



Delft University of Technology

Document Version

Final published version

Licence

CC BY

Citation (APA)

Colos, M. M., Marote, L. C. D., Coelho, G. P., Abreu, A. E. S. D., Pande, S., Han, S., & Lucas, M. C. (2026). Revealing spatial inequalities in flood exposure across São Paulo State, Brazil. *Frontiers in Water*, 8, Article 1857531. <https://doi.org/10.3389/frwa.2026.1857531>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership. Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.



OPEN ACCESS

EDITED BY

Marlies Barendrecht,
King's College London, United Kingdom

REVIEWED BY

Elton Almeida,
São Paulo State University, Brazil
Richard De Assis Moreira,
Federal University of Minas Gerais, Brazil

*CORRESPONDENCE

Saket Pande
✉ S.Pande@tudelft.nl

RECEIVED 16 April 2026

REVISED 10 July 2026

ACCEPTED 17 July 2026

PUBLISHED 31 August 2026

CITATION

Colos MM, Marote LCD, Coelho GP, Abreu AES, Pande S, Han S and Lucas MC (2026) Revealing spatial inequalities in flood exposure across São Paulo State, Brazil. *Front. Water* 8:1857531. doi: 10.3389/frwa.2026.1857531

COPYRIGHT

© 2026 Colos, Marote, Coelho, Abreu, Pande, Han and Lucas. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Revealing spatial inequalities in flood exposure across São Paulo State, Brazil

Manuel Mendoza Colos¹, Luca Calçada Dolim Marote¹,
Guilherme Palermo Coelho¹, Ana Elisa Silva de Abreu²,
Saket Pande^{3*}, Shasha Han⁴ and Murilo Cesar Lucas¹

¹School of Technology, Universidade Estadual de Campinas, Limeira, São Paulo State, Brazil, ²Institute of Geosciences, Universidade Estadual de Campinas, Campinas, São Paulo State, Brazil, ³Department of Water Management, Delft University of Technology, Delft, Netherlands, ⁴School of Geography, Earth and Environmental Sciences and Birmingham Institute for Sustainability and Climate Action, University of Birmingham, Birmingham, United Kingdom

Floods remain among the most frequent and destructive natural hazards, requiring spatially explicit tools to support risk-reduction planning and emergency preparedness. This study develops an integrated flood susceptibility and exposure assessment for São Paulo State, Brazil, combining machine learning, explainable artificial intelligence, and census-based demographic information. Flood Susceptibility Maps (FSMs) were generated at 30 m spatial resolution using the Random Forest (RF) algorithm, while the contribution of each conditioning factor was interpreted using SHapley Additive exPlanations (SHAP). Human exposure was assessed by integrating the susceptibility map with the spatial granularity of the 2022 Brazilian Institute of Geography and Statistics (IBGE) Census. Based on 15 conditioning factors and a balanced dataset of 2,000 flood-prone and 2,000 non-flood-prone samples, the RF model achieved high predictive performance (accuracy = 0.939; AUC = 0.982), with river proximity, slope, Curve Number, and altitude emerging as the most influential conditioning factors. Integrating susceptibility patterns with census data revealed marked spatial inequalities in flood exposure: the 6.2% of the state territory classified as Very High flood susceptibility concentrated 22.3% of the population, corresponding to approximately 9.89 million people, and 23.5% of all housing units. The municipal-level analysis further showed that 24.6% of people younger than 17 years and 26.8% of people older than 65 years were located in municipalities classified as High exposure. To support exposure-informed planning, a Combined Prioritization Index (CPI) was developed by integrating dominant municipal susceptibility, relative population exposure, and absolute population exposure. The CPI identified higher-priority municipalities not only along the coastal region but also in inland areas, including Guaruj, Rio Claro, Santos, Itanham, Praia Grande, Cubatão, São Vicente, and Caraguatatuba. Overall, the findings reveal spatial inequalities in flood exposure by showing where elevated flood susceptibility overlaps with concentrated human exposure, providing evidence for more transparent and anticipatory flood-risk planning.

KEYWORDS

flood susceptibility, machine learning, population exposure, spatial inequalities, urban flood resilience, water security

1 Introduction

Globally, floods are among the most frequent and destructive natural hazards, imposing serious consequences on infrastructure, economic stability, and human security (Waleed and Sajjad, 2025a; Yang et al., 2025). More than 1.81 billion people, approximately 23% of the world's population, currently live in areas exposed to significant flood hazard, and this number may reach 2.3 billion by 2050 (Rentschler et al., 2022). This exposure is unevenly distributed from a socioeconomic perspective. An estimated 89% of people exposed to flooding live in low- and middle-income countries, indicating that flood exposure is closely linked to broader patterns of inequality, urbanization, and limited access to safer land (Rentschler et al., 2022; Paszkowski et al., 2025).

Flood susceptibility reflects the spatial propensity of an area to experience flooding based on the combined influence of environmental, topographic, hydrological, and anthropogenic conditioning factors (Bui et al., 2020). In urban regions, susceptibility is shaped not only by natural terrain and hydrological conditions, but also by land-use change, surface sealing, and the occupation of flood-prone areas (Amiri et al., 2024; El baida et al., 2024). Rapid urban expansion, inadequate land-use planning, and increasing surface impermeability can intensify runoff generation and place growing pressure on drainage infrastructure and disaster-risk management systems (Huu Duy et al., 2024; Li et al., 2026; Zhu et al., 2024). These processes are particularly relevant in tropical and rapidly urbanizing regions, where high-intensity rainfall events may interact with dense occupation and limited planning capacity (Knighton et al., 2020; Bui et al., 2020).

The consequences of flooding, however, cannot be understood through physical susceptibility alone. Flood-prone areas often overlap with densely occupied riverbanks, floodplains, coastal zones, and informal or rapidly expanding urban settlements (Zhao et al., 2024). In these settings, exposure is shaped by settlement patterns, population density, housing concentration, and the presence of demographic groups that may face greater difficulty during evacuation and emergency response. Children, older adults, and persons with disabilities are frequently recognized as socially sensitive groups in disaster-risk studies because they may experience greater mobility constraints and dependence on assistance during flood events (Khaldi et al., 2025; Khan et al., 2025). Therefore, integrating susceptibility information with demographic data is essential for identifying where flood-prone conditions become socially relevant.

Although flood susceptibility maps are widely used to describe the spatial distribution of flood-prone areas, their planning value increases when they are combined with information on human exposure. Physical susceptibility models identify where flooding is more likely to occur under favorable environmental and anthropogenic conditions, but they do not indicate how many people or housing units are located in those areas. Integrating susceptibility maps with fine-scale demographic data allows the identification of locations where elevated susceptibility coincides with concentrated population and residential exposure (Knighton et al., 2020; Goto et al., 2022; Khaldi et al., 2025). By revealing where elevated flood susceptibility overlaps with concentrated population and housing exposure, this integration

supports the identification of spatial inequalities in flood exposure and provides evidence for more transparent prioritization in flood-risk planning (Paszkowski et al., 2025).

Recent studies have increasingly emphasized the importance of combining flood mapping with population exposure assessments (Rentschler et al., 2022; Waleed and Sajjad, 2025a). However, many analyses remain focused on national-scale estimates, individual cities, or the physical characterization of flood-prone areas. A gap remains for statewide, fine-resolution assessments that combine machine learning-based flood susceptibility mapping with census-tract-level demographic information and municipal-level prioritization. This gap is particularly relevant in regions such as São Paulo State, Brazil, where dense urbanization, coastal and riverine occupation, and strong socioeconomic contrasts create heterogeneous patterns of flood exposure.

To address this gap, flood susceptibility assessments should be integrated with high-resolution demographic data to quantify human exposure alongside physical susceptibility. Overlaying flood susceptibility maps with census-tract-level information can reveal where elevated susceptibility coincides with concentrated populations and housing units (Knighton et al., 2020). In addition, municipal-level demographic information can support the identification of age-sensitive population groups, particularly children and older adults, in areas where exposure to flood susceptibility is higher (Tran et al., 2024). Such information can support emergency preparedness, early warning strategies, land-use planning, and the prioritization of municipalities where susceptibility and exposure indicators converge (Waleed and Sajjad, 2025a).

This study has three objectives: (1) to develop a 30 m flood susceptibility model for São Paulo State, Brazil; (2) to interpret the relative importance and directional influence of conditioning factors using SHAP-based explainable artificial intelligence; and (3) to quantify population and housing exposure at the census-tract level, assess municipal-level exposure of age-sensitive demographic groups, and develop a municipal-level prioritization index based on flood susceptibility and relative and absolute population exposure.

To the best of our knowledge, this is the first statewide assessment in São Paulo to integrate machine learning-based flood susceptibility mapping with census-tract-level demographic exposure and municipal prioritization, explicitly identifying spatial inequalities in flood exposure where elevated flood susceptibility overlaps with population concentration and age-sensitive demographic groups.

2 Materials and methods

2.1 Study area

The study area is the State of São Paulo, located in southeastern Brazil, it covers an area of 248,219 km² and comprises 645 municipalities with a population exceeding 44 million inhabitants (Instituto Brasileiro de Geografia e Estatística (IBGE), 2022). It extends geographically between latitudes 19°46' S and 25°18' S and longitudes 44°10' W and 53°06' W, encompassing

distinct coastal, plateau, and inland regions (Figure 1). São Paulo State exhibits pronounced topographic variability, with elevations ranging from sea level to over 2,800 m. This elevation gradient directly modulates local hydrological processes by governing flow velocities, drainage network configurations, and water accumulation zones. Consequently, low-lying coastal plains and alluvial valleys systematically concentrate surface runoff, significantly elevating flood susceptibility where expanding urban infrastructure intersects natural drainage corridors. The study area is predominantly characterized by a humid subtropical regime (Cwa in the Köppen classification) (Alvares et al., 2013).

Given this high economic vitality and demographic concentration, São Paulo State serves as a critical area for multi-scale flood susceptibility and exposure modeling. Hence, to evaluate the interactions between flood susceptibility and population exposure, this study operates across two nested spatial scales:

- Macro-scale (State level): Flood susceptibility mapping is conducted across the entire administrative boundary of São Paulo State. This scale captures broad spatial variations in topography, land use, hydrography, and anthropogenic influences that shape flood susceptibility across metropolitan, coastal, and inland regions; and
- Micro-scale (census-tract level): Population exposure analyses are conducted at the census-tract level using the spatial boundaries of the 2022 IBGE Census (Instituto Brasileiro de Geografia e Estatística (IBGE), 2022). This scale supports the assessment of four complementary dimensions: (1) the spatial distribution of census tracts intersecting areas of elevated flood susceptibility; (2) population exposure, including age-sensitive groups such as children and older adults; (3) the area of land classified as highly susceptible within each census tract; and (4) housing exposure, measured as the number of private dwellings located within susceptible areas.

2.2 Methodology

Grounded in a spatial inequalities in flood exposure perspective, the methodological framework integrates a hydro-demographic framework that evaluates the intersection of physical flood susceptibility and human exposure across four sequential stages: (1) data pre-processing, (2) processing, (3) results analysis, and (4) application of results (Figure 2).

In the data pre-processing stage, the analytical baseline was established by compiling, harmonizing, and structuring 15 geo-environmental and anthropogenic conditioning variables (Figure 3). To reduce conditioning factor (predictor) redundancy and improve model interpretability, these spatial conditioning factors were subjected to multi-collinearity diagnostic via a Variance Inflation Factor (VIF) and correlation matrix analysis. Concurrently, a comprehensive authoritative flood-prone inventory was assembled. This inventory, comprising documented authoritative flood presence (occurrence) and absence (non-occurrence) coordinates,

was partitioned into independent datasets: 70% for model training and 30% for testing.

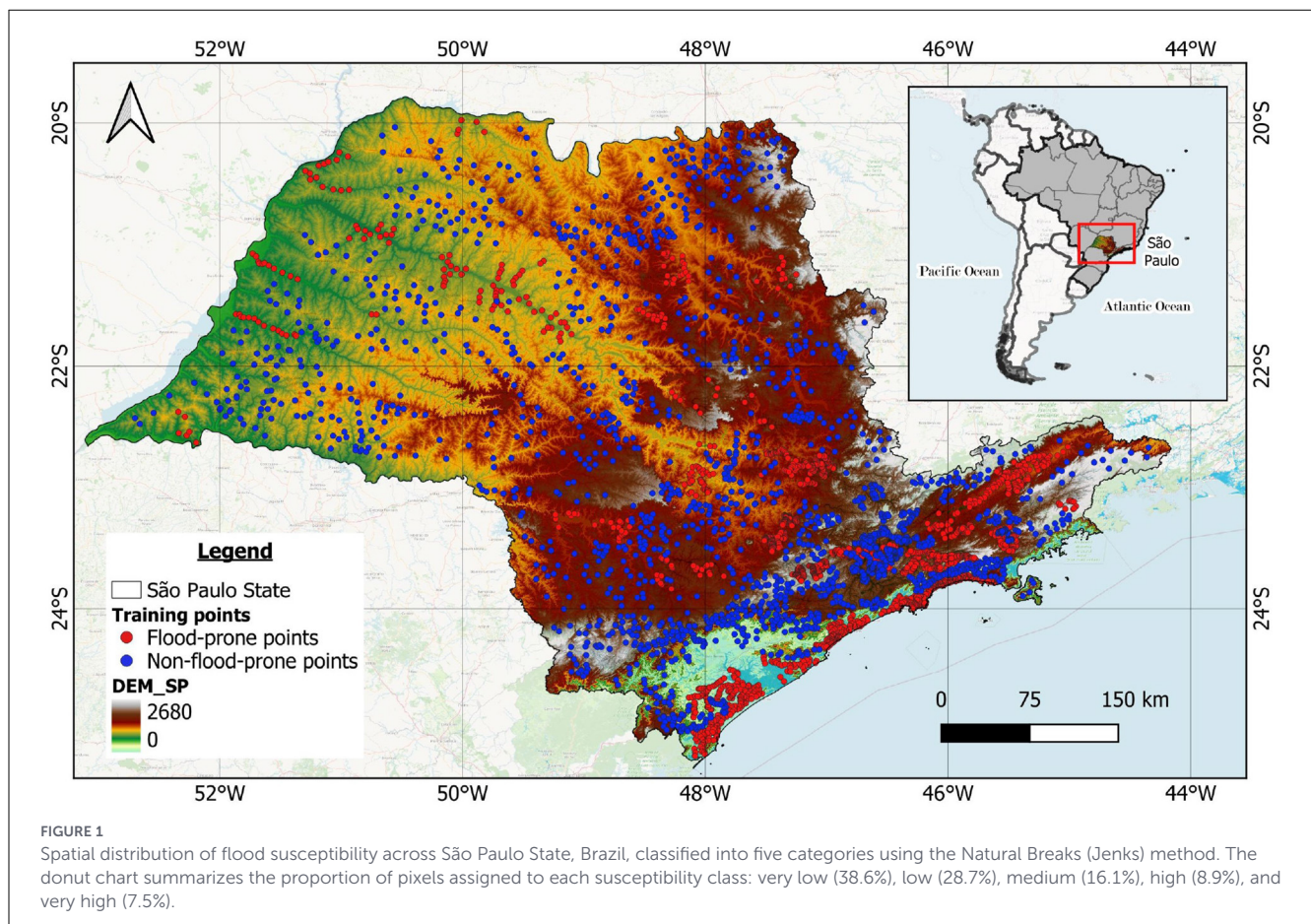
Flood susceptibility was modeled using the Random Forest (RF) algorithm. The model was trained using the prepared dataset and optimized through hyperparameter tuning to maximize predictive performance. Model performance was evaluated using the Receiver Operating Characteristic–Area Under the Curve (ROC–AUC), accuracy, recall, and F1-score. The resulting susceptibility map was then integrated with 2022 IBGE Census data (Instituto Brasileiro de Geografia e Estatística (IBGE), 2022) to quantify population and housing exposure across susceptibility classes. This integration also supported the municipal-level assessment of age-sensitive demographic exposure and the development of the Combined Prioritization Index (CPI), which identifies municipalities where elevated flood susceptibility coincides with greater population exposure.

This integrated approach extends conventional flood susceptibility mapping by incorporating spatially explicit population and housing exposure assessments. The resulting framework provides a quantitative basis for identifying and prioritizing locations where elevated flood susceptibility coincides with greater concentrations of exposed populations, particularly among age-sensitive demographic groups. By integrating physical susceptibility and demographic exposure, the approach supports a spatial inequality perspective in flood-risk planning while maintaining a transparent and reproducible analytical structure. Although developed for São Paulo State, the analytical logic may be adapted to other regions where comparable susceptibility, geospatial, and demographic datasets are available, supporting the incorporation of socio-environmental considerations into disaster planning and risk-reduction strategies.

2.2.1 Authoritative flood-prone inventory

Authoritative flood-prone records constitute the empirical foundation of the susceptibility model, enabling the characterization of spatial relationships between flood-prone areas and environmental conditioning factors (Arabameri et al., 2022; Khosravi et al., 2019a). Flood-prone area polygons were obtained from the Brazilian Geological Service (SGB) through the GeoSGB portal (<https://geoportal.sgb.gov.br/geosgb/>). These datasets were produced by the Companhia de Pesquisa de Recursos Minerais (CPRM) through technical mapping, modeling procedures, and field validation conducted between 2012 and 2019 at the national scale. The polygons delineate areas recognized as susceptible to flooding and were used as the reference dataset for generating point samples (Serviço Geológico do Brasil, SGB-CPRM).

To construct the presence class, 2,000 flood-prone points (class = 1) were randomly sampled within mapped flood-prone areas. The sampled points were predominantly located in floodplains and low-slope urbanized zones, where terrain morphology favors runoff accumulation and soil water saturation (Debnath et al., 2024; Zhao et al., 2024). To construct the absence class, 2,000 non-flood points (class = 0) were generated using restrictive criteria (Slope > 5%, and NDVI $\geq$ 0.5). These thresholds were adopted to reduce the likelihood of selecting locations potentially affected



by flooding, minimize ambiguous transition zones, and improve class separability between areas with contrasting flood susceptibility characteristics (Bui et al., 2023; Khosravi et al., 2019a).

2.2.2 Flood conditioning factors influencing susceptibility

Flood susceptibility reflects the interaction of topographic, hydrological, climatic, geological, and anthropogenic controls. Given the high sensitivity of ML models to input structure, the conditioning factors were carefully selected to ensure statistical reliability and physical coherence (Amiri et al., 2024; Arabameri et al., 2022; Khosravi et al., 2019a).

The conditioning factors were compiled from multiple sources and spatial scales (Supplementary Table S1), guided by recent literature and regional data availability. The variables were selected based on their relevance to runoff generation, flow concentration, and water accumulation processes.

Key topographic variables were derived from the 30 m DEM, including altitude, slope, and aspect. These topographic features influence flow direction and runoff velocity, with lower elevations and gentler slopes generally favoring water accumulation, particularly in floodplains and urbanized areas (Huu Duy et al., 2024; Saravanan and Abijith, 2022). Additionally, plan curvature and profile curvature were included to distinguish

concave from convex terrain, thereby representing conditions of flow convergence and runoff dispersion (Bui et al., 2020).

The Topographic Wetness Index (TWI) was selected to represent soil moisture potential and identify zones of water accumulation and soil saturation (Beven and Kirkby, 1979), as shown in Equation 1:

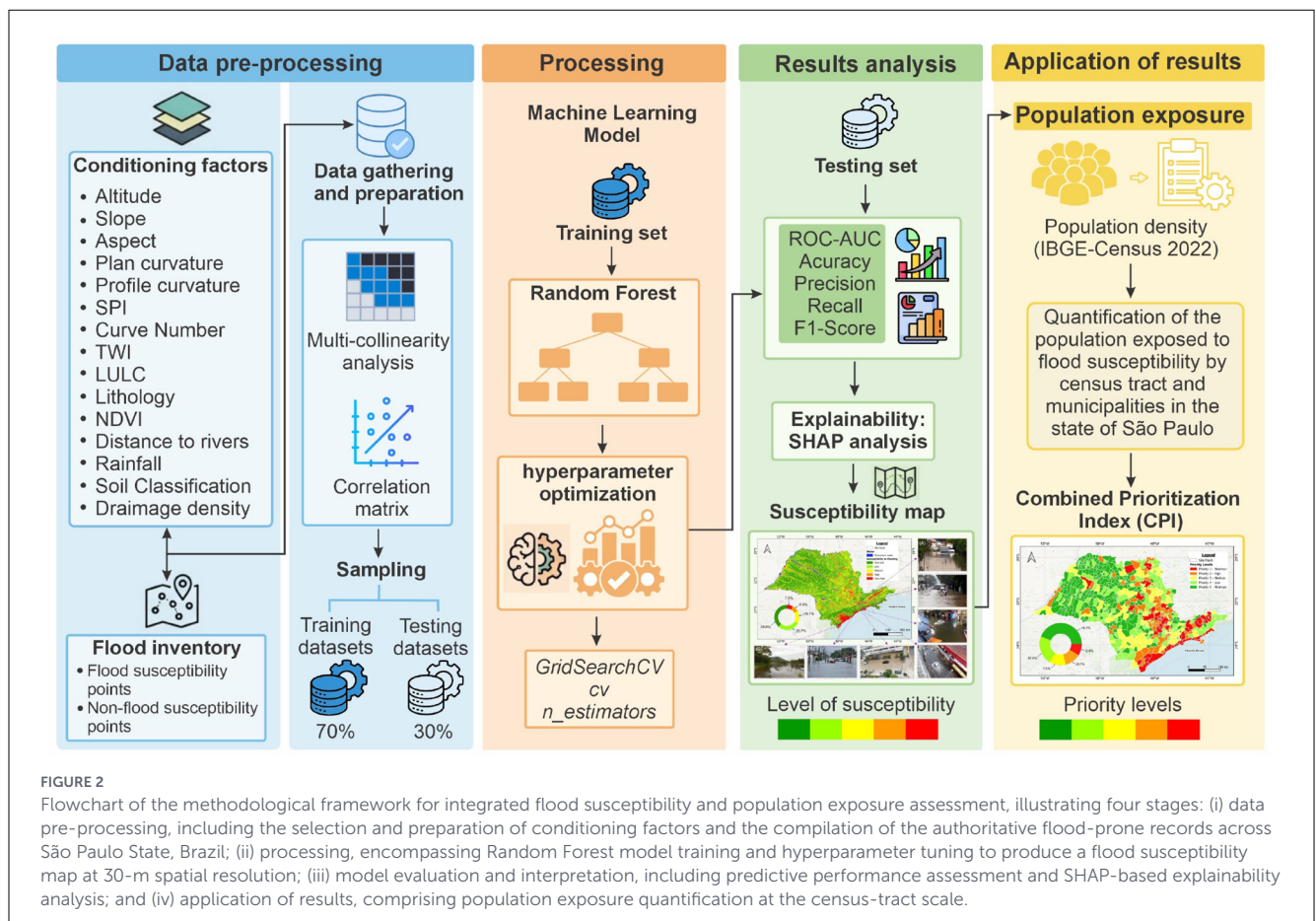
$$TWI = \ln \left(\frac{A_s}{\tan \theta} \right) \quad (1)$$

where A_s denotes the upslope contributing area per unit contour width and θ represents the slope angle in radians.

Furthermore, the Stream Power Index (SPI) was incorporated to capture the erosive potential of flow and its relationship to fluvial dynamics and capture the influence of flow accumulation and erosive power on hydrological processes (Beven and Kirkby, 1979), as shown in Equation 2:

$$SPI = A_s \cdot \tan \theta \quad (2)$$

Areas closer to drainage channels generally exhibit greater flood probability due to overflow and flow concentration (Khosravi et al., 2019a; Zhao et al., 2024). Hence, the hydrological proximity effect was represented by the distance to the river network, and drainage density was included as an indicator of terrain dissection and runoff efficiency; higher values generally correspond to increased susceptibility (Hussain et al., 2025).



The climatic forcing was characterized using mean accumulated precipitation derived from the CHIRPS dataset and processed in Google Earth Engine (GEE) (Paredes-Trejo et al., 2017; de Oliveira-Júnior et al., 2021). Precipitation constitutes the primary trigger of floods in tropical and subtropical systems, particularly during intense and concentrated rainfall events (Abdo et al., 2025; Waleed and Sajjad, 2025a).

Areas with sparse vegetation typically show higher flood susceptibility (Amiri et al., 2024). To capture vegetation density influence in water infiltration and surface retention, the Normalized Difference Vegetation Index (NDVI) was computed as shown in Equation 3:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (3)$$

where NIR and Red correspond to near-infrared and red spectral reflectance, respectively.

Finally, lithological data from the Environmental Data and Information Bank (BDiA) and land use and land cover (LULC) data from the MapBiomas Brasil project (MapBiomas Brasil, 2024). These variables influence soil permeability, storage capacity, and hydrological response, and thus play an important role in flood generation in both natural and urban systems (Ahmad et al., 2025).

All thematic layers were reprojected to a common coordinate reference system and resampled to a uniform spatial resolution of 30 m in raster format to ensure spatial consistency and comparability across variables (Figure 3). We then extracted the

values of each conditioning factor at the locations of presence (1) and non-flood (0) points (Figure 4). This procedure enabled us to construct the structured dataset used as input for the RF model.

2.2.3 Multi-collinearity analysis and explanatory variable selection

Multi-collinearity assessment was conducted to identify potential redundancy among explanatory variables (i.e., conditioning factors) and to facilitate model interpretation. Strong correlations among explanatory variables can affect the stability of variable importance measures and complicate the interpretation of conditioning factor contributions (Abdo et al., 2025; Chakraborty et al., 2022). The correlation matrix of the conditioning factors is presented in Supplementary Material S1. Multi-collinearity arises when two or more explanatory variables exhibit high linear association, which can compromise predictive performance and hinder interpretability (Arabameri et al., 2022).

Multi-collinearity was evaluated among the conditioning factors using the Tolerance index (TOL) and the VIF, both widely applied in machine learning and spatial hazard modeling (Arabameri et al., 2022). The TOL is defined as Khosravi et al. (2019a) in Equation 4:

$$TOL_j = 1 - R_j^2 \quad (4)$$

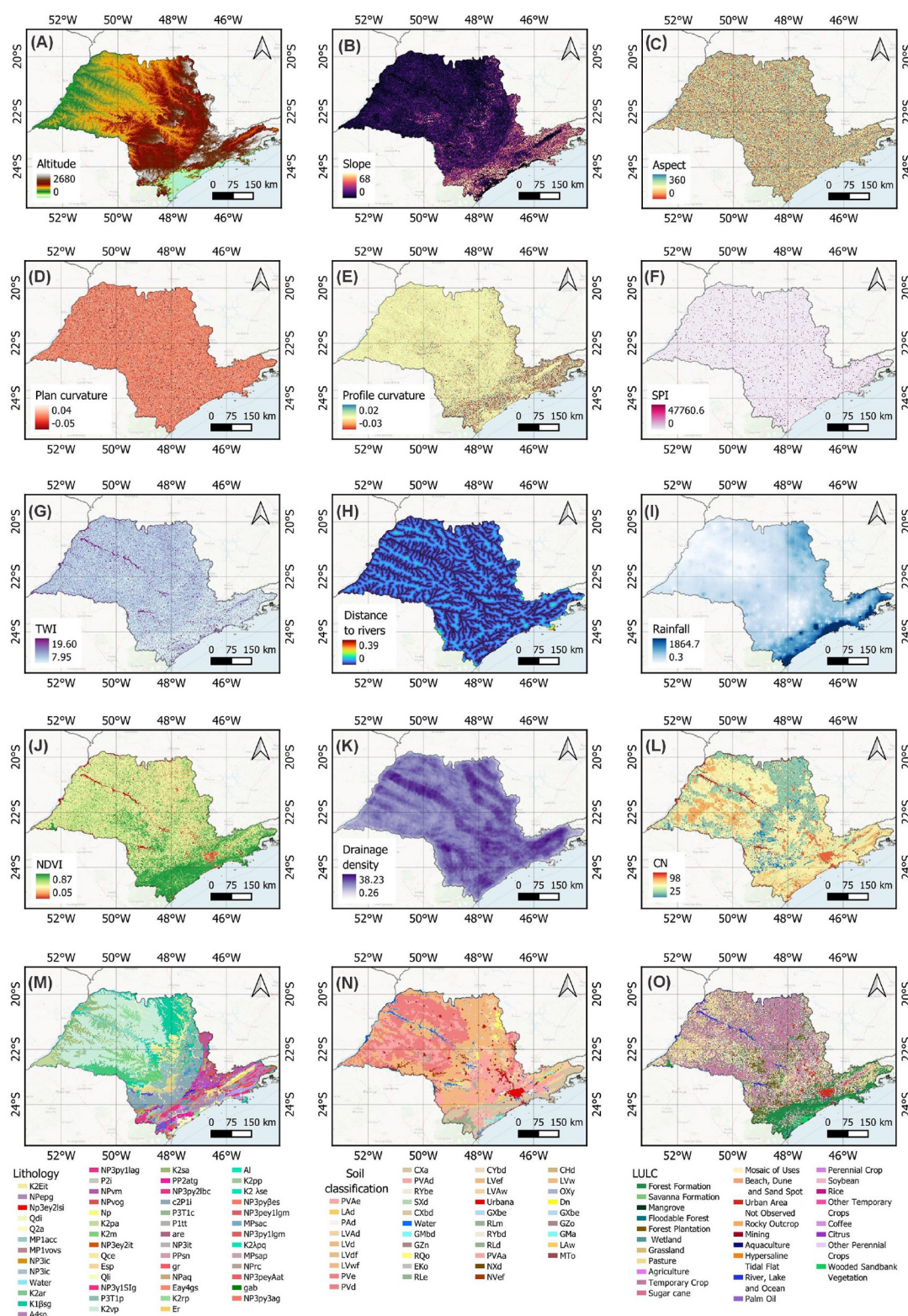
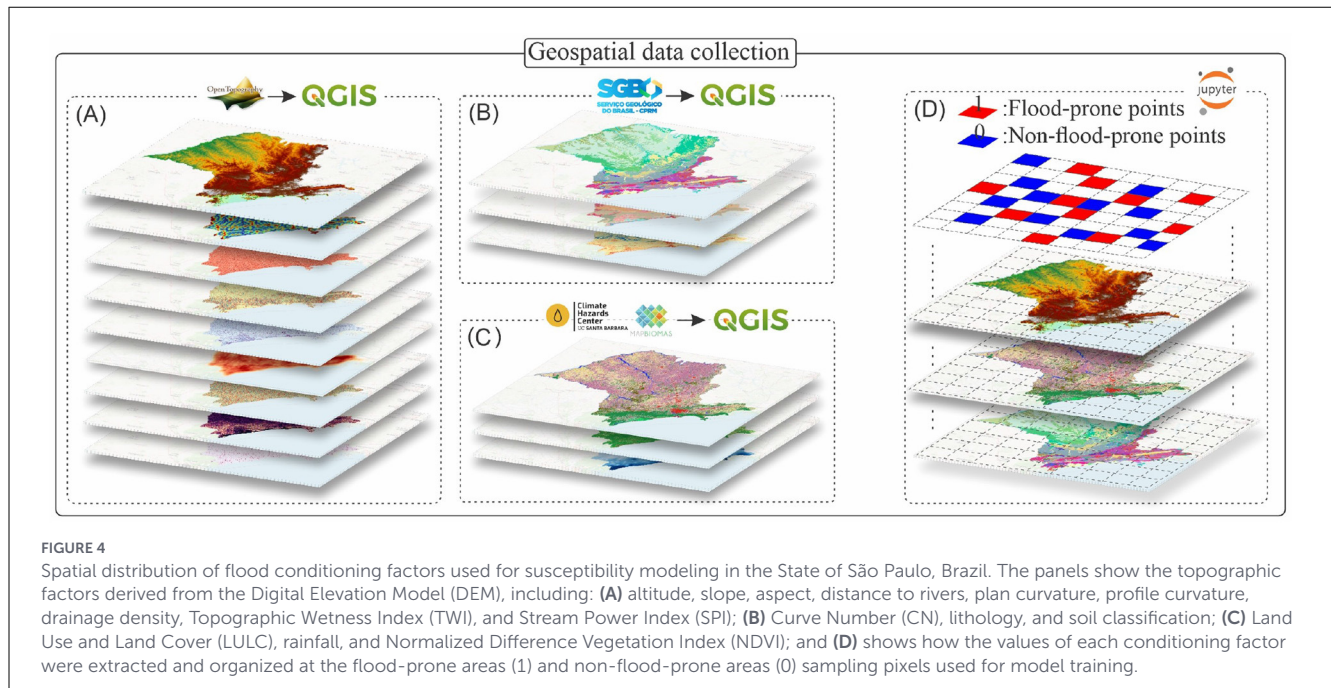


FIGURE 3

Spatial distribution of flood conditioning factors used for susceptibility modeling in the State of São Paulo, Brazil. (A) altitude; (B) slope; (C) aspect; (D) plan curvature; (E) profile curvature; (F) SPI; (G) TWI; (H) distance to rivers; (I) rainfall; (J) NDVI; (K) drainage density; (L) CN; (M) lithology; (N) soil classification; and (O) LULC.



and the VIF as [Khosravi et al. \(2019a\)](#) in Equation 5:

$$VIF_j = \frac{1}{TOL_j} \quad (5)$$

where R_j^2 represents the coefficient of determination obtained by regressing predictor j against all remaining explanatory variables in the dataset.

Following established thresholds in the literature, VIF values greater than 5 and TOL values lower than 0.1 were adopted as indicative of problematic multi-collinearity, requiring removal of the corresponding variable ([Abdo et al., 2025](#); [Ahmad et al., 2025](#)). All predictors satisfied these criteria ([Supplementary Table S2](#)) and were retained for subsequent modeling. This procedure confirmed the absence of problematic multi-collinearity among the selected predictors and supported the interpretability and robustness of the subsequent flood susceptibility modeling.

2.2.4 Random forest modeling and explainable artificial intelligence

The RF model was selected due to its strong performance in flood susceptibility assessment and its capacity to capture nonlinear interactions among multiple conditioning factors in complex environments ([Leggesse et al., 2024](#)). RF is generally more robust to overfitting than many alternative machine learning algorithms when applied to high-dimensional datasets and heterogeneous spatial contexts, and it has frequently shown improved predictive performance compared with traditional statistical approaches in flood susceptibility applications ([Breiman, 2001](#); [Kaur et al., 2025](#); [Yang et al., 2025](#)).

To optimize predictive performance and ensure statewide applicability across São Paulo State, hyperparameter tuning was

performed using Grid Search with 10-fold cross-validation on the training subset. It is important to mention that 324 parameter combinations were systematically evaluated, testing the following ranges: `n_estimators`: [50, 100, 200], `max_depth`: [None, 10, 20, 30], `min_samples_split`: [2, 5, 10], `min_samples_leaf`: [1, 2, 4], and `max_features`: ['sqrt', 'log2', None]. The optimal configuration as 200 trees (`n_estimators` = 200) with a maximum depth of 20 levels (`max_depth` = 20). This structure was selected to balance model complexity and generalization capacity, reducing the risk of overfitting while maintaining predictive performance on unseen data.

To further stabilize the ensemble, a minimum of 5 samples was required to split an internal node (`min_samples_split` = 5), and the square root of the total number of features was adopted as the feature selection criterion at each split (`max_features` = 'sqrt'), following [Breiman \(2001\)](#); [Huu Duy et al. \(2024\)](#). This strategy effectively reduces correlation among individual trees and improves ensemble robustness.

The dependent variable was coded as binary (1 = flood-prone; 0 = non-flood-prone). Conditioning factors were selected after performing correlation analysis and calculating the VIF to mitigate multi-collinearity, following [Bui et al. \(2020\)](#). This procedure allows the integration of morphometric, hydrological, climatic, and anthropogenic (e.g., LULC and CN) variables into a unified predictive framework without imposing parametric assumptions regarding data distribution ([Knighton et al., 2020](#)). A balanced dataset (1:1 ratio) eliminates potential class imbalance bias and supports stable model training ([Chen et al., 2019](#)). The flood-prone inventory was subsequently partitioned into training (70%) and testing (30%) subsets, ensuring an independent evaluation of predictive performance, as recommended by [Alwathaf et al. \(2025\)](#).

Following [Aydin and Iban \(2023\)](#), to improve model interpretability, SHAP analysis was applied. Grounded in game theory, SHAP quantifies both the relative importance of each

variable and the direction of its contribution to model predictions. This interpretative layer links statistical output to physical flood-generating processes and facilitates the interpretation of the environmental and anthropogenic drivers underlying flood susceptibility patterns (Waleed and Sajjad, 2025a).

2.2.5 Model validation

The predictive performance of the RF model was evaluated to assess the reliability of the generated flood susceptibility map. Model performance was assessed using statistical metrics commonly employed in binary classification and flood susceptibility studies (Ahmed et al., 2022). The trained model was applied to the independent testing subset (30% of the dataset) to evaluate its predictive capability.

The ROC curve and the corresponding AUC were used to assess the model's discrimination capacity. AUC values approaching 1 indicate excellent predictive performance, whereas values near 0.5 indicate classification performance comparable to random assignment (Alwathaf et al., 2025).

In addition, overall classification Accuracy, Recall, Precision, and F1-score were calculated to provide a comprehensive evaluation of model performance. Accuracy measures the proportion of correctly classified samples, recall quantifies the ability of the model to identify flood-prone areas, precision reflects the proportion of predicted flood-prone areas that were correctly classified, and the F1-score summarizes the balance between precision and recall (Sokolova and Lapalme, 2009). They are defined as Amiri et al. (2024) in Equations 6–9:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{F1-Score} = \frac{2TP}{2TP + FP + FN} \quad (9)$$

where TP denotes true positives, TN true negatives, FP false positives, and FN false negatives.

Together, these metrics provide a comprehensive evaluation of classification performance by quantifying different aspects of model behavior, including discrimination capacity, predictive accuracy, and the balance between omission and commission errors. Their combined use supports a robust assessment of model performance and facilitates comparison with previous flood susceptibility studies.

2.2.6 Flood susceptibility map

After training and validation, the optimized RF model was applied to the entire study area to generate a continuous flood

susceptibility surface. Each pixel was assigned a susceptibility score ranging from 0 to 1, indicating the relative propensity of that location to be associated with flood-prone conditions. This continuous representation captures spatial gradients of susceptibility and is consistent with established machine learning-based flood susceptibility mapping approaches (e.g., Abdo et al., 2025; Khaldi et al., 2025).

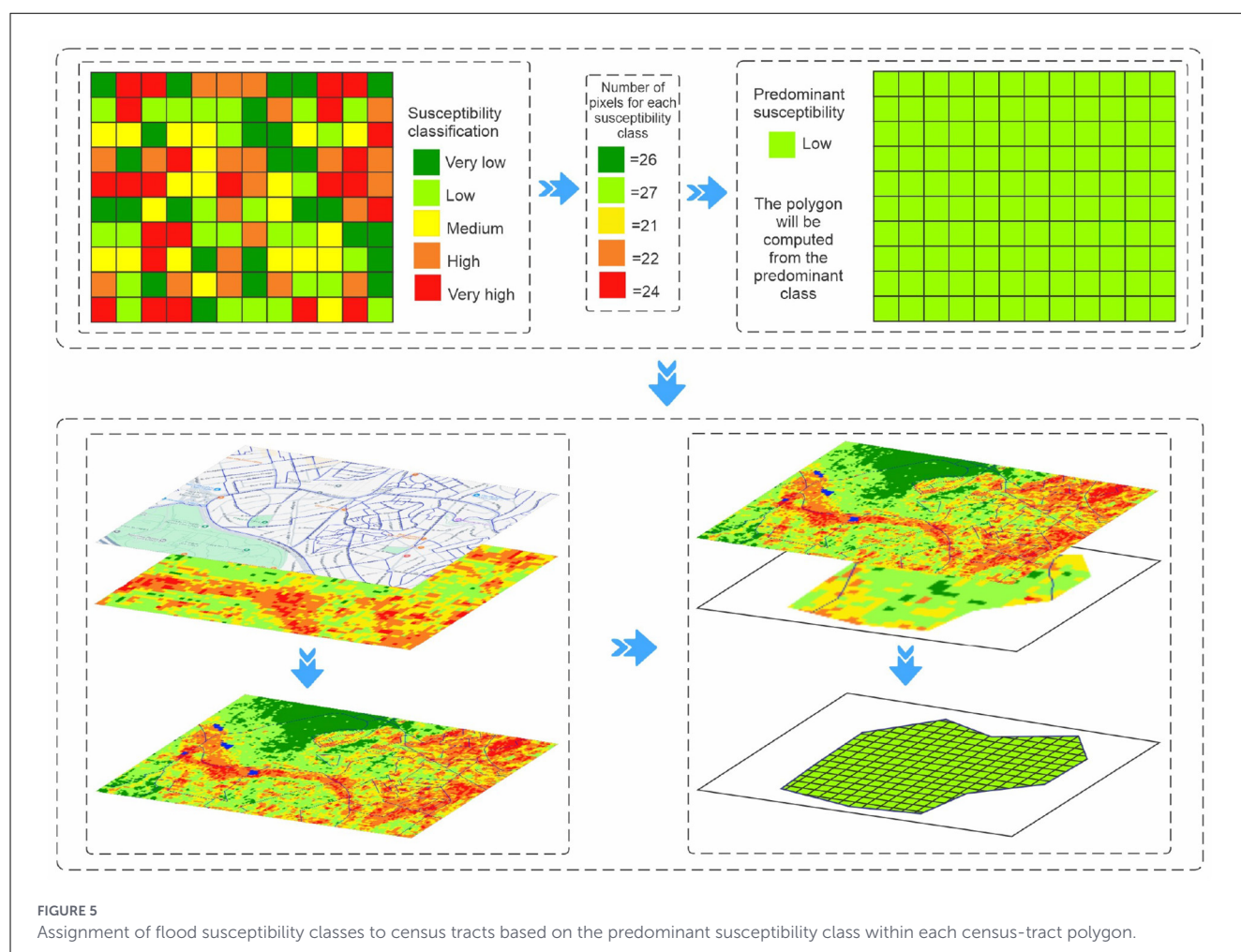
To facilitate interpretation, the continuous susceptibility output was classified into five categories: very low, low, medium, high, and very high susceptibility. The Natural Breaks (Jenks) method was adopted for classification, as it is widely used in flood susceptibility studies (e.g., Huu Duy et al., 2024; Kaur et al., 2025). This approach minimizes variance within classes and maximizes variance between classes, allowing the identification of data-driven class thresholds (Alwathaf et al., 2025). By structuring susceptibility values according to their inherent distribution, the procedure reduces arbitrariness in threshold definition and supports the consistent delineation of critical flood-prone areas (Khosravi et al., 2019a).

2.2.7 Assessment of population exposure

Demographic data from the 2022 IBGE Census (Instituto Brasileiro de Geografia e Estatística (IBGE), 2022) were integrated with the flood susceptibility map to support an assessment of human exposure. Census tracts, representing the smallest spatial units for which detailed demographic information is reported, were adopted as the basis for the analysis (Supplementary Figure S3.B). This integration enabled the identification of locations where elevated flood susceptibility coincides with high population concentrations and large numbers of housing units. By combining spatial susceptibility patterns with demographic information, the analysis extends beyond the characterization of physical susceptibility and incorporates its potential societal implications (Tran et al., 2024).

Following Knighton et al. (2020); Rentschler et al. (2022), census tracts were selected due to their high spatial resolution and demographic detail. Overlaying census data with the 30 m susceptibility map enabled the identification of spatial inequalities in exposure, where relatively small areas may concentrate disproportionately large populations (Tellman et al., 2021). A susceptibility class was assigned to each census tract based on the predominant (modal) susceptibility class among the pixels contained within the tract (Figure 5), thereby preserving the dominant susceptibility condition represented within each census unit. Population and housing-unit counts were subsequently aggregated according to susceptibility class to quantify human and residential exposure.

To complement the census-tract exposure analysis, a municipal-level assessment was conducted for age-sensitive demographic groups, defined here as people younger than 17 years and older than 65 years, using data from the 2022 IBGE Census (Instituto Brasileiro de Geografia e Estatística (IBGE), 2022). These groups were selected because they may face greater mobility constraints, dependency, and evacuation challenges during flood events (Tran et al., 2024). Rather than measuring vulnerability



directly, this analysis identifies municipalities where age-sensitive populations coincide with elevated flood susceptibility, thereby providing additional information for emergency planning and preparedness (e.g., Goto et al., 2022; Khaldi et al., 2025). Quantifying the exposure of these groups is relevant for designing evacuation strategies, shelter allocation, warning communication, and logistical planning that account for populations with potentially reduced autonomous mobility (e.g., Tran et al., 2024; Khan et al., 2025).

2.2.8 Combined prioritization index (CPI) equation

To translate flood susceptibility and population exposure into a municipal-level decision-support metric, a index (i.e., CPI) was developed. Municipalities were adopted as the analytical unit because they represent the primary administrative scale for territorial planning, emergency coordination, and local mitigation strategies (Quesada-Román, 2022).

The CPI integrates three indicators: (i) the dominant flood susceptibility class within each municipality, (ii) the percentage of the municipal population exposed to elevated flood susceptibility, and (iii) the absolute number of inhabitants exposed to elevated

flood susceptibility. In this study, elevated flood susceptibility refers to areas classified as high or very high susceptibility. This structure follows the logic of Multi-Criteria Decision Analysis (MCDA) frameworks, which are commonly used to combine environmental hazard or susceptibility information with human exposure indicators in spatial prioritization assessments (e.g., da Silva et al., 2020; Waleed and Sajjad, 2025b).

Each indicator was reclassified into an ordinal score ranging from 1 to 10. Higher scores were assigned to municipalities with higher dominant susceptibility classes, greater relative population exposure, and larger absolute exposed populations. The inclusion of both relative and absolute exposure allows the CPI to capture two complementary dimensions of prioritization: the proportional concentration of exposed residents within each municipality and the total number of people potentially affected. This scoring procedure was adopted to reflect increasing priority for flood-risk planning and emergency preparedness in municipalities where elevated susceptibility coincides with greater human exposure (e.g., Khaldi et al., 2025).

The CPI was computed using the Simple Additive Weighting (SAW) method (e.g., Meshram et al., 2020; Khosravi et al., 2019b). The absolute exposed population received the highest weight (40%), followed by the susceptibility score (35%) and relative population exposure (25%). These decision-oriented weights were

defined to emphasize, respectively, the potential magnitude of human exposure, the prevailing susceptibility condition, and the proportional concentration of exposed residents within each municipality. The index is defined in Equation 10:

$$\text{CPI} = 0.35S + 0.25E_r + 0.40E_a \quad (10)$$

where S is the municipal susceptibility score, E_r is the relative population exposure score, and E_a is the absolute population exposure score.

Finally, the composite CPI values were classified into five priority levels: priority 1—maximum, priority 2—high, priority 3—medium, priority 4—low, and priority 5—minimum. This classification enables the identification and mapping of municipalities requiring prioritized attention for flood-risk planning and emergency preparedness. Detailed scoring thresholds and the weighting scheme are provided in Supplementary Figure S5.

3 Results

3.1 Performance and conditioning factor interpretation

The RF model demonstrated high discriminatory capacity for flood susceptibility classification, effectively separating flood-prone and non-flood-prone samples, with an AUC of 0.982 (Supplementary Figure S2) and an overall accuracy of 0.939. Detailed performance metrics are presented in Supplementary Table S3. These results indicate strong predictive performance on the independent testing dataset and support the reliability of the subsequent spatial and demographic analyses.

The SHAP analysis (Figure 6) provides insight into the relative importance and directional influence of conditioning factors. Distance to rivers emerged as the most influential predictor. Areas closer to drainage networks corresponded to positive SHAP contributions and higher flood susceptibility, highlighting fluvial proximity as the primary conditioning factor.

Topographic variables also exerted substantial influence on flood susceptibility. For instance, slope and altitude exhibited inverse relationships with susceptibility contributions; low altitudes and gentle slopes produced positive SHAP contributions, whereas steep slopes and higher altitudes produced negative contributions. These patterns are consistent with the expected accumulation of runoff in low-gradient environments and floodplain settings.

The surface-related conditioning factors also demonstrated strong predictive importance for flood susceptibility (Figure 6), with CN and LULC ranking as the third and fifth most influential predictors, respectively. Higher CN values were associated with positive SHAP values, indicating that surface conditions characterized by lower infiltration capacity and greater runoff-generation potential were associated with increased flood susceptibility. LULC also contributed to the model predictions, indicating that land-cover characteristics played an important role in shaping susceptibility patterns across the state.

Drainage density, TWI, rainfall, and soil classification showed secondary contributions, suggesting that local runoff concentration, wetness potential, and substrate characteristics also influenced model predictions, although less strongly than distance to rivers, slope, CN, and altitude.

Precipitation showed a moderate but consistent positive contribution, indicating that greater rainfall depth is associated with higher flood susceptibility, although its influence occurs in conjunction with terrain characteristics and land-cover conditions. In contrast, profile curvature, plan curvature, and aspect displayed SHAP distributions concentrated around zero, suggesting comparatively limited influence within the statewide modeling framework.

Overall, the SHAP results indicate that flood susceptibility across São Paulo State is primarily associated with the combined influence of fluvial proximity, topographic conditions, and anthropogenic surface characteristics.

3.2 Spatial distribution of flood susceptibility

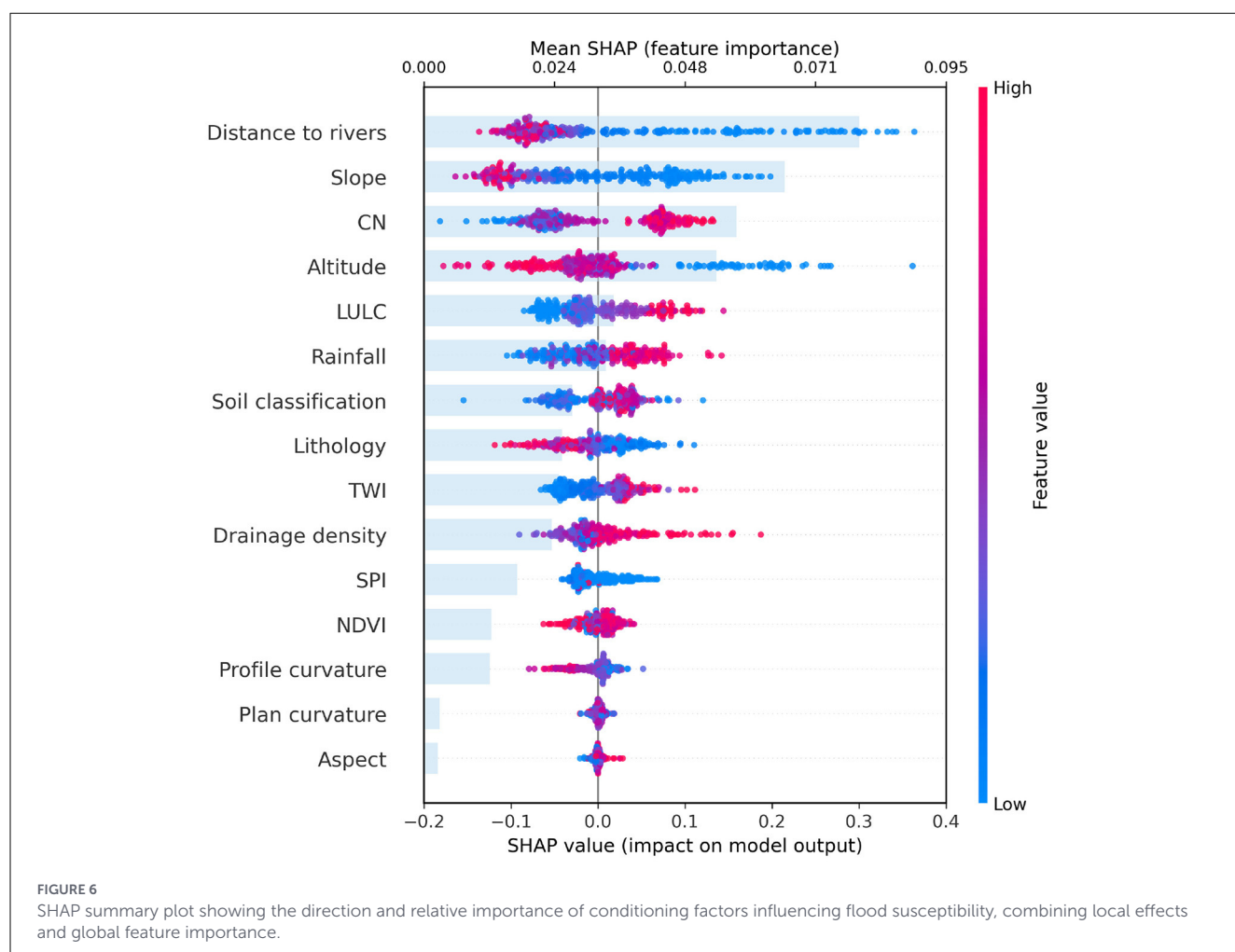
The spatial distribution of flood susceptibility across São Paulo State revealed marked spatial heterogeneity, with higher susceptibility levels predominantly concentrated along the coastal zone. The resulting susceptibility map was qualitatively compared to the occurrence of flood events reported in digital media and news, providing additional evidence of agreement between areas classified as highly susceptible and locations where flooding has been documented (Figure 7).

Using the Natural Breaks (Jenks) classification method, the continuous susceptibility scores were categorized into five classes: very low, low, medium, high, and very high susceptibility (Figure 7). The distribution was dominated by the very low and low classes, which together accounted for two-thirds of the mapped area (67.3%), whereas high and very high susceptibility zones occupied a comparatively smaller proportion of the state (16.4%). These patterns indicate a pronounced spatial structuring of flood susceptibility, with highly susceptible areas concentrated in specific territorial settings rather than being uniformly distributed across São Paulo State.

High and very high susceptibility areas were predominantly concentrated in the coastal region. Detailed examination of selected high-susceptibility areas (Supplementary Figure S4) indicates that these zones coincide with densely urbanized environments and extensive built-up surfaces. Low altitudes, gentle slopes, and pronounced anthropogenic land-cover characteristics were consistently observed within these areas, suggesting the combined influence of topographic and anthropogenic conditioning factors on flood susceptibility.

3.3 Population exposure to flood susceptibility

The evaluation of population exposure reveals a marked spatial mismatch between the territorial extent of flood susceptibility



classes and the distribution of people and housing units across São Paulo State. Although most of the mapped area was classified as very low or low susceptibility, population and housing units were disproportionately concentrated in areas classified as high and very high susceptibility (Figure 8 and Supplementary Table S4).

The detailed analysis shown in Figure 8 indicates substantial spatial inequality in exposure. Areas classified as very high susceptibility represented only 6.2% of the state territory, but contained 22.3% of the total population, corresponding to approximately 9.89 million people. When the high and very high susceptibility classes were combined, they accounted for only 9.2% of the total analyzed area (22,816 km²), but concentrated 44.6% of the population and 45.2% of housing units. Of the 19,637,377 recorded housing units, approximately 8,877,707 were located in high and very high susceptibility zones, whereas 7,247,024 were located in low and very low susceptibility classes. These results indicate that exposure is not proportional to the spatial extent of susceptible areas, with a relatively small fraction of the territory concentrating a large share of the population and residential structures.

This spatial configuration is also reflected in the internal structure of the census tracts and the distribution of the population across susceptibility classes. Among the census tracts analyzed, the high and very high classes accounted for 42.7% of the total,

corresponding to approximately 43,716 tracts, while the low and very low classes represented 40.0% (or approximately 40,952 tracts). In terms of population, nearly 19.80 million people were located in areas classified as high and very high susceptibility, representing 44.6% of the total population analyzed. By contrast, 16.55 million people were located in low and very low classes (37.3%), and 8.08 million in the medium class (18.2%). These results show that population exposure is not proportional to territorial extent alone, but depends on the spatial overlap between susceptible areas and densely populated census tracts.

A complementary municipal-level analysis was conducted to assess the exposure of age-sensitive demographic groups to flood susceptibility (Figure 9). At this scale, exposure was summarized into three levels—low, medium, and high—to improve the interpretation of demographic exposure patterns. For the population younger than 17 years (Figure 9A), totaling 12.4 million people, 64.0% were classified within low exposure, 11.5% within medium exposure, and 24.6% within high exposure. Although most individuals in this group were located in low-exposure municipalities, approximately one-quarter were associated with high exposure to flood susceptibility.

For the population older than 65 years (Figure 9B), totaling 5.3 million people, 62.8% were classified within low exposure, 10.4% within medium exposure, and 26.8% within high exposure.

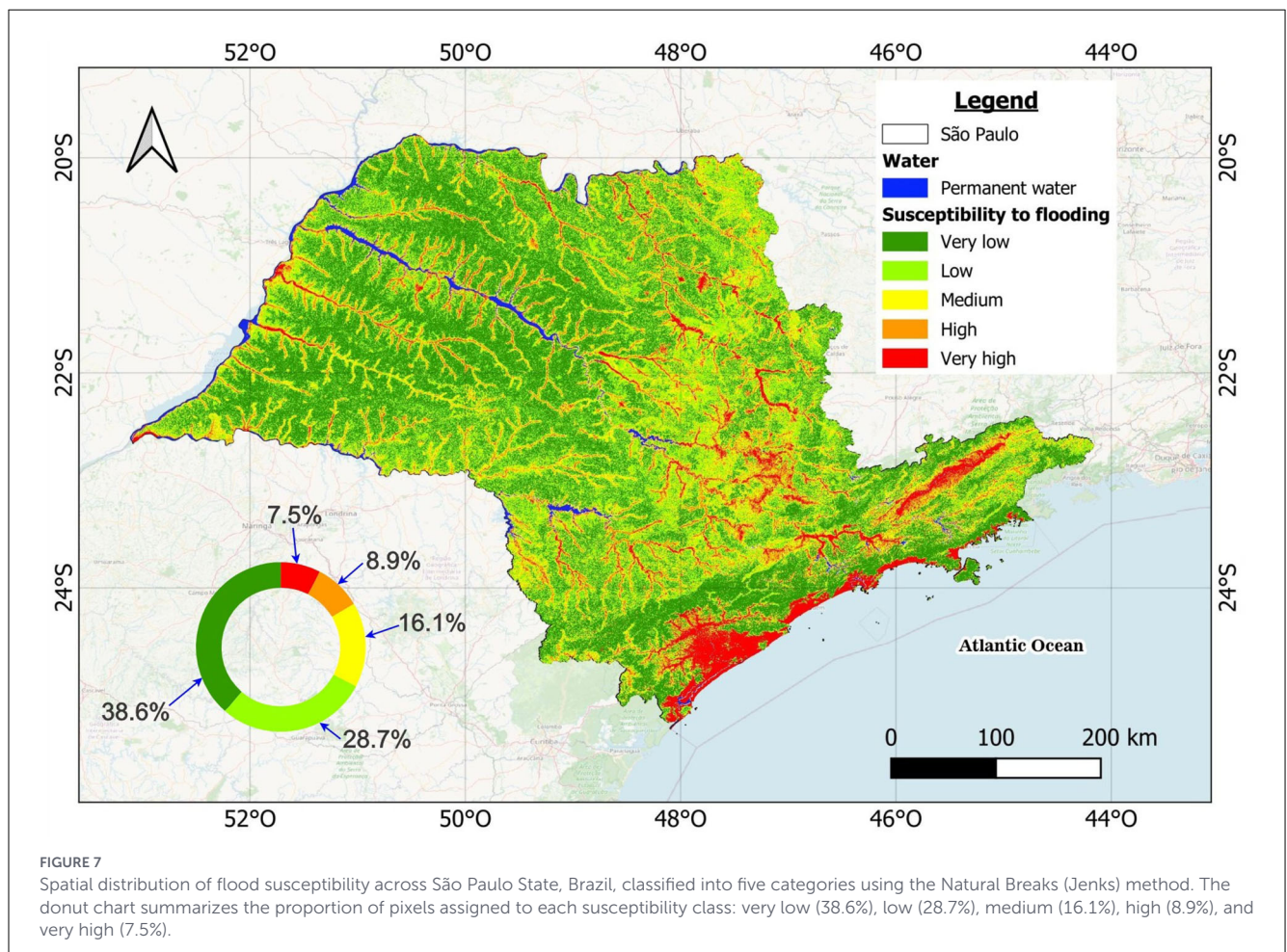


FIGURE 7

Spatial distribution of flood susceptibility across São Paulo State, Brazil, classified into five categories using the Natural Breaks (Jenks) method. The donut chart summarizes the proportion of pixels assigned to each susceptibility class: very low (38.6%), low (28.7%), medium (16.1%), high (8.9%), and very high (7.5%).

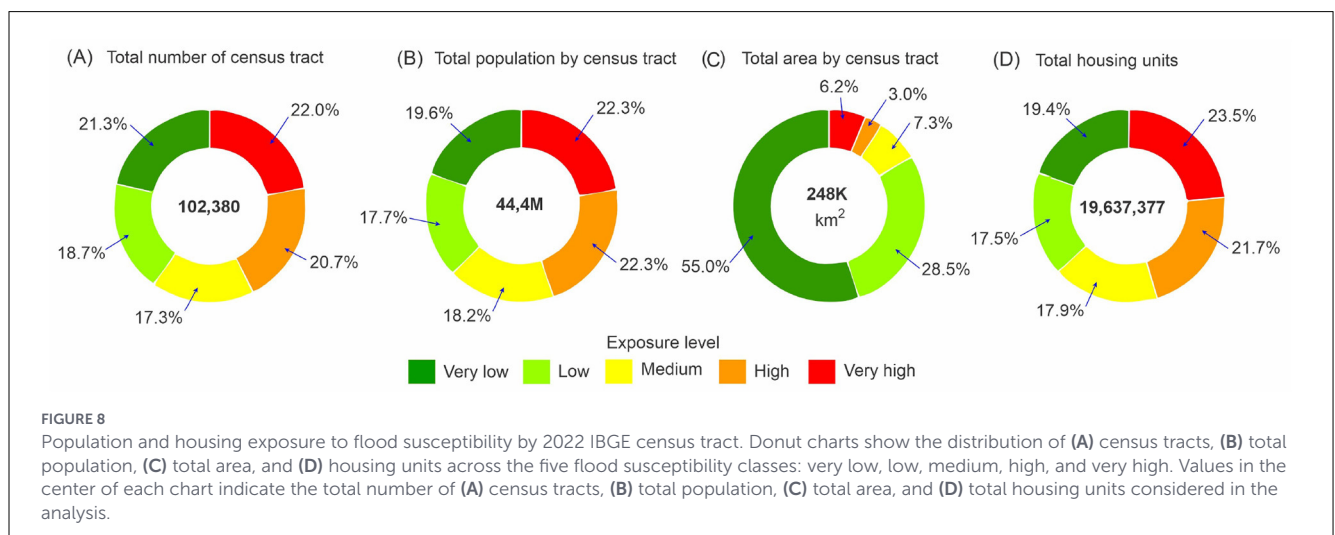


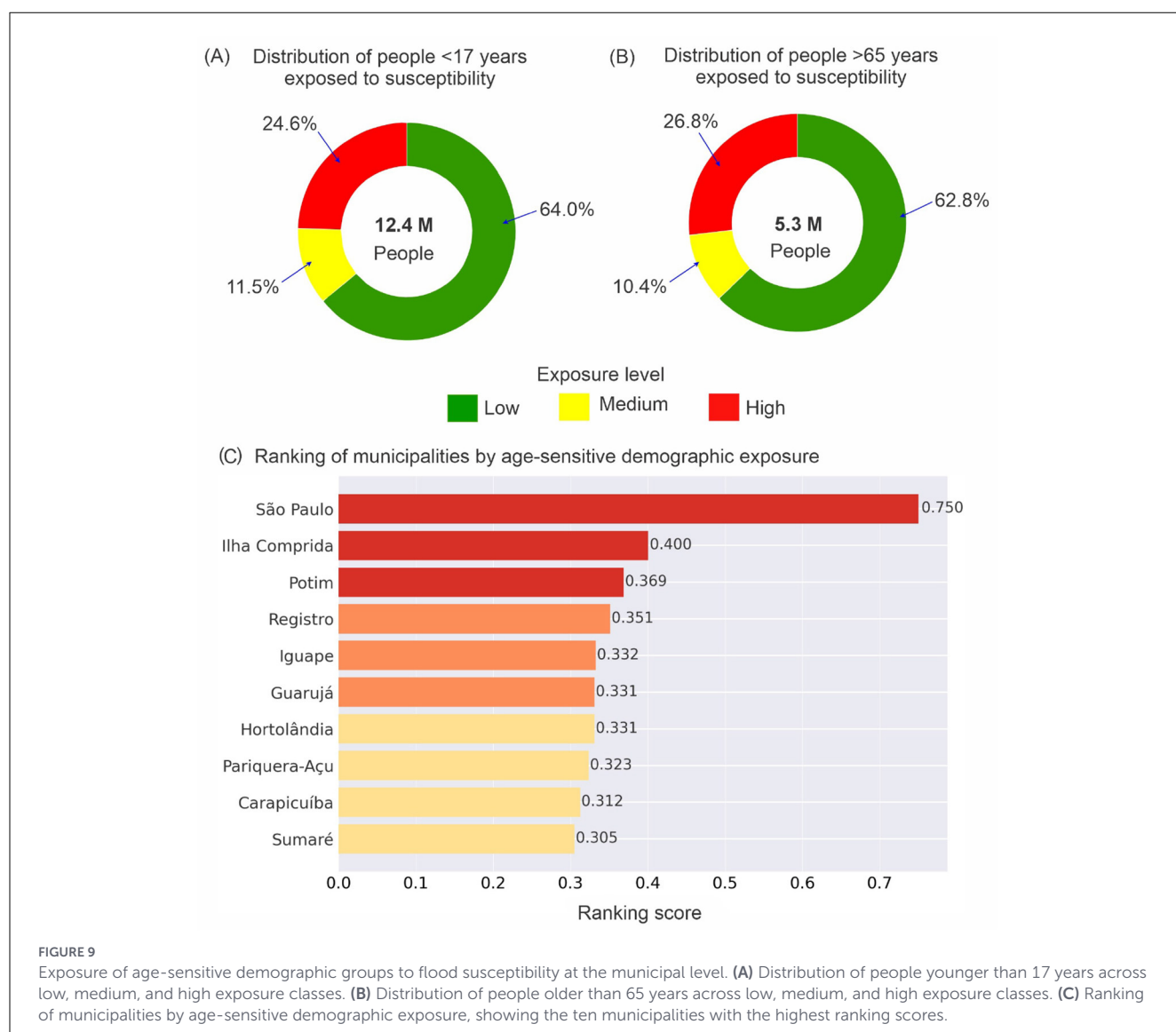
FIGURE 8

Population and housing exposure to flood susceptibility by 2022 IBGE census tract. Donut charts show the distribution of (A) census tracts, (B) total population, (C) total area, and (D) housing units across the five flood susceptibility classes: very low, low, medium, high, and very high. Values in the center of each chart indicate the total number of (A) census tracts, (B) total population, (C) total area, and (D) total housing units considered in the analysis.

Compared with the younger age group, older adults showed a slightly higher proportion in the high-exposure class, indicating a greater relative concentration of this age-sensitive group in municipalities with elevated flood susceptibility.

For the age-sensitive demographic groups, a municipal-level ranking score was estimated on a normalized scale from 0 to 1, where values closer to 0 indicate lower ranking scores and

values closer to 1 indicate higher ranking scores for demographic exposure in flood-risk planning. Figure 9C identified São Paulo as the highest-ranking municipality, with a ranking score of 0.750, followed by Ilha Comprida (0.400) and Potim (0.369). The remaining municipalities among the ten highest-ranking municipalities were Registro (0.351), Iguape (0.332), Guarujá and Hortolândia (0.331), Pariqueira-Açu (0.323), Carapicuíba (0.312),



and Sumaré (0.305). These results identify municipalities where elevated flood susceptibility coincides with larger concentrations of age-sensitive population groups, providing complementary information for flood-risk planning.

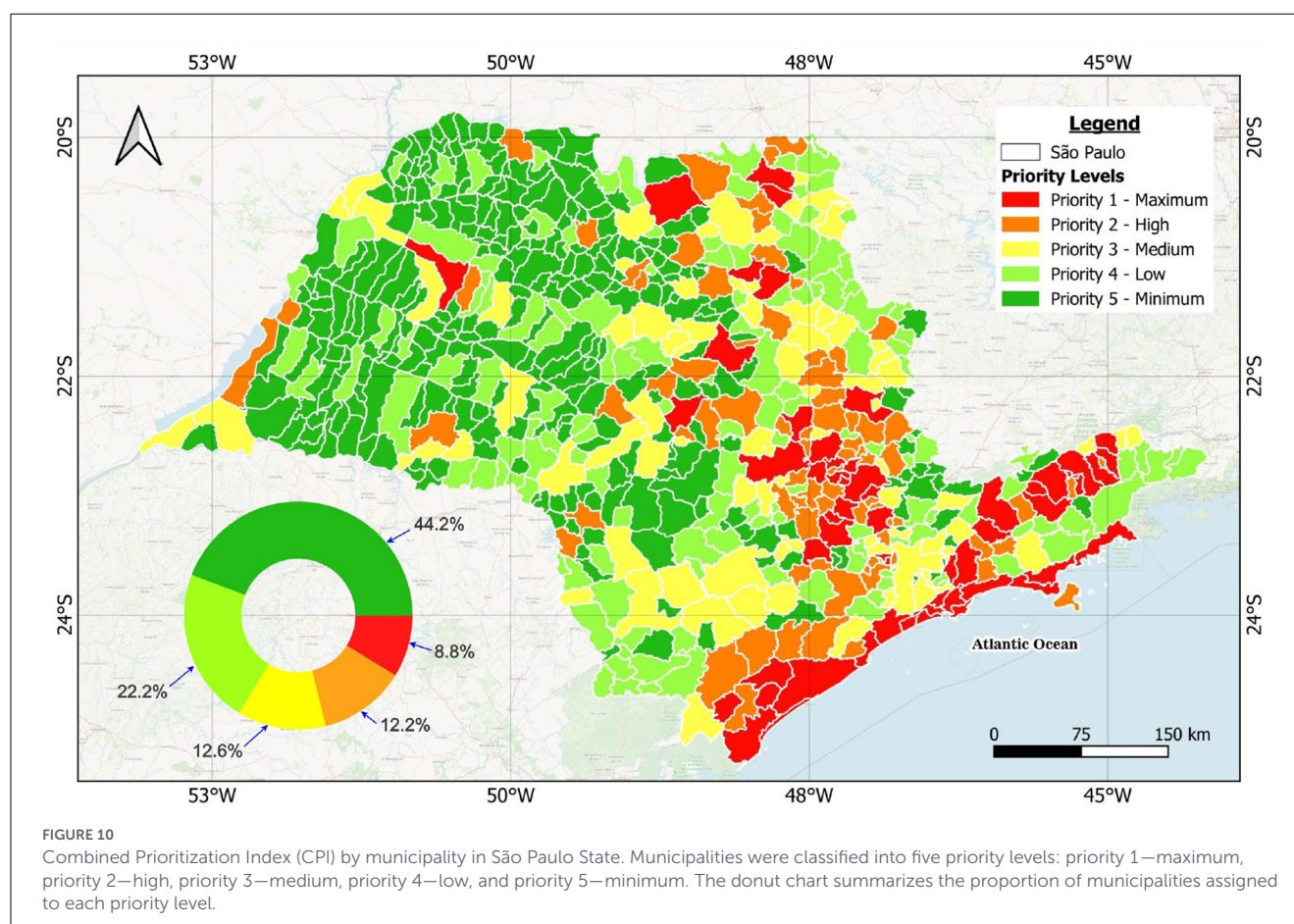
3.4 Municipal prioritization based on flood susceptibility and population exposure

Because municipalities represent the primary administrative units for local risk management and territorial planning, the distribution of municipalities across dominant flood susceptibility classes was assessed before constructing the CPI. Among the 645 municipalities of São Paulo State, 63 (9.8%) were classified as very low susceptibility, 261 (40.5%) as low, 180 (27.9%) as medium, 108 (16.7%) as high, and 33 (5.1%) as very high. This distribution indicates that, although most municipalities fall within the low and medium susceptibility classes, a substantial number

are characterized by elevated susceptibility levels that may coincide with population exposure.

The CPI revealed a spatially heterogeneous distribution of priority levels across the state, with higher-priority municipalities occurring not only along the coastal region but also in inland regions (Figure 10). Overall, 8.8% of municipalities were classified as priority 1—maximum, 12.2% as priority 2—high, 12.6% as priority 3—medium, 22.2% as priority 4—low, and 44.2% as priority 5—minimum. Thus, priority levels 1, 2, and 3 together accounted for 33.6% of all municipalities. The spatial pattern indicates that prioritization is not restricted to the coastal susceptibility corridor, as several inland municipalities also exhibit combinations of susceptibility and population exposure indicators that place them among the higher-priority classes. The municipalities with the highest CPI values are presented in Table S5.

Higher-priority municipalities were not restricted to the coastal region. Several inland municipalities, including Rio Claro, Sumaré, Jacareí, Pindamonhangaba, Taubaté, Guaratinguetá, Paulínia, Jundiaí, Valinhos, Ribeirão Preto, Piracicaba, Santa



Bárbara d'Oeste, and Hortolândia, were also classified among the higher-priority classes. These municipalities exhibited elevated CPI values because elevated flood susceptibility coincided with substantial relative and absolute population exposure. This pattern indicates that municipal priority is not determined by geographic location alone, but also by the spatial overlap between flood susceptibility and population exposure.

4 Discussion

The SHAP analysis indicates that flood susceptibility in São Paulo State is not explained solely by static topographic conditions. Instead, the model suggests that susceptibility emerges from the combined influence of fluvial proximity, terrain configuration, and urbanization-related surface modification.

Distance to rivers and slope were the most influential conditioning factors, consistent with Amiri et al. (2024), who highlight the importance of relief and low-gradient floodplain environments in flood susceptibility patterns. In addition, CN ranked as the leading non-topographic conditioning factor, highlighting the relevance of runoff-generating surface conditions in the model. This result is consistent with Zhu et al. (2024) and Debnath et al. (2024), who show that impervious and disturbed surfaces can substantially increase direct runoff

generation in urbanized basins. Together, these findings indicate that flood susceptibility in São Paulo State is shaped not only by fluvial proximity and terrain configuration, but also by anthropogenic surface conditions that favor runoff generation.

The integration of the susceptibility map with 2022 IBGE Census data revealed a marked spatial asymmetry in human exposure across São Paulo State. Areas classified as very high susceptibility occupy only 6.2% of the territory, yet they contain 22.3% of the total population and 23.5% of housing units. This disproportion indicates that exposure is not proportional to the spatial extent of susceptible areas. Instead, a relatively small fraction of the state concentrates a large share of the population and residential structures. This pattern is consistent with the global trends described by Tellman et al. (2021) and Rentschler et al. (2022), who show that flood-prone areas often remain attractive for settlement because of economic activity, accessibility, and settlement patterns commonly observed in coastal and riverine areas. By using census tracts for population and housing exposure, this study makes this imbalance explicit and shows how flood susceptibility becomes socially relevant when it overlaps with densely occupied territories.

Beyond overall exposure, the municipal-level analysis highlighted the distribution of age-sensitive demographic groups. The results showed that 24.6% of people younger than 17 years and 26.8% of people older than 65 years lived in municipalities

classified as having high exposure. This concentration is particularly evident in municipalities such as São Paulo, Ilha Comprida, and Potim. These groups are commonly recognized as age-sensitive in disaster-risk assessment because they may face greater mobility constraints, dependency, and evacuation challenges during flood events (Knighton et al., 2020; Tran et al., 2024). Mortality patterns reported in other flood disasters reinforce the relevance of considering age structure in exposure assessments. For example, during the 2022 floods in Khyber Pakhtunkhwa, Pakistan, authorities recorded 306 deaths, including 116 children, who accounted for 38% of the total (Khan et al., 2025). Although the contexts differ, such evidence supports the need to incorporate age-sensitive groups into flood-risk planning and emergency preparedness.

The CPI proposed in this study provides a context-specific decision-support tool for identifying municipalities where flood susceptibility and population exposure indicators converge. By linking elevated susceptibility with relative and absolute population exposure, the assessment extends beyond the physical characterization of flood-prone areas and supports a spatial inequalities in flood exposure perspective in municipal decision-making. Rather than simply classifying municipalities into broad priority categories, the CPI provides a structured basis for identifying where anticipatory planning and risk-reduction measures may be most needed.

The main contribution of this study lies in connecting physically based susceptibility patterns with demographic exposure and municipal prioritization. In this sense, the analysis moves beyond conventional flood susceptibility mapping by showing where model-derived susceptibility becomes socially relevant through its overlap with population concentration and age-sensitive demographic groups. This integration is particularly important for socio-hydrological applications because flood impacts are not determined by physical susceptibility alone, but by the interaction between hydrological conditions, urban occupation, and exposed populations. The proposed approach therefore provides a bridge between machine learning-based susceptibility assessment and planning-oriented information that can support more transparent and equity-oriented flood-risk governance.

The analysis shows that higher-priority municipalities are not restricted to the coastal region of São Paulo State. Several inland municipalities also exhibited elevated CPI values because elevated flood susceptibility coincided with substantial relative and absolute population exposure. This finding indicates that priority is not determined by geographic location alone or by historically flood-prone coastal settings, but by the spatial overlap between susceptibility and population exposure indicators. Therefore, flood-risk planning should consider both coastal and inland municipalities where population exposure is concentrated in highly susceptible areas.

From a spatial inequality perspective, these results show that flood susceptibility becomes socially relevant when it overlaps with concentrated population exposure. In this study, spatial inequalities in flood exposure is not treated as a directly measured outcome, but as an interpretive lens for identifying spatial inequalities in exposure and supporting more transparent prioritization in flood-risk planning.

As argued by Bui et al. (2023) and Khaldi et al. (2025), integrating machine learning outputs into decision-support frameworks can help standardize criteria and support more transparent prioritization before, during, and after disasters. In the present study, the finding that high and very high susceptibility classes occupy only 9.2% of the analyzed area but concentrate 44.6% of the population and 45.2% of housing units provides a strong empirical basis for targeted planning. Such information can support the prioritization of early warning systems, land-use planning in flood-prone corridors, and emergency preparedness measures, particularly in municipalities where exposure indicators are concentrated.

Despite these contributions, the results should be interpreted with caution. The susceptibility model was developed using mostly static conditioning factors and does not explicitly represent the temporal dynamics of rainfall events, drainage-system performance, or flood propagation processes. In addition, the flood-prone area inventory may reflect mapping and reporting biases, particularly in more urbanized or better-monitored municipalities. The age-sensitive demographic analysis was conducted at the municipal level, which limits its spatial detail compared with the census-tract-level assessment of population and housing exposure. Finally, the CPI weights are decision-oriented and should be interpreted as a transparent prioritization scheme rather than as an objective measure of flood risk or vulnerability.

5 Conclusion

This study developed a flood susceptibility and exposure assessment for São Paulo State, Brazil, integrating machine learning, explainable artificial intelligence, and census-based demographic information. By combining a Random Forest susceptibility model with SHAP interpretation and 2022 IBGE Census data, the analysis provides an integrated view of where flood-prone conditions coincide with population and housing exposure. The findings reveal spatial inequalities in flood exposure by showing that exposure to flood susceptibility is unevenly distributed across São Paulo State.

A marked spatial imbalance was identified. Areas classified as Very High susceptibility occupy only 6.2% of the state territory, yet they contain 22.3% of the total population and 23.5% of housing units. When High and Very High susceptibility classes are combined, they represent only 9.2% of the analyzed area but concentrate 44.6% of the population and 45.2% of housing units. These results demonstrate that human exposure is not proportional to the spatial extent of susceptible areas. Instead, a relatively small portion of the territory concentrates a large share of people and residential structures exposed to elevated flood susceptibility.

The SHAP analysis showed that flood susceptibility is shaped by the combined influence of fluvial proximity, terrain configuration, and anthropogenic surface conditions. Distance to rivers and slope were the most influential conditioning factors, while Curve Number (CN) emerged as the leading non-topographic factor. These results indicate that topography

alone does not explain susceptibility patterns across São Paulo State. Surface conditions associated with runoff generation, land-cover modification, and urbanization-related imperviousness also play an important role in shaping areas more prone to flooding.

The exposure assessment further showed that age-sensitive demographic groups are unevenly distributed across municipalities with elevated flood susceptibility. At the municipal level, 24.6% of people younger than 17 years and 26.8% of people older than 65 years were located in municipalities classified as High exposure. This finding reinforces the need to consider demographic structure in flood-risk planning, particularly because these groups may face greater mobility constraints and evacuation challenges during flood events.

The Combined Prioritization Index (CPI) provides a context-specific decision-support tool for identifying municipalities where flood susceptibility and population exposure indicators converge. The results showed that higher-priority municipalities are not restricted to the coastal region. Several inland municipalities also exhibited elevated CPI values because elevated susceptibility coincided with substantial relative and absolute population exposure. This finding indicates that municipal priority is not determined by geographic location alone, but by the spatial overlap between susceptibility and population exposure indicators. The CPI therefore offers a structured basis for supporting more transparent and exposure-informed prioritization in flood-risk planning.

Several limitations should be acknowledged. The analysis relied primarily on static datasets, and the 30 m spatial resolution of the DEM may smooth fine-scale urban variability relevant to local flooding processes. In addition, the authoritative flood-prone area inventory may reflect mapping and reporting biases, particularly in more urbanized or better-monitored areas. The model does not explicitly incorporate temporal variability in conditioning factors, nor does it simulate the stochastic behavior of extreme rainfall events. Moreover, age-sensitive demographic exposure was assessed at the municipal level, which limits the spatial detail of this component compared with the census-tract-level analysis of population and housing exposure. Finally, the CPI weights are decision-oriented and should be interpreted as a transparent prioritization scheme rather than as an objective measure of flood risk or vulnerability.

Future research should incorporate temporally dynamic predictors, higher-resolution terrain and drainage information, and event-based hydrometeorological data to improve the representation of flood-generating processes. Deep learning architectures and spatiotemporal modeling approaches may further support the analysis of climate variability, land-use change, and evolving exposure patterns. Overall, this study demonstrates that flood susceptibility mapping becomes more relevant for planning when combined with demographic exposure information. The findings reveal spatial inequalities in flood exposure by showing where susceptible areas overlap with concentrated human exposure. Urban flood resilience therefore depends not only on infrastructure, but also on recognizing spatially unequal exposure patterns and using this

information to guide more anticipatory, transparent, and inclusive territorial planning.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: the datasets generated from open-access products for this study can be found in the Zenodo repository <https://doi.org/10.5281/zenodo.19455781>.

Author contributions

MC: Conceptualization, Methodology, Investigation, Data curation, Software, Formal analysis, Validation, Visualization, Resources, Writing – original draft, Writing – review & editing. LM: Conceptualization, Methodology, Investigation, Data curation, Software, Formal analysis, Validation, Visualization, Resources, Writing – review & editing. GC: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Validation, Visualization, Writing – review & editing. AA: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Validation, Visualization, Writing – review & editing. SP: Conceptualization, Methodology, Formal analysis, Validation, Visualization, Writing – review & editing. SH: Conceptualization, Methodology, Formal analysis, Validation, Visualization, Writing – review & editing. ML: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Validation, Visualization, Resources, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was funded by the São Paulo Research Foundation (FAPESP) (Grant No. 25/16161-0). The first author acknowledges the National Council for Scientific and Technological Development (CNPq) for awarding a master scholarship. The authors acknowledge financial support from the Fundo de Apoio ao Ensino, à Pesquisa e à Extensão (FAEPEX) of the State University of Campinas (UNICAMP) for supporting the visit of Professor Saket Pande (Grant No. 2448/26).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors SP, ML declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

References

- Abdo, H. G., Richi, S. M., Zeng, T., Alqadhi, S., Prasad, P., Nguyen-Huy, T., et al. (2025). Machine learning-based assessment of flood susceptibility in the eastern mediterranean: a case study of Baniyas river basin. *Geomat. Natur. Hazards Risk* 16:2524417. doi: 10.1080/19475705.2025.2524417
- Ahmad, A., Chen, J., Chen, X., Khadka, N., Khan, M. U., Wang, C., et al. (2025). Flood risk modelling by the synergistic approach of machine learning and best-worst method in Indus Kohistan, western Himalaya. *Geomat. Natur. Hazards Risk* 16:2469766. doi: 10.1080/19475705.2025.2469766
- Ahmed, I. A., Talukdar, S., Shahfahad Parvez, A., Rihan, M., Baig, M. R. I., and Rahman, A. (2022). Flood susceptibility modeling in the urban watershed of Guwahati using improved metaheuristic-based ensemble machine learning algorithms. *Geocarto Int.* 37, 12238–12266. doi: 10.1080/10106049.2022.2066200
- Alvares, C. A., Stape, J. L., Sentelhas, P. C., de Moraes Goncalves, J. L., and Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorol. Zeitschrift* 22, 711–728. doi: 10.1127/0941-2948/2013/0507
- Alwathaf, Y., Al-Areeq, A. M., Al-Masnay, Y. A., Al-Aizari, A. R., and Al-Areeq, N. M. (2025). Enhancing flood susceptibility mapping in Sanaa, Yemen with random forest and extreme gradient boosting algorithms. *Geocarto Int.* 40:2482707. doi: 10.1080/10106049.2025.2482707
- Amiri, A., Soltani, K., Ebtehaj, I., and Bonakdari, H. (2024). A novel machine learning tool for current and future flood susceptibility mapping by integrating remote sensing and geographic information systems. *J. Hydrol.* 632:130936. doi: 10.1016/j.jhydrol.2024.130936
- Arabameri, A., Seyed Danesh, A., Santosh, M., Cerda, A., Chandra Pal, S., Ghorbanzadeh, O., et al. (2022). Flood susceptibility mapping using meta-heuristic algorithms. *Geomat. Natur. Hazards Risk* 13, 949–974. doi: 10.1080/19475705.2022.2060138
- Aydin, H. E., and Iban, M. C. (2023). Predicting and analyzing flood susceptibility using boosting-based ensemble machine learning algorithms with shapley additive explanations. *Natural Hazards* 116, 2957–2991. doi: 10.1007/s11069-022-05793-y
- Beven, K. J., and Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrol. Sci. J.* 24, 43–69. doi: 10.1080/02626667909491834
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Bui, D. T., Hoang, N.-D., Martínez-Álvarez, F., Ngo, P.-T. T., Hoa, P. V., Pham, T. D., et al. (2020). A novel deep learning neural network approach for predicting flash flood susceptibility: a case study at a high frequency tropical storm area. *Sci. Total Environ.* 701:134413. doi: 10.1016/j.scitotenv.2019.134413
- Bui, Q. D., Luu, C., Mai, S. H., Ha, H. T., Ta, H. T., and Pham, B. T. (2023). Flood risk mapping and analysis using an integrated framework of machine learning models and analytic hierarchy process. *Risk Anal.* 43, 1478–1495. doi: 10.1111/risa.14018
- Chakraborty, R., Chandra Pal, S., Rezaie, F., Arabameri, A., Lee, S., Roy, P., et al. (2022). Flash-flood hazard susceptibility mapping in Kangsabati river basin, India. *Geocarto Int.* 37, 6713–6735. doi: 10.1080/10106049.2021.1953618
- Chen, W., Hong, H., Li, S., Shahabi, H., Wang, Y., Wang, X., et al. (2019). Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *J. Hydrol.* 575, 864–873. doi: 10.1016/j.jhydrol.2019.05.089
- da Silva, L. B. L., Humberto, J. S., Alencar, M. H., Ferreira, R. J. P., and de Almeida, A. T. (2020). Gis-based multidimensional decision model for enhancing flood risk prioritization in urban areas. *Int. J. Disast. Risk Reduct.* 48:101582. doi: 10.1016/j.ijdrr.2020.101582
- de Oliveira-Júnior, J. F., da Silva Junior, C. A., Teodoro, P. E., Rossi, F. S., Blanco, C. J. C., Lima, M., et al. (2021). Confronting chirps dataset and in situ stations in the detection of wet and drought conditions in the Brazilian midwest. *Int. J. Climatol.* 41, 4478–4493. doi: 10.1002/joc.7080
- Debnath, J., Debbarma, J., Debnath, A., Meraj, G., Chand, K., Singh, S. K., et al. (2024). Flood susceptibility assessment of the agartala urban watershed, India, using machine learning algorithm. *Environ. Monit. Assess.* 196:110. doi: 10.1007/s10661-023-12240-3
- El baida, M., Boushaba, F., Chourak, M., Hosni, M., and Sabar, H. (2024). Classification machine learning models for urban flood hazard mapping: case study of Zaio, ne morocco. *Natur. Hazards* 120, 10013–10041. doi: 10.1007/s11069-024-06596-z
- Goto, E. A., Suarez, A. M., and Ye, H. (2022). Spatial analysis of social vulnerability in São paulo city, Brazil. *Appl. Geograp.* 140:102653. doi: 10.1016/j.apgeog.2022.102653
- Hussain, M. A., Chen, Z., Zhou, Y., Ullah, H., and Ying, M. (2025). Spatial analysis of flood susceptibility in coastal area of pakistan using machine learning models and SAR imagery. *Environ. Earth Sci.* 84:120. doi: 10.1007/s12665-025-12129-z
- Huu Duy, N., Tuan Pham, L., Xuan Linh, N., Van Truong, T., Dang, D. K., Quang Hai, T., et al. (2024). Flood risk assessment using machine learning, hydrodynamic modelling, and the analytic hierarchy process. *J. Hydroinform.* 26, 1852–1882. doi: 10.2166/hydro.2024.033
- Instituto Brasileiro de Geografia e Estatística (IBGE) (2022). 2022 population census: Preliminary results. Available online at: <https://www.ibge.gov.br> (Accessed June 30, 2026).
- Kaur, B., Binns, A., McBean, E., Sandink, D., Castro, K., and Gharabaghi, B. (2025). Urban flood susceptibility mapping for Toronto, Canada, using supervised regression and machine learning models. *J. Flood Risk Manag.* 18:e70051. doi: 10.1111/jfr3.70051
- Khalidi, L., Elabed, A., and El Khanchoufi, A. (2025). Multidimensional risk assessment based on flood susceptibility mapping and multiple socioeconomic variables under climate change. *Sci. African* 29:e02834. doi: 10.1016/j.sciaf.2025.e02834
- Khan, T. U., Nabi, G., Ullah, S., Akbar, A., Omifolaj, J. K., Achakzai, J. K., et al. (2025). Mapping flood resilience: a comprehensive geospatial insight into regional vulnerabilities. *Front. Water* 7:1465505. doi: 10.3389/frwa.2025.1465505
- Khosravi, K., Melesse, A. M., Shahabi, H., Shirzadi, A., Chapi, K., and Hong, H. (2019a). “Flood susceptibility mapping at ningdu catchment, China using bivariate and data mining techniques,” in *Extreme hydrology and climate variability* (Elsevier), 419–434. doi: 10.1016/B978-0-12-815998-9.00033-6
- Khosravi, K., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., et al. (2019b). A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. *J. Hydrol.* 573, 311–323. doi: 10.1016/j.jhydrol.2019.03.073
- Knighton, J., Buchanan, B., Guzman, C., Elliott, R., White, E., and Rahm, B. (2020). Predicting flood insurance claims with hydrologic and socioeconomic

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frwa.2026.1857531/full#supplementary-material>

- demographics via machine learning: exploring the roles of topography, minority populations, and political dissimilarity. *J. Environ. Manage.* 272:111051. doi: 10.1016/j.jenvman.2020.111051
- Leggesse, E. S., Derseh, W. A., Zimale, F. A., Tilahun, S. A., and Meshesha, M. A. (2024). Urban flash flood hazard mapping using machine learning, bahir dar, ethiopia. *J. Hydroinform.* 26, 2124–2145. doi: 10.2166/hydro.2024.277
- Li, Y., Fang, Z., Liu, J., Lu, Z., Zhou, H., Yin, W., et al. (2026). A comparative study of urban pluvial flood susceptibility assessment based on multi-machine learning algorithm. *Water Resour. Manag.* 40:35. doi: 10.1007/s11269-025-04414-0
- MapBiomass Brasil (2024). *Annual land use and land cover maps of Brazil*. Available online at: https://storage.googleapis.com/mapbiomas-public/initiatives/brasil/collection_10/lulc/coverage/brazil_coverage_2024.tif (Accessed July 1, 2026).
- Meshram, S. G., Alvandi, E., Meshram, C., Kahya, E., and Fadhil Al-Quraishi, A. M. (2020). Application of saw and toposim in prioritizing watersheds. *Water Resour. Manag.* 34, 715–732. doi: 10.1007/s11269-019-02470-x
- Paredes-Trejo, F. J., Barbosa, H., and Kumar, T. L. (2017). Validating chirps-based satellite precipitation estimates in northeast brazil. *J. Arid Environ.* 139, 26–40. doi: 10.1016/j.jaridenv.2016.12.009
- Paszkowski, A., Tiggeloven, T., Borgomeo, E., and Hall, J. W. (2025). Disparities in exposure to hydrogeomorphic hazards in bangladesh. *Nat. Commun.* 16:10208. doi: 10.1038/s41467-025-64920-y
- Quesada-Román, A. (2022). Flood risk index development at the municipal level in costa rica: A methodological framework. *Environ. Sci. Policy* 133, 98–106. doi: 10.1016/j.envsci.2022.03.012
- Rentschler, J., Salhab, M., and Jafino, B. A. (2022). Flood exposure and poverty in 188 countries. *Nat. Commun.* 13:3527. doi: 10.1038/s41467-022-30727-4
- Saravanan, S., and Abijith, D. (2022). Flood susceptibility mapping of northeast coastal districts of tamil nadu india using multi-source geospatial data and machine learning techniques. *Geocarto Int.* 37, 15252–15281. doi: 10.1080/10106049.2022.2096702
- Serviço Geológico do Brasil (SGB-CPRM) (2012–2019). *Geosgb: Áreas com suscetibilidade a inundação*. Available online at: <https://geoportal.sgb.gov.br/geosgb/> (Accessed July 1, 2026).
- Sokolova, M., and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Inf. Proc. Manag.* 45, 427–437. doi: 10.1016/j.ipm.2009.03.002
- Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., et al. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature* 596, 80–86. doi: 10.1038/s41586-021-03695-w
- Tran, T. X., Liu, S., Ha, H., Bui, Q. D., Nguyen, L. Q., Nguyen, D. Q., et al. (2024). A spatial landslide risk assessment based on hazard, vulnerability, exposure, and adaptive capacity. *Sustainability* 16:9574. doi: 10.3390/su16219574
- Waleed, M., and Sajjad, M. (2025a). 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:e13047. doi: 10.1111/jfr.13047
- Waleed, M., and Sajjad, M. (2025b). High-resolution flood susceptibility mapping and exposure assessment in pakistan: an integrated artificial intelligence, machine learning and geospatial framework. *Int. J. Disast. Risk Reduct.* 121:105442. doi: 10.1016/j.ijdr.2025.105442
- Yang, J., Chen, S., Duan, Z., Tang, Y., and Lu, P. (2025). Enhancing flood susceptibility predictions by using certainty factor in non-flood selection: a case study of guangdong province with four tree-based machine learning models. *Stochastic Environ. Res. Risk Assess.* 39, 3123–3146. doi: 10.1007/s00477-025-03012-1
- Zhao, J., Zhang, C., Wang, J., Abbas, Z., and Zhao, Y. (2024). Machine learning and shap-based susceptibility assessment of storm flood in rapidly urbanizing areas: a case study of Shenzhen, China. *Geomat. Natur. Hazards Risk* 15:2311889. doi: 10.1080/19475705.2024.2311889
- Zhu, X., Guo, H., and Huang, J. J. (2024). Urban flood susceptibility mapping using remote sensing, social sensing and an ensemble machine learning model. *Sustain. Cities Soc.* 108:105508. doi: 10.1016/j.scs.2024.105508