

PAPER • OPEN ACCESS

## Flood Susceptibility Mapping Using Machine Learning: Spatial Risk Assessment and Implications for the Green Riyadh Project

To cite this article: Zheng Cong *et al* 2026 *IOP Conf. Ser.: Earth Environ. Sci.* **1615** 012066

View the [article online](#) for updates and enhancements.

### You may also like

- [Spatial Modelling of Landscape Disaster \(Flood\) Prevention in Bekasi City](#)  
Hafiz Agam Hidayat and Alinda Fitriany  
Malik Zain
- [Monitoring and prediction of flood risks in Tipar Watershed through land use change analysis using remote sensing and GIS techniques](#)  
Albertus Erico Jerry Krisna Nugroho, Imam Mahdi, Ike Fitri Samsiyah *et al.*
- [Mapping flood vulnerability in Wasit province, Iraq, using enhanced AHP-GIS techniques](#)  
Rabab H. Abood, Iman A. Hameed and Abduladheem S. Alrikabi

# Flood Susceptibility Mapping Using Machine Learning: Spatial Risk Assessment and Implications for the Green Riyadh Project

Zheng Cong<sup>1,2</sup>, Mirza Waleed<sup>1,2,3</sup>, Furqan Tahir<sup>1,2</sup> and Sami G. Al-Ghamdi<sup>1,2\*</sup>

<sup>1</sup> Environmental Science and Engineering Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.

<sup>2</sup> KAUST Climate and Livability Initiative, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.

<sup>3</sup> Department of Geography, Hong Kong Baptist University, Hong Kong, China

\*E-mail: sami.alghamdi@kaust.edu.sa

**Abstract.** Flash flooding presents a critical challenge for sustainable development in arid cities such as Riyadh, where rapid urbanization and complex geomorphology exacerbate hydrological risks. This study integrates high-resolution remote sensing, geospatial analysis, and machine learning (Light Gradient Boosting Machine, LGBM) to produce a 30-meter flood susceptibility map for Riyadh, complemented by spatial analysis of vegetation distribution using NDVI clustering. The results indicate that the overall flood risk in the studied sub-catchment is low, but highly concentrated in low-lying drainage corridors, especially along the Wadi Hanifa watershed. Notably, spatial overlap between green infrastructure clusters and flood-prone zones is observed, suggesting that green spaces are often situated in hydrologically sensitive areas. A detailed assessment of three major urban parks under construction as part of the Green Riyadh Project—King Salman Park, Wadi Hanifa Park, and Al-Uruba Park—reveals substantial variability in flood risk, primarily governed by local topography and hydrological connectivity. These findings highlight the necessity of integrating targeted green infrastructure planning with flood risk management and underscore the importance of adaptive, data-driven approaches for future urban resilience in arid environments.

## 1. Introduction

Flash flooding in urban environments arises from the complex interplay of climatic variability, geographical features, and rapid urban development [1,2]. In Riyadh, the city's topography—surrounded by valleys and marked by steep slopes—significantly increases its vulnerability to flash floods [3]. The region's arid climate is characterized by sporadic but intense rainfall events, which further magnify the threat to urban infrastructure and public safety [4–6]. These challenges are further exacerbated by ongoing urbanization, which alters natural drainage patterns and increases surface runoff.

Recent advances in modeling and monitoring, particularly the use of high-resolution terrain data and remote sensing technologies, have become indispensable for effective flood risk



assessment and management in urban contexts [7]. Managing flood risk in Riyadh requires comprehensive planning and disaster management strategies specifically tailored to the city's unique environmental and urban characteristics. Urban planning must prioritize the integration of flood risk mitigation into development agendas, emphasizing sustainable practices that are adaptable to evolving climate realities [6,8]. Flood susceptibility mapping has emerged as a critical tool in flood risk management. By leveraging machine learning algorithms and geospatial technologies, these maps identify areas most prone to flooding, thereby enhancing both the understanding of flood risks and the effectiveness of decision-making in disaster management and urban planning. Advanced machine learning models, such as Light Gradient Boosting Machine (LGBM), have demonstrated strong performance in flood susceptibility assessment, owing to their ability to process large, heterogeneous datasets and uncover complex relationships among risk factors, including topography, land use, and climatic conditions [9,10]. The practical application of flood susceptibility maps enables municipalities to prioritize infrastructural improvements and allocate resources more efficiently in anticipation of flood events.

The Green Riyadh Project is a major urban greening initiative aimed at transforming Riyadh by planting 7.5 million trees across public spaces, institutions, and thoroughfares, thereby increasing per capita green space from 1.7 to 28 square meters and expanding total green areas to around 9% of the city. This initiative aligns with the broader adoption of green infrastructure principles, aiming to enhance environmental resilience, mitigate flood risks, and promote ecological benefits within urban settings. Incorporating data-driven flood susceptibility mapping into the evaluation of mega-green initiatives is crucial for effective future flood risk management, as it enables targeted planning, adaptive implementation, and long-term resilience in urban environments.

### *1.1 Study area*

Riyadh, the capital and largest city of Saudi Arabia, is centrally located on the Arabian Peninsula (24° N, 46° E). In recent decades, the city has undergone rapid urban expansion, placing new strains on sustainable urban development. At the same time, increases in both the frequency and intensity of rainfall—trends tied to climate change—have heightened flood risk, underscoring the challenge of flood risk management under arid environments. A key geographical and hydrological feature influencing flood dynamics in Riyadh is the Wadi Hanifah watershed. Extending approximately 120 km from the north (near Al Uyaynah) to the south (around Al-Hair city) [11]. The watershed comprises ephemeral streams that typically flow during and shortly after rainfall events, shaping patterns of surface runoff and influencing the spatial distribution of flash floods. In this study, the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) was employed to delineate the sub-catchment of Wadi Hanifah. This delineation ensures that both upstream runoff dynamics and their interactions with urbanized zones are appropriately captured, thereby improving the reliability and spatial relevance of flood susceptibility assessments.

## **2. Data and methods**

### *2.1 Data sources and acquisition*

The overall methodology of this study is divided into five main stages: data acquisition, feature preparation, machine learning (ML) modeling, accuracy assessment, and flood susceptibility evaluation. The research begins by collecting data from a range of sources, including the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM), Climate Hazards Group

InfraRed Precipitation with Station data (CHIRPS) rainfall estimates, Sentinel-2 multispectral imagery, and vector datasets representing hydrological and infrastructural features such as river networks and catchment boundaries. These datasets are systematically processed to derive a suite of flood-influencing factors, including elevation, slope, aspect, rainfall, and various remote sensing indices. Flood inventory data were obtained from historical satellite imagery interpretation, official government disaster records, and documented local flood events in Riyadh. The prepared features were subsequently used to construct training and validation datasets for machine learning model development. In addition, the Green Riyadh Project Plan was sourced from the publicly available official website and used as a reference to identify three major ongoing park developments: King Salman Park, Al-Uruba Park, and Wadi Hanifa Park.

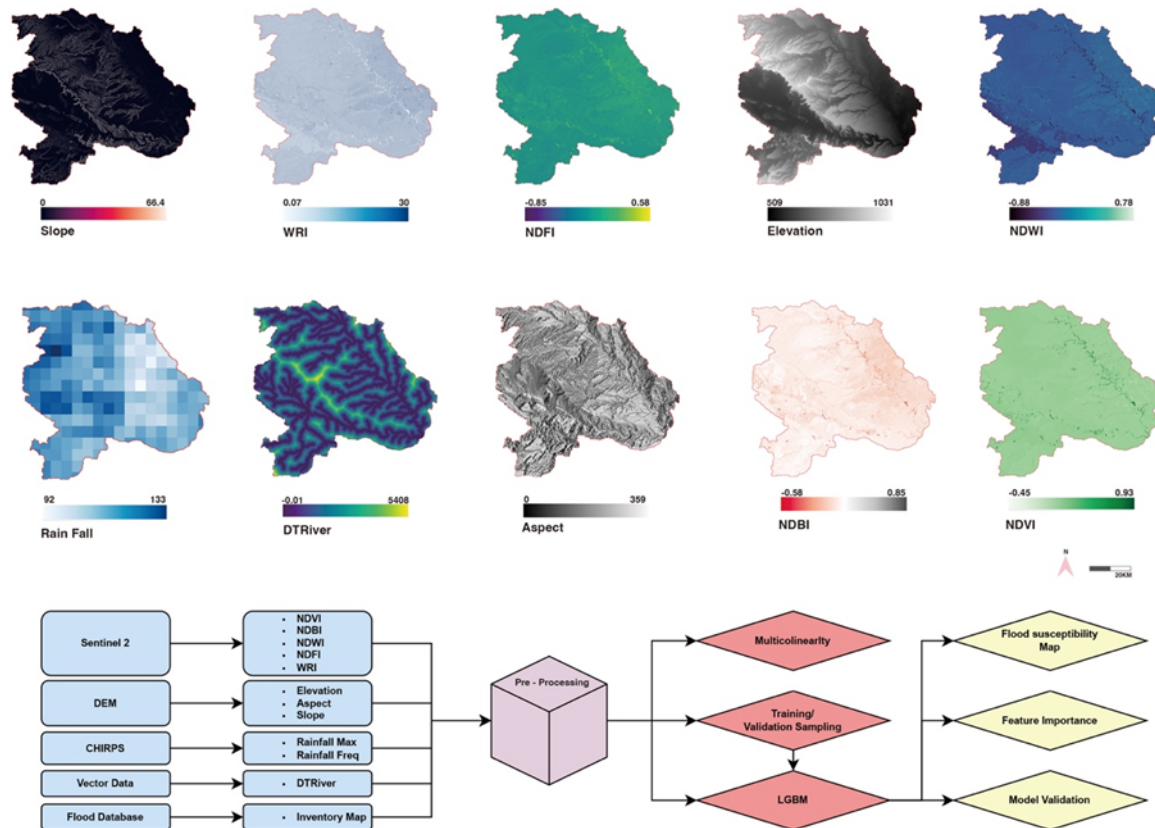
### *2.2 Feature extraction and data preprocessing*

As shown in Figure 1, A total of 10 features were extracted based on their relevance to flash flood processes in arid urban environments, as confirmed by both literature and local context. Topographic variables—including elevation, slope, and aspect—were derived from the DEM, while distance to river channels (DtRiver) and other proximity measures were calculated using Euclidean distance analysis in GIS. Meteorological features were represented by total rainfall and rainfall frequency, obtained from the processed CHIRPS dataset. Land cover and urbanization were characterized using indices such as the Normalized Difference Built-up Index (NDBI), Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Normalized Difference Flood Index (NDFI), Water Ratio Index (WRI).

All input feature layers, originally in GeoTIFF format, were systematically converted into one-dimensional arrays using NumPy and then reorganized into two-dimensional matrices, with each row representing a spatial observation and each column denoting a distinct feature variable. Data preprocessing included a logarithmic transformation to mitigate skewness and diminish the impact of extreme values; a scaling factor was introduced when necessary to ensure all values remained positive. Subsequent normalization, implemented via either min-max scaling or z-score standardization, standardized the feature ranges and enhanced model convergence. The processed arrays were saved in a compressed NumPy format to optimize storage efficiency and streamline subsequent data loading.

### *2.3 Machine learning modeling and validation*

The LGBM algorithm was employed for flash flood susceptibility mapping owing to its efficiency and strong performance in handling heterogeneous, high-dimensional data [12]. Positive (flood) and negative (non-flood) samples were derived from flood inventory records using stratified random sampling to maintain class balance. The dataset was split into training (67%) and testing (33%) subsets, with all relevant features included as model inputs. Model hyperparameters were optimized through grid search in conjunction with cross-validation to maximize generalizability and minimize overfitting. The trained LGBM model generated a continuous 30-meter resolution susceptibility surface, quantifying the probability of flood occurrence for each grid cell. Model performance was rigorously evaluated using the hold-out test set, with both flooded and non-flooded classes proportionally represented. Key metrics—including accuracy, precision, recall, and F1-score—were calculated to assess classification quality.



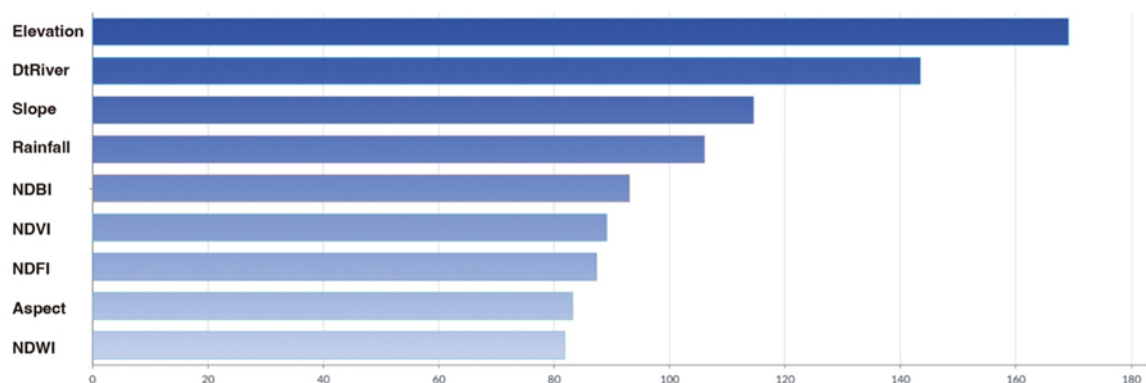
**Figure 1.** Data sources, feature layers, and workflow for flood susceptibility mapping

### 3. Results

#### 3.1 Feature importance and model performance

Feature importance analysis, derived from the LGBM's built-in algorithm, provides valuable insights into the relative contribution of each variable to flood risk prediction. As shown in Figure 2, elevation emerged as the most influential feature, reflecting the dominant role of topography in controlling surface runoff patterns and water accumulation in arid urban environments. Distance to river (DtRiver) and slope were also ranked highly, consistent with the understanding that proximity to drainage channels and land surface gradients are key determinants of flood pathways and velocity. Rainfall variables, including total precipitation and rainfall frequency from the CHIRPS dataset, contributed substantially, highlighting how episodic yet intense rainfall events drive flash flood hazards in Riyadh's climate.

Urbanization and land cover features—captured by indices such as the NDBI and NDVI—also showed moderate importance. The NDBI reflects the extent of impervious surfaces, which accelerate runoff and reduce infiltration, while NDVI indicates vegetation's role in modulating hydrological responses. The LGBM model demonstrated strong predictive performance, achieving an accuracy of 0.88, precision of 0.90, recall of 0.86, and an F1-score of 0.88 on the hold-out test set. These metrics confirm the model's robustness and generalizability for identifying both flood-prone and safe areas.



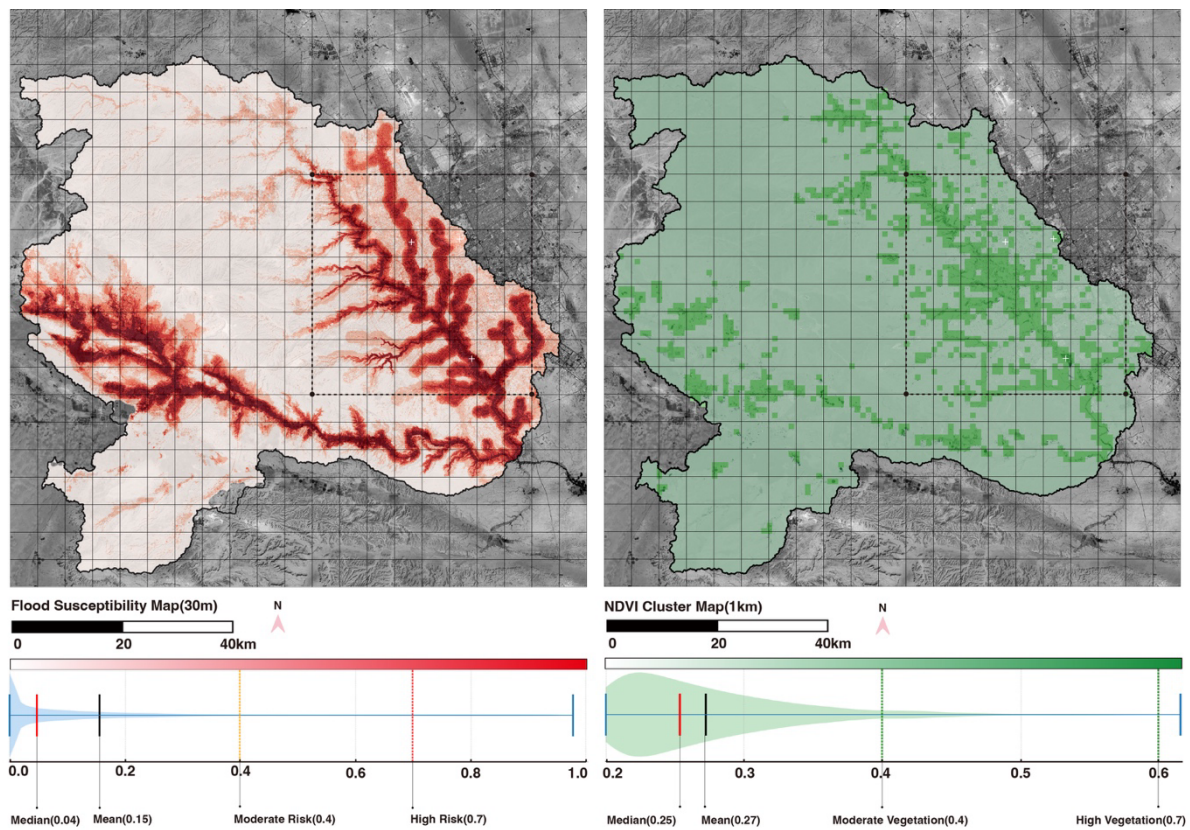
**Figure 2.** Feature importance ranking for flood susceptibility modeling using the LGBM algorithm.

### 3.2 Spatial distribution in flood susceptibility

The final flood susceptibility map, produced at a 30-meter spatial resolution, exhibits values ranging from 0.00 to 0.98. The mean flood susceptibility is 0.16, with a median of 0.05, indicating that the overall flood risk across Riyadh is relatively low. The majority of study areas (86.43%) are categorized as low risk (<0.40), while 7.00% are classified as high risk (>0.70) and 6.57% as moderate risk (0.40–0.70). The standard deviation of 0.24 further reflects moderate spatial variability within the study area.

These findings are consistent with the climatic and environmental characteristics of Saudi Arabia, particularly the Riyadh region, which is defined by its arid conditions and infrequent high-intensity rainfall events, as substantiated by the CHIRPS dataset. The distinctly polarized distribution of flood susceptibility—where most areas are at very low risk and only a small proportion exhibit extremely high risk—emphasizes the necessity for event-based flood modeling in arid environments. This pattern suggests that, while widespread flooding is rare, localized flash flood hazards may be severe and therefore demand a detailed site-specific investigation.

In terms of spatial distribution, flood risk demonstrates pronounced heterogeneity across the region. As illustrated in Figure 3, the flood susceptibility map, high-risk zones (dark red) are predominantly located along the Wadi Hanifah and its tributaries, forming a characteristic dendritic drainage pattern. These high-risk areas are generally situated in low-lying regions with convergent topography and elevated runoff accumulation, making them particularly vulnerable to localized flooding during extreme precipitation events. Moderate-risk zones (light red) are mainly distributed as transitional areas surrounding the high-risk corridors. Overall, the spatial pattern of flood risk is strongly governed by topographic and hydrological convergence factors.



**Figure 3.** Spatial distribution of flood susceptibility and green space clusters in Riyadh.

### 3.3 Spatial distribution of green cluster

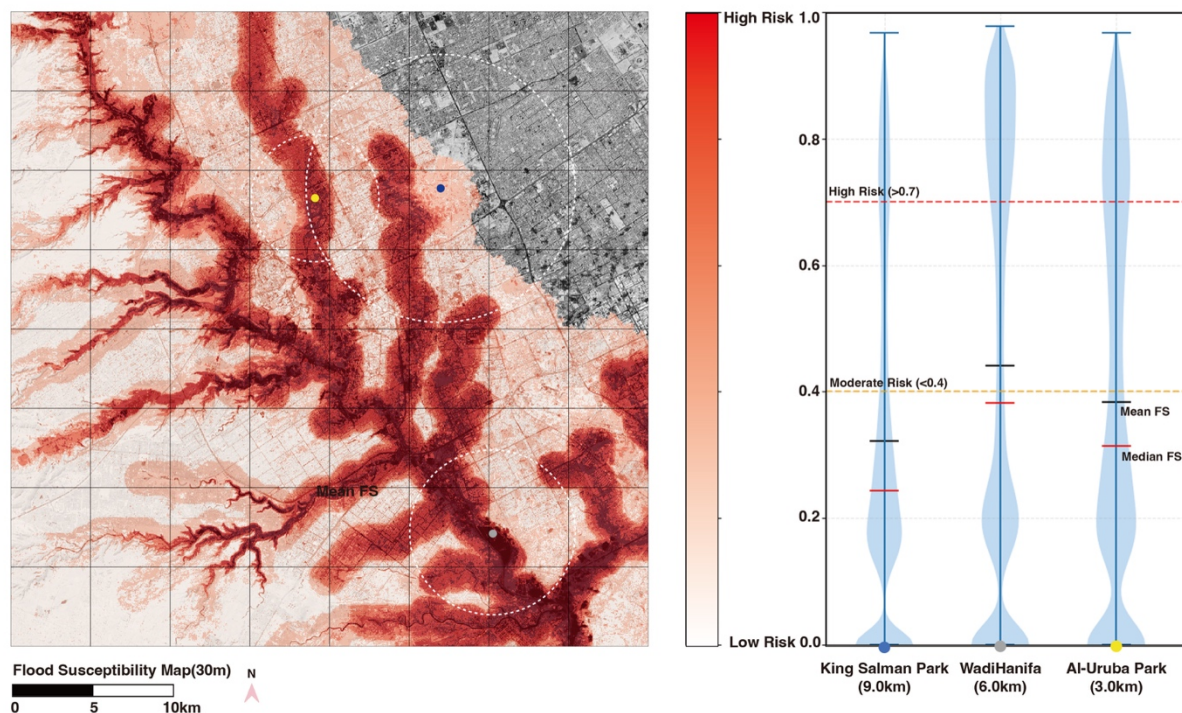
To further elucidate the relationship between the Green Riyadh Project and flood risk, and to prioritize the influence of green infrastructure on flood mitigation, we conducted an in-depth spatial analysis of the distribution of vegetation using the Normalized Difference Vegetation Index (NDVI). The analysis began with a histogram evaluation of NDVI values across the study area, which revealed that the majority of values clustered between -0.03 and 0.20, indicating generally sparse vegetation, with only a small proportion of pixels exhibiting higher NDVI values characteristic of denser vegetation. In light of this distribution and considering the local climatic and environmental context, a threshold of  $NDVI > 0.20$  was established to define areas with significant vegetation cover. A binary mask was subsequently generated to isolate pixels exceeding this threshold, representing potential green infrastructure. To identify spatially contiguous patches of vegetation, we applied an 8-connected neighborhood clustering algorithm using the `scipy.ndimage.label` function in Python. This approach considers pixels contiguous if they share either an edge or a corner, enabling the extraction of coherent green infrastructure units across the landscape. To facilitate regional-scale comparison and analysis, the resulting clusters were resampled from the original 30-meter spatial resolution to a coarser 1-kilometer grid and formed an NDVI cluster map, as shown in Figure 3.

The results show NDVI values ranging from 0.20 to 0.62, with a mean of 0.27 and a median of 0.25, further confirming the overall sparsity of vegetation cover in the region. Specifically, 93.81% of the grid cells fall within the low vegetation coverage category ( $<0.40$ ), 6.10% are classified as moderate (0.40–0.60), and only 0.10% exhibit high coverage ( $>0.60$ ). The spatial comparison

reveals a notable overlap between areas of high flood susceptibility and regions characterized by elevated NDVI cluster values. As illustrated in Figure 3, the spatial distribution of green spaces tends to coincide with the main drainage corridors and low-lying zones that also exhibit high flood risk. The co-location of green space clusters and flood-prone areas can be attributed to the underlying hydrological and geomorphological settings. Channels and depressions not only serve as conduits for surface runoff during rainfall events but also provide localized moisture conditions that facilitate vegetation growth in an otherwise arid landscape. As a result, these green patches emerge as ecologically significant zones that both benefit from and influence hydrological processes.

### 3.4 Flood risk assessment of major urban parks

Building on the demonstrated spatial overlap between green infrastructure clusters and flood-prone corridors at the city scale, a detailed, site-specific flood susceptibility assessment was undertaken for three major urban parks designated in the Green Riyadh Project master plan: King Salman Park, Wadi Hanifa Park, and Al-Uruba Park. To account for the varying spatial extents and landscape influences of each park, customized buffer zones proportional to park size were applied—9 km for King Salman Park, 6 km for Wadi Hanifa Park, and 3 km for Al-Uruba Park—ensuring a context-sensitive and equitable evaluation of flood risk. The analysis reveals pronounced heterogeneity in flood risk profiles among the three parks (see Figure 4).



**Figure 4.** Localized flood susceptibility assessment for major urban parks in Riyadh.

Wadi Hanifa Park exhibits the highest level of flood susceptibility, with a mean value of 0.44 and 35.2% of its area classified as high risk ( $FS > 0.7$ ). This elevated vulnerability aligns with previous findings that major vegetated clusters frequently coincide with natural drainage corridors. The park's location along the Wadi Hanifa wadi system exposes it to rapid runoff

accumulation and channelized flood flows during intense rainfall events, underscoring the dual role of these corridors as both ecological assets and areas of heightened hydrological risk.

By contrast, Al-Uruba Park and King Salman Park show substantially lower average flood susceptibility values (0.38 and 0.32, respectively), with a greater proportion of their buffered areas falling into the low-risk category (56.6% and 68.6%). Both parks are situated in relatively elevated or hydrologically less connected zones, reducing their exposure to direct flood hazards despite the presence of vegetation. This distinction illustrates how local topography and hydrological connectivity can mediate the effectiveness and vulnerability of green infrastructure in urban flood risk management.

The spatial correlation between flood susceptibility and NDVI clusters underscores the critical role of riparian and floodplain vegetation as multifunctional infrastructure. While such vegetation can enhance infiltration, reduce runoff, and facilitate sediment retention, its exposure to extreme flood events necessitates careful and adaptive management to sustain its ecological and protective functions.

In summary:

- Parks embedded within or adjacent to major Wadi systems (e.g., Wadi Hanifa Park) are subject to higher flood risk, even with significant green infrastructure.
- Parks located on higher ground or in less hydrologically connected settings (e.g., King Salman Park, Al-Uruba Park) benefit from both lower inherent flood susceptibility and the hydrological advantages provided by vegetation.

### *3.5 Implementation for future development*

The findings of this study provide a strategic framework for integrating flood risk assessment with green infrastructure planning in Riyadh's future urban development. The pronounced spatial overlap between high flood susceptibility zones and vegetated areas presents both opportunities and challenges for multifunctional landscape design—where green infrastructure can simultaneously deliver ecological benefits and perform flood mitigation functions.

For effective implementation, urban planners should prioritize the placement and enhancement of green infrastructure within or adjacent to major drainage corridors identified as high flood risk [13,14]. This requires a nuanced, context-driven approach: the selection of flood-tolerant plant species, the design of landscape elements that promote infiltration and delay surface runoff, and the construction of green spaces resilient to periodic inundation. The routine integration of high-resolution, dynamically updated flood susceptibility maps into urban planning and the Green Riyadh Project's monitoring framework will enable adaptive management. Such a data-driven approach ensures that green infrastructure strategies remain effective amid changes in land use, urban expansion, and climatic variability. Collaboration across sectors—including municipal authorities, ecologists, engineers, and local communities—is essential. Public engagement initiatives can increase awareness of the multifunctional benefits of green infrastructure, while participatory planning can ensure that local knowledge and needs are incorporated into implementation. Establishing robust monitoring and evaluation protocols—such as remote sensing-based vegetation health assessments and post-flood performance reviews—will support the continuous optimization of green infrastructure, maximizing both environmental and social co-benefits.

These insights reinforce the necessity of tailoring green infrastructure strategies to local geomorphological and hydrological contexts. Integrating these findings into the future planning and implementation of the Green Riyadh Project will be critical for maximizing the resilience and multifunctionality of urban green spaces, particularly in arid urban environments.

#### 4. Discussion

While the study leverages advanced remote sensing and machine learning techniques, several limitations remain. First, the reliance on historical flood inventory data and satellite-based rainfall estimates may introduce uncertainties, particularly in the characterization of rare or undocumented flood events. The spatial resolution mismatch between flood susceptibility mapping (30m) and NDVI clustering (1km) could also affect the precision of spatial correlation analyses. Additionally, the static nature of input features does not capture temporal variability in land use, vegetation phenology, or rainfall intensity, which could influence dynamic flood risk patterns.

Future research should aim to incorporate time-series remote sensing data and near-real-time hydrological modeling to better capture the dynamic interplay between urban development, vegetation growth, and flood risk. The integration of socio-economic data, such as population density and infrastructure value, would also enable more comprehensive risk assessments. Experimental work on nature-based flood mitigation strategies, including the design and monitoring of green infrastructure interventions in high-risk corridors, is recommended to validate and refine the predictive models. Finally, participatory approaches involving local communities and stakeholders could enhance the relevance and effectiveness of green infrastructure planning in rapidly growing arid cities.

#### 5. Concluding remarks

This research demonstrates the effectiveness of integrating high-resolution geospatial analysis, remote sensing, and advanced machine learning for flood susceptibility assessment in arid environments. The LGBM model, informed by a comprehensive set of environmental and anthropogenic variables, successfully delineates areas of heightened flood risk, particularly along major drainage networks such as the Wadi Hanifa corridor.

A key finding is the significant spatial overlap between green infrastructure clusters and flood-prone areas, indicating both opportunities and challenges for urban resilience. While vegetated zones can play a crucial role in mitigating runoff and enhancing ecological connectivity, their co-location with high-risk areas necessitates adaptive, context-specific design and management. The detailed analysis of major urban parks within the Green Riyadh Project underscores the importance of considering local topography and hydrological connectivity in green space planning. Parks situated within drainage corridors face higher flood vulnerability, reinforcing the need for tailored approaches to vegetation selection and landscape design. Overall, this study highlights the value of data-driven, adaptive flood risk management in shaping the future of sustainable urban development. The integration of flood susceptibility mapping with green infrastructure planning provides a pathway for Riyadh—and similar arid megacities—to enhance urban resilience, ecological health, and the well-being of their populations in the face of ongoing climatic and urbanization pressures.

#### References

- [1] Serdar M Z, Ajjur S B and Al-Ghamdi S G 2022 Flood Susceptibility Assessment in Arid Areas: A Case Study of Qatar *Sustainability* **14** 9792
- [2] Tahir F, Ajjur S B, Serdar M Z, Al-Humaiqani M M A, Kim D, Al-Thani S K and Al-Ghamdi S G 2021 *Qatar Climate Change Conference 2021: A platform for addressing key climate change topics facing Qatar and the world* ed S G Al-Ghamdi and S K Al-Thani (Doha, Qatar: Hamad bin Khalifa University Press (HBKU Press))
- [3] Rahman M T, Aldosary A S, Nahiduzzaman K M and Reza I 2016 Vulnerability of flash flooding in Riyadh, Saudi

- Arabia *Nat. Hazards* **84** 1807–30
- [4] Ali A, Ullah W, Khan U A, Ullah S, Ali A, Jan M A, Bhatti A S and Jan Q 2024 Assessment of multi-components and sectoral vulnerability to urban floods in Peshawar – Pakistan *Nat. Hazards Res.* **4** 507–19
- [5] Al-Humaiqani M and Al-Ghamdi S 2023 Assessing the built environment’s reflectivity, flexibility, resourcefulness, and rapidity resilience qualities against climate change impacts from the perspective of different stakeholders *MDPI Sustain.*
- [6] Alharbi T 2024 A Weighted Overlay Analysis for Assessing Urban Flood Risks in Arid Lands: A Case Study of Riyadh, Saudi Arabia *Water* **16** 397
- [7] Sharif H O, Al-Juaidi F H, Al-Othman A, Al-Dousary I, Fadda E, Jamal-Uddeen S and Elhassan A 2016 Flood hazards in an urbanizing watershed in Riyadh, Saudi Arabia *Geomatics, Nat. Hazards Risk* **7** 702–20
- [8] Truu M, Annus I, Roosimägi J, Kändler N, Vassiljev A and Kaur K 2021 Integrated Decision Support System for Pluvial Flood-Resilient Spatial Planning in Urban Areas *Water* **13** 3340
- [9] Seydi S T, Kanani-Sadat Y, Hasanlou M, Sahraei R, Chanussot J and Amani M 2022 Comparison of Machine Learning Algorithms for Flood Susceptibility Mapping *Remote Sens.* **15** 192
- [10] Waleed M and Sajjad M 2025 Advancing flood susceptibility prediction: A comparative assessment and scalability analysis of machine learning algorithms via artificial intelligence in high-risk regions of Pakistan *J. Flood Risk Manag.* **18**
- [11] AlAli A M, Salih A and Hassaballa A 2023 Geospatial-Based Analytical Hierarchy Process (AHP) and Weighted Product Model (WPM) Techniques for Mapping and Assessing Flood Susceptibility in the Wadi Hanifah Drainage Basin, Riyadh Region, Saudi Arabia *Water* **15** 1943
- [12] Waleed M and Sajjad M 2025 High-resolution flood susceptibility mapping and exposure assessment in Pakistan: An integrated artificial intelligence, machine learning and geospatial framework *Int. J. Disaster Risk Reduct.* **121** 105442
- [13] Al-Humaiqani M M and Al-Ghamdi S G 2023 Integrating Green–Blue–Gray Infrastructure for Sustainable Urban Flood Risk Management: Enhancing Resilience and Advantages *Sustainable Cities in a Changing Climate* (Wiley) pp 207–26
- [14] Azmeer A, Tahir F and Al-Ghamdi S G 2024 Progress on green infrastructure for urban cooling: Evaluating techniques, design strategies, and benefits *Urban Clim.* **56** 102077