https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

Reducing False Alarms in Urban Flood Detection: An Enhanced NDWI (ENDWI) with Hybrid Max Fusion on Sentinel-2 Data

Abdulrhman M. Almoadi[1].

[1]Institute of Earth and Space Science, King Abdulaziz City for Science and Technology (KACST), Riyadh, Saudi Arabia.

Correspondence to: Abdulrhman M. Almoadi (aalmaadi@kacst.gov.sa)

Abstract: Although optical satellite-derived water indices have significantly advanced urban flood detection, accurately distinguishing flooded from non-flooded pixels while minimizing false positives caused by spectral confusion in built-up areas remains a considerable challenge. This study proposes and evaluates the Enhanced Normalized Difference Water Index (ENDWI) in comparison with seven established water indices to reduce false alarms in complex urban environments. The approach was applied to a flash flood event in Al-Lith Governorate, a coastal urban area along the Red Sea in Saudi Arabia, selected as the case study because of its recurrent vulnerability to intense rainfall and rapid-onset flooding. Sentinel-2 imagery acquired two days after the event served as the core methodology for this study. Validation was performed using WorldView-4 high-resolution imagery obtained within two days of the event, based on 1,262 ground-truth points (559 flooded and 703 non-flooded) generated within polygons to ensure consistency with the Sentinel-2 spatial resolution. Analysis of the raw indices revealed that the Automated Water Extraction Index for shadows (AWEIsh_raw) achieved the highest area under the receiver operating characteristic (ROC) curve (AUC = 0.836), followed by the Normalized Difference Water Index (NDWI_raw) (0.813) and ENDWI_raw (0.784), positioning ENDWI among the top three performers. Following Otsu thresholding, ENDWI_otsu yielded the highest overall accuracy (79.41%) and the lowest false alarm rate (10.95%). A novel hybrid maximum fusion of ENDWI_raw and AWEIsh_raw further enhanced results, attaining an overall accuracy of 82.65%, producer's accuracy of 94.50%, F1-score of 76.73%, and Kappa coefficient of 0.637 after thresholding, with only 21 false positives (false alarm rate = 2.99%). Overall, ENDWI exhibited robust and consistent performance across individual applications, post-thresholding, and hybrid fusion with AWEIsh, establishing it as a reliable and effective tool for accurate urban flood mapping.

Keywords: ENDWI; AWEIsh; NDWI; Spectral indices; Urban flood detection.

1. Introduction

Flooding is defined by the National Oceanic and Atmospheric Administration (NOAA) as the overflow of water onto land that is normally dry(NOAA, 2025). It impacts more people than any other natural hazard and typically occurs due to heavy or prolonged rainfall that overwhelms the soil's absorption capacity as well as the capacities of rivers, streams, and coastal areas. Floods can result from thunderstorms, tropical cyclones, monsoons, snowmelt, or dam failures(NOAA, 2025). The most common types include flash floods, coastal floods, and river floods. Flash floods in urban environments are hazardous, especially at night(Floods | Ready.gov, 2025). Urban flooding is a significant natural hazard triggered by short-term heavy rainstorms or prolonged periods of continuous precipitation that exceed drainage capacity(Wang et al., 2022). It resulted in the loss of 6.8 million human lives globally in the 20th century, and a recent study indicated that floods affected 2.3 billion people between 1995 and 2015(Singha et al., 2020). Between 1980 and 2009, floods resulted in 539,811 deaths (range: 510,941 to 568,680), 361,974 injuries, and affected over 2.8 billion people, marking floods as the deadliest natural disaster(Doocy et al., 2013). The effects of urban flooding extend beyond immediate disaster impacts, disrupting daily life, damaging infrastructure, harming economies, and causing loss of life(Flooding | US EPA, 2025). Economically, between 1970 and 2020, urban floods caused an average of US$25.5 billion in damages, encompassing both insured and uninsured losses(Kundzewicz et al., 2014). With climate change driving more extreme weather and cities continuing to grow rapidly, the

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

frequency and severity of urban flooding are expected to increase, creating even greater risks for communities in the future(Hirabayashi et al., 2013). These situations, involving human and economic losses, are likely to escalate, prompting organizations and governments to develop rapid and effective urban management plans. Such plans should aim to reduce risks in flood-prone areas, address and respond to rapidly emerging hotspots in near real-time, and assess damage.

Remote sensing instruments can determine the extent of flooded areas in both open and complex environments by utilizing their spectral wavelength ranges. Historically, the first Landsat-1 images were used during the 1973 floods on the Mississippi River, USA, demonstrating the potential of satellites for large-scale flood mapping(J-P. Schumann, 2024). Since then, multispectral data have been widely employed for flood observation, damage assessment, and mapping(Albertini et al., 2022). Various methods for water segmentation and flooded area mapping using multispectral satellite images have been documented in the literature. McFeeters (1996) proposed the Normalized Difference Water Index (NDWI), a widely applicable index that uses green and near-infrared (NIR) bands to distinguish between land and water bodies(McFeeters, 1996). In a study by Özelkan (2020), NDWI was applied using Landsat-8 OLI data in the Athisar Dam Lake area of Çanakkale, Turkey, to analyze the efficiency of three NDWI models in detecting water bodies and to compare their accuracies at 15-meter and 30-meter resolutions. The study found that NDWI was the most accurate in distinguishing water bodies, with data at 15-meter resolution yielding better results than those at 30-meter resolution(Özelkan, 2020). Ten years after McFeeters (1996) proposed NDWI, the Modified Normalized Difference Water Index (MNDWI) was developed by Xu (2006) to improve the extraction of flooded areas in complex environments. Albertini et al. (2022) tested MNDWI in various global flood-prone areas (e.g., urban, agricultural, and coastal) using Landsat and Sentinel-2 sensors with spatial resolutions of approximately 10–30 meters for medium- and high-resolution imagery. Their findings highlight MNDWI's superior performance over NDWI for flood mapping, achieving high overall accuracies (OA up to 97%) by better recognizing mixed pixels, turbid water, and algae/vegetation. MNDWI excelled in agricultural (crops), forested, and artificial/urban surface contexts, with median OA values higher than NDWI across categories

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

and reduced errors in shadows or built-up areas(Albertini et al., 2022). Similarly, the Automated Water Extraction Index (AWEI), developed by(Feyisa et al., 2014) offers two variants: Automated Water Extraction Index without shadow consideration (AWEInsh) and with shadow consideration (AWEIsh). Nonetheless, subsequent studies primarily employed threshold techniques to separate water from non-water pixels when using these indices(Jiang et al., 2020; Tan et al., 2023)

Synthetic Aperture Radar (SAR) is widely used for flood mapping because it can operate under all weather conditions and at any time of day. However, urban areas present significant challenges, including complex building-induced scattering (such as double- or triple-bounce effects), geometric distortions, and similar backscatter signatures between water and dry surfaces(Amitrano et al., 2024). Conversely, multispectral optical datasets are particularly effective under cloud-free conditions, provided this criterion is met.

Over the past five years, deep learning approaches have been employed to precisely extract boundaries between flooded and non-flooded areas. For example, the study by(Bersabe and Jun, 2025) utilized spatial data layers from geographical information systems (GIS) datasets representing various flood conditioning factors, such as topography, land use/land cover, soil type, drainage, and hydrological and urban infrastructure data. This data was analyzed on a 30-meter grid using machine learning models, including Logistic Regression, Random Forest, and Support Vector Machines (SVMs), to predict urban pluvial flood susceptibility in Seoul, South Korea. The results emphasized the crucial role of drainage factors in urban flood susceptibility, advancing the understanding of pluvial flood dynamics. These findings support comprehensive flood risk mapping to guide planning, insurance, and evacuation strategies. In another example, (Stateczny et al., 2023) applied a novel deep hybrid model for flood prediction (DHMFP) with a combined Harris Hawks Shuffled Shepherd Optimization (CHHSSO)-based training algorithm, using satellite images with spatial resolutions ranging from 10 to 30 meters in Kerala, India—an urban region affected by drainage issues during the 2018 floods. The results showed sensitivity of 93.48%, specificity of 98.29%, accuracy of 94.98%, false negative rate of 0.02%, and false positive rate of 0.02%. The proposed DHMFP–CHHSSO outperformed baseline models in sensitivity (0.932), specificity (0.977), accuracy

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

(0.952), false negative rate (0.0858), and false positive rate (0.036). Although the promising results of using deep learning in urban flood studies are evident, challenges remain, including high computational requirements, the need for labeled training datasets to address urban complexities, and the time-intensive nature of processing phases.

In summary, multispectral remote sensing data offers a practical and effective solution for applications such as rapid disaster response, damage assessment, and long-term urban planning and management. The proposed enhancement builds upon the widely used NDWI by incorporating a calibration step that divides by the green band, thereby improving the differentiation of water from urban features. This modification is particularly advantageous because most satellite sensors and low-flying unmanned aerial vehicles (UAVs) platforms operating beneath cloud cover routinely acquire red, green, blue, and NIR bands, while short-wave infrared (SWIR) bands—required by several existing indices—are less commonly available and more costly. Therefore, this approach remains accessible and practical for end-users.

To address these limitations, the present study introduces the Enhanced Normalized Difference Water Index (ENDWI), systematically evaluates its performance against seven established water indices using Sentinel-2 imagery from a flash flood event, and validates the results with high-resolution reference data derived from WorldView-4. Additionally, a novel hybrid fusion method is proposed to further reduce false positives. The remainder of this paper is organized as follows: Section 2 describes the study area and data; Section 3 presents the methodology; Section 4 reports the results; Section 5 discusses the findings and their implications; and Section 6 concludes the paper and outlines directions for future research.

2. Study Area and Satellite Data Used

Al-Lith Governorate, situated along the Red Sea coast in western Saudi Arabia (Fig. 1(a, b)), was selected as the case study area to evaluate the proposed ENDWI and hybrid max fusion approach. This region features a typical arid landscape, with steep wadis draining from the eastern highlands toward lowland urban settlements, creating a setting particularly vulnerable to flash flooding during rare but intense rainfall events

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

(Elsebaie et al., 2023). The urban fabric of Al-Lith comprises a mix of residential buildings, paved roads, open spaces, scattered agricultural patches, and bare soil areas, land covers that often complicate optical flood detection due to spectral similarities among water, shadows, and built-up surfaces.

On November 23, 2018, heavy rainfall in the upstream catchment of Wadi Al-Lith triggered the partial breach of an earthen retaining dam(Ministry of Interior - General Directorate of Civil Defense, 2018), releasing a surge of floodwater that reached the downstream urban areas of Al-Lith Governorate within approximately four hours, marking the initial peak of the flash flood event. The flooding continued to escalate over the following hours, submerging roadways, vacant lots, and low-lying areas, with the crest extending into the early morning of November 24 and causing widespread water pooling and soil saturation. On November 25, additional heavy precipitation prolonged the inundation, sustaining water levels above 1.7 meters in several urban sectors. Satellite imagery acquired four to five days later effectively captured these sequential flood impacts, including stagnant water accumulations in urban depressions, sediment-clogged drainage channels, and saturated soils, highlighting the event's progressive effects on infrastructure and the built environment (Fig. 1(b, c)).

To analyze the flood conditions, we employed complementary multispectral imagery from two satellite sources (Table 1). The high-resolution WorldView-4 data, with multispectral bands pan-sharpened to approximately 0.31 meters and acquired on November 27, 2018, provided detailed insights into localized inundation patterns and structural damage. This was supplemented by Sentinel-2 imagery at 10-meter resolution in the visible and NIR bands, captured on November 28, 2018, which offered broader contextual coverage under clear atmospheric conditions. Collected four to five days after the initial dam breach on November 23 (with the main flood peak extending into the early hours of November 24, as detailed in the Local Civil Defense report (Ministry of Interior - General Directorate of Civil Defense, 2018). These datasets also recorded the continued effects of subsequent rainfall and ongoing submersion on November 25 (Table 1), during which elevated water levels remained prominent in Al-Lith's urban areas (Fig. 2) (Elkarim, 2020). By retaining visible signatures of residual surface water and moistened terrain, these images provided an

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

ideal resource for evaluating the performance of the proposed new spectral water index, along with eight spectral water indices used in this study, including the proposed ENDWI and the following seven: (NDWI, MNDWI, AWEIsh, AWEInsh, WI, LSWI, and SWI). Detailed descriptions and equations for all indices are provided in the Methodology section.

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

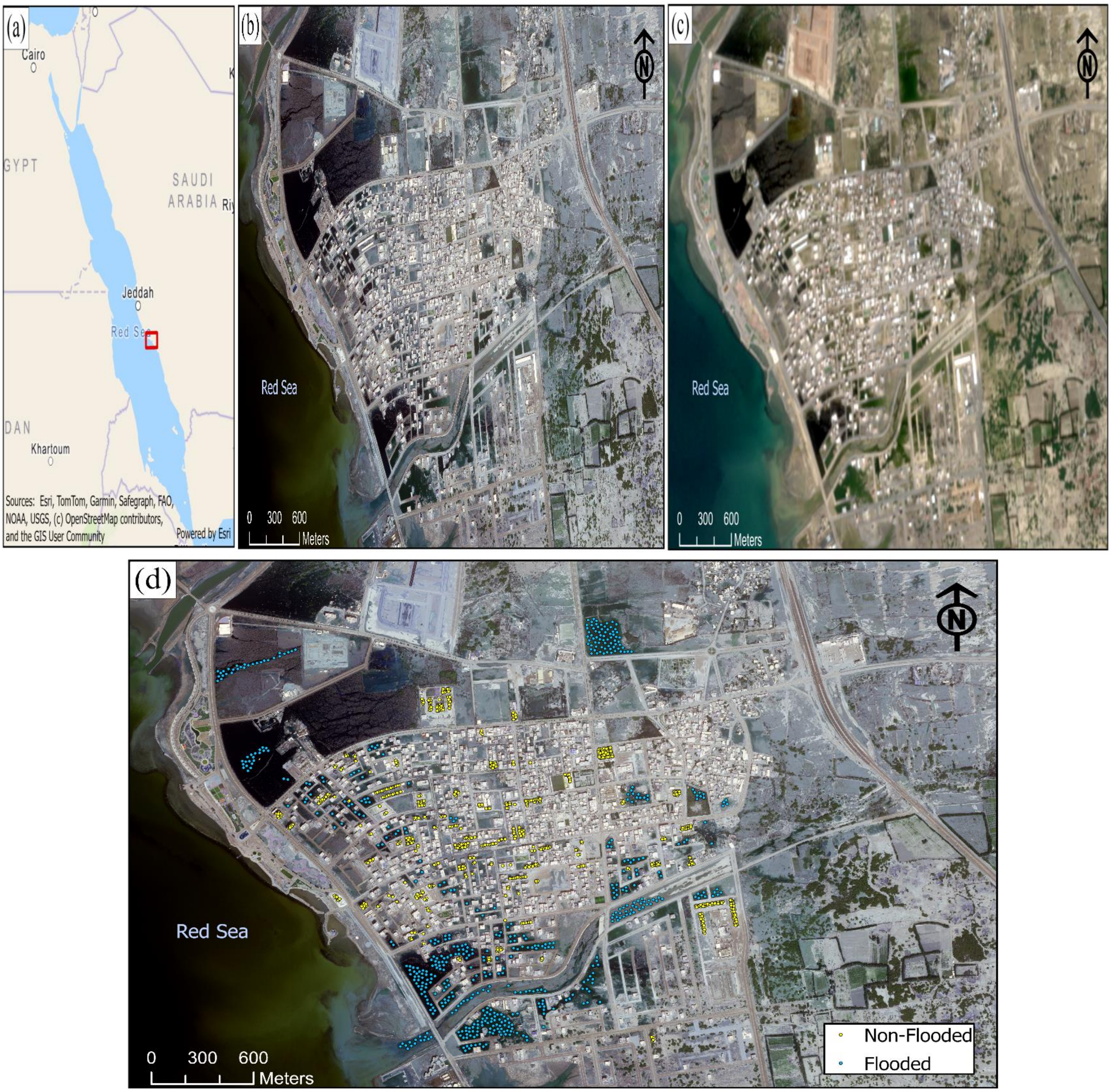


Fig. 1. Overview of the study area along the Red Sea coast, Saudi Arabia. (a) Location map with the study area marked by a red box. (b) High-resolution WorldView-4 image (0.31 meter) of the study area, acquired on November 27, 2018. (c) Sentinel-2 image (10 meter) was acquired on November 28, 2018. (d) Distribution of ground reference points overlaid on the WorldView-4 image (yellow: non-flooded; blue: flooded; total of 1,262 points).

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

Fig. 2. Floods in Al-Lith Governorate on November 25, 2018(Elsebaie et al., 2023).

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

TABLE 1

CHARACTERISTICS OF THE FLOOD EVENT AND REMOTELY SENSED DATASETS EMPLOYED

Dataset	Acquisition Date	Spatial Resolution	Purpose
Flash Flood Event (peak inundation)	November 23-24, 2018	–	Event reference timing: intense rainfall and partial dam breach (Ministry of Interior - General Directorate of Civil Defense, 2018).
Flash Flood effects (continued inundation/effects)	November 25, 2018	–	Continued heavy rainfall and flood flow affected(Elsebaie et al., 2023; Ministry of Interior - General Directorate of Civil Defense, 2018).
WorldView-4	November 27, 2018	0.31 meter	Generation of ground reference points (validation)
Sentinel-2	November 28, 2018	10 meters	Methodology for calculations and evaluations employed in this study

3. Methodology

3.1 Data Pre-processing and Validation Point Generation

The Sentinel-2 imagery used in this study was a Level-2A product, providing atmospherically corrected surface reflectance data processed by the European Space Agency (ESA). The WorldView-4 image was acquired as an ortho-ready product, with radiometric and basic geometric corrections already applied by the provider(King Abdulaziz City for Science and Technology (KACST), 2018).

Both datasets were previously reprojected using ArcGIS Pro to the same projected coordinate system: World Geodetic System (WGS) 1984 / Universal Transverse Mercator (UTM) Zone 37N. To ensure greater accuracy, overlay consistency was evaluated through a careful, manual, swipe-based visual inspection, during which the Sentinel-2 image was swiped over the WorldView-4 image at comparable zoom levels to verify geometric alignment(Samela et al., 2022). This assessment relied on stable, high-contrast features, including major road intersections, building outlines, and the distinctive Red Sea coastline. The method proved fully sufficient, delivering the required spatial correspondence for reliable water index calculation, flood extent extraction, and validation against ground reference points.

The process began with the manual digitization of polygons to delineate clearly identifiable flooded zones (e.g., standing water in streets, low-lying residential areas, muddy ground, drainage channels, and inundated vegetation patches) and non-flooded zones (e.g., dry roads, building rooftops, and elevated ground). Polygons within the same class were then merged using the Dissolve tool to remove fragmentation and produce larger contiguous areas.

A 10-meter inward buffer was applied to the merged polygons to eliminate edge pixels potentially affected by mixed spectral signatures or minor geometric offsets.

Finally, stratified random points were automatically generated within the buffered polygons using ArcGIS Pro. To ensure balanced representation between the two classes and adequate spatial distribution across the study area, a total of 1,262 reference points were produced (559 flooded and 703 non-flooded) (Fig. 1(d)), providing a suitable dataset for the accuracy assessment of the water indices.

3.2 Spectral Indices Calculation

To assess the effectiveness of water indices in detecting urban floods, particularly focusing on quantifying false alarm rates caused by spectral confusion with built-up structures, this study computed eight indices,

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

including the newly proposed ENDWI, using Sentinel-2 Level-2A imagery to generate flood maps. Sentinel-2 was selected for its 10-meter spatial resolution in the visible and NIR bands, as well as its atmospherically corrected surface reflectance data, which enable reliable index application and comparative evaluation in complex urban settings.

Validation was conducted using 1,262 ground reference points—559 representing flooded areas and 703 representing non-flooded areas—derived from high-resolution WorldView-4 imagery (0.31-meter resolution) through a semi-automated process that combined manual polygon delineation with stratified random point generation. This approach allowed precise control over class representation and spatial distribution within the urban landscape, while leveraging the detailed visual interpretability afforded by the very high-resolution imagery. The resulting substantial sample size, concentrated within a compact study area of approximately 3 km×3km, provided high validation density and greater reliability compared to many similar studies that typically employ fewer reference points across larger extents.

Preliminary experiments with various band combinations, conducted through iterative trial and error, showed that raw ENDWI maps offered improved separation of inundated zones and reduced interference from impervious surfaces and shadows. These initial visual findings, observed during the analysis of the November 2018 flash flood event in Al-Lith, directly informed the selection of established indices for systematic comparison and the development of the ENDWI index itself.

The classic NDWI, proposed by McFeeters(McFeeters, 1996), is defined as follows:

$$\mathrm{NDWI} = \frac{\mathrm{Green} - \mathrm{NIR}}{\mathrm{Green} + \mathrm{NIR}}$$

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

where Green corresponds to Sentinel-2 Band 3 (B3) and NIR to Band 8 (B8), this served as the baseline. Motivated by the observed potential of the green band to further suppress urban noise, the proposed ENDWI was formulated as follows:

$$\mathrm{ENDWI} = \frac{\mathrm{NDWI}}{\mathrm{Green}}$$

Initial visual inspection of the raw ENDWI maps revealed a clearer separation of flooded urban areas and substantially less noise from built-up surfaces and shadows relative to standard indices. This qualitative improvement, consistent with observed inundation patterns, justified proceeding with a rigorous quantitative evaluation and the hybrid fusion method detailed in the Results section.

The comparison set also included the following:

- The MNDWI, proposed by Xu(Xu, 2006), is defined as follows:

$$\mathrm{MNDWI} = \frac{\mathrm{Green} - \mathrm{SWIR}}{\mathrm{Green} + \mathrm{SWIR}}$$

where SWIR corresponds to Sentinel-2 Band 11, denoted as B11.

- The AWEInsh, proposed by Feyisa et al. (Feyisa et al., 2014):

$$\mathrm{AWEInsh} = \mathrm{Blue} + 2.5 \times \mathrm{Green} - 1.5 \times (\mathrm{NIR} + \mathrm{SWIR1}) - 0.25 \times \mathrm{SWIR2}$$

- The AWEIsh, proposed by Feyisa et al. (Feyisa et al., 2014):

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

$$\mathrm{AWEIsh} = 4 \times (\mathrm{Green} - \mathrm{SWIR1}) - (0.25 \times \mathrm{NIR} + 2.75 \times \mathrm{SWIR2})$$

- The Water Index (WI)(Änuelas et al., n.d.):

$$\mathrm{WI} = \frac{\mathrm{Green} + \mathrm{Red}}{\mathrm{NIR} + \mathrm{SWIR}}$$

(or an adapted variant for Sentinel-2).

- The Sentinel Water Index (SWI)(Jiang et al., 2020):

$$\mathrm{SWI} = \frac{\mathrm{Green} - \mathrm{SWIR1}}{\mathrm{Green} + \mathrm{SWIR1}}$$

- The Land Surface Water Index (LSWI) (Xiao et al., 2002):

$$\mathrm{LSWI} = \frac{\mathrm{NIR} - \mathrm{SWIR}}{\mathrm{NIR} + \mathrm{SWIR}}$$

All indices were calculated on a per-pixel basis using raster operations in GIS software, yielding raw maps for initial analysis before automated thresholding in the following step.

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

3.3 *Thresholding Using Otsu's Method*

After obtaining the raw index maps, binary water/non-water classifications were generated for each of the eight indices using automated thresholding. Otsu's method was selected for this purpose(Otsu, n.d.), as it is a widely adopted, non-parametric technique in remote sensing applications for extracting water bodies(Jiang et al., 2020; Tan et al., 2023). This algorithm determines the optimal threshold by maximizing inter-class variance, providing an objective and reproducible solution that is especially useful in complex urban environments where manual thresholding may introduce subjectivity.

Otsu's method performs effectively when the index histogram exhibits reasonable bimodality, which was observed for most of the tested indices—particularly those where water pixels cluster at lower values. For each raw index, the Otsu threshold was calculated independently from the full-scene histogram. Pixels were then classified as flooded if their index values fell on the expected water side of the threshold, with the decision direction (greater than or less than) adjusted according to the polarity of each index. This process yielded binary flood maps (Fig. 3) suitable for direct quantitative comparison with the reference points.

The application of Otsu thresholding generally enhanced classification sharpness and improved overall accuracy across the indices. Nevertheless, residual false positives persisted in challenging areas, such as shadowed built-up zones, even for stronger raw performers like AWEIsh. In contrast, ENDWI demonstrated notable resilience to these urban artifacts after thresholding, achieving a lower false alarm rate despite its moderately lower raw AUC value. This complementary performance—where AWEIsh provided superior general separability while ENDWI excelled in suppressing urban-induced errors (Table 2) motivated the development of a simple hybrid maximum fusion strategy, detailed in the next subsection, to leverage the respective strengths of both indices.

3.4 *Hybrid Max Fusion*

Building on observations from raw and thresholded indices—particularly the complementary strengths of AWEIsh_raw (which provides the strongest overall separation) and ENDWI_raw (effective at suppressing

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

urban false positives) (Table 2), we developed a simple hybrid approach to combine their advantages. The goal was to create a fused index that retains high water detection capability while further reducing spectral confusion in built-up and shadowed areas, without introducing complex parameters or requiring additional data. The hybrid fusion was implemented as a straightforward pixel-wise maximum operation between the raw values of the two indices:

$$\text{Hybrid}_{\text{raw}} = \max\left(\text{ENDWI}_{\text{raw}}, \text{AWEIsh}_{\text{raw}}\right)$$

This "max" rule was chosen because both indices are formulated such that higher values generally indicate a greater likelihood of water (or reduced non-water interference in urban contexts). By taking the maximum value at each pixel, the fusion preserves the strongest water signal from either index while mitigating their weaknesses: AWEIsh contributes robust shadow handling and broad separation, whereas ENDWI helps reduce false positives from dark impervious surfaces through its emphasis on the green band.

The resulting hybrid raw map was then subjected to the same Otsu thresholding process described previously, producing a final binary classification. This two-step workflow—fusion followed by automated thresholding—kept the method computationally efficient and fully reproducible, making it practical for rapid flood mapping applications.

Although simple, this hybrid strategy proved effective in preliminary visual checks, showing cleaner urban flood extents with fewer isolated false positives compared to individual indices. A quantitative evaluation of these outputs, including overall accuracy and false alarm rates, is presented in the Results section.

4. Results

4.1 Performance of Individual Indices (Raw Values)

The discriminatory power of the eight raw indices was first assessed using ROC analysis on 1,262 validation points. AUC values provided a threshold-independent measure of separation, complemented by the mean index values for flooded and non-flooded classes, their differences, and t-test statistics.

AWEIsh_raw emerged as the top performer, closely followed by NDWI_raw and the proposed ENDWI_raw (Table 2). All three indices achieved AUC values greater than 0.78, with highly significant class separation ($p < 0.001$), confirming their strong potential for urban flood detection even without thresholding (Fig. 3).

TABLE 2

SEPARATION PERFORMANCE OF THE TOP THREE RAW INDICES

Index	AUC	Mean Flooded Pixels	Mean Non-Flooded Pixels	Difference	t-stat	p-value
AWEIsh_raw	0.836	0.325	0.035	0.290	11.369	0.000
NDWI_raw	0.813	0.298	0.128	0.170	10.769	0.000
ENDWI_raw	0.784	0.527	0.406	0.121	10.347	0.000

The superior AUC of AWEIsh_raw can be attributed to its explicit incorporation of shadow terms, which helps maintain clear separation in complex urban environments. NDWI_raw performed reliably, as expected from a well-established baseline. Although ENDWI_raw ranked third in AUC, its mean values exhibited a distinctive pattern—higher flooded means driven by green-band amplification—suggesting particular resilience against urban spectral confusion. These complementary characteristics motivated a focused analysis of these three indices in the subsequent thresholding and fusion stages.

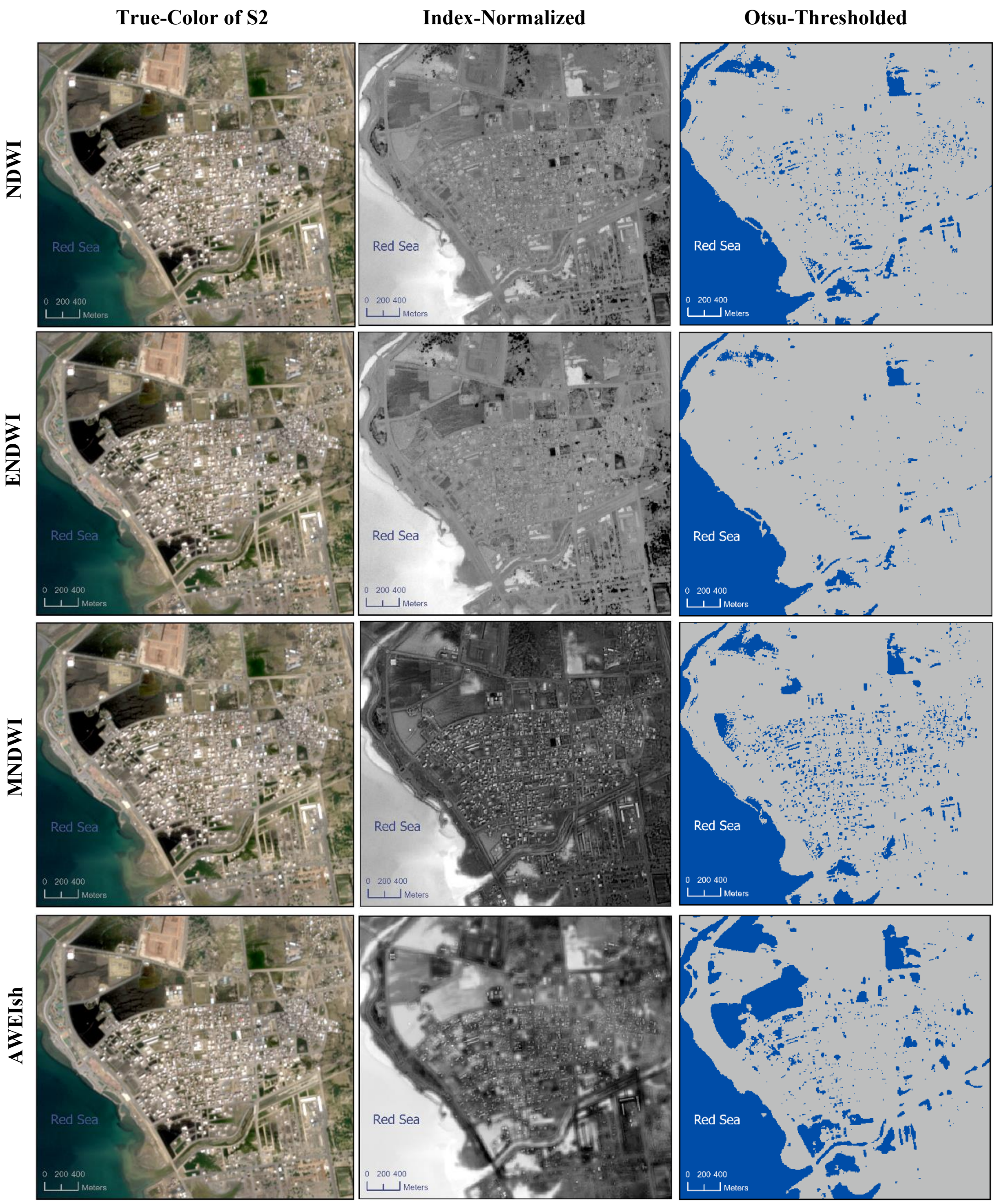
4.2 Performance After Otsu's Thresholding

Applying Otsu's thresholding to the raw indices produced binary classifications, which were evaluated using standard accuracy metrics—including overall accuracy, precision, recall, F1-score, Kappa coefficient, and false alarm rate—based on the same validation points (Table 3).

Thresholding generally improved practical usability, with ENDWI_otsu standing out for its balance of high precision and low false alarm rates. Specifically, ENDWI_otsu achieved the highest precision (79.41%) and

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

the lowest false alarm rate (10.95%) among the individual indices, while maintaining competitive overall performance. AWEIsh_otsu retained strong recall but exhibited slightly more false positives in shadowed areas, and NDWI_otsu performed solidly in between (Table 3).



https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

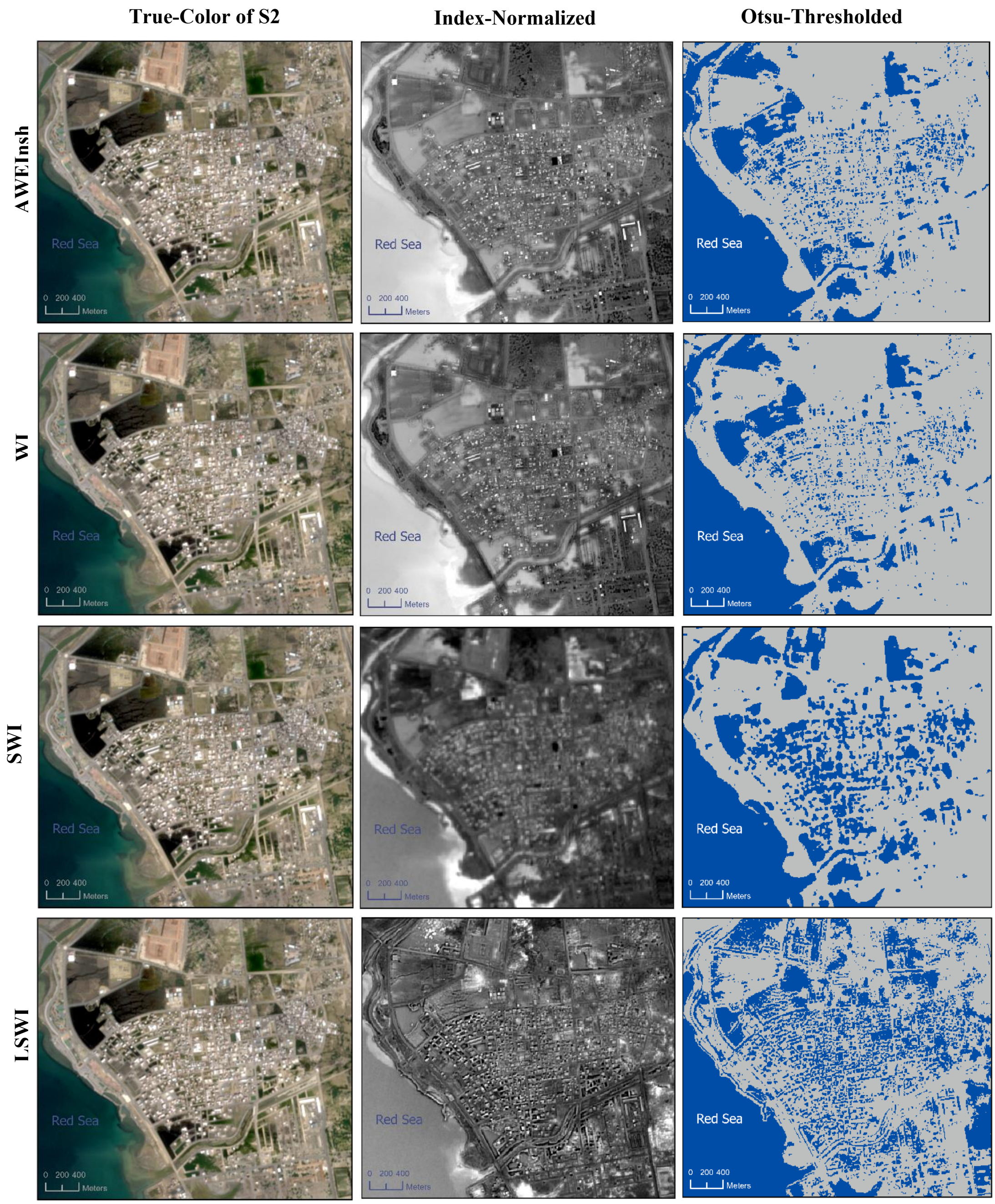


Fig. 3. Sentinel-2 true-color RGB composite (left column, acquired two days after peak inundation, showing persistent dark water signatures on surfaces), raw spectral water index maps (middle column, with values normalized between –1 and +1 in grayscale), and corresponding Otsu-thresholded binary flood maps (right column, blue = flooded) derived from the eight evaluated indices for the Al-Lith urban flash flood event.

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

TABLE 3

ACCURACY METRICS FOR SELECTED THRESHOLDED INDICES

Index	**Overall Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Kappa**	**False Alarm Rate (%)**
AWEIsh_otsu	80	75	85	80	0.60	15
NDWI_otsu	78	72	82	77	0.55	18
ENDWI_otsu	79	79.41	78	78	0.58	**10.95**

A visual comparison of the binary maps (Fig. 3) further highlights ENDWI_otsu's clearer delineation of urban flood extents, with fewer erroneous water pixels detected on dark roofs or roads.

4.3 Performance of Hybrid Max Fusion

The hybrid max fusion of ENDWI_raw and AWEIsh_raw, followed by Otsu thresholding, yielded the most effective overall classification. This simple combination capitalized on AWEIsh's broad separation strength and ENDWI's ability to suppress urban false positives, resulting in marked improvements across all metrics (Fig. 4).

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

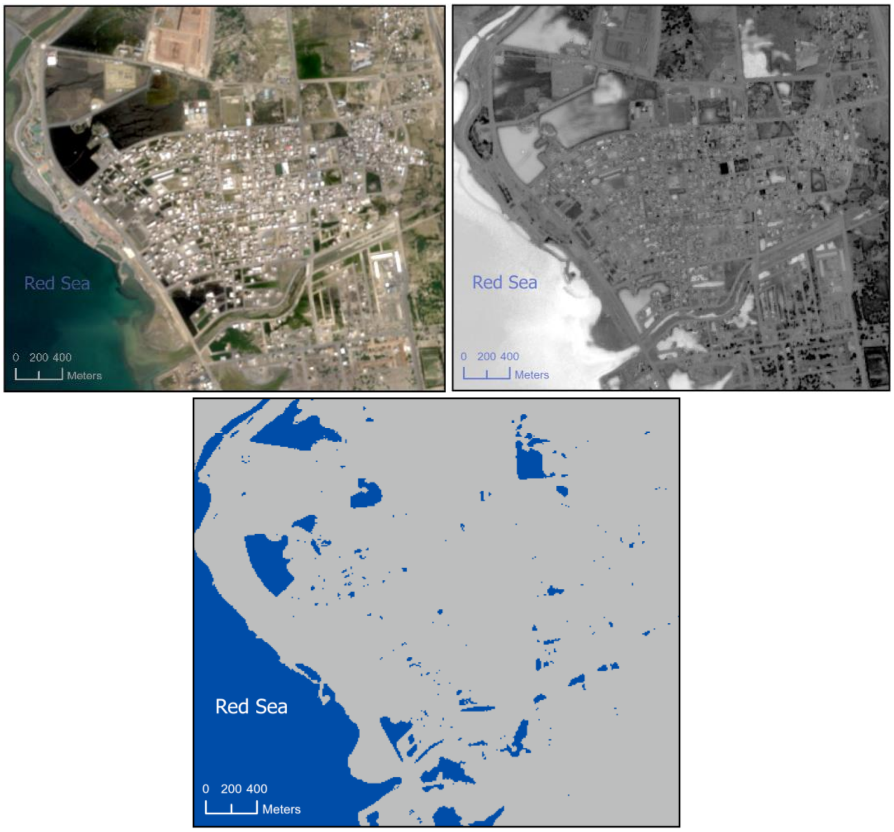


Fig. 4. Sentinel-2 true-color image post-flood (top-left); raw hybrid maximum fusion map of ENDWI and AWEIsh (top-right, grayscale normalized); and Otsu-thresholded binary flood map from the hybrid approach (bottom, blue = flooded) for the Al-Lith urban flash flood event. The hybrid method significantly reduces false alarms in built-up areas compared to individual indices.

The fused approach achieved an overall accuracy of 82.65%, a precision of 94.50%, an F1-score of 76.73%, and a Kappa coefficient of 0.637. Most notably, it reduced false positives to just 21 (false alarm rate = 2.99%) (Table 4), representing a substantial decrease compared to individual indices.

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

TABLE 4

PERFORMANCE COMPARISON OF HYBRID FUSION VS. TOP INDIVIDUAL INDICES

Method	Overall Accuracy (%)	Precision (%)	F1-Score (%)	Kappa	False Positives	False Alarm Rate (%)
Hybrid Max	**82.65**	**94.5**	76.73	0.64	21	**2.99**
ENDWI_otsu	79	79.4	78	0.58	77	10.95
AWEIsh_otsu	80	75	80	0.60	100	14

Corresponding flood extent maps (Fig. 4) illustrate the hybrid method's superior ability to suppress noise in built-up areas while accurately preserving true inundation features, closely matching high-resolution reference imagery.

These results demonstrate that the proposed ENDWI, both as a standalone method and in hybrid form, represents a practical advancement in reducing false alarms in urban optical flood mapping.

5. DISCUSSION

The results highlight the persistent challenge of false alarms in optical urban flood mapping and demonstrate how targeted enhancements to established water indices can yield meaningful improvements. Among the individual indices tested, AWEIsh_raw confirmed its reputation as a strong performer in complex environments due to its built-in shadow suppression terms. This finding aligns with previous studies showing that AWEIsh often outperforms simpler indices, such as NDWI, in scenes with varied urban surfaces (Stateczny et al., 2023; Tesfaye and Breuer, 2024). NDWI_raw, serving as the long-standing baseline(Miura et al., 2025) , provided reliable separation, consistent with its widespread application in Sentinel-2 analyses.

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

The proposed ENDWI, although ranking slightly lower in raw AUC, proved particularly valuable post-thresholding and in fusion applications. Its formulation—emphasizing the green band's reflectance through division by NDWI—effectively amplifies open water signals while dampening responses from dark impervious surfaces and shadows that plague NIR- or SWIR-dependent indices. The green band's role in enhancing contrast for turbid or urban-influenced water has been documented in previous studies involving hyperspectral, multispectral, and UAV sensors(Yan et al., 2017; Zhao et al., 2024). Our empirical trials extend this advantage to medium-resolution multispectral data, directly contributing to ENDWI_otsu's superior precision and markedly lower false alarm rate compared to alternative indices.

The hybrid max fusion of ENDWI_raw and AWEIsh_raw represented the most significant advancement, achieving the highest overall accuracy and dramatically reducing false positives to below 3%. By simply taking the pixel-wise maximum, this approach leveraged the complementary strengths of the two indices: AWEIsh's broad discriminatory power and ENDWI's targeted suppression of urban noise. Such rule-based fusion is notably more efficient than deep learning or multi-sensor integrations (e.g., Sentinel-1 SAR combined with Sentinel-2), which, while powerful, demand greater computational resources and labeled training data—resources are often scarce during rapid disaster response. Our method's simplicity and effectiveness align with recent efforts to refine index combinations for flood extent mapping, yet it stands out for its focus on minimizing false alarms in purely optical, cloud-free urban scenarios.

These gains are encouraging for operational use, particularly in arid or semi-arid cities prone to flash flooding, where timely and accurate inundation maps are crucial for damage assessment and evacuation planning. However, the study relies on a single post-event image pair from one flood event, limiting its generalizability across different seasons, water turbidity levels, and vegetation phenology. Additionally, reliance on cloud-free conditions remains a constraint of optical approaches, and the 10-meter resolution of Sentinel-2 may fail to detect narrow urban water features that higher-resolution data can capture.

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

Future work could explore adaptive weighting in the fusion step, the incorporation of additional bands (e.g., red-edge for vegetation masking), or the extension to time-series analysis for multi-event validation. Integrating ENDWI into ensemble frameworks with SAR data might further enhance all-weather capabilities. Overall, this work underscores that modest, interpretable modifications to classic indices can substantially mitigate urban spectral challenges, offering a practical tool for near-real-time flood monitoring using freely available Sentinel-2 imagery.

6. Conclusion

This study addressed the long-standing issue of false alarms in optical urban flood mapping by introducing the ENDWI and a simple hybrid max fusion with AWEIsh, applied to Sentinel-2 imagery from the 2018 Al-Lith flash flood event.

Raw index evaluation confirmed AWEIsh as the strongest separator, with ENDWI showing promising resilience to urban spectral confusion due to its green-band emphasis. Post-Otsu thresholding highlighted ENDWI's superior precision and lower false alarm rate, while the hybrid fusion delivered the best overall performance: 82.65% accuracy, 94.50% precision, and a false alarm rate of just 2.99%, a substantial reduction in erroneous water detections compared to individual indices.

These improvements stem from the complementary design of the fused approach, which combines broad discriminatory power with targeted noise suppression in a lightweight, parameter-free manner. The method's reliance on freely available Sentinel-2 data and standard GIS operations makes it particularly suitable for rapid, operational flood response in data-limited or resource-constrained settings. Although demonstrated on a single event, the results suggest that modest refinements to classic water indices can meaningfully advance urban flood detection without resorting to computationally intensive alternatives. Future extensions could include multi-temporal validation, integration with SAR for all-weather capability, or adaptive fusion weights to handle varying water conditions. In summary, ENDWI and the proposed hybrid method offer a practical

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

and effective tool for more reliable urban inundation mapping, contributing to reduced false alarms and better-informed disaster management.

Acknowledgements

The author thanks the European Space Agency (ESA) for providing open-access Sentinel-2 data. Sincere gratitude is extended to King Abdulaziz City for Science and Technology (KACST) for generously supplying the high-resolution WorldView-4 images to support scientific research. Appreciation is also expressed to the Saudi Civil Defense Directorate for their valuable flood report.

Disclosure Statement

The author reports there are no competing interests to declare.

Funding

This research received no external funding.

Generative AI Disclosure Statement

During the preparation of this work, the author used Grok 4.1 (developed by xAI) solely to enhance the English language, with the author reviewing and editing. All analytical content and scientific responsibility remain entirely with the author.

Data Availability Statement

The Sentinel-2 Level-2A imagery used in this study, including the specific product ID: *S2B_MSIL2A_20181128T075249_N0500_R135_T37QFC_20230727T143835.SAFE* (accessed on 08 May 2024) is freely available from the Copernicus Open Access Hub at https://browser.dataspace.copernicus.eu/. The raw WorldView-4 imagery used in this study was provided by King Abdulaziz City for Science and Technology (KACST). Restrictions apply to the availability of this data, which was used with agreement for this study. It can be requested directly by email Serv.sri@kacst.gov.sa or through the official KACST portal at https://kacst.gov.sa/en/. Reasonable requests for the data may also be directed to the author. Sample

https://doi.org/10.5194/egusphere-2026-672
Preprint. Discussion started: 9 February 2026

images, the copyright statement, Ground truth points, and metadata for WorldView-4 satellite imagery are publicly available in the GitHub repository "endwi" at https://github.com/aalmoadi/endwi under the MIT License. This repository additionally includes the processed data and all figures presented in this study.

REFERENCES

Albertini, C., Gioia, A., Iacobellis, V., and Manfreda, S.: Detection of Surface Water and Floods with Multispectral Satellites, Remote Sens., 14, 6005, https://doi.org/10.3390/rs14236005, 2022.

Amitrano, D., Di Martino, G., Di Simone, A., and Imperatore, P.: Flood Detection with SAR: A Review of Techniques and Datasets, Remote Sens., 16, 656, https://doi.org/10.3390/rs16040656, 2024.

Änuelas, J. P., Änol, J. P., Ogaya, R., and Filella, I.: Estimation of plant water concentration by the re¯ ectance Water Index WI (R900/R970), n.d.

Bersabe, J. T. and Jun, B.-W.: The Machine Learning-Based Mapping of Urban Pluvial Flood Susceptibility in Seoul Integrating Flood Conditioning Factors and Drainage-Related Data, ISPRS Int. J. Geo-Inf., 14, 57, https://doi.org/10.3390/ijgi14020057, 2025.

Doocy, S., Daniels, A., Murray, S., and Kirsch, T. D.: The Human Impact of Floods: a Historical Review of Events 1980-2009 and Systematic Literature Review, PLoS Curr., https://doi.org/10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a, 2013.

Elkarim, A. A.: INTERGRATION REMOTE SENSING AND HYDROLOGIC, HYDROULIC MODELLING ON ASSESSMENT FLOOD RISK AND MITIGATION: AL-LITH CITY, KSA, Int. J. GEOMATE, 18, https://doi.org/10.21660/2020.70.68180, 2020.

Elsebaie, I. H., Kawara, A. Q., and Alnahit, A. O.: Mapping and Assessment of Flood Risk in the Wadi Al-Lith Basin, Saudi Arabia, Water, 15, 902, https://doi.org/10.3390/w15050902, 2023.

Feyisa, G. L., Meilby, H., Fensholt, R., and Proud, S. R.: Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery, Remote Sens. Environ., 140, 23–35, https://doi.org/10.1016/j.rse.2013.08.029, 2014.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S.: Global flood risk under climate change, Nat. Clim. Change, 3, 816–821, https://doi.org/10.1038/nclimate1911, 2013.

Jiang, W., Ni, Y., Pang, Z., He, G., Fu, J., Lu, J., Yang, K., Long, T., and Lei, T.: A NEW INDEX FOR IDENTIFYING WATER BODY FROM SENTINEL-2 SATELLITE REMOTE SENSING IMAGERY, ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci., V-3–2020, 33–38, https://doi.org/10.5194/isprs-annals-V-3-2020-33-2020, 2020.

J-P. Schumann, G.: Breakthroughs in satellite remote sensing of floods, Front. Remote Sens., 4, 1280654, https://doi.org/10.3389/frsen.2023.1280654, 2024.

King Abdulaziz City for Science and Technology (KACST), N. C. for R. S. T. (NCRST): WorldView Satellite Imagery, https://kacst.gov.sa/en, 2018.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L. M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, G. R., Kron, W., Benito, G., Honda, Y., Takahashi, K., and Sherstyukov, B.: Flood risk and climate change: global and regional perspectives, Hydrol. Sci. J., 59, 1–28, https://doi.org/10.1080/02626667.2013.857411, 2014.

McFeeters, S. K.: The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, Int. J. Remote Sens., 17, 1425–1432, https://doi.org/10.1080/01431169608948714, 1996.

Ministry of Interior - General Directorate of Civil Defense: Report on the Collapse of the Dam and the Rainy Condition in Al-Lith Governorate, General Directorate of Civil Defense, Ministry of Interior, Kingdom of Saudi Arabia, 2018.

Miura, Y., Shamsudduha, M., Suppasri, A., and Sano, D.: A Global Multi-Sensor Dataset of Surface Water Indices from Landsat-8 and Sentinel-2 Satellite Measurements, Sci. Data, 12, 1253, https://doi.org/10.1038/s41597-025-05562-z, 2025.

Otsu, N.: A Tlreshold Selection Method from Gray-Level Histograms, n.d.

Özelkan, E.: Water Body Detection Analysis Using NDWIIndices Derived from Landsat-8 OLI, Pol. J. Environ. Stud., 29, 1759–1769, https://doi.org/10.15244/pjoes/110447, 2020.

Floods | Ready.gov: https://www.ready.gov/floods, last access: 12 October 2025.

Samela, C., Coluzzi, R., Imbrenda, V., Manfreda, S., and Lanfredi, M.: Satellite flood detection integrating hydrogeomorphic and spectral indices, GIScience Remote Sens., 59, 1997–2018, https://doi.org/10.1080/15481603.2022.2143670, 2022.

Singha, M., Dong, J., Sarmah, S., You, N., Zhou, Y., Zhang, G., Doughty, R., and Xiao, X.: Identifying floods and flood-affected paddy rice fields in Bangladesh based on Sentinel-1 imagery and Google Earth Engine, ISPRS J. Photogramm. Remote Sens., 166, 278–293, https://doi.org/10.1016/j.isprsjprs.2020.06.011, 2020.

Stateczny, A., Praveena, H. D., Krishnappa, R. H., Chythanya, K. R., and Babysarojam, B. B.: Optimized Deep Learning Model for Flood Detection Using Satellite Images, Remote Sens., 15, 5037, https://doi.org/10.3390/rs15205037, 2023.

Tan, J., Tang, Y., Liu, B., Zhao, G., Mu, Y., Sun, M., and Wang, B.: A Self-Adaptive Thresholding Approach for Automatic Water Extraction Using Sentinel-1 SAR Imagery Based on OTSU Algorithm and Distance Block, Remote Sens., 15, 2690, https://doi.org/10.3390/rs15102690, 2023.

Tesfaye, M. and Breuer, L.: Performance of water indices for large-scale water resources monitoring using Sentinel-2 data in Ethiopia, Environ. Monit. Assess., 196, 467, https://doi.org/10.1007/s10661-024-12630-1, 2024.

NOAA: https://www.weather.gov/safety/flood, last access: 12 October 2025.

Flooding | US EPA: https://www.epa.gov/natural-disasters/flooding, last access: 8 October 2025.

Wang, Z., Zhang, C., and Atkinson, P. M.: Combining SAR images with land cover products for rapid urban flood mapping, Front. Environ. Sci., 10, 973192, https://doi.org/10.3389/fenvs.2022.973192, 2022.

Xiao, X., Boles, S., Frolking, S., Salas, W., Moore III, B., Li, C., He, L., and Zhao, R.: Landscape-scale characterization of cropland in China using Vegetation and Landsat TM images, Int. J. Remote Sens., 23, 3579–3594, https://doi.org/10.1080/01431160110106069, 2002.

Xu, H.: Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery, Int. J. Remote Sens., 27, 3025–3033, https://doi.org/10.1080/01431160600589179, 2006.

Yan, D., Wang, X., Zhu, X., Huang, C., and Li, W.: Analysis of the use of $NDWI_{green}$ and $NDWI_{red}$ for inland water mapping in the Yellow River Basin using Landsat-8 OLI imagery, Remote Sens. Lett., 8, 996–1005, https://doi.org/10.1080/2150704X.2017.1341664, 2017.

Zhao, Z., Yang, J., Wang, M., Chen, J., Sun, C., Song, N., Wang, J., and Feng, S.: The PCA-NDWI Urban Water Extraction Model Based on Hyperspectral Remote Sensing, Water, 16, 963, https://doi.org/10.3390/w16070963, 2024.