A Data-Driven Framework to Anticipate Urban Flooding in Mumbai Using Verified Hydrological and Land-Use Indicators

Maitreyi Joglekar
Department of Information Technology
Vidyalankar School of Information Technology, Mumbai, India

OPEN ACCESS

Volume: 13

Special Issue: 2

Month: January

Year: 2026

E-ISSN: 2582-0397

P-ISSN: 2321-788X

Citation:
Joglekar, Maitreyi. "A Data-Driven Framework to Anticipate Urban Flooding in Mumbai Using Verified Hydrological and Land-Use Indicators." *Shanlax International Journal of Arts, Science and Humanities*, vol. 13, no. 2, 2026, pp. 7–13.

DOI:
https://doi.org/10.34293/sijash.v13iS2-i2-Jan.10503

Abstract

Mumbai's recent monsoon seasons, including several events in 2025, have shown how short intense spells of rain can quickly disrupt the city, especially in low-lying and densely built areas. Many earlier studies discuss heavy rainfall, terrain and land-use change, but usually treat them separately, which makes it difficult to get a clear, citywide view of flood susceptibility. This paper presents a data-driven framework that combines three sets of routinely available information: daily rainfall from the Colaba and Santacruz IMD stations, elevation and slope derived from a digital elevation model, and land-use patterns indicating where surfaces are mostly built-up or permeable. From these datasets, basic indicators of rainfall stress, terrain setting and surface response are created and examined together. Looking at these indicators in combination helps to identify locations where intense rain, low ground levels and hard urban surfaces overlap. The framework brings out existing hotspots and helps identify places where vulnerability may be increasing. The framework offers a practical basis for ward-level preparedness and can be updated annually as new rainfall and land-use datasets become available.

Keywords: Urban Flooding, Flood Susceptibility, Rainfall Indicators, Impervious Surfaces, Data-driven Flood Framework, Hydrometeorological Indicators

Introduction

Mumbai is routinely affected by intense monsoon rainfall that disrupts transport, damages infrastructure and, in extreme cases, paralyses large parts of the city. The July 2005 floods were a turning point: they exposed how a combination of very heavy rain, high tide and limited drainage capacity can quickly overwhelm an already dense coastal metropolis. Since then, national agencies have repeatedly highlighted Mumbai as a key example of urban flood risk in India and have called for city-specific flood management frameworks rather than treating urban floods as a minor extension of riverine flooding [1], [2].

Recent observations show that such extremes are no longer rare outliers. In 2025, Mumbai experienced an unusually active and spatially uneven monsoon. IMD data analysed for May 2025 indicate that the Colaba observatory recorded about 456.5 mm of rain and the Santacruz observatory 342 mm between 1 and 27 May, exceeding a century-old May rainfall record for the city [3]. The monsoon strengthened considerably in August 2025. Santacruz recorded 1,188.8 mm for the month — over twice its long-term average and making it one of the wettest Augusts in decades. Nearly 70% of this total fell within a four-day spell (15–19 August). Colaba registered 686 mm during the same month. IMD advisories and news reports described this period as an unusually intense rainfall episode with several days exceeding 200 mm.

The rainfall charts prepared by the Regional Meteorological Centre, Mumbai, for the 2025 monsoon season illustrate this behavior clearly. Fig. 1 shows the cumulative rainfall plot for Colaba from 1 June to the end of September. The daily rainfall chart for Colaba and Santacruz reveals multiple high-intensity events scattered through the season, including several days above 100 mm and peaks around 220–240 mm in mid-August as shown in Fig. 2. These patterns match IMD's classification of "very heavy" and "extremely heavy" rainfall events and are typical of conditions that generate rapid surface runoff and waterlogging in dense urban areas [3], [4].

While rainfall is the immediate trigger, Mumbai's flood response is shaped by terrain, land use and drainage infrastructure. Municipal technical presentations on urban flooding note that large parts of the storm-water network were designed decades ago for much lower design intensities and, in many locations, are constrained by flat gradients, siltation and tidal backwater effects [5]. Geospatial studies of Greater Mumbai further show that low-lying pockets, high built-up density and reduced permeable surfaces coincide with many of the city's recurrent flood hotspots [6]. National land-use assessments by the National Remote Sensing Centre confirm a steady expansion of built-up land and provide mapped information on impervious cover at urban scales.

National disaster management guidelines emphasize that effective urban flood management should combine such hydrometeorological, topographic and land-use information into a city-specific decision framework, rather than relying only on historical experience or single-source indicators [1]. However, in practice, rainfall statistics, elevation data and land-cover information are often examined independently and their combined influence on localized flooding is not fully utilized for anticipatory planning. This paper responds to that gap by outlining a data-driven framework to anticipate urban flooding in Mumbai using three verified indicator groups:

- rainfall intensity and accumulation derived from IMD observations;
- elevation and terrain attributes from digital elevation models; and
- land-use and impervious surface information from national land-cover products.

The framework is intended not as a full operational model but as a structured basis that civic agencies and researchers can adapt to identify ward-level flood susceptibility, explore "what-if" rainfall scenarios, and support more proactive flood preparedness in Mumbai.

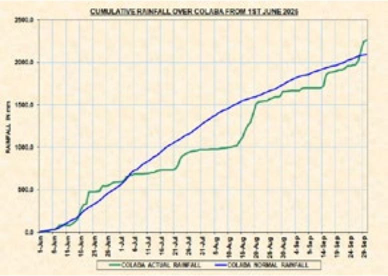

Figure 1 Cumulative Rainfall Plot for Colaba, 1 June to End of September 2025

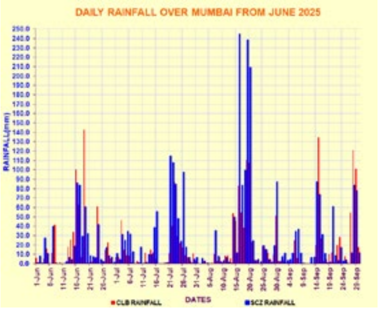

Figure 2 Daily Rainfall Chart for Colaba and Santacruz 2025

Related Work

Research on urban flooding in Mumbai and other Indian cities has evolved from descriptive event analyses to more formal, data-driven susceptibility mapping. Early work on Mumbai, including case studies of the 2005 flood, concentrated on documenting rainfall extremes, drainage limitations and land-use change, and highlighted the need for systematic assessment tools but did not develop explicit indicator-based models [7], [8].

A more structured approach was introduced by Ramesh and Iqbal, who developed a flood susceptibility zonation map for Greater Mumbai using GIS-based statistical models [6]. Their work used a flood inventory and twelve spatial factors such as elevation, slope, land-use, topographic indices and proximity to drainage networks. These layers were combined through frequency ratio, fuzzy gamma and evidential belief function models. While the study effectively mapped high-risk zones, it relied mainly on static spatial factors and did not offer a simple indicator framework that could be updated easily with new rainfall or land-use data.

Other Mumbai-focused studies have adopted hydraulic or network-based perspectives. A hydrodynamic modelling study on the city's road network, for instance, used detailed road geometry, rainfall inputs and overland flow simulation to identify sections of the urban road system most prone to flooding [9]. Technical and planning reports linked to the BRIMSTOWARD project and subsequent updates have examined storm-water drainage capacity, outfall levels and backwater effects in detail, often recommending engineering augmentation of the network. These works offer important insight into infrastructure constraints, but they are typically infrastructure-centric and do not integrate broader hydrometeorological and land-surface indicators into a unified susceptibility framework [7], [8].

An urban watershed study in Guwahati constructed a flood inventory and a set of flood-influencing factors, then applied ensemble machine learning models such as random forest combined with metaheuristic optimization (e.g., ant search, bee colony, grey wolf optimizers) to improve prediction performance [10]–[12]. Internationally, several works have integrated digital elevation models, slope, curvature, wetness indices, vegetation indices (NDVI), land use, soil and rainfall data with statistical and machine learning methods to derive flood susceptibility maps, often reporting good performance for random forest and related ensemble models [13], [14]. A recent systematic review of urban flood susceptibility mapping emphasizes that most successful models rely on combinations of hydrometeorological variables, terrain characteristics and land-use information derived from remote sensing and GIS products [15], [16].

Taken together, this literature shows three clear patterns. First, for Mumbai, there is already evidence that combining multiple spatial indicators in GIS can produce meaningful susceptibility zonation, but existing work is largely model-focused and does not define a simple, transferable indicator framework grounded in nationally available datasets [6], [7]. Second, studies in other Indian cities such as Chennai and Guwahati demonstrate that machine learning models can effectively exploit combinations of rainfall, elevation and land-surface factors when suitable flood inventories and geospatial data are available [12], [17]. Third, recent reviews underline the importance of systematically structuring these indicators rather than treating them as ad hoc model inputs. The present work builds on these insights but takes a different emphasis: rather than developing another machine-learning model or hydraulic simulation, it sets out a simple data-driven framework that brings together rainfall, elevation and land-use indicators from reliable national and municipal sources.

Proposed Framework

The framework developed in this study organizes rainfall, elevation and land-use indicators into a structured assessment that highlights locations in Mumbai where flooding is more likely during periods of intense rainfall. Rather than building a predictive hydrodynamic or machine-learning model, the framework provides a transparent way to combine verified datasets that are consistently available for the city. Its purpose is to support anticipatory planning by identifying spatial conditions that elevate flood susceptibility under different rainfall situations.

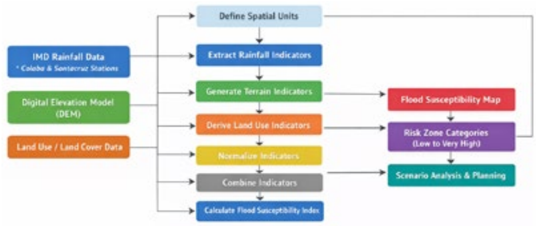

Figure 3 Overall Structure of the Framework

Rainfall Indicator Layer

The rainfall indicator layer acts as the primary forcing input in the framework. Daily observations from the Colaba and Santacruz stations are analyzed to extract numerical descriptors that characterize the distribution, concentration and extremes of the 2025 monsoon. These descriptors include cumulative seasonal rainfall, the intensity of the most severe daily events and the frequency with which IMD-defined heavy or very heavy rainfall days occurred.

Elevation and Terrain Indicator Layer

Terrain plays a major role in determining where water tends to collect. Using a digital elevation model, elevation and slope values are extracted for each area. These terrain indicator (T) values help to indicate whether an area allows water to drain away easily or whether it is more likely to hold water. Mumbai's landscape makes this step important because many parts of the city have very little slope and are also affected by tide levels which slow down drainage.

The terrain values are normalized. Lower areas and places with gentle slopes are given higher values, since they are more prone to waterlogging during strong rainfall.

Land-use and Surface Indicator Layer

The structure and utilization of land largely influence the behavior of rainwater. Using land-cover datasets, each area is assessed for proportion of built-up surfaces, paved areas, vegetation and open areas. Based on this information, the degree of imperviousness is estimated. Areas that are mostly concrete and asphalt allow very little water to soak in the ground, while lands with vegetation and open ground allow more water to infiltrate.

These land-use summaries are converted into values that indicate the ability of an area to absorb rainwater. Built-up regions received higher susceptibility scores because they tend to speed up runoff and reduce the time available for drainage.

Combined Indicator Assessment

Rainfall (R), terrain (T) and land-use (L) indicators are normalized (0–1 range) for uniformity. The Flood Susceptibility Index (FSI) is then calculated for each spatial unit using equation (1) as shown below-

$$FSI = 1/3\ (R + T + L). \qquad (1)$$

FSI value is used for comparison across different areas or wards and forms the basis for mapping susceptibility and examining shifts under stronger rainfall conditions.

This combined view is useful because it highlights areas where several risk factors overlap. For example, a place with only moderate rainfall might still become a concern if it sits at a lower height and has many built-up surfaces. On the other hand, an area with heavy rain but better slope and more open ground may show lower susceptibility. This type of combined behavior is often seen in real flood situations and helps give the framework practical meaning.

Figure 3 presents the overall structure of the framework. It brings together rainfall characteristics, terrain information and land-use patterns through a set of simple indicator steps that can be applied across the city. By placing these indicators on a common scale, the framework helps interpret how different factors combine to shape local flood susceptibility under typical as well as intense monsoon conditions.

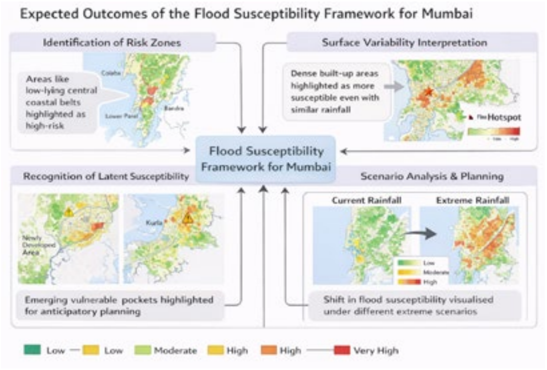

Figure 4 Expected Outcomes of the Proposed Data-Driven Flood Susceptibility Framework for Mumbai

The design is intentionally flexible. Each dataset can be replaced or updated as newer IMD rainfall records, revised DEMs or updated land-use layers become available. This allows the framework to be refreshed for each monsoon season and supports the exploration of extreme-rainfall scenarios without altering the underlying approach.

Expected Outcomes and Discussions

The structure allows clear inference about the types of patterns and outputs that would emerge when the indicators are combined.

The proposed framework is expected to produce a clear picture of how flood susceptibility varies across Mumbai. When the rainfall indicators are assigned to different parts of the city, some areas stand out immediately. Localities that receive intense spells and lie on lower or flatter terrain are likely to appear more vulnerable, which matches long-observed patterns in the central low-lying belt and several coastal stretches.

A second outcome is the ability to distinguish areas that get similar rainfall but respond very differently. Areas having built-up and paved surfaces tend to convert rain directly into runoff, while places with more vegetated land or open areas show lower susceptibility under the same conditions. This helps to explain why adjacent wards show very different levels of waterlogging.

The proposed method can also explore new or emerging risk zones. Recently developed areas, reclaimed lands and small depressions are often overlooked in existing approaches. These points can be more sensitive to intense rainfall. Highlighting such locations can support ward-level planning and early preparedness.

Another useful outcome comes from different scenarios of testing. By adjusting rainfall inputs to represent extreme one-day events or prolonged heavy spells, the proposed approach can show how susceptibility patterns might shift if the city experiences stronger monsoon behavior. This will help authorities to think beyond past events and prepare for conditions that are becoming more common.

The proposed framework is very practical to use. The datasets involved are easy to process and the indicator extraction steps are straightforward. Once the inputs are prepared, the susceptibility index can be calculated quickly. While the study does not quantify accuracy, the relationships reflected in the indicators are well established. Intense rainfall, low terrain and hard surfaces consistently relate to higher flood occurrence. This makes the proposed method a reliable early-screening tool even without detailed modelling.

In summary, the framework shows how rainfall, terrain and land-use information can be viewed together to form a realistic picture of flood susceptibility in Mumbai. Its strength lies in its simplicity, use of easily available datasets and ability to be updated for each monsoon season. Figure 4 illustrates how these indicators help identify highly vulnerable areas, new risk pockets and possible changes under different rainfall scenarios. Maps shown in the figure are schematic illustrations created for conceptual representation and do not depict actual GIS spatial outputs.

Conclusion

This study uses three basic datasets that are easily available: rainfall records, elevation data and land-use information. When these datasets are viewed together, they give a clearer idea of where Mumbai is likely to face flooding. Areas that often flood usually have a similar mix of conditions. They receive strong or repeated monsoon showers, lie on lower ground and have many hard, built-up surfaces. The proposed method helps to show these patterns and explains why some places flood more than others.

As the datasets are updated every year, the algorithm can be refreshed without much effort. It also allows authorities to test how flood-prone areas might change if the city receives heavier or more uneven rainfall in the future. As monsoon behavior in Mumbai continues to change, this kind of testing scenario becomes important. If additional information such as drainage capacity or real-time rainfall is added later, the proposed approach can become more detailed.

In its current form, the framework gives municipal agencies an easy way to read flood vulnerability using information they already have. It encourages a more connected understanding of the city's terrain, surfaces and rainfall patterns. It can also support better planning and preparedness as Mumbai's monsoon continues to evolve.

References

1. National Disaster Management Authority. https://ndma.gov.in/Natural-Hazards/Urban-Floods (accessed Dec. 12, 2025).
2. National Disaster Management Guidelines.
3. India Meteorological Department, Ministry of Earth Sciences. https://mausam.imd.gov.in/ (accessed Dec. 11, 2025).
4. Data.gov.in. Rainfall data, Mumbai city, Maharashtra, 01 June 2025–13 August 2025. https://www.data.gov.in/resource/rainfall-data-mumbai-city-maharashtra-01st-june-2025-13-august-2025 (accessed Dec. 12, 2025).
5. Mumbai Climate Action Plan. https://mcap.mcgm.gov.in/urban-flooding-and-water-resource-management/ (accessed Dec. 12, 2025).
6. V. Ramesh and S. S. Iqbal, Urban flood susceptibility zonation mapping using evidential belief function,

frequency ratio and fuzzy gamma operator models in GIS: a case study of Greater Mumbai, Maharashtra, India. Geocarto International, 37(2), 581–606. Jan. 2022. doi: 10.1080/10106049.2020.1730448.

7. K. Gupta, Urban floods: case study of Mumbai. Disaster Development, 3(2), 1–14. 2009.
8. S. Bhattacharjee, Storm water drainage in Mumbai, pp. 147–165. 2014.
9. TERI, Identification of flood risk on urban road network using hydrodynamic model: case study of Mumbai floods. New Delhi: Energy Resources Institute, p. 105. 2016.
10. S.-D. Jang, J.-H. Yoo, Y.-S. Lee, and B. Kim, Flood prediction in urban areas based on machine learning considering the statistical characteristics of rainfall. Progress in Disaster Science, 26, 100415. 2025. doi: 10.1016/j.pdisas.2025.100415.
11. Z. Fan, J. Zhang, Y. Chen, and H. Xu, Urban flood prediction model based on Transformer-LSTM-Sparrow Search Algorithm. 2025.
12. I. A. Ahmed, S. Talukdar, A. Parvez, M. Rihan, M. R. I. Baig, and A. Rahman, Flood susceptibility modeling in the urban watershed of Guwahati using improved metaheuristic-based ensemble machine learning algorithms. Geocarto International, 37(26), 12238–12266. Dec. 2022. doi: 10.1080/10106049.2022.2066200.
13. R. Al-Ruzouq, A. Shanableh, R. Jena, M. B. A. Gibril, N. A. Hammouri, and F. Lamghari, Flood susceptibility mapping using a novel integration of multi-temporal Sentinel-1 data and eXtreme deep learning model. Geoscience Frontiers, 15(3), 101780. 2024. doi: 10.1016/j.gsf.2024.101780.
14. M. Feizbahr, N. Brake, H. Arbabkhah, H. Hariri Asli, and K. Woods, Flood susceptibility mapping using machine learning and geospatial-Sentinel-1 SAR integration for enhanced early warning systems. Remote Sensing, 17(20), 3471. 2025. doi: 10.3390/rs17203471.
15. T. Islam, E. B. Zeleke, M. Afroz, and A. M. Melesse, A systematic review of urban flood susceptibility mapping: remote sensing, machine learning, and other modeling approaches. Remote Sensing, 17(3), 524. 2025. doi: 10.3390/rs17030524.
16. J. Sanyal and X. X. Lu, Application of remote sensing in flood management with special reference to monsoon Asia: a review. Natural Hazards, 33(2), 283–301. 2004. doi: 10.1023/B:NHAZ.0000037035.65105.95.
17. K. R. Deepak and P. Sabu, Urban flood susceptibility map of Chennai — GIS and random forest method. International Research Journal of Engineering and Technology, 09(10), 545–557. 2022.