed urban features, and their inherent errors (Fisher & Tate, 2006; Wechsler, 2007). The DL-derived DEM achieved an $RMSE$ of 0.86 (m) and an MAE of 0.47 (m), representing an 74.7% and 81.4% improvement over the $RMSE$ (3.41 m) and MAE (2.54 m) values of the best-performing GDEM (NASADEM), respectively. In terms of flood modeling, the overall accuracy of the DL-derived DEM is 71%, indicating a significant improvement over the GDEMs. Specifically, the POD shows a 12% increase, reflecting a more reliable identification of flood-prone areas. Additionally, the FAR decreased by 13%, suggesting a reduction in the misclassification of non-flooded areas as flooded.

Although the proposed CNN U-Net model demonstrates significant improvements in DEM resolution and accuracy, its generalizability is likely limited to areas with characteristics similar to the study area, particularly those with similar topography, elevation ranges, and land cover types. Addressing this limitation requires expanding the training datasets to include a broad range of topographic conditions, elevation ranges, and land cover types and leveraging transfer learning techniques to enhance the model's adaptability. Additionally, future research can focus on refining the deep learning models to further reduce prediction errors and enhance their applicability. One promising direction is the use of hydrologic conditioning processes, similar to those employed by USGS, to refine stream networks and improve hydrologic accuracy. Finally, exploring hybrid machine learning algorithms to post-process the improved DEM can also offer further refinement and is likely to reduce overestimation issues.

Appendix A: CNN U-Net Model Configuration and Performance

In this Appendix, we provide an overview of the configuration and architectural details of the CNN U-Net model discussed in Section 3.2. Table A1 outlines the key hyperparameters and their best values that have proven most effective during model training. Figure A1 illustrates the training and validation loss curves, showcasing the model's performance over 150 epochs.

During the configuration process, we evaluated several ranges for specific parameters to identify the best-performing settings. For the activation function, we tested ReLU, LeakyReLU, and ELU, ultimately selecting ReLU for its stable and accurate results in processing elevation data. The learning rate, a hyperparameter, was fine-tuned in the range of 0.0001 to 0.001. Our analysis determined that a learning rate of 0.00015 provided an optimal balance between convergence speed and model performance. Additionally, we adjusted the number of encoding blocks and base filters, with four encoding blocks and a base filter size of 64 yielding the highest accuracy for DEM enhancement. These optimized settings were determined using metrics such as RMSE and MAE to guide the selection of the most effective values. Upon implementation on the test dataset (the entire area enclosed within the red box in Figure 3-b), the model demonstrated high precision, achieving an average $RMSE$ of 0.032 and an average MAE of 0.0199. These metrics, derived from normalized data ranging from 0 to 1, ensure that the error assessments are consistent and comparable across different datasets and conditions.

Table A1. Key configuration settings of the CNN U-Net model for DEM enhancement. This table outlines the primary settings used, including activation functions, learning parameters, and architectural choices. The values in bold represent the best-performing settings based on hyperparameter tuning.

Hyperparameter	Best Value
Input Channels	39
Activation Function	ReLU
Learning Rate	0.00015
Number of Epochs	150
Loss Function	RMSE
Optimizer	Adam
Patch Size	128
Batch Size	64
Overlap (%)	10
Number of encoding blocks (depth)	4
Base Filter	64
Kernel Size	3
Pool Size	2

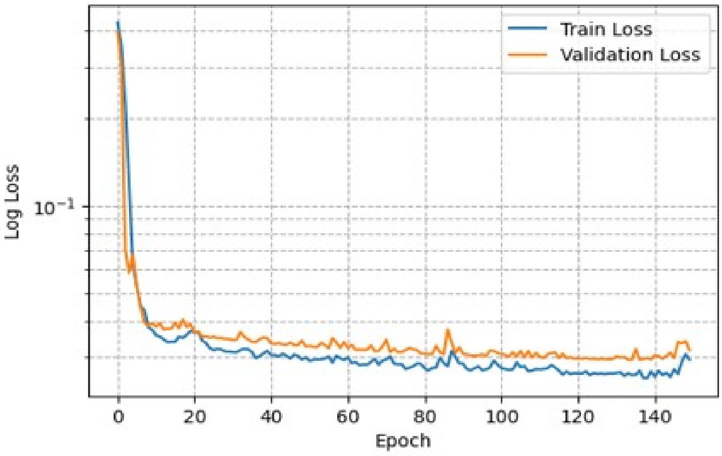


Figure A1. Training and Validation Loss Curves for the CNN U-Net Model: This graph displays the logarithmic scale loss over 150 epochs, with the blue line representing the training loss and the orange line depicting the validation loss.

Appendix B: Supplementary material

The supplementary material document provides supplementary information related to the flood model (Text S1), error analysis (Text S2), comparative analysis of elevation profiles (Text S3), model performance outside of the case study area (Text S4), and the datasets used in this study (Text S5).

Acronyms

DEM	Digital Elevation Model
GDEMs	Global Digital Elevation Models
CNN	Convolutional Neural Network
ALOS DEM	Advanced Land Observing Satellite DEM
SRTM DEM	Shuttle Radar Topography Mission DEM
ASTER DEM	Advanced Spaceborne Thermal Emission and Reflection Radiometer DEM
OSM	OpenStreetMap
DL-derived DEM	Deep Learning-derived DEM
Reference DEM	USGS DEM 1m, resampled to 5m in this study
CS1–CS6	Cross Sections 1 through 6
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MPE	Mean Percentage Error
POD	Probability of Detection
FAR	False Alarm Ratio

CRediT authorship contribution statement

Zanko Zandsalimi: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. **Sergio A. Barbosa**: Investigation, Writing – review & editing. **Negin Alemazkoor**: Conceptualization, Methodology, Supervision, Writing – review & editing. **Jonathan L. Goodall**: Conceptualization, Methodology, Supervision, Writing – review & editing. **Majid Shafiee-Jood**: Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

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.

Acknowledgements

J.L. Goodall acknowledges support from the United States National Science Foundation (award number 2209139). S. Barbosa acknowledges support from the University of Virginia's Environmental Institute.

Data availability

The datasets used in this study, including GDEMs, remotely sensed products, and the USGS DEM, are publicly available. We provide links to all the data used in our study in the supplementary materials. Additionally, the code used in this study is accessible on GitHub at `https://github.com/Zanko-WRM/CNN-U-Net4DEM-Improvement`.

References

Alam, A., Bhat, M. S., & Maheen, M. (2020). Using Landsat satellite data for assessing the land use and land cover change in Kashmir Valley. *GeoJournal*, *85*(6), 1529–1543.

Aristizabal, F., Chegini, T., Petrochenkov, G., Salas, F., & Judge, J. (2024). Effects of high-quality elevation data and explanatory variables on the accuracy of flood inundation mapping via Height Above Nearest Drainage. *Hydrology and Earth System Sciences*, *28*(6), 1287–1315.

assessing lulc changes and lst through ndvi and ndbi spatial indicators: A case of bengaluru, india. (n.d.).

Bettiol, G. M., Ferreira, M. E., Motta, L. P., Cremon, É. H., & Sano, E. E. (2021). Conformity of the NASADEM_HGT and ALOS AW3D30 dem with the altitude from the brazilian geodetic reference stations: A case study from Brazilian Cerrado. *Sensors*, *21*(9), 2935.

Biswal, S., Sahoo, B., Jha, M. K., & Bhuyan, M. K. (2023). A hybrid machine learning-based multi-DEM ensemble model of river cross-section extraction: Implications on streamflow routing. *Journal of Hydrology*, *625*. doi: 10.1016/j.jhydrol.2023.129951

Chang, L., & Chen, Y.-T. (2024). Performance Evaluation and Improvement of Shoreline Detection Using Sentinel-1 SAR and DEM Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Chudasama, B., Ovaskainen, N., Tamminen, J., Nordbäck, N., Engström, J., & Aaltonen, I. (2024). Automated mapping of bedrock-fracture traces from UAV-acquired images using U-Net convolutional neural networks. *Computers & Geosciences*, *182*, 105463.

Courty, L. G., Soriano-Monzalvo, J. C., & Pedrozo-Acuña, A. (2019). Evaluation of open-access global digital elevation models (AW3D30, SRTM, and ASTER) for flood modelling purposes. *Journal of Flood Risk Management*, *12*. doi: 10.1111/jfr3.12550

Crippen, R., Buckley, S., Agram, P., Belz, E., Gurrola, E., Hensley, S., … Tung, W. (2016). Nasadem global elevation model: Methods and progress. In (Vol. 41, p. 125-128). International Society for Photogrammetry and Remote Sensing.

doi: 10.5194/isprsarchives-XLI-B4-125-2016

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., … others (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment*, *120*, 25–36.

ESA. (2024). *Sentinel-2: Resolution and Swath.* `https://sentinel.esa.int/web/sentinel/missions/sentinel-2`. (Accessed on: 17 April 2024)

Fahrland, E., Paschko, H., Jacob, P., & Kahabka, D. (2020). Copernicus DEM Product Handbook. *Airbus Defence and Space, Version*, *2*. Retrieved from `https://spacedata.copernicus.eu/documents/20126/0/GEO1988-CopernicusDEM-SPE-002_ProductHandbook_I1.00.pdf`

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., … Alsdorf, D. E. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, *45*. doi: 10.1029/2005RG000183

Fereshtehpour, M., & Karamouz, M. (2018). DEM Resolution Effects on Coastal Flood Vulnerability Assessment: Deterministic and Probabilistic Approach. *Water Resources Research*, *54*, 4965-4982. doi: 10.1029/2017WR022318

Fernández, A., Najafi, M. R., Durand, M., Mark, B. G., Moritz, M., Jung, H. C., … others (2016). Testing the skill of numerical hydraulic modeling to simulate spatiotemporal flooding patterns in the Logone floodplain, Cameroon. *Journal of Hydrology*, *539*, 265–280.

Feyisa, G. L., Meilby, H., Fensholt, R., & Proud, S. R. (2014). Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote sensing of environment*, *140*, 23–35.

Fisher, P. F., & Tate, N. J. (2006). Causes and consequences of error in digital elevation models. *Progress in physical Geography*, *30*(4), 467–489.

Ford, A., Barr, S., Dawson, R., Virgo, J., Batty, M., & Hall, J. (2019). A multi-scale urban integrated assessment framework for climate change studies: A flooding application. *Computers, Environment and Urban Systems*, *75*, 229-243. doi: 10.1016/j.compenvurbsys.2019.02.005

Fukushima, K. (1988). Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural networks*, *1*(2), 119–130.

Hawker, L., Rougier, J., Neal, J., Bates, P., Archer, L., & Yamazaki, D. (2018). Implications of Simulating Global Digital Elevation Models for Flood Inundation Studies. *Water Resources Research*, *54*, 7910-7928. doi: 10.1029/2018WR023279

Hinkel, J., Feyen, L., Hemer, M., Cozannet, G. L., Lincke, D., Marcos, M., … Wolff, C. (2021). Uncertainty and Bias in Global to Regional Scale Assessments of Current and Future Coastal Flood Risk. *Earth's Future*, *9*. doi: 10.1029/2020EF001882

Hirt, C., Filmer, M. S., & Featherstone, W. E. (2010). Comparison and validation of the recent freely available ASTER-GDEM ver1, SRTM ver4.1 and GEODATA DEM-9s ver3 digital elevation models over Australia. *Australian Journal of Earth Sciences*, *57*, 337-347. doi: 10.1080/08120091003677553

Islam, M. T., & Meng, Q. (2022). An exploratory study of Sentinel-1 SAR for rapid urban flood mapping on Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation*, *113*, 103002.

Jarihani, A. A., Callow, J. N., McVicar, T. R., Niel, T. G. V., & Larsen, J. R. (2015). Satellite-derived Digital Elevation Model (DEM) selection, preparation and correction for hydrodynamic modelling in large, low-gradient and data-sparse catchments. *Journal of Hydrology*, *524*, 489-506. doi: 10.1016/j.jhydrol.2015.02.049

Kasi, V., Yeditha, P. K., Rathinasamy, M., Pinninti, R., Landa, S. R., Sangamreddi, C., … Dandu Radha, P. R. (2020). A novel method to improve vertical accuracy of cartosat dem using machine learning models. *Earth Science Informatics*, *13*, 1139–1150.

Kebede, T. A., Hailu, B. T., & Suryabhagavan, K. V. (2022). Evaluation of spectral built-up indices for impervious surface extraction using Sentinel-2A MSI imageries: A case of Addis Ababa city, Ethiopia. *Environmental Challenges*, *8*, 100568.

Kim, D. E., Liong, S. Y., Gourbesville, P., Andres, L., & Liu, J. (2020). Simple-Yet-Effective SRTM DEM improvement scheme for dense urban cities using ANN and remote sensing data: Application to flood modeling. *Water (Switzerland)*, *12*, 1-14. doi: 10.3390/w12030816

Kim, D. E., Liu, J., Liong, S. Y., Gourbesville, P., & Strunz, G. (2021). Satellite dem improvement using multispectral imagery and an artificial neural network. *Water (Switzerland)*, *13*. doi: 10.3390/w13111551

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90.

Kulp, S. A., & Strauss, B. H. (2018). CoastalDEM: A global coastal digital elevation model improved from SRTM using a neural network. *Remote Sensing of Environment*, *206*, 231-239. doi: 10.1016/j.rse.2017.12.026

Li, F., Yan, X. F., & Duan, H. F. (2019). Sustainable Design of Urban Stormwater Drainage Systems by Implementing Detention Tank and LID Measures for Flooding Risk Control and Water Quality Management. *Water Resources Management*, *33*, 3271-3288. doi: 10.1007/s11269-019-02300-0

Li, Y., Fu, H., Zhu, J., Wu, K., Yang, P., Wang, L., & Gao, S. (2022). A Method for SRTM DEM Elevation Error Correction in Forested Areas Using ICESat-2 Data and Vegetation Classification Data. *Remote Sensing*, *14*. doi: 10.3390/rs14143380

Lidberg, W., Nilsson, M., Lundmark, T., & Ågren, A. M. (2017). Evaluating preprocessing methods of digital elevation models for hydrological modelling. *Hydrological Processes*, *31*(26), 4660–4668.

Liu, H. Q., & Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE transactions on geoscience and remote sensing*, *33*(2), 457–465.

Liu, Y., Bates, P. D., Neal, J. C., & Yamazaki, D. (2021). Bare-Earth DEM Generation in Urban Areas for Flood Inundation Simulation Using Global Digital Elevation Models. *Water Resources Research*, *57*. doi: 10.1029/2020WR028516

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 3431–3440).

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, *152*, 166–177.

Magruder, L., Neuenschwander, A., & Klotz, B. (2021). Digital terrain model elevation corrections using space-based imagery and ICESat-2 laser altimetry. *Remote Sensing of Environment*, *264*. doi: 10.1016/j.rse.2021.112621

Masafu, C., & Williams, R. (2024). Satellite video remote sensing for flood model validation. *Water Resources Research*, *60*(1), e2023WR034545.

McClean, F., Dawson, R., & Kilsby, C. (2020). Implications of Using Global Digital Elevation Models for Flood Risk Analysis in Cities. *Water Resources Research*, *56*. doi: 10.1029/2020WR028241

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*, *17*(7), 1425–1432.

Meadows, M., & Wilson, M. (2021). A comparison of machine learning approaches to improve free topography data for flood modelling. *Remote Sensing*, *13*, 1-28. doi: 10.3390/rs13020275

Muench, R., Cherrington, E., Griffin, R., & Mamane, B. (2022). Assessment of open access global elevation model errors impact on flood extents in southern niger. *Frontiers in Environmental Science*, *10*, 880840.

Nandam, V., & Patel, P. (2024). A framework to assess suitability of global digital elevation models for hydrodynamic modelling in data scarce regions. *Journal of Hydrology*, *630*, 130654. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S0022169424000489` doi: 10.1016/j.jhydrol.2024.130654

OSM. (2024). *Openstreetmap.* Online Map. Retrieved from `https://www.openstreetmap.org/#map=7/41.377/2.615` (Accessed: 2024-04-19)

Ouyang, Z., Zhou, C., Xie, J., Zhu, J., Zhang, G., & Ao, M. (2023). SRTM DEM Correction Using Ensemble Machine Learning Algorithm. *Remote Sensing*, *15*. doi: 10.3390/rs15163946

O'Donnell, E. C., & Thorne, C. R. (2020). Drivers of future urban flood risk. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *378*. doi: 10.1098/rsta.2019.0216

Pasquier, U., He, Y., Hooton, S., Goulden, M., & Hiscock, K. M. (2019). An integrated 1d–2d hydraulic modelling approach to assess the sensitivity of a coastal region to compound flooding hazard under climate change. *Natural Hazards*, *98*, 915-937. doi: 10.1007/s11069-018-3462-1

Passalacqua, P., Belmont, P., Staley, D. M., Simley, J. D., Arrowsmith, J. R., Bode, C. A., ... others (2015). Analyzing high resolution topography for advancing the understanding of mass and energy transfer through landscapes: A review. *Earth-Science Reviews*, *148*, 174–193.

Pearson, R. A., Smart, G., Wilkins, M., Lane, E., Harang, A., Bosserelle, C., ... Measures, R. (2023). GeoFabrics 1.0. 0: An open-source Python package for automatic hydrological conditioning of digital elevation models for flood modelling. *Environmental Modelling & Software*, *170*, 105842.

Preety, K., Prasad, A. K., Varma, A. K., & El-Askary, H. (2022). Accuracy Assessment, Comparative Performance, and Enhancement of Public Domain Digital Elevation Models (ASTER 30 m, SRTM 30 m, CARTOSAT 30 m, SRTM 90 m, MERIT 90 m, and TanDEM-X 90 m) Using DGPS. *Remote Sensing*, *14*. doi: 10.3390/rs14061334

QGIS. (2024). *Quantum geographic information system (qgis).* Retrieved from `https://qgis.org/en/site/` (Accessed: 2024-04-19)

Rentschler, J., Salhab, M., & Jafino, B. A. (2022). Flood exposure and poverty in 188 countries. *Nature communications*, *13*(1), 3527.

Robinson, N., Regetz, J., & Guralnick, R. P. (2014). EarthEnv-DEM90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *87*, 57-67. doi: 10.1016/j.isprsjprs.2013.11.002

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–miccai 2015: 18th international conference, munich, germany, october 5-9, 2015, proceedings, part iii 18* (pp. 234–241).

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., et al. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ*, *351*(1), 309.

Roy, D. P., Wulder, M. A., Loveland, T. R., Woodcock, C. E., Allen, R. G., Anderson, M. C., ... others (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote sensing of Environment*, *145*, 154–172.

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., & Trigg, M. A. (2016). Perspectives on open access high resolution digital elevation models to produce global flood hazard layers. *Frontiers in Earth Science*, *3*. doi: 10.3389/feart.2015.00085

Schumann, G. J., & Bates, P. D. (2018). The Need for a High-Accuracy, Open-Access Global DEM. *Frontiers in Earth Science*, *6*. doi: 10.3389/feart.2018.00225

Shastry, A., & Durand, M. (2020). Water Surface Elevation Constraints in a Data Assimilation Scheme to Infer Floodplain Topography: A Case Study in the Logone Floodplain. *Geophysical Research Letters*, *47*. doi: 10.1029/2020GL088759

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*(11), 8558–8593.

Sima, S., Zandsalimi, Z., & Darzi, A. (2023). A review and uncertainty analysis of Lake Urmia's hypsometric relationships. *Journal of Great Lakes Research*, 102239.

Su, Y., Guo, Q., Ma, Q., & Li, W. (2015). SRTM DEM correction in vegetated mountain areas through the integration of spaceborne LiDAR, airborne LiDAR, and optical imagery. *Remote Sensing*, *7*, 11202-11225. doi: 10.3390/rs70911202

Syme, W. (2001). TUFLOW-Two & One dimensional unsteady flow Software for rivers, estuaries and coastal waters. In *Ieaust water panel seminar and workshop on 2d flood modelling, sydney.*

Tadono, T., Nagai, H., Ishida, H., Oda, F., Naito, S., Minakawa, K., & Iwamoto, H. (2016). Generation of the 30 M-MESH global digital surface model by alos prism. In (Vol. 41, p. 157-162). International Society for Photogrammetry and Remote Sensing. doi: 10.5194/isprsarchives-XLI-B4-157-2016

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., … others (2012). GMES Sentinel-1 mission. *Remote sensing of environment*, *120*, 9–24.

United Nations, Department of Economic and Social Affairs, Population Division. (2019). *World urbanization prospects: The 2018 revision* (Tech. Rep. No. ST/ESA/SER.A/420). New York: United Nations.

U.S. Army Corps of Engineers, Norfolk District. (2015). *City of Portsmouth, Virginia 2015 Floodplain Management and Repetitive Loss Plan Update* (Flood Plain Management Services Program Report). U.S. Army Corps of Engineers.

USGS. (2016). *Tnm download.* `https://www.sciencebase.gov/catalog/item/63abdd35d34e92aad3ca122a`. (Accessed: 17.04.2024)

Wechsler, S. (2007). Uncertainties associated with digital elevation models for hydrologic applications: a review. *Hydrology and Earth System Sciences*, *11*(4), 1481–1500.

Wendi, D., Liong, S. Y., Sun, Y., & doan, C. D. (2016). An innovative approach to improve SRTM DEM using multispectral imagery and artificial neural network. *Journal of Advances in Modeling Earth Systems*, *8*, 691-702. doi: 10.1002/2015MS000536

Wieland, M., & Martinis, S. (2019). A modular processing chain for automated flood monitoring from multi-spectral satellite data. *Remote Sensing*, *11*(19), 2330.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

Yan, K., Baldassarre, G. D., Solomatine, D. P., & Schumann, G. J. (2015). A review of low-cost space-borne data for flood modelling: topography, flood extent and water level. *Hydrological Processes*, *29*, 3368-3387. doi: 10.1002/hyp.10449

Yang, J., Zhou, K., Li, Y., & Liu, Z. (2021). Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334*.

Yue, L., Shen, H., Zhang, L., Zheng, X., Zhang, F., & Yuan, Q. (2017). High-quality seamless DEM generation blending SRTM-1, ASTER GDEM v2 and ICESat/GLAS observations. *ISPRS Journal of Photogrammetry and Remote Sensing*, *123*, 20-34. doi: 10.1016/j.isprsjprs.2016.11.002

Zandsalimi, Z., Feizabadi, S., Yazdi, J., & Salehi Neyshabouri, S. A. A. (2024). Evaluating the Impact of Digital Elevation Models on Urban Flood Modeling: A Comprehensive Analysis of Flood Inundation, Hazard Mapping, and Damage Estimation. *Water Resources Management*, 1–26.

Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International journal of remote sensing*, *24*(3), 583–594.

Zhang, K., Gann, D., Ross, M., Robertson, Q., Sarmiento, J., Santana, S., … Fritz, C. (2019). Accuracy assessment of ASTER, SRTM, ALOS, and TDX DEMs for Hispaniola and implications for mapping vulnerability to coastal flooding. *Remote Sensing of Environment*, *225*, 290-306. doi: 10.1016/j.rse.2019.02.028

Zhou, C., Zhang, G., Yang, Z., Ao, M., Liu, Z., & Zhu, J. (2020). An Adaptive Terrain-Dependent Method for SRTM DEM Correction over Mountainous Areas. *IEEE Access*, *8*, 130878-130887. doi: 10.1109/ACCESS.2020.3009851