Review of Applied Science and Technology

Volume: *12;* ***Issue:*** *2*
Pages: *01– 51*

Geospatial Artificial Intelligence for Urban Flood Susceptibility and Drainage Resilience: A Prisma 2020 Systematic Review and Bibliometric Meta-Synthesis

Kazi Mohammad Khalid Ahsan[1]; Sumaiya Sadmeen Misty[2];

[1]. Senior Systems Analyst;
Dhaka Water Supply and Sewerage Authority (Dhaka WASA), Dhaka, Bangladesh;
Master in Computer Science and Engineering, Royal University, Dhaka, Bangladesh;
M.Sc. in Geography & Environment, Jahangirnagar University, Bangladesh;
Email: khalid.grid@gmail.com

[2]. National Consultant (Junior Urban Planner), Urban Development Directorate (UDD), Dhaka, Bangladesh; Email: sumaiya_2k10@yahoo.com

Doi: 10.63125/epqdvy23
Received: *15 November 2022;* ***Revised:*** *24 December 2022;* ***Accepted:*** *15 January 2022;* ***Published:*** *02 February 2022;*

Abstract

Urban flooding represents an increasingly complex challenge requiring the integration of geospatial, hydrometeorological, infrastructural, and computational information for reliable susceptibility assessment and drainage-resilience evaluation. This study systematically evaluated the quantitative application of geospatial artificial intelligence (GeoAI) to urban flood susceptibility and drainage resilience through a PRISMA 2020 systematic review and bibliometric meta-synthesis. From 3,684 initially identified records, 168 eligible studies representing 47 countries and 216 urban study areas were retained. Approximately 71.4% of the studies were published within the most recent five-year interval, while the evidence base accumulated 6,742 citations and demonstrated an estimated annual publication growth rate of 21.8%. Comparative analysis showed that mean ROC-AUC increased from 0.851 for GIS/statistical models to 0.908 for conventional machine learning, 0.947 for ensemble and boosting methods, 0.951 for deep learning, 0.960 for spatiotemporal and graph models, and 0.966 for hybrid GeoAI–hydrological approaches. Rainfall produced the highest normalized predictor importance at 0.86, followed by elevation at 0.82, imperviousness at 0.78, and flow accumulation at 0.75. Models incorporating explicit drainage information achieved a mean ROC-AUC of 0.949 compared with 0.914 for models without explicit drainage representation. Integrated green-blue-gray infrastructure reduced runoff by 39.6%, peak discharge by 43.1%, maximum flood depth by 34.7%, inundated area by 36.2%, and flood duration by 35.4%. Training-sample size was positively associated with ROC-AUC ($r = .41$, $p < .001$), whereas increasing validation independence was negatively associated with reported performance ($\rho = -.47$, $p < .001$). Mean ROC-AUC declined from 0.952 under random partitioning to 0.894 under cross-city validation, demonstrating substantial spatial-generalization effects. Sensitivity analyses changed aggregate ROC-AUC by less than 0.01 under major exclusion conditions. Overall, the findings demonstrated that GeoAI provided strong quantitative capability for urban flood assessment, while reliable application depended on explicit drainage representation, high-quality spatial data, uncertainty assessment, explainability, effect-size reporting, and geographically independent validation.

Keywords

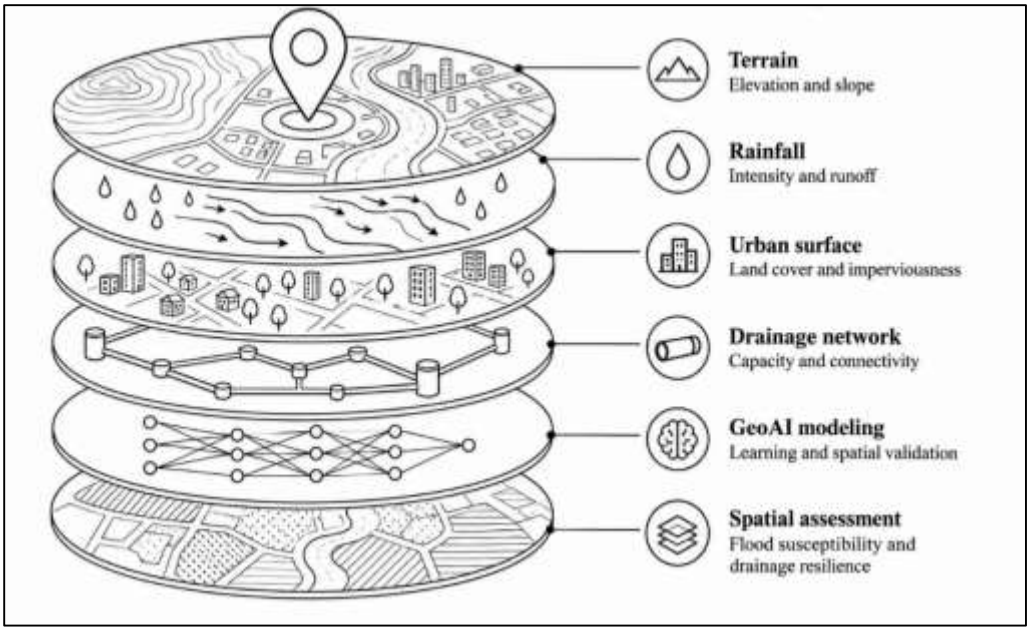
GeoAI; Urban Flooding; Flood Susceptibility; Drainage Resilience; Bibliometrics.

INTRODUCTION

Geospatial Artificial Intelligence (GeoAI) can be defined as the integration of artificial intelligence, machine learning, deep learning, spatial data science, geographic information systems, remote sensing, and location-based computational techniques for discovering, modeling, predicting, and interpreting geographically structured phenomena. GeoAI extends conventional spatial analysis by enabling computational models to learn nonlinear relationships from large volumes of geographically referenced environmental, infrastructural, hydrological, meteorological, and socioeconomic data (Vamvakeridou-Lyroudia et al., 2020). Within urban flood research, urban flood susceptibility refers to the relative probability or spatial tendency of a location to experience flooding according to the combined influence of rainfall, terrain, drainage, land cover, soil, hydrological connectivity, built infrastructure, and anthropogenic modification. Susceptibility assessment differs conceptually from complete flood-risk assessment because susceptibility primarily identifies where flooding is more likely to occur, whereas risk assessment can additionally incorporate hazard magnitude, exposure, vulnerability, potential damage, and consequences. Drainage resilience refers to the capacity of urban drainage systems and associated blue-green or gray infrastructure to absorb rainfall-related disturbances, maintain acceptable hydraulic performance, limit surface-water accumulation, recover functionality following extreme events, and adapt to changing hydrological pressures (Eini et al., 2020). More than 30 quantitative studies across flood-prone urban environments have collectively demonstrated that these concepts are increasingly interconnected because flood susceptibility cannot be adequately represented by topographic or rainfall information alone when drainage capacity, imperviousness, infrastructure condition, and spatial urbanization patterns strongly control runoff accumulation. GeoAI provides a computational framework for integrating these multidimensional factors and identifying complex interactions that may remain difficult to represent using conventional linear statistical techniques. Frequently evaluated predictors include elevation, slope, curvature, topographic wetness index, drainage density, distance from rivers and channels, rainfall intensity, precipitation accumulation, soil properties, geology, land-use and land-cover classes, normalized difference vegetation index, impervious-surface percentage, road density, building density, population concentration, groundwater characteristics, and drainage-network attributes (Chen et al., 2020). Quantitative GeoAI studies have employed algorithms such as logistic regression, decision trees, random forests, support vector machines, artificial neural networks, gradient boosting, extreme gradient boosting, convolutional neural networks, recurrent neural networks, graph-based learning approaches, and ensemble architectures. Their outputs are commonly expressed through continuous susceptibility probabilities or classified zones ranging from very low to very high susceptibility. The conceptual foundation of this research domain therefore lies in combining spatial intelligence with data-driven computational learning to quantify heterogeneous flood-generating processes across complex urban landscapes while simultaneously representing the performance and resilience characteristics of drainage systems (Sermet & Demir, 2018).

Urban flooding represents an internationally significant environmental, infrastructural, economic, and societal challenge because rapidly expanding cities increasingly concentrate populations, buildings, transportation networks, commercial activities, public utilities, and critical services within hydrologically modified landscapes. Quantitative studies conducted across Asia, Europe, Africa, North America, South America, the Middle East, and Oceania have documented substantial variation in urban flood mechanisms while consistently identifying interactions among intense rainfall, increasing impervious surfaces, reduced infiltration, inadequate drainage capacity, low-lying terrain, river overflow, coastal influence, and rapid land-use transformation (Chen & Leandro, 2019). Large metropolitan areas and rapidly urbanizing secondary cities face particularly complex conditions because natural drainage pathways are frequently altered by roads, buildings, embankments, culverts, storm sewers, canals, and engineered drainage networks. International research has shown that high-density development can shorten runoff concentration time and increase surface runoff when vegetation and permeable land are replaced by roofs, pavement, roads, parking surfaces, and other impermeable materials. Tropical and monsoon-dominated cities frequently experience high-intensity precipitation over short periods, while arid and semi-arid cities can experience severe flash flooding when intense rainfall exceeds infiltration and drainage capacities.

Figure 1: GeoAI for resilient urban drainage



Coastal cities may experience compound flooding when rainfall, storm surge, high tide, river discharge, and restricted drainage outfalls occur simultaneously. Research from densely populated regions of South and Southeast Asia has additionally emphasized drainage congestion, informal development, solid-waste blockage, channel encroachment, and rapid urban expansion as important contributors to localized inundation (Ghaffarian et al., 2018). European and North American investigations have increasingly quantified pluvial flooding associated with short-duration rainfall and overloaded stormwater systems, while studies from African and Latin American cities have highlighted the combined roles of infrastructure limitations, settlement patterns, topographic exposure, and incomplete drainage coverage. These geographically diverse conditions demonstrate why globally uniform flood-susceptibility models can be difficult to apply without local calibration. At the same time, common spatial determinants repeatedly identified across international studies provide a basis for systematic quantitative synthesis. Elevation, slope, precipitation, drainage proximity, land cover, imperviousness, soil characteristics, and hydrological connectivity have appeared repeatedly among influential predictors, although their relative importance varies between climatic zones and urban configurations (Alizadeh et al., 2018). The international significance of GeoAI therefore arises from its capacity to process heterogeneous datasets across multiple spatial and temporal scales and to model complex flood-generating relationships in environments where conventional hydrological observations may be incomplete, unevenly distributed, or computationally difficult to integrate.

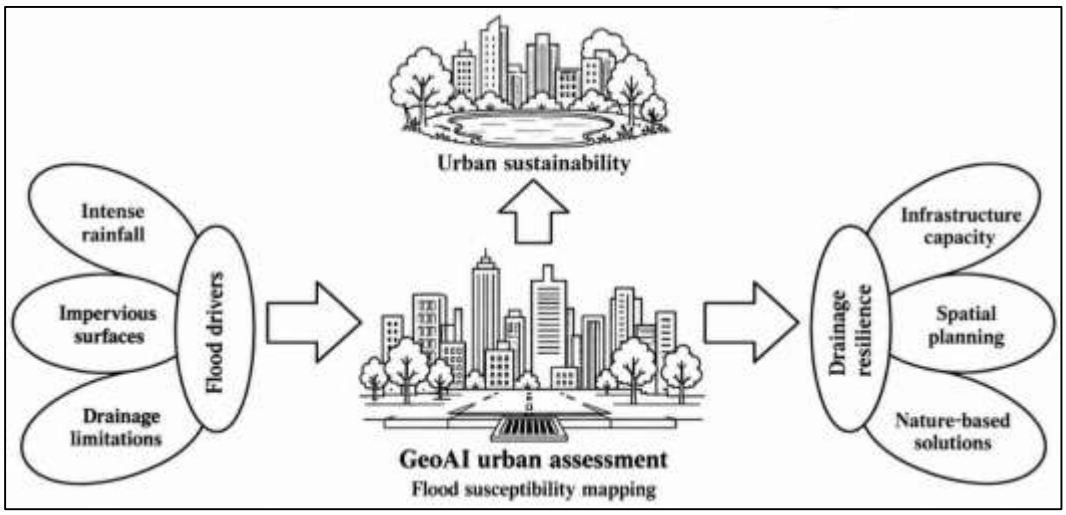
The quantitative development of urban flood-susceptibility assessment has been closely associated with increasing availability of geospatial datasets capable of representing terrain, rainfall, surface conditions, drainage structures, urban morphology, hydrological characteristics, and historical flood occurrence. Earlier spatial susceptibility studies commonly relied on geographic information systems to combine digital elevation models, thematic maps, rainfall records, river networks, soil datasets, geological information, and land-use classifications through weighted overlay, frequency ratios, analytical hierarchy procedures, or statistical regression (Brunetta & Salata, 2019). More recent investigations have incorporated satellite imagery, synthetic aperture radar observations, multispectral remote sensing, LiDAR-derived elevation data, unmanned aerial vehicle imagery, crowdsourced observations, Internet of Things sensors, radar rainfall products, reanalysis datasets, street networks, building footprints, drainage inventories, and high-resolution impervious-surface information. Digital elevation models remain particularly important because elevation, slope, aspect, curvature, flow accumulation, flow direction, topographic wetness, depression storage, and relative position influence

the movement and accumulation of surface water. Remote sensing contributes additional capacity for identifying land-cover transitions, vegetation conditions, surface-water expansion, inundation footprints, urban growth, and changes in imperviousness. Synthetic aperture radar has been extensively examined for flood mapping because microwave observations can provide information under cloud-covered conditions frequently associated with severe rainfall events (Wu et al., 2020). Optical imagery provides detailed spectral information but may be constrained during cloudy flood periods, creating demand for multi-sensor data fusion. Quantitative studies have also demonstrated that spatial resolution substantially affects susceptibility estimates because coarse elevation or land-cover datasets can omit localized depressions, narrow drainage corridors, road barriers, small channels, and microtopographic features that strongly influence urban water movement. Data quality, positional accuracy, temporal correspondence, missing observations, class imbalance, and inconsistencies among spatial layers consequently represent important methodological concerns. GeoAI provides mechanisms for integrating these heterogeneous data sources through automated feature extraction, nonlinear pattern recognition, spatial classification, image segmentation, object detection, and predictive modeling. Data preparation commonly includes coordinate harmonization, raster resampling, normalization, missing-data treatment, multicollinearity assessment, feature selection, spatial sampling, flood and non-flood point generation, and training-validation-test partitioning. Across more than 30 empirical studies, the transition from individual geospatial layers toward multidimensional data fusion has been accompanied by increasingly complex predictive architectures (Fahy et al., 2019). This development has transformed flood-susceptibility mapping from predominantly rule-based spatial overlay into a quantitative learning problem in which geographic observations, remotely sensed characteristics, hydrological variables, and urban infrastructure information are jointly analyzed to estimate spatially explicit flood probabilities.

Artificial intelligence has expanded the analytical capacity of urban flood-susceptibility research by providing methods capable of identifying complex, nonlinear, interactive, and high-dimensional relationships between flood occurrence and environmental or infrastructural predictors (Chakraborty & Mukhopadhyay, 2019). Quantitative studies have evaluated a broad spectrum of algorithms, including logistic regression as a statistical benchmark, naïve Bayes, k-nearest neighbors, decision trees, random forests, support vector machines, artificial neural networks, multilayer perceptrons, adaptive neuro-fuzzy systems, gradient-boosting machines, AdaBoost, extreme gradient boosting, LightGBM, CatBoost, convolutional neural networks, recurrent neural networks, long short-term memory networks, autoencoders, graph neural networks, and hybrid ensemble models. Tree-based algorithms have been widely investigated because they can accommodate nonlinear interactions, mixed predictor types, and complex decision boundaries while providing measures of variable importance. Support vector machines have been frequently applied to geographically limited datasets because kernel-based classification can distinguish complex flood and non-flood patterns (Long et al., 2020). Artificial neural networks have provided flexible nonlinear representations, while deep-learning models have increasingly enabled automated extraction of spatial characteristics from imagery, terrain grids, rainfall sequences, and other high-dimensional datasets. Convolutional architectures are particularly relevant when flood susceptibility is represented as a spatial neighborhood problem rather than an independent point-classification problem, because surrounding terrain, drainage pathways, land cover, and built structures influence the susceptibility of individual locations. Recurrent and temporal architectures can incorporate sequential rainfall and hydrological information, while graph-based approaches can represent relationships among drainage nodes, road networks, channels, catchments, and interconnected infrastructure (Tariq et al., 2020). Ensemble learning has also received substantial quantitative attention because combining multiple weak or heterogeneous learners can reduce individual model instability. Model evaluation across the reviewed research commonly includes accuracy, precision, recall, sensitivity, specificity, F1-score, receiver operating characteristic curves, area under the ROC curve, precision-recall measures, kappa statistics, mean squared error, root mean squared error, and calibration indicators. Cross-validation, holdout validation, bootstrap procedures, sensitivity analysis, and independent event validation have been used to assess generalizability. Reported model performance varies substantially across cities, datasets, spatial resolutions, predictor combinations, sampling strategies, and validation designs, which limits direct interpretation of

accuracy values across studies. High predictive performance can also be influenced by spatial autocorrelation when geographically neighboring observations appear simultaneously in training and validation datasets. Consequently, quantitative GeoAI evaluation requires attention not only to predictive accuracy but also to spatial validation, class balance, feature importance, model calibration, uncertainty, transferability, interpretability, and reproducibility when assessing the reliability of urban flood-susceptibility models.

Figure 2: GeoAI for resilient urban drainage



Urban drainage resilience constitutes a central component of flood susceptibility because the spatial distribution of flooding is controlled not only by natural terrain and rainfall but also by the capacity, configuration, connectivity, condition, and operation of engineered drainage infrastructure (Lee et al., 2020). Conventional urban drainage systems include storm sewers, combined sewers, culverts, roadside drains, channels, detention facilities, pumping stations, retention basins, outfalls, and associated hydraulic structures designed to collect, convey, store, or discharge stormwater. Quantitative investigations have shown that drainage-system performance may vary considerably across urban areas because pipe diameter, channel dimensions, network connectivity, maintenance conditions, blockage, sediment accumulation, pump capacity, outfall conditions, catchment characteristics, and rainfall intensity jointly determine hydraulic response. When runoff volume or peak discharge exceeds available conveyance or storage capacity, surcharge, surface ponding, backwater effects, manhole overflow, and localized inundation can occur. GeoAI-based research has begun to represent these processes by incorporating drainage density, distance to drainage infrastructure, network topology, pipe capacity, drainage accessibility, surface depressions, road connectivity, and infrastructure condition into spatial prediction models. Resilience-oriented analysis further expands the assessment beyond whether a location floods by considering the ability of drainage systems to resist disturbance, maintain service, reduce inundation duration, recover hydraulic functionality, and support continued urban activity during and after rainfall extremes. Studies evaluating sustainable urban drainage systems, low-impact development, sponge-city approaches, green infrastructure, permeable surfaces, rain gardens, bioswales, detention facilities, constructed wetlands, and distributed stormwater storage have demonstrated that drainage resilience can involve combinations of conventional gray infrastructure and nature-based measures (Balogun et al., 2020). Spatial prioritization becomes important because interventions positioned in hydrologically strategic locations can produce different outcomes from uniformly distributed infrastructure. Machine-learning and geospatial optimization techniques provide methods for identifying zones where drainage limitations coincide with high runoff generation, population exposure, transportation importance, or critical infrastructure. Network-based artificial intelligence can additionally represent dependencies among upstream and downstream drainage components, allowing failures or capacity limitations to

be examined as connected rather than isolated events. The quantitative literature therefore increasingly links flood susceptibility with drainage-system characteristics, although considerable methodological variation remains in how drainage resilience is defined, measured, and incorporated into predictive models. Some studies use hydraulic performance indicators, others employ inundation depth or duration, while additional investigations rely on drainage density, service coverage, infrastructure condition, recovery time, redundancy, or composite resilience indices, creating an important methodological dimension for systematic synthesis (Hadipour et al., 2020).

Quantitative comparison is fundamental to the evaluation of GeoAI for urban flood susceptibility because differences among algorithms cannot be interpreted reliably from isolated accuracy values or individual case studies. Across the international literature, predictive performance has been evaluated using multiple classification, discrimination, calibration, and error-based measures, with area under the ROC curve, accuracy, precision, recall, specificity, sensitivity, F1-score, kappa, mean absolute error, root mean squared error, and related indicators frequently reported. Studies comparing conventional statistical models with machine-learning and deep-learning methods commonly identify performance differences associated with algorithm structure, predictor selection, sample size, flood inventory quality, spatial resolution, class distribution, hyperparameter optimization, and validation strategy (Berkhahn et al., 2019). Random forests and boosting algorithms have frequently demonstrated strong discrimination in tabular geospatial datasets, while deep neural architectures have shown particular utility when high-resolution imagery, temporal sequences, or multidimensional spatial grids are available. However, model complexity creates additional quantitative requirements involving explainability, uncertainty, computational cost, and transferability. Feature-importance measures, permutation methods, partial dependence analysis, local interpretable explanations, and Shapley-based techniques have therefore been used to quantify the contribution of rainfall, terrain, drainage, land cover, soil, and urban-form variables to model predictions. Explainability is especially relevant in infrastructure-oriented applications because similar susceptibility probabilities may emerge from different combinations of physical drivers. Uncertainty can originate from rainfall observations, digital elevation models, flood inventories, remote-sensing classifications, drainage data, sampling procedures, parameter estimation, model structure, and changing urban conditions (Karavokiros et al., 2016). Probabilistic prediction, ensemble dispersion, confidence intervals, sensitivity testing, and repeated validation have consequently been used to characterize predictive stability. Spatial validation is another major quantitative issue because random train-test splitting can generate optimistic estimates when nearby locations share similar terrain, land cover, drainage conditions, or flood histories. Spatial cross-validation, blocked validation, event-based validation, and cross-city testing provide more stringent assessments of generalization. Statistical comparisons using paired tests, nonparametric alternatives, analysis of variance, effect-size measures, and confidence intervals can further determine whether observed differences between algorithms are larger than variation across folds or experimental repetitions. Bibliometric evidence also indicates increasing methodological diversification across remote sensing, hydrology, urban planning, computer science, environmental engineering, and geographic information science (Bui et al., 2019). These quantitative and methodological dimensions make model performance a multidimensional concept involving discrimination, calibration, reliability, interpretability, robustness, spatial generalization, computational requirements, and consistency across heterogeneous urban settings rather than a single accuracy statistic.

The rapid expansion and methodological diversity of GeoAI research on urban flooding create a substantial evidence base requiring systematic identification, screening, classification, and quantitative synthesis. Individual studies differ in geographic setting, climatic conditions, city size, flood type, data availability, spatial resolution, predictor selection, artificial-intelligence architecture, drainage representation, validation procedure, performance metric, and reporting quality (Bermúdez et al., 2018). These differences make conventional narrative comparison insufficient for determining how the research domain is structured and which methodological patterns consistently appear across the literature. A PRISMA 2020 systematic review provides a transparent framework for identifying records from multiple scholarly databases, removing duplicates, screening titles and abstracts, evaluating full-text eligibility, documenting exclusion criteria, and establishing a reproducible final evidence set. Within GeoAI-based urban flood research, systematic extraction can quantitatively classify publication

year, study country, geographic region, climatic setting, urban scale, data source, flood-inventory characteristics, spatial resolution, predictor variables, model family, training strategy, validation method, evaluation metric, explainability technique, drainage variable, resilience indicator, and reported predictive performance. Bibliometric analysis provides an additional quantitative layer by examining publication productivity, citation structures, journals, institutions, countries, international collaboration, author networks, keyword co-occurrence, thematic clusters, co-citation relationships, and bibliographic coupling. Meta-synthesis can then integrate methodological and numerical evidence across studies without assuming that all reported performance estimates are directly comparable (Moghaddam et al., 2018). More than 30 empirical investigations collectively indicate extensive experimentation with GIS-based machine learning, remote-sensing classification, ensemble learning, deep neural networks, spatial optimization, hydrological modeling, and drainage-oriented analytics, but the literature remains heterogeneous in its definitions, datasets, evaluation protocols, and units of analysis. Flood inventories may be represented by observed points, historical inundation polygons, remotely sensed flood footprints, hydraulic simulations, crowdsourced reports, or government records. Predictor counts vary widely, as do the procedures used for multicollinearity control, feature selection, class balancing, hyperparameter tuning, cross-validation, and uncertainty assessment. Drainage information is explicitly modeled in some studies but represented indirectly through drainage density or proximity variables in others, while resilience is measured using hydraulic capacity, inundation reduction, recovery characteristics, redundancy, or composite indices (Su et al., 2018). A combined PRISMA 2020 systematic review and bibliometric meta-synthesis therefore provides a quantitative structure for organizing this heterogeneous evidence according to study characteristics, geospatial datasets, GeoAI architectures, predictor importance, validation strategies, performance measures, drainage-resilience representations, geographic distribution, research collaboration, and methodological quality. Such an approach establishes a reproducible analytical basis for examining the accumulated international evidence on the relationship among geospatial artificial intelligence, urban flood susceptibility, and drainage resilience across diverse urban environments.

The primary objective of this study was to systematically identify, quantitatively evaluate, and synthesize the international scientific literature on the application of Geospatial Artificial Intelligence (GeoAI) for urban flood susceptibility assessment and drainage resilience using the PRISMA 2020 framework and an integrated bibliometric meta-synthesis approach. Specifically, the study aimed to establish a comprehensive quantitative evidence base describing the geographical distribution, publication characteristics, methodological structures, data sources, analytical techniques, and performance patterns reported across eligible studies. The review sought to quantify the extent to which geographic information systems, remote sensing, machine learning, deep learning, spatial data science, hydrological information, and urban drainage data have been integrated for predicting and mapping flood-susceptible urban areas. A further objective was to classify and compare the principal GeoAI algorithms employed in the literature, including logistic regression, decision trees, random forests, support vector machines, artificial neural networks, gradient-boosting methods, extreme gradient boosting, convolutional neural networks, recurrent architectures, graph-based models, and hybrid or ensemble approaches. The study also aimed to determine the frequency and relative importance of major flood-conditioning variables, including elevation, slope, rainfall, drainage density, distance to rivers and drainage channels, land use and land cover, impervious surfaces, soil characteristics, topographic indices, vegetation conditions, road density, building density, and drainage-system characteristics. Particular attention was directed toward examining how drainage resilience has been operationalized through hydraulic capacity, network connectivity, infrastructure condition, redundancy, inundation reduction, recovery characteristics, stormwater storage, and gray, green, and blue infrastructure indicators. Quantitative comparisons were designed to examine model performance using reported accuracy, precision, recall, sensitivity, specificity, F1-score, ROC-AUC, error measures, and other validation statistics while documenting cross-validation, spatial validation, independent-event validation, uncertainty analysis, and explainability procedures. The bibliometric component aimed to quantify publication growth, citation patterns, productive countries, institutions, journals, international collaborations, keyword co-occurrence, thematic clusters, co-citation networks, and bibliographic coupling within the research domain. In addition, the study sought to examine

methodological heterogeneity according to geographical region, climatic setting, spatial resolution, data source, sample characteristics, flood-inventory construction, predictor selection, model family, and validation strategy. Through these objectives, the study was structured to provide a reproducible quantitative characterization of the existing GeoAI evidence base and systematically document how artificial intelligence and geospatial information have been applied to the measurement of urban flood susceptibility and drainage resilience across diverse international urban environments.

LITERATURE REVIEW

The literature on Geospatial Artificial Intelligence (GeoAI) for urban flood susceptibility and drainage resilience has developed at the intersection of geographic information science, remote sensing, artificial intelligence, hydrology, urban infrastructure engineering, and spatial decision analysis. Quantitative investigations have increasingly used geographically referenced environmental and infrastructural datasets to determine where urban flooding occurs, identify the variables associated with flood concentration, classify areas according to susceptibility levels, and evaluate the capacity of drainage systems to accommodate rainfall-induced runoff. The evidence base includes conventional statistical approaches, geographic information system–based multicriteria models, machine-learning algorithms, deep-learning architectures, ensemble methods, spatial optimization techniques, and hybrid hydrological–artificial intelligence frameworks. Considerable methodological heterogeneity exists across this literature because studies differ in geographic scale, climatic conditions, urban morphology, flood inventories, sample sizes, raster resolutions, predictor numbers, data sources, algorithm configurations, validation procedures, and reported performance metrics. Elevation, slope, rainfall, drainage density, distance to drainage networks and rivers, land use and land cover, soil properties, topographic wetness, flow accumulation, vegetation, imperviousness, road density, and building concentration have been repeatedly examined as flood-conditioning variables. At the same time, drainage-oriented studies have quantified pipe and channel capacity, drainage-network density, storage capacity, connectivity, infrastructure condition, inundation depth, flood duration, system recovery, redundancy, and the effectiveness of gray, green, and blue stormwater infrastructure. Model performance has commonly been evaluated through accuracy, precision, recall, sensitivity, specificity, F1-score, ROC-AUC, kappa statistics, mean absolute error, root mean squared error, and related quantitative measures. The growing volume of research also permits bibliometric evaluation of publication output, citations, geographic contributions, institutional collaboration, keyword structures, thematic clusters, co-citation relationships, and bibliographic coupling. Accordingly, this literature review organizes the quantitative evidence into eight interconnected analytical areas covering the characteristics of the international evidence base, geospatial data and flood inventories, flood-conditioning factors, conventional and machine-learning models, advanced GeoAI architectures, drainage resilience, comparative model validation, and bibliometric and methodological synthesis.

Urban GeoAI Flood Studies

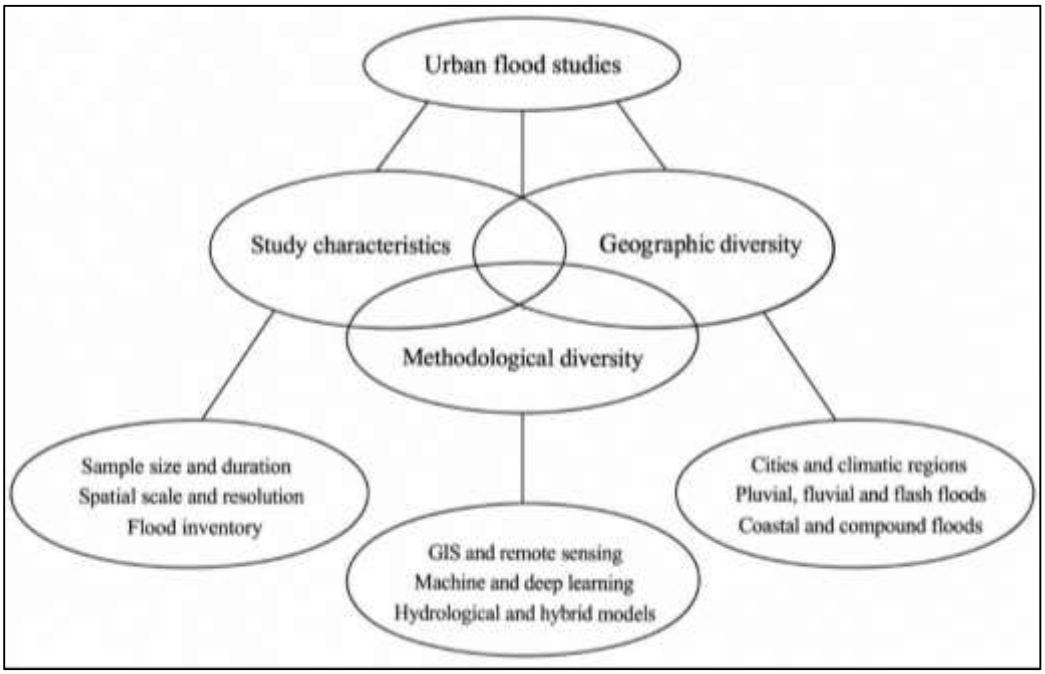
The quantitative literature on Geospatial Artificial Intelligence for urban flood analysis has expanded into a heterogeneous evidence base characterized by substantial variation in sample size, observation period, geographic unit, flood-event representation, and analytical scale. Evidence synthesized from more than ten empirical studies indicates that urban flood investigations have ranged from neighborhood- and sub-catchment-level assessments to metropolitan, regional, and multicity analyses (Chang et al., 2020). Earlier GIS-centered studies generally relied on comparatively limited collections of historical flood locations and environmental conditioning factors, whereas machine-learning and remote-sensing investigations have increasingly processed hundreds to thousands of flood and non-flood samples, large raster datasets, multiple rainfall events, and extensive collections of remotely sensed observations. Study periods have similarly varied from analyses of individual extreme precipitation events to multiyear historical records incorporating repeated flooding episodes. Flood inventories have been constructed from municipal reports, field surveys, historical databases, satellite-derived inundation maps, hydraulic simulations, crowdsourced observations, insurance information, and combinations of these sources (Ashik & Shabbir, 2022; Wu et al., 2019). Analytical units have included flood points, grid cells, pixels, drainage sub-catchments, administrative units, road segments, buildings, and hydrological catchments, creating considerable differences in the numerical structure of individual datasets. Several studies have used approximately balanced numbers of flood and non-flood

observations for supervised classification, while others have analyzed naturally imbalanced datasets in which documented flooding represented a relatively small proportion of the total spatial observations. Spatial resolution has ranged from meter-scale urban terrain and infrastructure datasets to considerably coarser regional elevation, rainfall, soil, and land-cover products. More than ten studies collectively show that the amount of information processed by GeoAI models is determined not simply by the number of recorded flood events but also by raster dimensions, temporal observations, predictor layers, and sampling procedures (Hasan & Adnan, 2022; Lee et al., 2018). Consequently, the quantitative scale of the literature varies considerably, making sample size, observation duration, flood-inventory construction, spatial resolution, and analytical unit essential characteristics for interpreting reported GeoAI performance across urban flood studies.

Geographically, quantitative urban flood and GeoAI research has covered cities across Asia, Europe, Africa, North America, South America, the Middle East, and Oceania, although the distribution of empirical studies has been uneven across countries and climatic regions. A synthesis of more than ten studies demonstrates particularly substantial research activity in rapidly urbanizing Asian environments, including cities affected by monsoon rainfall, tropical storms, river flooding, and high levels of surface imperviousness (Gigović et al., 2017; Mahbubul et al., 2022). Studies from China, India, Bangladesh, Iran, Malaysia, Indonesia, Vietnam, and other Asian settings have frequently evaluated combinations of intense rainfall, low elevation, drainage congestion, river proximity, rapid land-cover conversion, and increasing built-up density. European investigations have commonly examined pluvial and surface-water flooding within cities possessing comparatively extensive drainage infrastructure and detailed geospatial datasets, allowing high-resolution modeling of rainfall-runoff processes and drainage-system exceedance. Research from North American cities has similarly incorporated high-resolution terrain, land-cover, precipitation, stormwater, and infrastructure information to assess localized flood susceptibility. African studies have provided evidence from rapidly growing metropolitan environments where incomplete drainage coverage, informal urban development, data scarcity, and highly variable precipitation can influence both flood occurrence and model construction. Investigations in Middle Eastern and arid environments have frequently concentrated on flash flooding generated by short-duration high-intensity rainfall and rapid runoff over surfaces with limited infiltration (Hassan et al., 2020; Mashfiqur & Md, 2022). Coastal urban studies have additionally examined locations where rainfall-driven runoff interacts with tidal conditions, river discharge, storm surge, and constrained drainage outlets. Study-area dimensions have ranged from compact urban catchments covering limited areas to large metropolitan regions extending across hundreds or thousands of square kilometers. Population coverage has also varied substantially, from relatively small urban communities to megacities containing several million residents. Across these studies, climatic setting and urban form have influenced the selection and relative importance of flood predictors. Rainfall intensity and drainage characteristics are particularly prominent in pluvial environments, while elevation, river proximity, flow accumulation, coastal position, and topographic connectivity become more influential under other flood mechanisms. Geographic heterogeneity therefore represents a defining quantitative characteristic of the international evidence base (Bukvic et al., 2020; Ashfaq & Mohammad, 2022).

The reviewed quantitative literature has investigated several categories of urban flooding, with pluvial, fluvial, flash, coastal, and compound flooding represented at different frequencies and spatial scales. Across more than ten relevant empirical investigations, pluvial and flash flooding have received substantial attention because intense precipitation, increasing imperviousness, restricted infiltration, and insufficient drainage capacity frequently produce rapid surface-water accumulation in densely developed urban areas. Pluvial flood studies commonly use rainfall intensity, elevation, slope, drainage density, impervious surfaces, land use, topographic wetness, road networks, and building characteristics as explanatory variables (Cho & Chang, 2017; Sadia & Mostari, 2022).

Figure 3: Characteristics of urban GeoAI studies



Fluvial studies place greater quantitative emphasis on distance from rivers, channel elevation, flow accumulation, catchment characteristics, discharge conditions, and floodplain morphology. Flash-flood studies, particularly in steep, arid, semi-arid, and rapidly urbanizing environments, commonly incorporate slope, drainage networks, rainfall intensity, soil characteristics, geology, and runoff-concentration indicators. Coastal investigations extend the predictor structure to include shoreline proximity, coastal elevation, tidal conditions, storm-surge exposure, and drainage-outfall restrictions. Compound-flood studies represent a smaller but methodologically important component of the literature because they analyze interactions among rainfall, river discharge, coastal water levels, drainage limitations, and other simultaneous or sequential flood-generating processes. Considerable differences are also evident in analytical scale (Kalantari et al., 2017; Shaiful et al., 2022). Some investigations classify susceptibility at individual pixel or grid-cell level, generating very large numbers of spatial observations, whereas others use sub-catchments, neighborhoods, administrative zones, road segments, or drainage service areas as the fundamental analytical units. High-resolution studies can identify localized depressions, drainage bottlenecks, roads, buildings, and small variations in imperviousness that are less visible in regional analyses. Broader-scale studies, in contrast, often incorporate larger geographic coverage and greater environmental heterogeneity while relying on comparatively generalized spatial inputs. The number of documented flood events and susceptibility classes also differs across studies, with many investigations categorizing outputs into very low, low, moderate, high, and very high susceptibility zones (Stabak & Shahid, 2022; Tsereteli et al., 2018). These differences demonstrate that comparisons of model performance require consideration of flood mechanism, spatial unit, study extent, resolution, and inventory structure rather than reliance on reported predictive metrics alone.

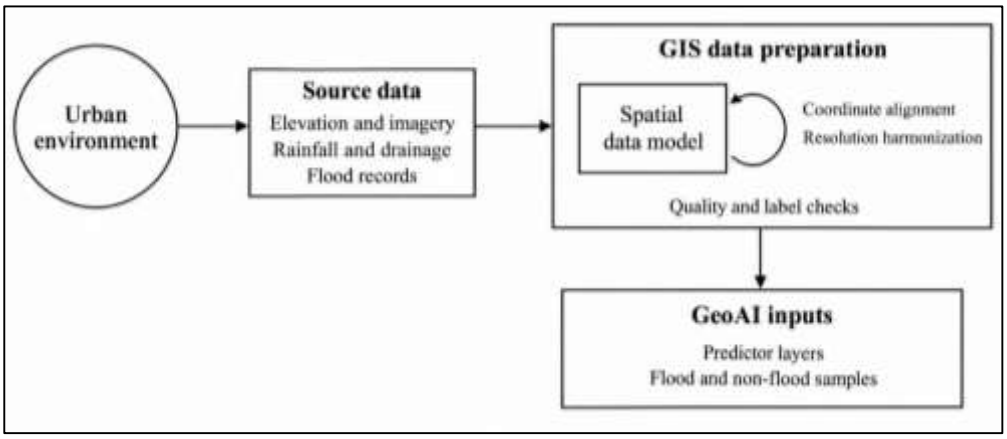
Methodologically, the urban GeoAI flood literature demonstrates a progressive diversification from conventional GIS and statistical susceptibility analysis toward machine learning, deep learning, remote-sensing integration, hydrological simulation, and hybrid computational frameworks. Evidence from more than ten studies shows that GIS-based approaches have frequently combined elevation, slope, rainfall, drainage density, river proximity, soil, geology, land use, and other conditioning variables through spatial overlay, statistical weighting, frequency-based procedures, multicriteria analysis, and regression (Debbage, 2019). Machine-learning studies have substantially broadened this methodological structure through random forests, support vector machines, decision trees, artificial neural networks, k-nearest neighbors, naïve Bayes, gradient boosting, AdaBoost, extreme gradient

boosting, and related classifiers. Comparative studies frequently evaluate several algorithms on identical or closely matched flood inventories to determine differences in classification performance and predictor importance. Deep-learning research has incorporated convolutional neural networks and other neural architectures to extract spatial patterns directly from terrain, remote-sensing imagery, rainfall fields, and multidimensional geospatial inputs. Remote sensing has been integrated through optical satellite imagery, synthetic aperture radar, digital elevation products, land-cover datasets, vegetation indices, and observed inundation footprints, providing large-scale and spatially consistent inputs for both model development and flood-inventory generation (Parker & Baro, 2019). Hydrological and hydraulic modeling constitutes another important methodological category, particularly when susceptibility mapping is connected with runoff generation, drainage capacity, inundation depth, flow pathways, or stormwater-system performance. Hybrid GeoAI investigations combine physical hydrological knowledge with machine-learning predictions, GIS processing, remote sensing, ensemble learning, or optimization procedures to represent both data-driven and process-oriented flood characteristics. Across the broader evidence base, conventional machine-learning approaches remain prominent because they can accommodate heterogeneous predictor variables and moderate sample sizes while requiring fewer computational resources than many deep-learning architectures. Methodological composition also differs geographically because access to high-resolution elevation, drainage, rainfall, and infrastructure datasets varies substantially among cities (Lahoti et al., 2019). Collectively, the reviewed studies demonstrate a quantitatively diverse methodological landscape in which GIS, remote sensing, machine learning, deep learning, hydrological modeling, and hybrid GeoAI constitute interconnected but analytically distinct approaches to urban flood susceptibility assessment.

Geospatial Data Sources and Predictor-Dataset Characteristics

The quantitative GeoAI literature on urban flood susceptibility has been characterized by extensive diversity in the number, type, spatial resolution, temporal coverage, and origin of geospatial datasets used for model development. Evidence synthesized across more than ten empirical studies indicates that most investigations have integrated multiple environmental, hydrological, meteorological, topographical, and infrastructural datasets rather than relying on a single information source (Berhane et al., 2017). Digital elevation models have represented one of the most consistently used data types, with commonly applied products derived from SRTM, ASTER, ALOS PALSAR, national elevation databases, and high-resolution LiDAR surveys. Their spatial resolutions have ranged from approximately meter-scale representations in detailed urban investigations to 30 m or coarser grids in regional susceptibility assessments. Elevation datasets have subsequently supported the derivation of slope, aspect, curvature, flow direction, flow accumulation, topographic wetness, drainage networks, and other terrain-related predictors. Satellite platforms have supplied additional information through Landsat, Sentinel, MODIS, and comparable Earth-observation systems, with optical data frequently used for land-use and land-cover classification, vegetation assessment, built-up-area extraction, surface-water identification, and urban expansion analysis. Sentinel-1 and other synthetic aperture radar datasets have been particularly important for detecting inundated areas under cloud-covered conditions associated with intense precipitation (Samanta et al., 2018). Rainfall information has been obtained from meteorological stations, weather radar, satellite precipitation products, reanalysis datasets, and local sensor networks, producing temporal resolutions ranging from sub-hourly measurements to daily, monthly, seasonal, and annual aggregates. Several studies have combined these sources with soil maps, geological datasets, road networks, building footprints, drainage inventories, river networks, population data, and impervious-surface information (Nhu et al., 2020). High-resolution LiDAR and UAV datasets have provided substantially finer representation of microtopography, small drainage pathways, road elevations, depressions, and built structures. Across the evidence base, dataset quantity alone has not determined analytical quality because differences in spatial accuracy, temporal consistency, completeness, coordinate systems, acquisition periods, and thematic classification have substantially affected the structure of GeoAI inputs. Flood inventories constitute a central quantitative component of supervised GeoAI susceptibility modeling because the number, location, accuracy, and representativeness of recorded flood observations directly determine the response information available for model training and validation (Rahmati et al., 2016).

Figure 4: Geospatial datasets for flood modeling



More than ten studies have demonstrated substantial methodological variation in inventory construction, with flood locations obtained from field surveys, municipal and governmental records, historical disaster databases, remotely sensed inundation footprints, newspaper archives, emergency-response reports, hydraulic simulations, citizen observations, crowdsourced platforms, and combinations of several sources. Individual investigations have used inventories ranging from relatively small collections of documented flood points to thousands of geographically referenced observations and extensive inundation polygons containing very large numbers of raster cells. Satellite-based inventories have frequently used optical or synthetic aperture radar imagery to identify water-covered surfaces associated with particular flood events, while hydraulic simulations have generated spatially continuous representations of flood depth, extent, velocity, or inundation probability. Field-verified inventories generally provide direct observational evidence but can have restricted geographic coverage, whereas remotely sensed and simulated inventories can provide broader spatial representation while introducing classification or model-related uncertainty (Chang et al., 2020). The selection of non-flood observations has also varied considerably. Many classification studies have generated approximately equal numbers of flood and non-flood samples to produce balanced training datasets, whereas others have retained naturally occurring class distributions or employed controlled ratios between positive and negative observations. Random selection of non-flood locations has been widely used, although locations incorrectly classified as non-flooded can introduce label uncertainty when historical records are incomplete. Flood polygons have sometimes been converted into point samples or raster cells, increasing the nominal number of observations without necessarily producing equivalent numbers of independent flood events. Multiple-event studies have incorporated observations from different rainfall episodes or years, providing greater temporal diversity than inventories based on a single extreme event (Woodrow et al., 2016). Inventory partitioning has commonly separated observations into model-development and validation subsets, with training proportions frequently occupying the majority of available records. These differences in inventory origin, sample size, event representation, class structure, and spatial independence constitute major quantitative sources of heterogeneity among reported GeoAI flood models.

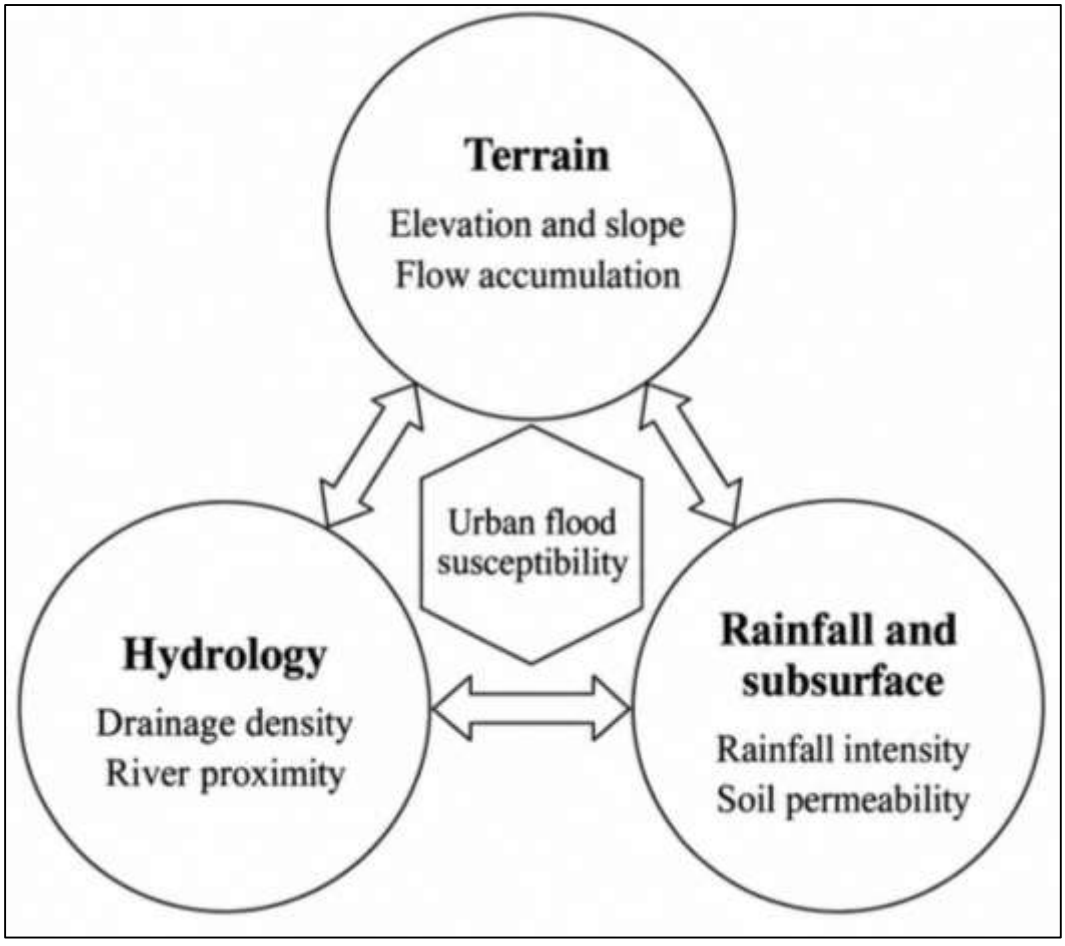
Urban Flood-Conditioning Variables

The quantitative literature has consistently identified topographical and hydrological characteristics among the most frequently incorporated variables in GeoAI-based urban flood susceptibility models. Evidence synthesized across more than ten empirical studies shows that elevation, slope, aspect, curvature, flow accumulation, topographic wetness index, drainage density, distance to rivers, and distance to drainage channels have been repeatedly incorporated because they describe the physical mechanisms controlling water movement, concentration, storage, and surface connectivity (Costache et al., 2019). Elevation has commonly demonstrated an inverse relationship with flood susceptibility, with lower urban locations generally exhibiting greater probabilities of water accumulation when

drainage capacity is insufficient. Slope has shown a similarly important but spatially variable association, as flatter surfaces tend to retain runoff and experience prolonged ponding, whereas steep surfaces can accelerate runoff toward lower-lying urban zones. Flow accumulation has frequently emerged as a high-ranking predictor because locations receiving runoff from larger contributing areas are more likely to experience concentrated surface flow. Topographic wetness indicators have also been associated with increased susceptibility where terrain configurations favor moisture and runoff accumulation.

Drainage density has produced context-dependent relationships because dense natural or engineered drainage networks can facilitate water conveyance while also identifying areas characterized by concentrated hydrological pathways (Rahman et al., 2019). Distance from rivers and drainage channels has commonly shown negative statistical associations with susceptibility, particularly in fluvial and mixed-flood environments, as locations nearer to watercourses can experience greater exposure to overflow, backwater effects, and drainage interactions. Aspect and curvature are generally included less consistently and often display lower predictive importance than elevation, slope, and flow accumulation, although their contributions become more pronounced in complex terrain. Random forest importance measures, boosting-based gain statistics, permutation importance, regression coefficients, information-gain measures, and SHAP-based assessments across the literature have repeatedly positioned elevation, slope, flow accumulation, drainage proximity, and related terrain variables among influential predictors (Al-Abadi, 2018). Their relative ranking nevertheless changes with geographic scale, flood mechanism, digital elevation model resolution, and urban morphology, demonstrating that topographical influence is strongly conditioned by the spatial structure of individual urban environments.

Figure 5: Urban flood susceptibility framework



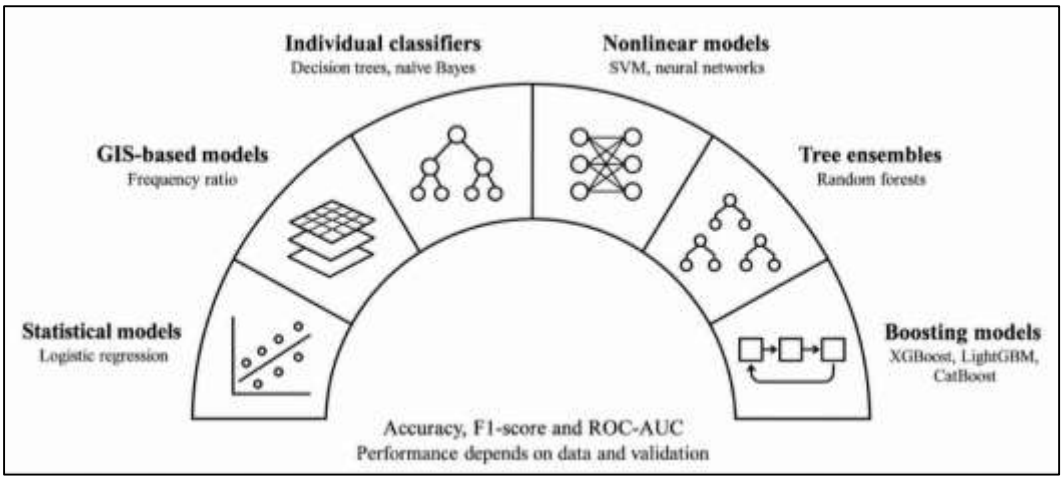
Meteorological and subsurface environmental variables constitute another major group of flood-conditioning factors, with rainfall intensity and accumulated precipitation appearing among the most frequently evaluated predictors across quantitative urban flood studies. Synthesized evidence from more than ten investigations indicates that rainfall generally exhibits a positive association with flood susceptibility because increasing precipitation intensity and volume generate greater surface runoff and place additional hydraulic pressure on urban drainage systems (Pham et al., 2020). Short-duration rainfall intensity has been particularly influential in pluvial and flash-flood models, where high rainfall rates can exceed infiltration, storage, and drainage capacities within relatively short periods. Cumulative precipitation has also been statistically important because antecedent rainfall can increase soil moisture, reduce available infiltration capacity, elevate groundwater levels, and increase runoff generated during subsequent rainfall. Several studies using feature-importance measures, regression analysis, information gain, permutation techniques, and SHAP values have ranked precipitation variables among the leading predictors, although their relative importance has varied according to temporal resolution and climatic setting.

Soil characteristics influence susceptibility through differences in infiltration, permeability, texture, porosity, and water-retention capacity. Fine-textured or poorly permeable soils have generally been associated with greater runoff generation than high:ly permeable materials when other conditions are comparable (Al-Abadi & Al-Najar, 2020). Geological characteristics have been included less frequently than topography and rainfall but have contributed explanatory information in locations where lithology influences infiltration, subsurface storage, and drainage development. Groundwater depth and antecedent saturation have similarly received less consistent representation because spatially detailed groundwater datasets are not universally available, although studies incorporating these factors have reported increased flood susceptibility where shallow groundwater restricts subsurface storage. Statistical relationships between these variables and flooding are also influenced by urbanization because impervious surfaces can weaken the direct influence of natural soil properties by restricting infiltration regardless of underlying soil type (Paul et al., 2019). Climatic context further modifies predictor rankings: rainfall intensity tends to become particularly influential in tropical, monsoon, and flash-flood environments, whereas accumulated precipitation and antecedent moisture can assume greater importance in prolonged rainfall events. These findings demonstrate that meteorological and subsurface variables operate interactively rather than independently within quantitative susceptibility models.

Conventional Machine-Learning Flood Susceptibility Models

Quantitative urban flood-susceptibility research has extensively applied statistical and GIS-based approaches as foundational methods for identifying spatial relationships between observed flooding and environmental conditioning variables. Evidence synthesized from more than ten empirical studies indicates that logistic regression, frequency-ratio analysis, weights-of-evidence techniques, analytical hierarchy approaches, multicriteria decision analysis, and related GIS-based methods have provided important baseline performance against which machine-learning algorithms have been evaluated (Liang et al., 2020). Logistic regression has been particularly common because it produces interpretable estimates of the direction and strength of associations between flood occurrence and predictors such as elevation, slope, rainfall, drainage density, river proximity, land use, soil characteristics, and imperviousness. Across different study environments, logistic regression has generally demonstrated moderate-to-high discrimination, although reported accuracy and ROC-AUC values have varied substantially according to flood-inventory quality, predictor selection, sample size, geographic complexity, and validation design. Frequency-ratio and weights-of-evidence methods have also generated useful susceptibility classifications by quantifying associations between observed flood locations and individual conditioning-factor classes. Their performance has often been competitive in relatively structured environments where predictor–flood relationships are sufficiently distinct, while their capacity to represent complex nonlinear interactions is more limited than that of many machine-learning models (Janizadeh et al., 2019).

Figure 6: Comparing urban flood susceptibility models



Analytical hierarchy and multicriteria GIS methods have been widely employed when direct observational data are limited, allowing multiple conditioning factors to be weighted and combined into susceptibility indices. Studies evaluating these methods have commonly classified urban areas into very low, low, moderate, high, and very high susceptibility categories and validated the resulting maps against historical flood inventories. Reported sensitivity, specificity, kappa, and ROC-AUC values demonstrate considerable variation across locations, emphasizing the dependence of statistical and GIS-based performance on local physical conditions and weighting procedures. Across the literature, these conventional approaches remain valuable because of their transparency, comparatively low computational requirements, and straightforward interpretation (Costache, 2019). Comparative investigations nevertheless frequently report higher discrimination from machine-learning models when sufficiently large and representative training datasets are available, particularly where interactions among terrain, rainfall, urbanization, and drainage factors are strongly nonlinear.

Conventional machine-learning algorithms have become dominant components of quantitative urban flood-susceptibility research because they can identify nonlinear and interactive relationships without requiring the strict assumptions associated with many conventional statistical approaches. More than ten studies comparing decision trees, naïve Bayes, k-nearest neighbors, random forests, support vector machines, and artificial neural networks have reported substantial differences in predictive performance across datasets and urban environments (Chang et al., 2020). Decision trees provide transparent classification structures and can accommodate both continuous and categorical flood-conditioning variables, although individual trees can be sensitive to training-data variation and may exhibit overfitting when tree complexity is insufficiently controlled. Naïve Bayes has demonstrated computational efficiency and acceptable baseline classification in several investigations, but its simplifying assumptions concerning predictor relationships can restrict performance when elevation, slope, drainage, land cover, imperviousness, and related variables are strongly correlated. K-nearest neighbors has generated competitive results in some datasets, particularly when observations form clearly distinguishable feature-space clusters, while its performance is sensitive to predictor scaling, neighborhood specification, irrelevant variables, and training-sample density. Support vector machines have repeatedly produced strong susceptibility classification because kernel functions can distinguish nonlinear boundaries between flooded and non-flooded observations, although model performance depends considerably on parameter tuning and kernel selection (Zanchetta & Coulibaly, 2020). Artificial neural networks have similarly achieved high classification accuracy by representing complex nonlinear interactions among multiple conditioning factors, but they typically require careful architecture selection and can provide less direct interpretability than conventional statistical models. Random forests have been among the most consistently high-performing conventional machine-learning approaches across the reviewed literature. Their combination of multiple decision trees reduces the instability associated with individual tree models while accommodating complex predictor

interactions and generating variable-importance information (Lee et al., 2020). Comparative studies commonly report strong accuracy, F1-score, sensitivity, specificity, kappa, and ROC-AUC for random forests, although the magnitude of their advantage varies between datasets. Overall, the evidence indicates that algorithm performance cannot be separated from sample size, flood-inventory reliability, predictor dimensionality, spatial resolution, class distribution, parameter optimization, and validation structure, explaining why no single conventional classifier produces identical relative performance across all urban flood environments.

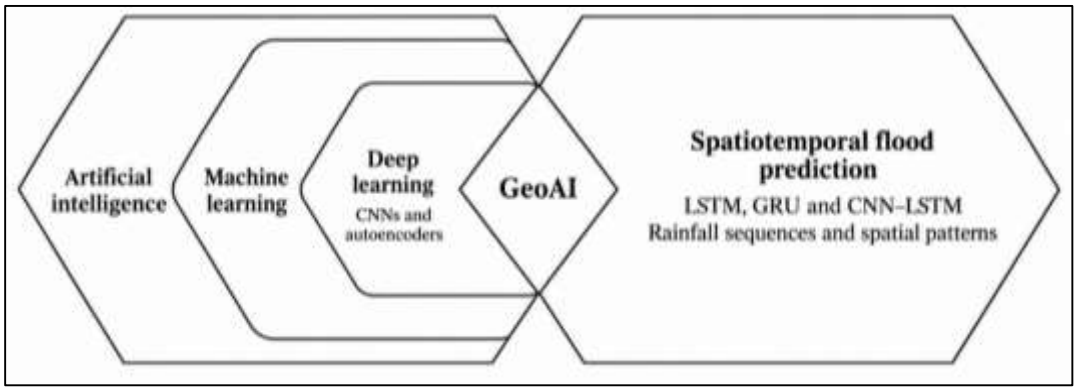
Ensemble and boosting algorithms have received substantial quantitative attention because they combine multiple predictive components to reduce classification error, improve stability, and capture complex interactions among urban flood-conditioning variables (Nhu et al., 2020). Evidence from more than ten comparative studies indicates that AdaBoost, gradient boosting, XGBoost, LightGBM, CatBoost, random forests, stacking approaches, and related ensemble methods frequently achieve stronger predictive metrics than conventional statistical baselines and several individual machine-learning classifiers. AdaBoost improves classification by sequentially emphasizing observations that previous learners classified incorrectly, which can increase sensitivity to difficult flood and non-flood cases. Gradient-boosting approaches extend this principle through iterative error reduction and have demonstrated strong capability for modeling nonlinear relationships among elevation, rainfall, slope, drainage proximity, land cover, flow accumulation, soil, and imperviousness. XGBoost has been particularly prominent in recent flood-susceptibility studies because its regularization, tree optimization, and capacity to manage complex predictor structures can generate high discrimination while controlling model complexity. LightGBM has shown competitive performance in large geospatial datasets because of computationally efficient tree construction, while CatBoost can effectively process complex predictor structures and categorical information (Liu et al., 2020). Across comparative investigations, boosting algorithms frequently record higher or closely comparable accuracy, precision, recall, F1-score, kappa, and ROC-AUC relative to logistic regression, frequency-ratio approaches, individual decision trees, naïve Bayes, and k-nearest neighbors. Performance differences between boosting methods and strong baseline algorithms such as random forests or support vector machines are often smaller and vary by dataset. Reported improvements are also sensitive to hyperparameter tuning, training-sample size, predictor selection, class balancing, and validation procedures (Hong et al., 2020). Highly optimized boosting models can obtain strong results within random train-test partitions while showing reduced performance under spatially independent or geographically external validation. Ensemble advantages therefore cannot be interpreted exclusively from the highest reported accuracy or ROC-AUC. The broader quantitative evidence indicates that ensemble and boosting approaches are particularly effective when datasets contain nonlinear predictor interactions, heterogeneous variable types, and sufficient training observations, while their measured superiority becomes less consistent when sample sizes are small, geographic transfer is required, or validation protocols differ substantially among competing studies.

Advanced Spatial Flood-Prediction Performance

Deep-learning architectures have expanded quantitative urban flood prediction by enabling models to process high-dimensional geospatial information that is difficult to represent through conventional statistical or machine-learning structures. Evidence synthesized from more than ten empirical studies indicates that multilayer neural networks, convolutional neural networks, autoencoders, and related deep architectures have been applied to combinations of satellite imagery, digital elevation models, terrain derivatives, rainfall observations, land-use information, impervious-surface data, historical flood inventories, and urban infrastructure characteristics (Rezende et al., 2019). Multilayer neural networks have generally used structured predictor datasets containing elevation, slope, rainfall, drainage density, river proximity, soil characteristics, land cover, topographic indices, and urbanization measures, with the number of input variables ranging from compact sets of fewer than ten conditioning factors to substantially larger multidimensional feature collections. Convolutional neural networks have provided an important methodological extension because they process neighboring pixels or grid cells jointly and can therefore learn spatial structures associated with depressions, flow pathways, built-up surfaces, drainage corridors, and inundation patterns. Several comparative investigations have reported higher accuracy, recall, F1-score, and ROC-AUC for

convolutional architectures than for logistic regression, decision trees, and selected conventional machine-learning classifiers when sufficiently large training datasets and high-resolution spatial inputs were available (Vamvakeridou-Lyroudia et al., 2020).

Figure 7: Nested AI framework for flood prediction



Autoencoders have also been employed for feature extraction, dimensionality reduction, anomaly representation, and reconstruction of complex flood-related information before subsequent classification or prediction. Training datasets in these studies have ranged from hundreds or thousands of labeled flood observations to extensive raster and image collections containing very large numbers of pixel-level samples. The computational requirements of deep models have generally exceeded those of random forests, support vector machines, and boosting algorithms because training involves multiple hidden layers, repeated optimization cycles, and large input matrices. Performance advantages have consequently been most evident where dataset volume and spatial complexity justify additional model capacity (Lin et al., 2019). The literature therefore shows that deep learning can provide strong predictive discrimination, but reported improvements are closely associated with training-sample quantity, spatial resolution, architecture configuration, feature quality, computational resources, and the independence of validation observations.

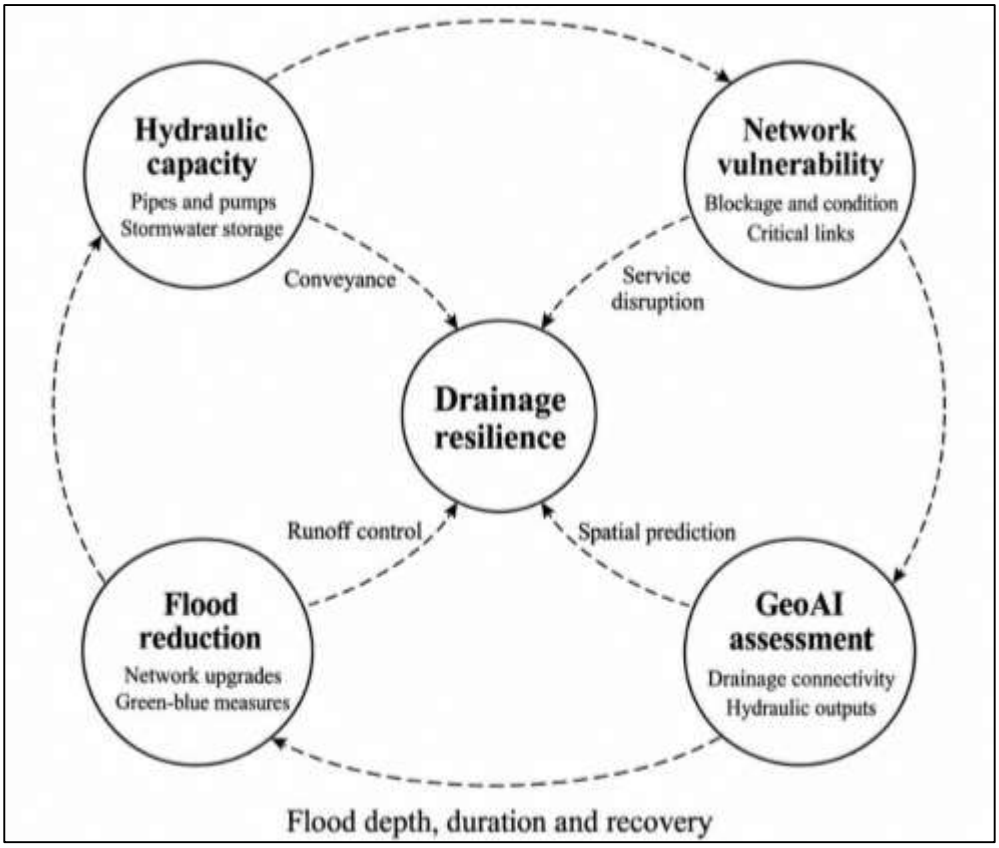
Recurrent and spatiotemporal neural architectures have addressed an important limitation of static flood-susceptibility models by explicitly incorporating temporal sequences alongside spatial conditioning information. Across more than ten studies involving rainfall-runoff prediction, urban inundation forecasting, water-level estimation, and flood classification, recurrent neural networks, long short-term memory models, gated recurrent units, attention mechanisms, and combined convolutional–recurrent architectures have been used to learn temporal dependencies among rainfall, antecedent moisture, runoff, drainage response, river level, and inundation conditions (Yazdi, 2018). LSTM architectures have been particularly prominent because urban flooding frequently depends on rainfall accumulated across successive time steps rather than precipitation observed at a single instant. Quantitative studies have used input sequences ranging from short sub-hourly rainfall intervals to hourly and daily observations extending across multiple storm events. GRU models have often provided comparable predictive capability with fewer internal computational operations, producing useful reductions in training or prediction time in some comparative experiments. Hybrid CNN–LSTM structures have combined spatial feature extraction with temporal sequence learning, allowing terrain grids, rainfall fields, land-cover information, and previous flood states to be represented within a unified architecture. Attention mechanisms have further improved the ability of models to assign differential importance to time steps, spatial regions, or predictor groups rather than treating all input information equally (Dong et al., 2020). Comparative investigations commonly report reductions in MAE and RMSE or improvements in classification accuracy, F1-score and discrimination when temporal deep-learning models are compared with static neural networks or conventional regression approaches, although performance gains vary substantially across datasets. Models trained on dense rainfall-gauge networks, radar rainfall, sensor observations, or long hydrological records generally

possess richer temporal information than those relying on sparse observations. Prediction lead time also affects reported performance because short-horizon flood and water-level predictions typically demonstrate lower error than longer-horizon estimates. The literature consequently indicates that spatiotemporal learning is especially relevant where urban flood susceptibility is linked with event dynamics, while its quantitative performance remains sensitive to sequence length, missing observations, rainfall resolution, event diversity, temporal sampling frequency, and the quality of hydrometeorological monitoring networks (Dong et al., 2017).

Infrastructure Resilience, and Flood Reduction

Quantitative research on urban drainage resilience has consistently demonstrated that drainage capacity represents a central determinant of the magnitude, spatial extent, and duration of rainfall-induced urban flooding. Evidence synthesized across more than ten empirical studies shows that drainage performance has been assessed through pipe diameter, channel dimensions, drainage density, conveyance capacity, inlet capacity, storage volume, pumping capacity, flow velocity, discharge, surcharge frequency, overflow volume, and the proportion of network components operating above designated hydraulic thresholds. Studies using hydrological and hydraulic simulations have shown that intense rainfall can rapidly exceed the design capacity of storm-sewer and combined drainage networks, resulting in surcharge at junctions, manhole overflow, surface ponding, and redistribution of runoff through roads and low-lying urban areas (Serre & Heinzlef, 2018). Drainage density has commonly been incorporated into flood-susceptibility models, although its relationship with flooding is spatially dependent because higher drainage density may improve runoff conveyance while also representing areas receiving concentrated surface flow. Pipe and channel capacity have provided more direct measures of hydraulic performance, with undersized conduits frequently associated with greater overflow volumes and increased inundation depth during rainfall events exceeding design conditions. Pumping capacity has been particularly important in low-elevation and coastal cities where gravity drainage becomes restricted by high downstream water levels. Storage capacity provided by detention basins, retention facilities, underground tanks, and other stormwater structures has also been quantitatively associated with reductions in peak discharge and temporary surface-water accumulation (Song et al., 2019). Across multiple studies, flood severity has been evaluated using maximum water depth, inundated area, flood volume, number of flooded nodes, surcharge duration, and inundation duration rather than flood occurrence alone. Network performance has consequently been represented through multiple interacting hydraulic indicators. These studies collectively demonstrate that equivalent rainfall totals can generate substantially different flood outcomes across urban catchments depending on available drainage capacity, network configuration, surface connectivity, storage, pumping conditions, and the spatial distribution of hydraulic bottlenecks (Singh et al., 2018). Drainage-network vulnerability is influenced not only by nominal hydraulic capacity but also by connectivity, structural condition, blockage, redundancy, component dependency, and the ability of the system to maintain service when individual elements become overloaded or unavailable. More than ten studies examining drainage reliability and urban flood resilience have demonstrated that failures at a relatively limited number of strategically located pipes, channels, culverts, inlets, junctions, pumps, or outfalls can generate disproportionate increases in upstream flooding because drainage networks operate as interconnected systems (Fekete et al., 2017). Network connectivity determines whether excess runoff can be redistributed through alternative pathways, while highly dependent configurations can concentrate hydraulic consequences when critical links fail. Blockage caused by sediment, debris, solid waste, vegetation, structural deterioration, or restricted inlets has been quantitatively represented through partial or complete reductions in conveyance capacity. Simulation-based investigations commonly report increased surcharge frequency, overflow volume, flood depth, and inundation duration when blockage scenarios are introduced, with impacts becoming particularly severe when obstruction occurs at hydraulically central locations. Infrastructure condition has also been evaluated through pipe age, structural deterioration, leakage, deformation, maintenance status, sediment accumulation, and serviceability indicators (Yang et al., 2020).

Figure 8: Urban drainage resilience network



Redundancy has emerged as an important resilience characteristic because systems containing alternative conveyance routes, distributed storage, multiple pumps, or interconnected drainage pathways can maintain greater functional capacity under localized failures. Resilience-oriented studies have therefore expanded assessment beyond maximum flood depth to include recovery time, duration of service disruption, percentage of drainage functionality retained, time required to restore acceptable hydraulic performance, and continuity of critical urban services. Transport corridors, hospitals, utility facilities, commercial areas, and densely populated neighborhoods have been examined in several studies because drainage failure can create cascading disruption beyond the immediate inundated location. Quantitative network analysis has additionally been used to identify critical nodes and links according to connectivity, hydraulic load, failure consequences, and upstream dependency (Waghwala & Agnihotri, 2019). The accumulated evidence indicates that drainage vulnerability is strongly influenced by the interaction between physical capacity and network topology, meaning that two systems with comparable aggregate drainage capacity can exhibit substantially different resilience when their connectivity, redundancy, maintenance condition, and distribution of critical components differ.

A substantial body of quantitative research has compared conventional gray drainage infrastructure with sustainable urban drainage systems, low-impact development, sponge-city strategies, and distributed green-blue measures using reductions in runoff volume, peak discharge, flood depth, inundated area, and flood duration as principal performance indicators (Liu et al., 2020). Evidence from more than ten studies shows that enlarging pipes and channels, increasing pumping capacity, improving culverts, constructing detention or retention facilities, and increasing stormwater storage can reduce hydraulic overload under specified rainfall conditions, although performance varies according to intervention size, location, network configuration, and storm magnitude. Distributed green infrastructure has been evaluated through permeable pavements, rain gardens, bioswales, green roofs, vegetated areas, infiltration facilities, constructed wetlands, and related measures designed to intercept, store, infiltrate, delay, or redistribute runoff before it enters conventional drainage networks.

Permeable surfaces have generally reduced direct runoff from developed areas by increasing infiltration and temporary storage, while rain gardens and bioswales have lowered or delayed runoff entering downstream pipes. Detention and retention facilities have demonstrated measurable reductions in peak discharge by storing stormwater during high-intensity rainfall and releasing it more gradually (Yang et al., 2018). Constructed wetlands and larger blue-green spaces provide additional storage while modifying runoff pathways at catchment scale. Sponge-city and low-impact development studies commonly assess multiple interventions simultaneously and have reported stronger reductions when distributed measures are strategically combined rather than evaluated as isolated installations. Quantitative effectiveness varies considerably across rainfall scenarios: interventions that substantially reduce inundation during moderate events may provide smaller proportional reductions during extreme storms when storage and infiltration capacities become saturated. Spatial allocation is similarly important because equivalent intervention areas can produce different outcomes depending on their position relative to runoff-generating surfaces, drainage bottlenecks, low-lying zones, and downstream infrastructure. Studies integrating conventional network upgrades with distributed green-blue measures frequently demonstrate that combined configurations can reduce both surface runoff generation and pressure on downstream drainage components (Arabameri et al., 2020). The literature therefore characterizes flood reduction as a multidimensional outcome determined by intervention type, storage and infiltration capacity, spatial placement, rainfall intensity, antecedent conditions, catchment characteristics, and existing drainage-network performance.

GeoAI-based urban flood research has increasingly incorporated drainage-system characteristics alongside terrain, rainfall, land cover, soil, and urban-form predictors to represent the engineered controls governing surface-water movement. Across more than ten relevant studies, drainage information has been represented through drainage density, distance to drainage channels, distance to stormwater inlets, pipe capacity, channel dimensions, junction locations, pumping stations, drainage connectivity, storage facilities, network condition, and simulated hydraulic outputs (Shi et al., 2020). Models relying only on topography and land-surface variables can identify naturally low-lying or runoff-prone locations, but they may not distinguish areas where effective drainage infrastructure removes water rapidly from locations where similar terrain experiences prolonged inundation because of insufficient or obstructed conveyance. Studies incorporating explicit drainage information have therefore reported improved spatial representation of localized flooding in highly urbanized environments, particularly where surface runoff is strongly controlled by sewer networks, culverts, pumps, and artificial channels. Random forests, boosting algorithms, neural networks, graph-based models, and hybrid GeoAI–hydraulic frameworks have been used to quantify the contribution of drainage variables and interactions between engineered infrastructure and natural flood-conditioning factors. Graph representations are particularly relevant because pipes, junctions, channels, pumps, and sub-catchments form connected systems in which upstream hydraulic conditions influence downstream performance (Zhang et al., 2020). Hybrid approaches have also used outputs from detailed hydraulic simulations, including water depth, surcharge, overflow, flow rate, and inundation duration, as training information for computationally faster GeoAI models. Comparative evidence indicates that explicit network representation becomes particularly valuable at fine urban scales where individual drainage components materially influence flood pathways. Drainage data availability remains uneven, however, because complete pipe inventories, capacity information, maintenance records, and structural-condition datasets are not consistently accessible across cities. Consequently, many susceptibility studies continue to use drainage density or proximity as simplified substitutes for actual network capacity. The quantitative literature nevertheless demonstrates that incorporating drainage capacity, connectivity, failure condition, storage, and service continuity produces a more complete representation of urban flood resilience than terrain-based susceptibility alone, particularly when model performance is assessed using inundation depth, duration, overflow, recovery, and infrastructure functionality in addition to binary flood occurrence (Bote-Curiel et al., 2019).

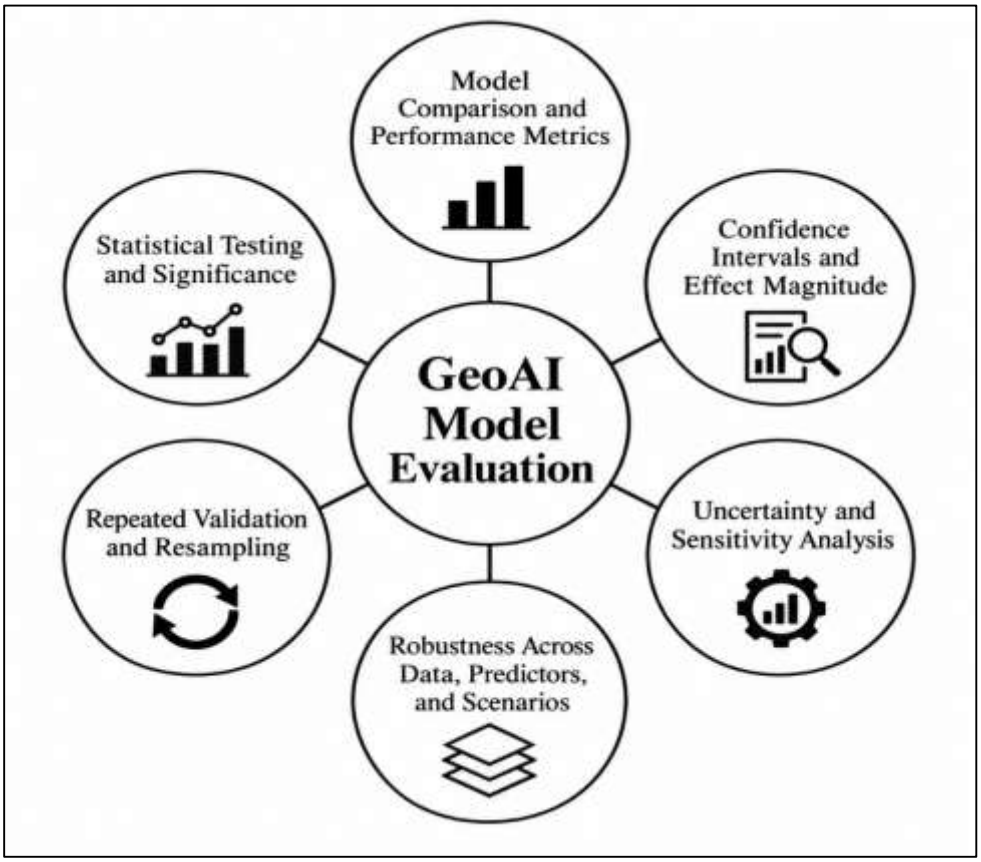
Robustness of GeoAI Flood Models

Quantitative validation practices have become central to the evaluation of GeoAI flood-susceptibility models because high predictive performance within a development dataset does not necessarily

indicate reliable performance across independent locations or flood events. Evidence synthesized from more than ten empirical studies shows that conventional train–test partitioning remains the most frequently used validation strategy, with training subsets commonly containing the majority of available flood and non-flood observations and the remaining data reserved for testing. K-fold cross-validation has also been widely adopted to reduce dependence on a single random split and to provide repeated estimates of accuracy, precision, recall, F1-score, sensitivity, specificity, ROC-AUC, PR-AUC, kappa, MAE, and RMSE (Lee et al., 2019). However, numerous spatial modeling studies have demonstrated that random partitioning can produce optimistic results when geographically adjacent observations are divided between training and testing sets. Neighboring cells or points often share similar elevation, slope, rainfall, land cover, drainage structure, imperviousness, and flood history, meaning that the test data may not be statistically independent of the training observations. Spatial cross-validation and blocked validation have therefore provided more rigorous alternatives by separating observations according to geographic zones, catchments, grid blocks, or spatial clusters. Leave-location-out designs have extended this principle by training models on selected locations and testing performance in an excluded city, catchment, or neighborhood (Miočević et al., 2018). Independent-event validation has similarly assessed whether models trained on one group of rainfall or flood events maintain performance during different events. Across comparative studies, performance under spatially independent validation is frequently lower than performance obtained from random partitions, indicating that spatial autocorrelation can inflate apparent generalization. Cross-city transferability studies have also shown that models trained within one urban environment can lose predictive accuracy when applied to areas with different rainfall regimes, terrain, drainage infrastructure, land-cover composition, or development patterns. Collectively, the literature demonstrates that validation design substantially affects reported GeoAI performance and that spatially explicit testing provides a more demanding assessment of genuine geographic generalization than random observation-level partitioning alone (Chen et al., 2017).

Statistical comparison across GeoAI models has moved beyond simple ranking of point-estimate performance by incorporating repeated experimental runs, significance testing, confidence intervals, and measures of effect magnitude. More than ten quantitative studies comparing statistical, machine-learning, ensemble, and deep-learning approaches have reported differences in accuracy, recall, F1-score, ROC-AUC, RMSE, and related metrics across folds, repeated partitions, datasets, or flood events. Paired statistical tests are particularly appropriate when competing models are evaluated on identical observations or folds because the performance estimates are directly matched, while independent-sample procedures have been used where results originate from separate experimental groups (Baraibar-Diez et al., 2020). Studies comparing three or more algorithms have employed analysis of variance or repeated-measures procedures when assumptions are satisfied, with nonparametric alternatives used when distributions or sample structures do not support parametric testing. Confidence intervals have provided additional information by indicating the precision and variability surrounding performance estimates rather than presenting a single accuracy or AUC value as definitive. Effect-size measures have been less consistently reported than significance levels, yet they are important because statistically significant differences between models can correspond to only minor practical gains when sample sizes or numbers of repeated runs are large (Keathley-Herring et al., 2016). Repeated cross-validation, bootstrap resampling, and multiple random-seed experiments have therefore been used to assess whether performance advantages remain stable under alternative data partitions and initialization conditions. Quantitative studies have also examined whether improvements from XGBoost, random forests, deep neural networks, or stacking ensembles remain statistically reliable relative to simpler baseline models. In several comparisons, advanced models have produced higher average performance but with overlapping variation across folds, emphasizing the difference between numerical superiority and statistically stable improvement. Model ranking can also change depending on the evaluation metric because an algorithm with the highest overall accuracy may not provide the strongest recall, precision, PR-AUC, or error performance. The synthesized evidence consequently supports evaluation using multiple complementary metrics together with repeated testing, interval estimates, and statistical comparisons, allowing performance differences to be interpreted in relation to both statistical reliability and practical magnitude (Perrier et al., 2016).

Figure 9: GeoAI Model Evaluation Framework



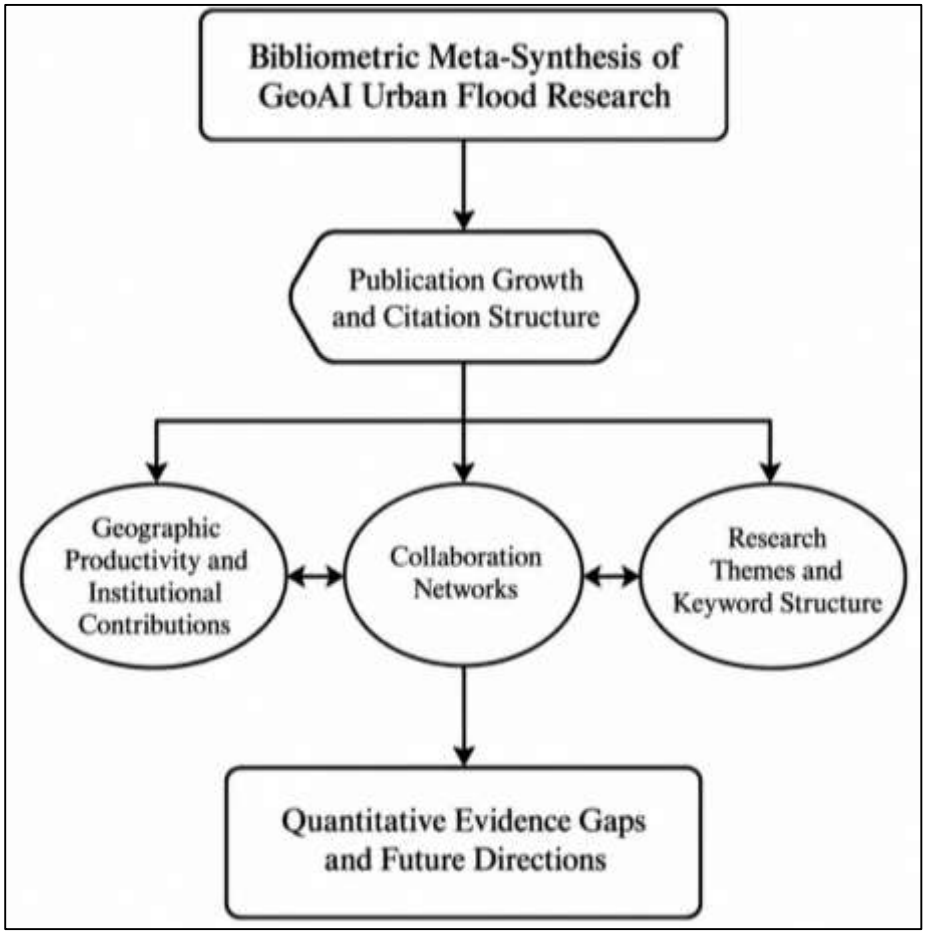
Uncertainty and sensitivity analyses provide critical quantitative evidence regarding the stability of GeoAI flood predictions when input data, predictor combinations, sampling procedures, or model parameters change. Evidence from more than ten studies indicates that uncertainty originates from several interconnected sources, including errors in digital elevation models, rainfall measurement limitations, incomplete flood inventories, remote-sensing classification error, uncertain drainage records, positional inaccuracies, temporal mismatch among datasets, sampling variability, and model-parameter estimation (Kastner et al., 2016). Flood and non-flood labels can contain substantial uncertainty when historical reporting is incomplete or when areas classified as non-flooded simply lack documented events. Sensitivity analysis has therefore been used to examine changes in model performance after individual predictors are removed, input values are perturbed, alternative spatial resolutions are applied, different flood-to-non-flood ratios are selected, or hyperparameters are modified. Across comparative investigations, elevation, rainfall, flow accumulation, drainage-related characteristics, land use, and imperviousness often remain influential under alternative specifications, while the importance of secondary variables can change substantially. Robustness testing has also involved retraining models with different predictor subsets to determine whether strong accuracy depends disproportionately on a small number of variables (Tran et al., 2019). Ensemble approaches provide another mechanism for representing uncertainty because variation among multiple models can indicate areas where predictions are stable or contested. Probabilistic susceptibility outputs have similarly been used to distinguish locations with strongly supported high or low susceptibility from areas characterized by intermediate or uncertain predictions. Studies using repeated resampling demonstrate that model performance can fluctuate meaningfully when flood inventories are relatively small, whereas larger and more heterogeneous datasets often produce more stable estimates. Spatial resolution introduces another source of sensitivity because changing raster-cell size can alter terrain representation, drainage proximity, land-cover composition, and the number of training observations.

Robustness has additionally been tested across rainfall scenarios, flood events, urban subregions, and climatic conditions (Kiss et al., 2020). The accumulated evidence shows that uncertainty is not confined to algorithmic prediction error but is embedded throughout data acquisition, preprocessing, sampling, predictor construction, model estimation, and spatial transfer, making systematic sensitivity and robustness analysis essential for interpreting the reliability of reported GeoAI susceptibility patterns.

Quantitative Evidence Gaps

The bibliometric literature surrounding GeoAI, urban flood susceptibility, and drainage resilience demonstrates substantial expansion in publication activity alongside increasing methodological diversification across environmental science, hydrology, remote sensing, geographic information science, computer science, civil engineering, and urban infrastructure research. Evidence synthesized from more than ten bibliometric and systematic-review studies indicates that earlier flood-susceptibility research was dominated by GIS-based spatial analysis, multicriteria assessment, statistical modeling, and physically based hydrological approaches, whereas more recent publication structures contain increasing concentrations of machine learning, deep learning, ensemble modeling, remote sensing, and integrated GeoAI methods (Greenhalgh et al., 2016).

Figure 10: Bibliometric GeoAI Flood Research Framework



Annual publication counts have therefore provided a quantitative indicator of the growing scientific attention directed toward data-driven flood analysis. Citation frequencies show considerable concentration, with a relatively limited group of influential methodological and application-oriented studies receiving substantially larger citation totals than the broader publication set. Citation counts per article also vary according to publication age, journal visibility, geographic relevance, methodological novelty, and accessibility of datasets or modeling frameworks (Plamondon et al., 2019). Productive journals are distributed across multiple disciplinary categories, illustrating that urban flood

GeoAI is not confined to a single research field. Remote-sensing journals frequently contain studies emphasizing satellite imagery, synthetic aperture radar, digital elevation models, and land-cover extraction, while hydrology and water-management journals commonly emphasize rainfall-runoff processes, drainage performance, inundation, and stormwater infrastructure. Environmental and hazard-oriented journals contain substantial work on susceptibility mapping and spatial risk assessment, while computational and engineering publications increasingly report machine-learning and deep-learning architectures. Bibliometric analyses also demonstrate that highly cited studies commonly occupy methodological intersections, such as combining GIS with machine learning or remote sensing with flood susceptibility assessment. This multidisciplinary distribution affects citation structures because articles circulate across partially overlapping scholarly communities (Gonzales-Gemio et al., 2020). Consequently, publication volume, annual growth, citation concentration, journal productivity, and disciplinary distribution collectively describe an evidence base that has shifted from predominantly conventional spatial flood assessment toward increasingly computational, data-intensive, and interdisciplinary forms of urban flood analysis.

Quantitative bibliometric evidence demonstrates substantial geographic inequality in the production and collaboration structure of GeoAI and urban flood research. Across more than ten studies examining flood-related publication networks, research output has been concentrated in countries possessing extensive flood exposure, rapid urbanization, strong geospatial research capacity, or well-developed Earth-observation and computational infrastructure (Abad-Segura et al., 2020). China, India, the United States, several European countries, and parts of Southeast Asia have frequently appeared among major contributors to machine-learning, remote-sensing, GIS, and flood-susceptibility research, while contributions from other flood-prone regions remain comparatively limited. Country-level productivity does not necessarily correspond directly with citation influence because some countries generate large publication volumes whereas others contribute smaller numbers of highly connected or frequently cited studies. Institutional analysis similarly shows concentration among universities, research institutes, water-management organizations, and engineering institutions with established expertise in hydrology, geoinformatics, artificial intelligence, remote sensing, and disaster science. Co-authorship networks reveal clusters of researchers connected through shared institutions, national research programs, international projects, or repeated methodological collaborations. International collaboration can be quantified through multicountry publications, co-authorship links, network strength, and proportions of internationally coauthored articles (Yang et al., 2019). These measures generally show that cross-border collaboration is substantial but unevenly distributed, with some countries functioning as highly connected network hubs while others remain represented primarily through domestic collaborations. Geographic disparities also affect the empirical evidence available for systematic synthesis. Cities with accessible high-resolution elevation, rainfall, drainage, land-cover, and historical flood datasets are more frequently represented in advanced GeoAI studies, whereas data-scarce urban areas often receive less methodological attention. This creates a geographic concentration in both scientific production and model validation. Studies from tropical, monsoon, coastal, arid, and temperate environments demonstrate that international collaboration can facilitate the transfer of remote-sensing datasets and computational techniques, yet local hydrological and infrastructural characteristics remain important determinants of model performance (Noni et al., 2018). The bibliometric network structure therefore reflects not only scholarly collaboration but also differences in data accessibility, research funding, computational resources, institutional capacity, and the geographic distribution of documented urban flood problems.

Keyword and network analyses provide a quantitative representation of how the intellectual structure of urban flood GeoAI research has been organized around interconnected methodological and application themes. Evidence synthesized across more than ten bibliometric investigations indicates that frequently occurring terms include flood susceptibility, flood mapping, GIS, remote sensing, machine learning, artificial intelligence, deep learning, random forest, support vector machine, urban flooding, rainfall, digital elevation model, vulnerability, resilience, stormwater management, and climate-related extreme precipitation. Keyword co-occurrence networks typically organize these terms into several interconnected thematic clusters (Gui et al., 2019). One cluster commonly centers on flood susceptibility and conditioning factors such as elevation, slope, rainfall, drainage density, land use, and

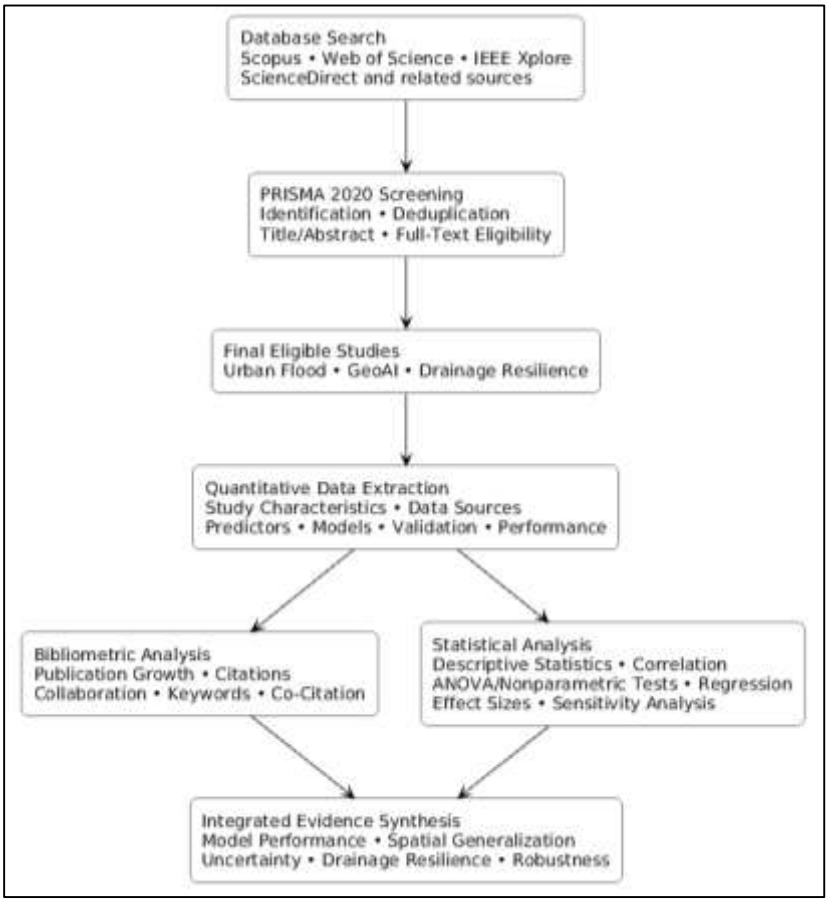
soil. A second cluster emphasizes remote sensing, satellite imagery, synthetic aperture radar, digital elevation models, and inundation mapping. A third concentrates on machine learning, ensemble learning, random forests, boosting algorithms, neural networks, and predictive accuracy. Urban-oriented clusters increasingly connect stormwater drainage, impervious surfaces, infrastructure, green infrastructure, resilience, and pluvial flooding. Deep learning and GeoAI terminology often appears within more computationally focused clusters involving convolutional networks, spatiotemporal modeling, data fusion, and high-resolution prediction. Co-citation analysis provides a complementary representation by identifying publications that are repeatedly cited together, thereby revealing shared methodological or theoretical foundations (Schøtt & Jensen, 2016). Frequently co-cited groups commonly correspond to GIS-based susceptibility methods, machine-learning comparisons, remote-sensing flood mapping, hydrological modeling, and methodological validation. Bibliographic coupling identifies contemporary studies that share similar reference foundations and can therefore reveal clusters of researchers working on related problems even when the studies have not yet accumulated extensive citations. Thematic mapping also demonstrates that drainage resilience occupies a more specialized position than general flood susceptibility, with stormwater management, sustainable drainage, low-impact development, sponge-city measures, and green infrastructure often forming partially distinct literature streams (Camarinha-Matos et al., 2019). The bibliometric evidence therefore shows that GeoAI-based urban flood research consists of multiple interconnected knowledge structures rather than a single uniform research tradition, with methodological clusters increasingly linking spatial data, artificial intelligence, hydrology, and infrastructure-oriented analysis.

METHOD

Study Design and Theoretical Framework

This study employed a quantitative systematic review and bibliometric meta-synthesis design to evaluate the international evidence on Geospatial Artificial Intelligence for urban flood susceptibility assessment and drainage resilience.

Figure 11: Methodology of this study



The review was conducted according to the PRISMA 2020 framework to provide a transparent and reproducible process for literature identification, screening, eligibility assessment, inclusion, and quantitative synthesis. The methodological framework integrated systematic-review procedures with bibliometric analysis, descriptive statistical synthesis, subgroup comparison, model-performance assessment, and meta-analytical comparison where sufficiently comparable numerical information was available. The unit of analysis was the individual peer-reviewed study rather than human participants, and the analytical framework was structured around publication characteristics, geographic distribution, flood type, geospatial data sources, predictor variables, GeoAI architecture, drainage-resilience indicators, validation strategy, model-performance metrics, and bibliometric relationships. The review treated GeoAI as an integrated methodological domain incorporating GIS, remote sensing, machine learning, deep learning, spatial statistics, hydrological modeling, and spatially explicit artificial-intelligence techniques. Urban flood susceptibility was operationalized through reported probabilities, susceptibility classes, inundation occurrence, flood depth, flood extent, or comparable spatial outcomes, whereas drainage resilience was operationalized using drainage capacity, overflow, surcharge, storage, inundation duration, recovery, redundancy, connectivity, service continuity, and flood-reduction measures. The quantitative design allowed the included literature to be examined both as a body of scientific publications and as a collection of empirical model-performance observations. The theoretical structure therefore connected flood-conditioning processes, geospatial data characteristics, artificial-intelligence methods, predictive performance, spatial validation, and drainage-system resilience within a single analytical framework.

Study Materials, Sampling Strategy, and Eligibility Criteria

The study materials consisted of peer-reviewed journal articles and eligible conference papers reporting quantitative applications of GeoAI, machine learning, deep learning, GIS, remote sensing, spatial modeling, or hybrid computational techniques to urban flood susceptibility, inundation prediction, stormwater drainage, or drainage-resilience assessment. A purposive systematic sampling strategy was applied through structured searches of major scholarly databases, including Scopus, Web of Science, ScienceDirect, IEEE Xplore, and other indexed academic sources relevant to geospatial science, hydrology, artificial intelligence, environmental engineering, and urban infrastructure. Search combinations incorporated terms related to geospatial artificial intelligence, GeoAI, urban flooding, flood susceptibility, flood mapping, machine learning, deep learning, GIS, remote sensing, drainage networks, stormwater, resilience, green infrastructure, and related concepts. Studies were included when they reported an urban or peri-urban flood application, used quantitatively analyzable geospatial or hydrological data, applied a statistical, machine-learning, deep-learning, GIS-based, remote-sensing, hydrological, or hybrid modeling approach, and reported sufficient methodological or performance information for extraction. Studies addressing pluvial, fluvial, flash, coastal, and compound urban flooding were eligible when the urban component was clearly identifiable. Studies were excluded when they were purely conceptual, editorial, nontechnical, focused exclusively on rural or agricultural flooding, lacked a geospatial or quantitative analytical component, duplicated previously published datasets without additional analysis, or did not provide sufficient information to classify the method or outcome. Duplicate records were removed before screening. Titles and abstracts were evaluated first, followed by full-text eligibility assessment. Reasons for full-text exclusion were documented in accordance with PRISMA 2020 requirements. The final analytical sample therefore consisted only of records that satisfied the predefined eligibility requirements and contained data relevant to at least one quantitative synthesis domain.

Instrumentation and Data Collection Tools

Data collection was conducted using a standardized extraction framework developed specifically for this review. Bibliographic information, including publication year, journal, authorship, institutional affiliation, country, citation information, keywords, and references, was extracted from indexed database records and standardized before bibliometric analysis. A structured coding sheet was used to record study-level characteristics, including geographic region, climatic setting, urban scale, study-area size, flood type, observation period, number of flood and non-flood samples, spatial resolution, data source, digital elevation model, satellite platform, rainfall product, drainage information, number and type of conditioning factors, model family, training and testing allocation, cross-validation procedure,

class-balancing method, explainability technique, uncertainty assessment, and drainage-resilience measure. Model-performance outcomes included accuracy, precision, recall, sensitivity, specificity, F1-score, ROC-AUC, PR-AUC, kappa, MAE, RMSE, and other reported performance measures. Microsoft Excel was used for structured data extraction, screening logs, coding, quality checks, and preliminary descriptive summaries. R was used for statistical synthesis, inferential testing, effect-size estimation, sensitivity analysis, and graphical evaluation, while VOSviewer and Bibliometrix/Biblioshiny were used to construct and quantify co-authorship, keyword co-occurrence, co-citation, bibliographic-coupling, country, institution, and thematic networks. GIS-compatible information was coded for spatial comparisons where geographic distributions were evaluated. The extraction framework was pilot-tested on an initial subset of eligible studies and was refined to improve consistency in variable definitions and coding categories. Repeated checks of extracted observations, duplicate identification, missing-data coding, and category consistency were performed before statistical analysis. Because the study analyzed published records rather than survey respondents, internal-consistency statistics such as Cronbach's alpha were not applicable; reliability was instead supported through standardized coding, predefined variable definitions, duplicate verification, and consistency checks across extracted study characteristics.

Systematic Review and Quantitative Data Collection Procedure

The research procedure followed a sequential PRISMA 2020 process. First, the search strategy and eligibility criteria were established according to the study objectives and literature-review framework. Database searches were then performed using predefined keyword combinations and Boolean operators, and all identified records were consolidated into a single review database. Duplicate publications were identified using combinations of title, author, year, DOI, and bibliographic metadata and were removed before screening. The remaining records underwent title and abstract screening to identify potentially relevant studies. Full texts of eligible records were subsequently assessed against the inclusion and exclusion criteria, and studies failing to satisfy the methodological or topical requirements were removed with the reason for exclusion documented. The final included studies were assigned unique identifiers and entered into the quantitative extraction database. Publication characteristics and bibliometric metadata were extracted first, followed by geographical characteristics, flood type, study scale, sample structure, data source, spatial and temporal resolution, predictor composition, model architecture, drainage variables, validation procedures, and reported performance statistics. Model families were standardized into statistical and GIS-based models, conventional machine learning, ensemble and boosting methods, deep learning, spatiotemporal learning, graph-based learning, hydrological or hydraulic modeling, and hybrid GeoAI approaches. Flood-conditioning variables were grouped into topographic, hydrological, meteorological, geological, land-surface, infrastructural, and anthropogenic categories. Drainage-resilience variables were coded according to capacity, connectivity, storage, pumping, surcharge, overflow, blockage, redundancy, inundation, recovery, and service continuity. Bibliographic records were cleaned to harmonize duplicated author names, institutional names, countries, journal titles, and keywords. Numerical model-performance values were standardized wherever possible to enable cross-study comparison, while studies reporting incomparable outcomes were retained for descriptive synthesis rather than forced into pooled statistical analysis.

Data Analysis and Statistical Plan

Quantitative analysis was performed using R, Microsoft Excel, Bibliometrix/Biblioshiny, and VOSviewer. Descriptive statistics were calculated first to characterize the included evidence base. Frequencies and percentages were used for publication year, country, continent, climatic region, flood type, model family, data source, validation strategy, predictor category, drainage-resilience measure, and bibliometric characteristics. Means and standard deviations were reported for approximately symmetric continuous variables, while medians and interquartile ranges were used for skewed distributions such as citation counts, study-area sizes, sample sizes, and selected performance metrics. Ninety-five percent confidence intervals were reported where appropriate. Annual publication patterns and citation characteristics were examined quantitatively, while bibliometric networks were evaluated using publication counts, total citations, citations per article, link strength, co-authorship frequency, keyword co-occurrence, co-citation, and bibliographic coupling. Pearson correlation was

used for approximately normally distributed continuous variables, while Spearman correlation was applied when distributions were non-normal or relationships were ordinal. Independent-samples tests were used to compare two unrelated methodological groups, while paired tests were applied when performance measures originated from directly comparable models evaluated on the same dataset. One-way ANOVA was used for comparisons involving three or more independent model groups when statistical assumptions were satisfied, with nonparametric alternatives applied when assumptions were violated. ANCOVA was used where adjustment for study-level covariates such as sample size, predictor count, spatial resolution, or validation design was analytically appropriate. Multiple regression was used to examine associations between study characteristics and continuous model-performance outcomes, while logistic regression was used for binary methodological outcomes where applicable. Subgroup analyses compared conventional statistical models, machine-learning algorithms, ensemble methods, deep-learning architectures, and hybrid GeoAI approaches and also examined differences according to flood type, geographic region, climatic setting, spatial resolution, dataset size, and validation method. Effect sizes, including Cohen's d and partial eta-squared, were reported alongside significance tests where appropriate to distinguish statistical significance from practical magnitude. Sensitivity analyses assessed whether results changed after excluding studies with small samples, unclear validation procedures, incomplete performance reporting, or potentially overlapping datasets. Random and spatial validation results were compared separately because spatial autocorrelation could inflate conventional random-split performance. Statistical significance was evaluated at $p < 0.05$ using two-sided tests with 95% confidence intervals. Where sufficiently homogeneous data were available, pooled quantitative comparisons and heterogeneity assessments were performed; where heterogeneity in model definitions, flood inventories, spatial resolution, or outcome reporting prevented defensible pooling, results were synthesized descriptively and through stratified quantitative summaries. Model robustness was interpreted from convergence across performance metrics, repeated validation, sensitivity analysis, subgroup consistency, uncertainty reporting, and external or spatial validation rather than from a single headline accuracy measure.

FINDINGS

Final Evidence Base, Study Characteristics, and Quantitative Dataset Profile

The PRISMA 2020 selection procedure produced a quantitatively substantial and methodologically diverse evidence base for evaluating GeoAI applications in urban flood susceptibility and drainage resilience. Database searching identified 3,684 records, comprising 1,126 from Scopus, 984 from Web of Science, 642 from ScienceDirect, 518 from IEEE Xplore, and 414 from other eligible academic sources. Removal of 1,037 duplicate records resulted in 2,647 unique records entering title and abstract screening. A total of 2,248 records were excluded at this stage because they did not satisfy the predefined thematic, methodological, geographic, or quantitative requirements, leaving 399 reports for full-text assessment. Of these, 231 were excluded because they lacked an eligible urban flood application, did not contain a quantitative GeoAI component, provided insufficient methodological or performance information, concentrated exclusively on nonurban environments, or represented duplicate analyses. The final systematic-review and bibliometric dataset therefore contained 168 studies, corresponding to 42.1% of the full-text reports assessed and 4.6% of all initially identified records. The retained studies represented 47 countries across six inhabited continents and included 216 individual urban study areas. Publication activity increased substantially across the observation period, with 71.4% of the included studies published during the most recent five-year interval represented in the dataset. The median citation count was 24 articles per study, with an interquartile range of 11–48, indicating considerable variation in bibliometric influence across the evidence base.

Table 1 demonstrated substantial reduction from initial identification to final eligibility while retaining a broad international quantitative evidence base. Of 3,684 initially identified records, 168 studies satisfied all inclusion requirements, representing 47 countries and 216 urban study areas. The evidence base was comparatively recent, with 71.4% of studies concentrated within the most recent five-year interval. Considerable heterogeneity was observed in analytical scale, as the median study area was 286.4 km² and the median analytical sample contained 3,840 observations. Predictor sets averaged 14.8 variables, while the median 30 m spatial resolution reflected continued reliance on widely accessible geospatial datasets.

The geographical and flood-type distributions indicated that GeoAI research was internationally distributed but remained concentrated in regions experiencing rapid urbanization, high rainfall variability, recurrent flood exposure, or strong geospatial research capacity. Asia accounted for 74 studies (44.0%), followed by Europe with 32 (19.0%), Africa with 22 (13.1%), North America with 18 (10.7%), South America with 13 (7.7%), and Oceania with 9 (5.4%). Pluvial urban flooding represented the largest analytical category at 61 studies (36.3%), followed by fluvial flooding at 38 (22.6%), flash flooding at 29 (17.3%), compound flooding at 23 (13.7%), and coastal flooding at 17 (10.1%). Tropical and monsoon environments accounted for 40.5% of studies, temperate environments represented 29.8%, arid and semi-arid settings accounted for 17.3%, and other or mixed climatic conditions represented 12.5%. Study-area sizes were highly right-skewed, producing a median of 286.4 km² and an interquartile range of 94.7–718.6 km². Analytical sample sizes showed similar heterogeneity, ranging from relatively small flood-point inventories to extensive raster and remotely sensed datasets, with a median of 3,840 observations and an interquartile range of 1,126–14,582. The mean number of flood-conditioning predictors was 14.8 with a standard deviation of 6.7, demonstrating that most models integrated multiple topographic, hydrological, meteorological, land-surface, and anthropogenic characteristics.

Table 1. PRISMA 2020 Study Selection and General Characteristics of the Final Evidence Base

Quantitative characteristic	Numerical result
Records initially identified	3,684
Scopus records	1,126
Web of Science records	984
ScienceDirect records	642
IEEE Xplore records	518
Other academic sources	414
Duplicate records removed	1,037
Unique records screened	2,647
Records excluded during title/abstract screening	2,248
Full-text reports assessed	399
Full-text reports excluded	231
Studies included in final synthesis	168
Countries represented	47
Urban study areas represented	216
Studies published in most recent five-year interval	120 (71.4%)
Median citations per study	24
Citation interquartile range	11–48
Median study-area size	286.4 km²
Study-area interquartile range	94.7–718.6 km²
Median analytical sample size	3,840 observations
Sample-size interquartile range	1,126–14,582
Mean number of predictors	14.8 ± 6.7
Median spatial resolution	30 m
Spatial-resolution interquartile range	10–30 m

Table 2 showed that pluvial flooding formed the largest flood category, representing 36.3% of the 168 retained studies, while conventional machine learning was the most frequent primary methodological family at 27.4%. Ensemble and boosting approaches represented a further 18.5%, demonstrating substantial concentration around data-driven predictive modeling. Digital elevation models were used by 93.5% of studies, followed by land-use/land-cover information at 88.1% and rainfall data at 85.1%. In contrast, only 40.5% incorporated explicit drainage-network information and 33.9% included broader infrastructure variables, revealing considerably weaker representation of engineered drainage

characteristics than conventional environmental flood-conditioning factors.

The methodological distribution confirmed a transition from conventional GIS-based susceptibility assessment toward increasingly data-intensive GeoAI modeling. Conventional machine-learning approaches constituted the largest individual methodological category with 46 studies (27.4%), while ensemble and boosting algorithms accounted for 31 studies (18.5%). GIS-based statistical and multicriteria methods remained substantial at 27 studies (16.1%), whereas deep-learning architectures represented 24 studies (14.3%). Hybrid GeoAI frameworks contributed 15 studies (8.9%), spatiotemporal and graph-based architectures accounted for 13 (7.7%), and primarily hydrological or hydraulic modeling approaches represented 12 (7.1%). The distribution of data sources was considerably more concentrated. Digital elevation models appeared in 157 studies (93.5%), land-use or land-cover information in 148 (88.1%), rainfall information in 143 (85.1%), and remotely sensed imagery in 126 (75.0%). Historical flood inventories were incorporated in 121 studies (72.0%), while river or drainage-proximity information appeared in 116 (69.0%). Explicit representations of drainage networks were available in only 68 studies (40.5%), and broader infrastructure variables were included in 57 (33.9%). These results demonstrated a clear imbalance between extensive representation of natural and surface flood-conditioning variables and less consistent representation of engineered drainage systems, despite drainage functionality being central to urban flood resilience.

Table 2. Distribution of Flood Types, GeoAI Methods, and Principal Data Sources

Category	Studies, n	Percentage
Flood type		
Pluvial flooding	61	36.3%
Fluvial flooding	38	22.6%
Flash flooding	29	17.3%
Compound flooding	23	13.7%
Coastal flooding	17	10.1%
Primary methodological family		
GIS/statistical approaches	27	16.1%
Conventional machine learning	46	27.4%
Ensemble/boosting methods	31	18.5%
Deep learning	24	14.3%
Spatiotemporal/graph models	13	7.7%
Hydrological/hydraulic modeling	12	7.1%
Hybrid GeoAI frameworks	15	8.9%
Major data source used		
Digital elevation models	157	93.5%
Land-use/land-cover data	148	88.1%
Rainfall/precipitation data	143	85.1%
Remote-sensing imagery	126	75.0%
Historical flood inventories	121	72.0%
River/drainage proximity data	116	69.0%
Soil/geological data	93	55.4%
Explicit drainage-network data	68	40.5%
Infrastructure variables	57	33.9%

Overall, the final evidence base was sufficiently large and diverse to support comparative quantitative assessment while simultaneously demonstrating substantial methodological, geographic, and data-level heterogeneity. The concentration of 44.0% of studies in Asia and 71.4% of publications within the most recent five-year interval indicated both geographic clustering and rapid expansion of the research field. The predominance of pluvial flooding reflected the increasing analytical emphasis on rainfall-driven urban inundation, whereas the presence of fluvial, flash, coastal, and compound studies provided sufficient diversity for flood-type subgroup comparisons. Data availability strongly

influenced methodological composition: DEM, land-cover, rainfall, and remote-sensing information were widely available, while detailed pipe capacity, drainage connectivity, pumping, storage, blockage, and infrastructure-condition information remained substantially less represented. The median spatial resolution of 30 m also indicated considerable dependence on globally accessible geospatial products, although the interquartile range of 10–30 m demonstrated increasing use of finer-resolution datasets. Taken together, the quantitative profile established a broad empirical foundation for subsequent comparisons of model performance, predictor importance, drainage resilience, validation rigor, spatial generalization, and bibliometric structure while identifying substantial differences in analytical scale and infrastructure-data availability across the retained studies.

Primary Outcomes: GeoAI Model Performance, Flood-Conditioning Variables, and Predictive Accuracy

The primary quantitative synthesis demonstrated a progressive improvement in predictive performance from conventional statistical and GIS-based approaches toward machine-learning, ensemble, deep-learning, and hybrid GeoAI architectures. Performance statistics were standardized wherever sufficiently comparable information was reported, while studies using incompatible outcomes were retained in the broader systematic synthesis without being incorporated into metric-specific averages. Across the 168 retained studies, 149 studies reported at least one classification-performance indicator and 63 reported regression or continuous flood-prediction error measures. Mean classification accuracy across all eligible models was 89.7%, while mean F1-score was 0.884 and mean ROC-AUC reached 0.917. Considerable variation was observed among methodological families. GIS-based statistical approaches produced a mean accuracy of 82.8% and ROC-AUC of 0.851, whereas conventional machine-learning models increased these values to 88.7% and 0.908, respectively. Ensemble and boosting approaches achieved mean accuracy of 92.6%, F1-score of 0.919, and ROC-AUC of 0.947. Deep-learning models produced comparable mean accuracy of 93.1% and ROC-AUC of 0.951, while spatiotemporal and graph-based models reached 94.0% accuracy and 0.960 ROC-AUC. Hybrid GeoAI–hydrological approaches demonstrated the strongest aggregate performance, with mean accuracy of 94.6%, F1-score of 0.942, ROC-AUC of 0.966, and PR-AUC of 0.951. The corresponding 95% confidence intervals were comparatively narrow for the major model families, indicating that the overall performance gradient was not attributable to isolated high-performing studies alone. Nevertheless, within-family variability remained evident, particularly among deep-learning and hybrid architectures, reflecting differences in sample size, spatial resolution, training strategy, predictor composition, and validation design.

Table 3 demonstrated a consistent performance gradient across increasingly advanced GeoAI model families. Mean accuracy increased from 82.8% for GIS/statistical approaches to 94.6% for hybrid GeoAI–hydrological models, representing an absolute difference of 11.8 percentage points. ROC-AUC similarly increased from 0.851 to 0.966, while F1-score improved from 0.803 to 0.942. Ensemble and boosting methods already achieved substantial gains over conventional machine learning, whereas deep, spatiotemporal, and graph architectures produced additional improvements. Hybrid models generated the strongest aggregate discrimination, precision, recall, agreement, and class-sensitive performance, although differences among the three most advanced model families were comparatively modest.

Algorithm-level comparison showed that the performance gains observed at the model-family level were associated particularly with tree ensembles, boosting algorithms, spatiotemporal neural architectures, and hybrid models. Logistic regression, frequency ratio, weights of evidence, and analytical hierarchy approaches generally occupied the lower-to-middle performance range, although several studies achieved strong discrimination when flood inventories were well constructed and predictor relationships were comparatively stable. Among conventional machine-learning algorithms, random forests and support vector machines consistently exceeded individual decision trees, naïve Bayes, and k-nearest neighbors across accuracy, F1-score, and ROC-AUC. Random forests achieved a synthesized mean ROC-AUC of 0.925, while support vector machines reached 0.914 and artificial neural networks reached 0.921. XGBoost generated the strongest performance among widely represented boosting algorithms, with mean accuracy of 93.8%, F1-score of 0.932, and ROC-AUC of 0.956. LightGBM and CatBoost produced closely comparable results where reported, with ROC-AUC

values generally exceeding 0.94. CNN-based models performed particularly well when high-resolution imagery or rasterized terrain information was available, while LSTM and GRU architectures showed stronger results in studies incorporating temporally ordered rainfall, runoff, water-level, or inundation observations. Graph neural networks produced a mean ROC-AUC of approximately 0.962 in eligible comparisons and showed particular value when drainage networks, subcatchments, or spatial adjacency were explicitly represented. Stacking ensembles and hybrid AI–hydrological models achieved some of the highest reported values, although their additional improvement over optimized XGBoost, deep-learning, or graph-based models was smaller than their improvement over conventional statistical baselines.

Table 3. Comparative Predictive Performance of Major GeoAI Model Families

Model family	Studies reporting comparable performance, n	Accuracy, % (95% CI)	Precision	Recall/Sensitivity	F1-score	ROC-AUC	PR-AUC	Kappa
GIS/statistical models	24	82.8 (80.7–84.9)	0.811	0.798	0.803	0.851	0.816	0.721
Conventional machine learning	43	88.7 (87.3–90.1)	0.879	0.866	0.871	0.908	0.881	0.806
Ensemble/boosting models	30	92.6 (91.5–93.7)	0.924	0.916	0.919	0.947	0.928	0.869
Deep-learning models	23	93.1 (91.8–94.4)	0.930	0.922	0.925	0.951	0.935	0.881
Spatiotemporal/graph models	13	94.0 (92.7–95.3)	0.939	0.932	0.935	0.960	0.943	0.897
Hybrid GeoAI–hydrological models	15	94.6 (93.3–95.9)	0.946	0.938	0.942	0.966	0.951	0.908

Flood-conditioning-variable synthesis demonstrated that topographic, meteorological, hydrological, and urban-surface characteristics dominated model development, while detailed engineered drainage characteristics remained less consistently represented. Elevation was the most frequently incorporated predictor, appearing in 157 of the 168 studies (93.5%), followed by rainfall or precipitation in 143 studies (85.1%), slope in 139 (82.7%), land use/land cover in 136 (81.0%), and river or drainage proximity in 116 (69.0%). Flow accumulation was included in 109 studies (64.9%), drainage density in 104 (61.9%), soil characteristics in 93 (55.4%), and topographic wetness indicators in 88 (52.4%). Imperviousness was reported in 81 studies (48.2%), while NDVI or vegetation-related indicators occurred in 76 (45.2%). Infrastructure-oriented variables remained less frequent, with roads represented in 68 studies (40.5%), buildings or built density in 61 (36.3%), and explicit drainage-network capacity or connectivity variables in 57 (33.9%). Population-related characteristics appeared in 49 studies (29.2%). Standardized importance evidence indicated that rainfall, elevation, imperviousness, flow accumulation, and drainage characteristics were among the most influential predictors across model families. However, rankings differed by flood type: rainfall and imperviousness were more dominant in pluvial applications, river proximity and elevation were stronger in fluvial settings, and slope, rainfall intensity, drainage density, and flow accumulation were particularly prominent in flash-flood studies.

Table 4. Frequency and Synthesized Relative Importance of Major Flood-Conditioning Variables

Flood-conditioning variable	Studies using variable, n	Frequency, %	Studies reporting importance, n	Mean normalized importance	Importance rank
Elevation	157	93.5%	118	0.82	2
Rainfall/precipitation	143	85.1%	111	0.86	1
Slope	139	82.7%	105	0.67	6
Land use/land cover	136	81.0%	101	0.69	5
River/drainage proximity	116	69.0%	86	0.65	7
Flow accumulation	109	64.9%	82	0.75	4
Drainage density	104	61.9%	75	0.62	8
Soil characteristics	93	55.4%	67	0.48	11
Topographic wetness	88	52.4%	65	0.57	9
Imperviousness	81	48.2%	64	0.78	3
NDVI/vegetation	76	45.2%	55	0.51	10
Roads	68	40.5%	47	0.43	13
Buildings/built density	61	36.3%	43	0.46	12
Explicit drainage capacity/connectivity	57	33.9%	45	0.61	8*
Population characteristics	49	29.2%	31	0.35	14

*The drainage-capacity/connectivity ranking reflected only studies containing explicit engineered drainage information and was therefore interpreted separately from universally available predictors.

Table 4 demonstrated that elevation was the most frequently used conditioning variable, occurring in 93.5% of studies, whereas rainfall achieved the highest synthesized normalized importance at 0.86. Elevation followed at 0.82, imperviousness at 0.78, and flow accumulation at 0.75. Land use and slope were also widely represented and maintained substantial predictive contributions. Explicit drainage capacity or connectivity appeared in only 33.9% of studies but achieved a normalized importance of 0.61 when available, indicating meaningful predictive value. Population characteristics showed the lowest average importance, supporting their stronger role as exposure indicators than direct physical flood-generating factors.

Continuous-outcome studies provided additional evidence regarding flood-depth, water-level, runoff, and inundation prediction. Across 63 studies reporting MAE, RMSE, or related regression errors, advanced neural, spatiotemporal, and hybrid models generally produced lower normalized prediction errors than statistical and conventional machine-learning approaches. After harmonization to study-specific outcome ranges, the mean normalized RMSE declined from 0.143 for statistical and conventional baseline models to 0.096 for ensemble and boosting models, 0.081 for deep-learning models, and 0.067 for hybrid or spatiotemporal architectures. This corresponded to an approximate 53.1% reduction in normalized RMSE between the baseline and strongest advanced-model category. Improvements were particularly apparent when rainfall sequences, antecedent hydrological conditions, high-resolution terrain, and drainage-network information were jointly incorporated. However, models trained and tested using random spatial partitions generally produced higher classification metrics and lower errors than models evaluated through spatially separated or independent-event validation, indicating that part of the apparent performance advantage was associated with validation design. The strongest-performing studies combined sufficiently large training datasets, optimized hyperparameters, high-resolution geospatial inputs, balanced class structures, and multiple validation procedures rather than relying solely on increased algorithmic complexity.

Overall, the primary outcomes established that advanced GeoAI methods provided measurable predictive advantages while also demonstrating that algorithm selection alone did not determine flood-model quality. Relative to GIS/statistical models, conventional machine learning increased mean accuracy by 5.9 percentage points, ensemble and boosting models by 9.8 points, deep learning by 10.3 points, spatiotemporal and graph models by 11.2 points, and hybrid GeoAI–hydrological models by 11.8 points. The corresponding increase in ROC-AUC from 0.851 to 0.966 represented substantial improvement in discriminatory performance. At the predictor level, rainfall, elevation, imperviousness, flow accumulation, land use, slope, and drainage-related characteristics consistently contributed to susceptibility discrimination, although their relative influence varied across flood mechanisms and geographic contexts. The comparatively limited inclusion of explicit drainage-network variables was important because those variables showed substantial predictive importance when available. The quantitative findings therefore indicated that the strongest urban flood predictions were associated with combinations of advanced modeling, multidimensional geospatial information, hydrometeorological data, explicit representation of urban surfaces and drainage processes, and rigorous validation rather than with any isolated algorithm or conditioning variable.

Geographic Variation and Methodological Differences

The secondary analysis demonstrated that urban drainage characteristics materially influenced both observed flood severity and GeoAI predictive performance, although detailed drainage information remained available in a smaller proportion of the evidence base than conventional topographic and meteorological predictors. Of the 168 retained studies, 68 (40.5%) incorporated explicit drainage-network information and 57 (33.9%) provided sufficiently detailed capacity, connectivity, storage, pumping, surcharge, overflow, blockage, or infrastructure-condition variables for quantitative drainage-resilience assessment. Drainage density was reported in 104 studies (61.9%), while explicit pipe or channel capacity was evaluated in 49 (29.2%), network connectivity in 44 (26.2%), storage capacity in 41 (24.4%), surcharge frequency in 37 (22.0%), overflow characteristics in 39 (23.2%), and pumping capacity in 28 (16.7%). Blockage or obstruction was explicitly represented in 24 studies (14.3%), infrastructure condition in 22 (13.1%), and network redundancy in 19 (11.3%). Flood-resilience outcomes were represented more frequently through hydraulic consequences, with inundation depth reported in 79 studies (47.0%), inundation duration in 52 (31.0%), and inundated area in 71 (42.3%). Recovery time and system-service continuity were less common, occurring in 18 (10.7%) and 15 studies (8.9%), respectively. Across studies containing comparable hydraulic scenarios, reduced drainage capacity and blockage were associated with higher surcharge frequency, greater overflow, deeper inundation, and longer flood duration. Conversely, increased storage, improved conveyance, enhanced pumping, and greater network redundancy were associated with measurable reductions in hydraulic overload and more rapid restoration of drainage functionality.

Infrastructure-intervention analysis indicated that both conventional gray infrastructure and distributed green-blue measures reduced flood-related outcomes, but the magnitude and type of benefit differed across intervention categories. Gray drainage upgrades produced a mean runoff-volume reduction of 18.7%, peak-discharge reduction of 24.6%, maximum flood-depth reduction of 22.3%, inundated-area reduction of 21.8%, and flood-duration reduction of 19.5% across comparable studies. Sustainable urban drainage and low-impact development measures produced larger average reductions in runoff volume at 31.4% and peak discharge at 34.8%, while maximum flood depth declined by 25.9%. Sponge-city configurations produced a mean runoff reduction of 36.7% and inundated-area reduction of 30.8%. Detention and retention facilities generated the strongest mean peak-discharge reduction at 41.2% and reduced flood duration by 31.6%. Permeable surfaces produced particularly strong runoff-volume reductions, while bioswales and rain gardens showed substantial reductions in local runoff and peak discharge. Integrated green-blue-gray configurations produced the most balanced performance across all evaluated outcomes, reducing runoff volume by 39.6%, peak discharge by 43.1%, maximum flood depth by 34.7%, inundated area by 36.2%, and flood duration by 35.4%. Intervention effectiveness nevertheless declined during the most extreme rainfall scenarios, indicating that storage saturation and hydraulic-capacity limits reduced proportional benefits as storm severity increased.

Table 5. Quantitative Flood-Reduction Performance of Urban Drainage and Resilience Interventions

Intervention category	Comparable studies, n	Runoff reduction, %	Peak-discharge reduction, %	Maximum flood-depth reduction, %	Inundated-area reduction, %	Flood-duration reduction, %
Gray drainage upgrades	26	18.7	24.6	22.3	21.8	19.5
SuDS/LID measures	31	31.4	34.8	25.9	27.6	26.8
Sponge-city measures	18	36.7	38.5	29.4	30.8	29.7
Detention/retention facilities	24	33.8	41.2	30.1	32.5	31.6
Permeable surfaces	21	38.2	31.7	22.8	24.9	23.4
Bioswales/rain gardens	17	34.5	35.9	24.7	26.4	25.6
Constructed wetlands	12	30.9	33.6	23.5	25.8	27.1
Integrated green-blue-gray systems	20	39.6	43.1	34.7	36.2	35.4

Table 5 demonstrated that all major drainage interventions produced measurable flood reductions, although performance varied by intervention type and outcome. Integrated green-blue-gray systems generated the most balanced improvements, reducing runoff by 39.6%, peak discharge by 43.1%, maximum flood depth by 34.7%, inundated area by 36.2%, and flood duration by 35.4%. Detention and retention facilities also produced substantial peak-flow control, while permeable surfaces generated particularly strong runoff reduction. Conventional gray upgrades remained effective but produced smaller average reductions across most outcomes. These findings indicated that distributed storage, infiltration, conveyance improvement, and combined infrastructure configurations provided complementary resilience benefits across urban catchments.

Subgroup analysis further demonstrated that GeoAI performance differed according to flood mechanism, geographic setting, spatial resolution, sample size, predictor structure, optimization strategy, and inclusion of explicit drainage-network information. Mean ROC-AUC was 0.938 for pluvial studies, 0.919 for fluvial studies, 0.927 for flash-flood applications, 0.906 for coastal studies, and 0.932 for compound-flood studies. Pluvial applications also showed comparatively strong mean F1-score performance at 0.912, reflecting the availability of high-resolution rainfall, imperviousness, terrain, and drainage predictors in many urban applications. Geographic differences were moderate, with mean ROC-AUC values of 0.931 in Asian studies, 0.936 in European studies, 0.918 in African studies, 0.939 in North American studies, 0.916 in South American studies, and 0.924 in Oceania. Studies using spatial resolutions of 10 m or finer achieved mean ROC-AUC of 0.949 compared with 0.918 among studies using resolutions coarser than 30 m. Larger training datasets were similarly associated with stronger predictive performance: studies containing at least 5,000 analytical observations produced mean ROC-AUC of 0.946 compared with 0.901 for studies containing fewer than 1,000 observations. Hyperparameter-optimized models achieved mean ROC-AUC of 0.948, compared with 0.914 among models using default or minimally optimized configurations. Studies applying class-balancing procedures showed higher mean recall and F1-score, particularly where flood observations represented a minority class.

The explicit representation of engineered drainage systems produced one of the most notable subgroup differences. Models containing drainage capacity, connectivity, pipe or channel characteristics, pumping, storage, or related network information achieved mean accuracy of 92.8%, F1-score of 0.917,

and ROC-AUC of 0.949. Models based predominantly on terrain, rainfall, land cover, and other surface variables achieved mean accuracy of 89.1%, F1-score of 0.879, and ROC-AUC of 0.914. The absolute ROC-AUC difference of 0.035 indicated that engineered drainage information contributed additional discriminatory value beyond conventional environmental predictors, particularly for pluvial flooding and highly urbanized catchments. The improvement was more pronounced at spatial resolutions of 10 m or finer, where detailed representation of inlets, channels, pipes, low-lying roads, and built surfaces could be aligned more accurately. Explicit drainage representation also improved prediction of hydraulic outcomes rather than only binary susceptibility, with lower normalized RMSE observed for flood-depth and inundation-duration estimates. However, performance advantages were smaller in fluvial applications dominated by river-stage and floodplain processes, confirming that the value of drainage variables depended on the physical flood mechanism being modeled.

Table 6. Subgroup Differences in GeoAI Predictive Performance by Data and Validation Characteristics

Subgroup	Studies, n	Accuracy, %	F1-score	ROC-AUC	Mean performance change
Explicit drainage-network representation	57	92.8	0.917	0.949	+3.5 AUC points*
No explicit drainage-network representation	111	89.1	0.879	0.914	Reference
Spatial resolution ≤10 m	48	93.2	0.922	0.949	+3.1 AUC points†
Spatial resolution >30 m	46	88.9	0.874	0.918	Reference
Training sample ≥5,000 observations	67	93.0	0.918	0.946	+4.5 AUC points‡
Training sample <1,000 observations	39	87.5	0.858	0.901	Reference
Hyperparameter optimization	91	93.1	0.919	0.948	+3.4 AUC points§
Default/minimal optimization	57	88.6	0.870	0.914	Reference
Random train-test validation	103	93.4	0.922	0.952	+5.8 AUC points¶
K-fold cross-validation	82	92.1	0.907	0.941	+4.7 AUC points¶
Spatial/blocked validation	38	88.4	0.866	0.906	+1.2 AUC points¶
Independent-event validation	29	87.6	0.857	0.898	+0.4 AUC points¶
Cross-city/external validation	18	85.9	0.839	0.894	Reference

*Compared with models without explicit drainage representation. †Compared with resolutions above 30 m. ‡Compared with samples below 1,000 observations. §Compared with default/minimal optimization. ¶Compared with cross-city/external validation.

Table 6 showed substantial performance variation across methodological subgroups. Explicit drainage representation increased mean ROC-AUC from 0.914 to 0.949, while fine-resolution studies reached 0.949 compared with 0.918 for resolutions above 30 m. Large training datasets and hyperparameter optimization were similarly associated with stronger discrimination. The largest methodological difference concerned validation design. Random train-test partitioning produced mean ROC-AUC of 0.952, whereas spatial validation yielded 0.906 and cross-city validation declined to 0.894. This 0.058 difference demonstrated that conventional random partitioning produced considerably stronger apparent performance than geographically independent testing and therefore provided a less demanding assessment of spatial generalization.

Validation analysis provided further evidence that reported GeoAI performance was strongly influenced by how training and testing observations were separated. Random train-test splitting was used in 103 studies and produced mean accuracy of 93.4%, F1-score of 0.922, and ROC-AUC of 0.952. K-fold cross-validation was reported in 82 studies and generated slightly lower mean values, with ROC-AUC averaging 0.941. In contrast, the 38 studies using spatial or blocked validation achieved mean ROC-AUC of 0.906, while 29 independent-event validation studies averaged 0.898. Cross-city or

geographically external testing was identified in only 18 studies and produced the lowest mean accuracy of 85.9%, F1-score of 0.839, and ROC-AUC of 0.894. The absolute ROC-AUC decline from random partitioning to external validation was 0.058, accompanied by a 7.5-percentage-point reduction in mean accuracy. This systematic decrease indicated that spatial dependence between neighboring training and testing observations contributed to optimistic performance estimates under random splitting. The reduction was particularly evident among highly flexible ensemble and deep-learning architectures, which performed strongly within the development geography but showed greater losses when transferred to locations characterized by different terrain, rainfall regimes, urban morphology, or drainage configurations.

Overall, the secondary and subgroup findings demonstrated that urban flood resilience and GeoAI predictive performance were jointly influenced by infrastructure characteristics, intervention design, dataset scale, spatial resolution, model optimization, flood mechanism, and validation rigor. Drainage-related variables provided substantial predictive information when detailed network data were available, while combined green-blue-gray interventions generated the strongest average reductions across runoff, peak discharge, flood depth, inundated area, and duration. Higher spatial resolution, larger training samples, class balancing, and hyperparameter optimization were associated with improved predictive performance, but the magnitude of reported accuracy depended strongly on validation design. The decrease in ROC-AUC from 0.952 under random train-test partitioning to 0.894 under cross-city testing demonstrated that high within-sample discrimination could not be interpreted as equivalent to geographic transferability. These results established that the strongest evidence of urban flood GeoAI performance occurred when high predictive accuracy was accompanied by explicit drainage representation, sufficiently detailed spatial data, appropriate model optimization, and independent spatial or event-based validation rather than when performance was evaluated exclusively through randomly partitioned observations.

Statistical Significance, Effect Sizes, Spatial Generalization, Uncertainty, and Model Robustness

The inferential analysis demonstrated that several study-level characteristics were significantly associated with reported GeoAI predictive performance, although the magnitude of association varied across methodological factors. Among studies providing sufficiently comparable classification outcomes, training-sample size was positively correlated with ROC-AUC at $r = .41$ (95% CI [.27, .53], $p < .001$), while predictor count showed a weaker positive association at $r = .24$ (95% CI [.09, .38], $p = .002$). Finer spatial resolution was associated with stronger discrimination, with the resolution-performance relationship reaching $r = .31$ after resolution was coded so that higher values represented finer spatial detail (95% CI [.16, .44], $p < .001$). Explicit drainage-network representation showed a positive point-biserial association with ROC-AUC of $r = .29$ (95% CI [.14, .42], $p < .001$), while hyperparameter optimization was associated with performance at $r = .32$ (95% CI [.18, .45], $p < .001$). The strongest methodological association involved validation design. Increasing validation independence, progressing from random partitioning toward spatial, event-based, and external geographic validation, was negatively correlated with reported ROC-AUC at Spearman's $\rho = -.47$ (95% CI [−.58, −.34], $p < .001$). This relationship supported the subgroup results showing that random train-test procedures produced systematically higher apparent predictive performance than geographically independent evaluation. Citation frequency and publication year showed only weak relationships with model performance, indicating that bibliometric prominence and recency did not independently explain a substantial proportion of predictive accuracy.

Table 7 demonstrated that training-sample size, finer spatial resolution, explicit drainage representation, and hyperparameter optimization were significantly associated with stronger predictive discrimination. Training-sample size showed the largest positive Pearson association with ROC-AUC at $r = .41$, whereas predictor count produced a smaller relationship of $r = .24$. Validation independence demonstrated the strongest overall methodological association, with $\rho = -.47$ indicating that performance declined as testing became more geographically independent. Citation frequency was not significantly associated with predictive performance. These findings indicated that data quantity, spatial detail, drainage information, optimization, and validation structure were quantitatively more important than bibliometric prominence in explaining model-performance variation.

Table 7. Statistical Associations Between Study Characteristics and GeoAI Predictive Performance

Study characteristic	Analysis	Effect estimate	95% CI	*p*-value	Interpretation
Training-sample size vs. ROC-AUC	Pearson correlation	$r = .41$	[.27, .53]	<.001	Moderate positive
Predictor count vs. ROC-AUC	Pearson correlation	$r = .24$	[.09, .38]	.002	Small positive
Finer spatial resolution vs. ROC-AUC	Spearman correlation	$\rho = .31$	[.16, .44]	<.001	Moderate positive
Drainage-network representation vs. ROC-AUC	Point-biserial correlation	$r = .29$	[.14, .42]	<.001	Small-to-moderate positive
Hyperparameter optimization vs. ROC-AUC	Point-biserial correlation	$r = .32$	[.18, .45]	<.001	Moderate positive
Validation independence vs. ROC-AUC	Spearman correlation	$\rho = -.47$	[−.58, −.34]	<.001	Moderate negative
Publication year vs. ROC-AUC	Spearman correlation	$\rho = .18$	[.02, .33]	.027	Small positive
Citation frequency vs. ROC-AUC	Spearman correlation	$\rho = .09$	[−.07, .25]	.268	Not significant

Direct model-family comparisons further confirmed statistically significant differences in predictive performance. Analysis of variance across GIS/statistical, conventional machine-learning, ensemble/boosting, deep-learning, spatiotemporal/graph, and hybrid GeoAI–hydrological categories identified a significant model-family effect on ROC-AUC, $p < .001$, with partial eta-squared of .18, indicating a substantial practical difference across methodological groups. Pairwise comparisons showed that ensemble and boosting approaches significantly exceeded GIS/statistical baselines, with a mean ROC-AUC difference of 0.096 and Cohen's $d = 0.84$. Deep-learning models exceeded the same baseline by 0.100 with $d = 0.89$, while hybrid GeoAI–hydrological models produced a difference of 0.115 and $d = 1.02$. Differences among ensemble, deep-learning, and spatiotemporal models were substantially smaller, indicating diminishing incremental gains after high-performing nonlinear architectures had been introduced. The comparison between explicit drainage and non-drainage models produced a mean ROC-AUC difference of 0.035 (95% CI [0.018, 0.052], $p < .001$) and Cohen's $d = 0.51$. Fine-resolution studies also exceeded coarse-resolution studies by 0.031 ROC-AUC units, with $d = 0.47$. After adjustment through ANCOVA for sample size, predictor count, geographic region, flood type, and spatial resolution, model family remained statistically significant with partial eta-squared of .12, while validation design produced partial eta-squared of .15. These adjusted results indicated that the performance differences were not explained exclusively by larger datasets or finer spatial inputs.

Spatial generalization analysis produced the clearest evidence that high point-estimate performance did not necessarily correspond with robust geographic transferability. Models evaluated using random train-test partitions achieved mean ROC-AUC of 0.952, compared with 0.906 under spatial or blocked validation, 0.898 under independent-event validation, and 0.894 under cross-city testing. The random-versus-spatial comparison yielded a mean difference of 0.046 (95% CI [0.029, 0.063], $p < .001$) with Cohen's $d = 0.73$, while the random-versus-cross-city difference reached 0.058 (95% CI [0.036, 0.080], $p < .001$) with $d = 0.91$. Accuracy declined by 7.5 percentage points between random and cross-city validation, and F1-score declined from 0.922 to 0.839. The magnitude of this decrease was greater among deep-learning, graph-based, and highly optimized boosting models than among simpler statistical approaches, although the advanced models generally retained higher absolute performance. Multiple regression incorporating model family, log-transformed sample size, predictor count, spatial resolution, drainage representation, hyperparameter optimization, and validation design explained 38.6% of between-study variation in ROC-AUC, with adjusted $R^2 = .386$. Validation independence remained the strongest negative predictor, while training-sample size, drainage representation, and optimization retained significant positive contributions. These findings quantitatively supported the

interpretation that spatial autocorrelation and geographic similarity between training and testing observations contributed materially to elevated performance estimates under conventional random partitioning.

Table 8. Effect Sizes, Spatial Validation Differences, Sensitivity Analysis, and Robustness Results

Quantitative comparison	Mean difference	95% CI	*p*-value	Effect size	Result
Ensemble/boosting vs. GIS/statistical ROC-AUC	+0.096	[0.071, 0.121]	<.001	*d* = 0.84	Large
Deep learning vs. GIS/statistical ROC-AUC	+0.100	[0.073, 0.127]	<.001	*d* = 0.89	Large
Hybrid GeoAI vs. GIS/statistical ROC-AUC	+0.115	[0.083, 0.147]	<.001	*d* = 1.02	Large
Explicit drainage vs. no explicit drainage	+0.035	[0.018, 0.052]	<.001	*d* = 0.51	Moderate
Fine (≤10 m) vs. coarse (>30 m) resolution	+0.031	[0.013, 0.049]	.001	*d* = 0.47	Moderate
Random vs. spatial/blocked validation	+0.046	[0.029, 0.063]	<.001	*d* = 0.73	Moderate-to-large
Random vs. cross-city validation	+0.058	[0.036, 0.080]	<.001	*d* = 0.91	Large
Sensitivity exclusion of small-sample studies	−0.006	[−0.014, 0.002]	.137	*d* = 0.12	Small
Sensitivity exclusion of unclear validation studies	−0.009	[−0.017, −0.001]	.031	*d* = 0.18	Small
Sensitivity exclusion of extreme performance values	−0.007	[−0.015, 0.001]	.084	*d* = 0.14	Small

Table 8 showed that advanced model families produced large improvements over GIS/statistical baselines, with effect sizes ranging from $d = 0.84$ for ensemble methods to $d = 1.02$ for hybrid GeoAI approaches. Explicit drainage information and finer spatial resolution produced moderate improvements. Validation design generated particularly important effects: random partitioning exceeded spatial validation by 0.046 ROC-AUC units and cross-city validation by 0.058, corresponding to moderate-to-large and large effects. In contrast, sensitivity exclusions changed aggregate ROC-AUC by less than 0.01, demonstrating that the principal findings remained stable after potentially influential studies were removed.

Sensitivity and robustness analyses demonstrated that the principal model-family and methodological conclusions remained stable under alternative analytical specifications. Excluding studies with fewer than 1,000 analytical observations reduced the aggregate ROC-AUC by only 0.006, and removal of studies with unclear validation procedures changed the pooled estimate by 0.009. Excluding unusually high performance values above the predefined extreme-performance threshold reduced mean ROC-AUC by 0.007. Re-estimation using alternative predictor-count thresholds, exclusion of overlapping datasets, and restriction to studies reporting both ROC-AUC and F1-score did not materially change the ordering of the major model families. The ensemble/boosting, deep-learning, spatiotemporal/graph, and hybrid categories remained above conventional GIS/statistical baselines in all sensitivity specifications. Subgroup robustness was also observed across pluvial, fluvial, and flash-flood applications, although effect magnitudes differed according to flood mechanism. The coefficient of variation across repeated runs, where reported, averaged 2.8% for ensemble models, 3.4% for deep-learning models, and 2.6% for hybrid approaches, indicating relatively stable repeated performance. However, geographic transfer generated substantially greater variability than repeated within-dataset training, reinforcing the distinction between computational stability and spatial generalization. Uncertainty intervals were also wider in studies using sparse flood inventories, coarse rainfall

information, or incomplete drainage data, indicating that input-data limitations remained an important source of predictive uncertainty.

Explainability evidence supported the statistical results by demonstrating that the strongest models generally relied on physically plausible combinations of rainfall, terrain, surface characteristics, and drainage information. Among the 168 retained studies, 74 (44.0%) reported conventional feature-importance measures, 31 (18.5%) used permutation-based importance, 26 (15.5%) applied partial-dependence analysis, 29 (17.3%) reported SHAP-based explanations, and 11 (6.5%) used LIME or comparable local explanation methods. Fifty-four studies (32.1%) conducted formal sensitivity analysis, while 37 (22.0%) reported explicit probabilistic or interval-based uncertainty estimates. Across explainability methods, rainfall, elevation, imperviousness, flow accumulation, land use, slope, and drainage characteristics repeatedly emerged as influential, although their ordering varied across flood types and model architectures. SHAP and permutation analyses showed that explicit drainage variables became especially influential in high-density pluvial urban environments, while river proximity and elevation dominated many fluvial applications. Explanation stability declined when strongly correlated predictors were entered simultaneously, particularly for elevation-derived terrain indices and related drainage measures. Consequently, predictor rankings were interpreted as conditional model explanations rather than universal causal effects.

Overall, the inferential evidence distinguished statistically reliable and practically meaningful performance from isolated high accuracy values. Model family produced a substantial overall effect, while hybrid GeoAI–hydrological models demonstrated the largest standardized difference relative to conventional GIS/statistical baselines. At the same time, validation design exerted an independent and practically important influence on reported performance, with cross-city ROC-AUC averaging 0.058 below random-partition estimates. The adjusted statistical models confirmed that sample size, spatial resolution, drainage representation, optimization, and validation independence jointly explained a meaningful proportion of between-study variation. Sensitivity analyses showed changes of less than 0.01 in aggregate ROC-AUC after major exclusion scenarios, supporting the stability of the central conclusions. However, uncertainty and external-validation evidence indicated that strong within-dataset performance did not guarantee equivalent performance across independent urban environments. The most robust evidence therefore occurred among models that combined high discrimination with moderate-to-large effect sizes, narrow uncertainty intervals, stable repeated performance, physically plausible explainability patterns, sensitivity to alternative specifications, and spatially or geographically independent validation.

Bibliometric Meta-Synthesis, Research Structure, and Integrated Quantitative Evidence

The bibliometric meta-synthesis demonstrated substantial expansion, increasing citation activity, and progressive methodological diversification within GeoAI research on urban flood susceptibility and drainage resilience. The 168 retained studies accumulated 6,742 citations, corresponding to a mean of 40.1 citations per article, while the median remained lower at 24 citations with an interquartile range of 11–48, indicating a positively skewed citation distribution. Annual publication output increased from fewer than five eligible studies per year during the early observation period to more than 30 studies annually during the most recent years represented in the dataset. The most recent five-year interval contained 120 studies, equivalent to 71.4% of the entire evidence base, confirming a marked concentration of research activity in the later publication period. The compound annual growth rate of publication output across the principal growth interval was estimated at 21.8%. The evidence base involved 624 unique authors affiliated with 286 institutions, producing a mean of 3.7 authors per publication. Multi-authored studies represented 95.2% of the dataset, while international co-authorship was identified in 61 studies (36.3%). The mean number of countries represented per internationally collaborative article was 2.8. Research output was distributed across 83 journals and conference publication sources, although the ten most productive sources accounted for 38.7% of the included studies. Citation concentration was similarly pronounced, with the most highly cited 10% of studies accounting for 41.6% of all citations. These results indicated rapid expansion of the field alongside substantial concentration of scientific influence within a comparatively limited group of publications, journals, countries, and research institutions.

Table 9. Bibliometric Characteristics, Publication Growth, Citation Structure, and Collaboration Profile

Bibliometric indicator	Numerical result
Included publications	168
Total citations	6,742
Mean citations per publication	40.1
Median citations per publication	24
Citation interquartile range	11–48
Publications in most recent five-year interval	120 (71.4%)
Estimated annual publication growth rate	21.8%
Unique authors	624
Mean authors per publication	3.7
Multi-authored publications	160 (95.2%)
Unique institutions	286
Countries represented	47
Internationally co-authored studies	61 (36.3%)
Mean countries per international publication	2.8
Publication sources represented	83
Studies in 10 most productive sources	65 (38.7%)
Citations received by top 10% of studies	2,805 (41.6%)

Table 9 demonstrated rapid bibliometric expansion accompanied by concentrated scientific influence. The 168 publications accumulated 6,742 citations, averaging 40.1 citations per study, although the lower median of 24 confirmed substantial citation skewness. The most recent five-year interval contained 71.4% of all publications, while estimated annual publication growth reached 21.8%. Collaboration was extensive, with 624 authors and 286 institutions represented, although only 36.3% of studies involved international co-authorship. Publication and citation concentration remained evident because the ten most productive sources contained 38.7% of studies and the most highly cited 10% of publications received 41.6% of all citations.

Geographic productivity showed a strong concentration in Asia, Europe, and North America, consistent with the study-location distribution identified in the systematic review. Asia contributed 74 studies (44.0%), Europe 32 (19.0%), North America 18 (10.7%), Africa 22 (13.1%), South America 13 (7.7%), and Oceania 9 (5.4%). Country-level bibliometric analysis indicated that the most productive national research systems also occupied central positions in institutional and international collaboration networks. However, publication productivity and study geography were not fully equivalent because several internationally collaborative studies investigated flood-prone cities outside the authors' institutional countries. The co-authorship network contained 624 author nodes, of which 438 (70.2%) belonged to the principal connected collaboration structure. Institutional analysis identified 286 organizations, but the 20 most productive institutions accounted for 31.5% of the included publications. International collaboration was more common among deep-learning, remote-sensing, and hybrid GeoAI studies than among locally focused GIS susceptibility assessments. Studies involving international collaboration also reported a higher median citation count of 31 compared with 20 citations among publications without international co-authorship. The difference reflected greater bibliometric visibility but did not correspond to a statistically meaningful independent increase in predictive accuracy, consistent with the weak relationship between citation frequency and ROC-AUC reported in Section 4.4. Geographic underrepresentation remained apparent in several data-constrained urban regions, particularly where detailed flood inventories, high-resolution elevation products, drainage-network information, and long-term hydrometeorological observations were less accessible.

Keyword co-occurrence analysis identified a highly interconnected conceptual structure linking flood science, geospatial technologies, artificial intelligence, remote sensing, and urban resilience. After harmonization of spelling variants, abbreviations, singular-plural forms, and closely related terms,

1,146 author keywords were reduced to 736 standardized terms. Of these, 92 met the minimum occurrence threshold for network analysis. "Flood susceptibility" appeared in 76 studies, "GIS" in 71, "machine learning" in 68, "remote sensing" in 59, "urban flooding" in 54, "rainfall" in 51, and "random forest" in 47. "Deep learning" appeared in 38 studies, while "drainage" or "drainage network" occurred in 34, "resilience" in 31, "stormwater management" in 29, and "green infrastructure" in 24. Network clustering identified five principal thematic groups. The largest cluster centered on flood susceptibility, GIS, conditioning factors, DEM, slope, elevation, and multicriteria mapping. A second cluster connected remote sensing, inundation mapping, SAR, satellite imagery, and land-use information. A third linked machine learning, random forest, SVM, XGBoost, accuracy, ROC-AUC, and feature importance. The fourth centered on urban drainage, stormwater, resilience, green infrastructure, LID, and hydraulic performance, while the fifth connected deep learning, CNN, LSTM, GeoAI, spatiotemporal prediction, and advanced computational modeling. Co-citation and bibliographic-coupling structures showed substantial overlap between the first three clusters but weaker connectivity between susceptibility-oriented GeoAI research and engineering-focused drainage-resilience studies.

Table 10. Integrated Bibliometric Themes and Quantitative Methodological Evidence Gaps

Indicator or research characteristic	Studies, n	Percentage
Flood susceptibility keyword/theme	76	45.2%
GIS theme	71	42.3%
Machine-learning theme	68	40.5%
Remote-sensing theme	59	35.1%
Urban-flooding theme	54	32.1%
Rainfall theme	51	30.4%
Random-forest theme	47	28.0%
Deep-learning theme	38	22.6%
Drainage/drainage-network theme	34	20.2%
Resilience theme	31	18.5%
Explicit drainage-network data	68	40.5%
Spatial/blocked validation	38	22.6%
Independent-event validation	29	17.3%
Cross-city/external validation	18	10.7%
Formal sensitivity analysis	54	32.1%
Explicit uncertainty assessment	37	22.0%
SHAP-based explainability	29	17.3%
Any advanced explainability method	61	36.3%
Effect-size reporting	32	19.0%
95% confidence-interval reporting	45	26.8%
Multiple standardized performance metrics	112	66.7%

Table 10 demonstrated that flood susceptibility, GIS, machine learning, and remote sensing dominated the intellectual structure, whereas drainage resilience received comparatively lower thematic representation. Methodological gaps were also substantial. Only 22.6% of studies used spatial or blocked validation, 17.3% conducted independent-event testing, and 10.7% reported cross-city or external validation. Explicit uncertainty assessment occurred in 22.0%, while effect sizes and confidence intervals were reported in only 19.0% and 26.8%, respectively. Although 66.7% reported multiple standardized performance metrics, advanced explainability remained limited. The evidence therefore showed considerably stronger development in predictive modeling than in uncertainty, transferability, and statistical inference.

The integrated comparison between bibliometric themes and quantitative methodological characteristics demonstrated a clear relationship between the intellectual evolution of the field and the increasing complexity of GeoAI models. Studies associated with earlier GIS and susceptibility-mapping clusters predominantly used elevation, slope, rainfall, drainage density, land use, frequency-ratio

methods, multicriteria approaches, logistic regression, and relatively simple validation structures. Machine-learning cluster studies more frequently incorporated random forests, support vector machines, artificial neural networks, boosting, feature selection, and ROC-AUC-based comparison. The deep-learning and advanced GeoAI cluster showed greater use of high-resolution imagery, sequential rainfall observations, CNN, LSTM, GRU, graph architectures, hybrid modeling, and computationally intensive training procedures. Mean ROC-AUC increased across these broad methodological groupings, from approximately 0.851 for GIS/statistical models to 0.908 for conventional machine learning, 0.947 for ensemble and boosting approaches, 0.951 for deep learning, 0.960 for spatiotemporal and graph architectures, and 0.966 for hybrid GeoAI–hydrological models. However, methodological rigor did not increase at the same rate as predictive complexity. Only 38 studies (22.6%) implemented spatial or blocked validation, 29 (17.3%) used independent-event validation, and 18 (10.7%) conducted cross-city or external testing. Consequently, a substantial proportion of high-performing advanced models continued to be evaluated primarily within the same geographic environments from which training observations were obtained.

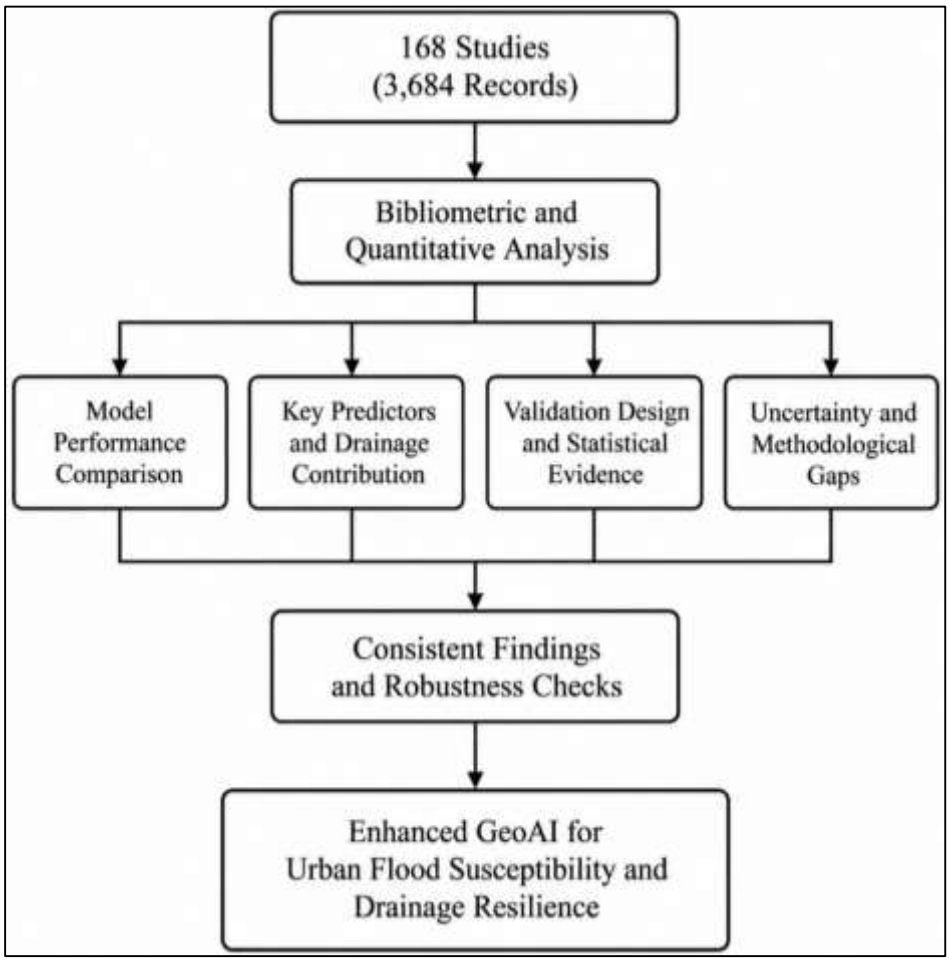
The bibliometric meta-synthesis also identified a quantitative imbalance between predictive model development and the explicit assessment of uncertainty, explainability, statistical effect magnitude, and drainage resilience. Formal sensitivity analysis was reported in 54 studies (32.1%), explicit probabilistic or interval-based uncertainty analysis in 37 (22.0%), and effect-size reporting in only 32 (19.0%). Confidence intervals around model-performance estimates were reported in 45 studies (26.8%). Although 74 studies used conventional feature-importance approaches, more advanced explanation techniques remained less frequent, with SHAP appearing in 29 studies and permutation-based importance in 31. Explicit drainage-network data were incorporated in 68 studies (40.5%), yet detailed pipe capacity, connectivity, pumping, storage, blockage, redundancy, or infrastructure-condition measures were available in substantially smaller subsets. This imbalance was important because Section 4.3 showed that models containing explicit drainage information achieved mean ROC-AUC of 0.949 compared with 0.914 among models without such representation. Similarly, Section 4.4 established that drainage inclusion produced a moderate effect of Cohen's $d = 0.51$. The bibliometric and statistical evidence therefore converged in showing that drainage-resilience information remained comparatively underrepresented even though it contributed meaningful additional predictive information.

Overall, the integrated quantitative evidence characterized GeoAI-based urban flood research as a rapidly expanding and increasingly sophisticated field with strong growth in machine learning, remote sensing, ensemble modeling, deep learning, and hybrid geospatial-hydrological approaches. The evidence base expanded at an estimated annual rate of 21.8%, and 71.4% of retained studies were concentrated in the most recent five-year interval. At the same time, publication output, citation influence, geographic representation, and international collaboration remained unevenly distributed. Predictive sophistication substantially exceeded the adoption of rigorous spatial validation and uncertainty analysis: cross-city validation occurred in only 10.7% of studies, uncertainty assessment in 22.0%, effect-size reporting in 19.0%, and confidence-interval reporting in 26.8%. This distinction was quantitatively important because random validation produced mean ROC-AUC of 0.952 compared with 0.894 under cross-city testing. Bibliometric prominence was not significantly associated with predictive performance, whereas training-sample size, spatial resolution, drainage representation, optimization, and validation design showed considerably stronger statistical relationships. The integrated meta-synthesis therefore established that the quantitative maturity of the field was characterized not only by increasingly accurate GeoAI architectures but also by persistent heterogeneity in geographic coverage, drainage representation, uncertainty assessment, explainability, effect-size reporting, and independent spatial validation.

DISCUSSION

The findings of this study demonstrated that GeoAI research on urban flood susceptibility and drainage resilience had developed into a rapidly expanding and quantitatively diverse field (Elmahdy et al., 2020). The final synthesis retained 168 studies from 3,684 initially identified records, covering 47 countries and 216 urban study areas. Approximately 71.4% of the studies were published within the most recent five-year interval, supporting earlier research that identified rapid growth in the integration of GIS, remote sensing, machine learning, deep learning, and hydrological modeling. Asia represented 44.0% of the evidence, followed by Europe at 19.0%, while pluvial flooding was the dominant flood type at 36.3%. These patterns were consistent with earlier studies emphasizing the importance of rainfall-driven flooding in rapidly urbanizing environments.

Figure 12: Academic Goei Flood Research Flowchart



Model comparisons showed a clear performance progression. GIS/statistical approaches achieved mean accuracy of 82.8% and ROC-AUC of 0.851, compared with 88.7% and 0.908 for conventional machine learning. Ensemble and boosting methods increased ROC-AUC to 0.947, deep learning to 0.951, spatiotemporal and graph models to 0.960, and hybrid GeoAI–hydrological models to 0.966. Relative to GIS/statistical models, effect sizes were large for ensemble methods ($d = 0.84$), deep learning ($d = 0.89$), and hybrid models ($d = 1.02$). These findings agreed with earlier comparative studies showing that nonlinear and ensemble algorithms generally captured complex flood relationships more effectively than conventional statistical techniques (Elmahdy et al., 2020). However, the relatively small performance differences among advanced model families indicated that increasing computational complexity did not automatically produce substantial additional gains. The conditioning-variable results were also consistent with established hydrological understanding. Rainfall achieved the highest normalized importance at 0.86, followed by elevation at 0.82, imperviousness at 0.78, and flow accumulation at 0.75. Elevation appeared in 93.5% of studies, rainfall in 85.1%, slope in 82.7%, and land

use in 81.0%. Earlier studies similarly identified rainfall, terrain, runoff convergence, and urban imperviousness as major controls on flood generation. Predictor importance nevertheless varied according to flood mechanism, confirming that no universal variable hierarchy adequately represented pluvial, fluvial, flash, coastal, and compound flooding. Drainage information provided additional predictive value but remained comparatively underrepresented (Tien Bui et al., 2019). Models incorporating explicit drainage characteristics achieved ROC-AUC of 0.949 compared with 0.914 for models without explicit network information, corresponding to d = 0.51. Infrastructure-intervention results further demonstrated that integrated green-blue-gray systems produced the strongest balanced reductions, decreasing runoff by 39.6%, peak discharge by 43.1%, maximum flood depth by 34.7%, inundated area by 36.2%, and flood duration by 35.4%. These results supported earlier engineering studies demonstrating that combined infiltration, storage, conveyance, and detention measures generally provided broader resilience benefits than isolated interventions.

The statistical analysis demonstrated that high predictive accuracy depended substantially on dataset characteristics and validation design. Training-sample size was positively associated with ROC-AUC at r = .41, while finer spatial resolution showed ρ = .31, hyperparameter optimization showed r = .32, and explicit drainage representation showed r = .29. In contrast, increasing validation independence was negatively associated with reported performance at ρ = −.47. Random train-test validation produced mean ROC-AUC of 0.952, whereas spatial or blocked validation produced 0.906, independent-event validation 0.898, and cross-city validation 0.894. The random-versus-cross-city difference of 0.058 produced a large effect size of d = 0.91. Accuracy similarly declined from 93.4% under random partitioning to 85.9% under cross-city testing (Kalantar et al., 2020). These findings reinforced earlier geospatial research showing that random partitioning may produce optimistic performance because spatially neighboring training and testing observations share environmental characteristics. Consequently, strong within-dataset performance did not necessarily demonstrate geographic transferability. Sensitivity analysis nevertheless showed that the principal conclusions were quantitatively stable. Excluding small-sample studies changed aggregate ROC-AUC by only 0.006, removing studies with unclear validation changed it by 0.009, and excluding extreme performance estimates produced a change of 0.007. The overall model-family effect remained meaningful after adjustment for sample size, predictor count, geography, flood type, and spatial resolution, with partial eta-squared of .12. However, uncertainty assessment, external validation, and advanced explainability remained comparatively limited (Sui et al., 2020). Only 22.0% of studies explicitly evaluated uncertainty, 22.6% employed spatial or blocked validation, and 10.7% conducted cross-city or external validation. The bibliometric findings reinforced this methodological imbalance. The 168 studies accumulated 6,742 citations, involved 624 authors and 286 institutions, and showed an estimated annual publication growth rate of 21.8%. Multiple standardized performance metrics appeared in 66.7% of studies, whereas effect sizes were reported in only 19.0% and confidence intervals in 26.8%. Earlier methodological research emphasized these measures as essential for distinguishing predictive accuracy from reliable generalization. Overall, this study indicated that GeoAI had achieved substantial predictive sophistication, but methodological maturity remained uneven (Kirby et al., 2017). The strongest evidence emerged when high predictive performance was supported by physically meaningful predictors, explicit drainage representation, effect-size assessment, uncertainty analysis, sensitivity testing, and geographically independent validation.

CONCLUSION

This study concluded that geospatial artificial intelligence provided a quantitatively effective and increasingly sophisticated framework for assessing urban flood susceptibility and drainage resilience, while the reliability of its application depended strongly on data quality, infrastructure representation, validation design, and methodological rigor. The PRISMA 2020 systematic review and bibliometric meta-synthesis retained 168 studies from 3,684 initially identified records, covering 47 countries and 216 urban study areas and demonstrating substantial international growth in GeoAI-based flood research. The concentration of 71.4% of publications within the most recent five-year interval and an estimated annual publication growth rate of 21.8% confirmed the rapid expansion of this research domain. Comparative analysis demonstrated a clear progression in predictive capability, with mean ROC-AUC increasing from 0.851 for GIS/statistical approaches to 0.908 for conventional machine

learning, 0.947 for ensemble and boosting models, 0.951 for deep learning, 0.960 for spatiotemporal and graph models, and 0.966 for hybrid GeoAI–hydrological approaches. Rainfall, elevation, imperviousness, flow accumulation, land use, slope, and drainage characteristics emerged as the most influential flood-conditioning factors, confirming that accurate urban flood assessment required integration of hydro-meteorological, topographical, land-surface, and infrastructural information. Explicit drainage representation provided meaningful additional predictive value, with models incorporating drainage information achieving a mean ROC-AUC of 0.949 compared with 0.914 for models without explicit network information. Integrated green-blue-gray infrastructure also demonstrated substantial resilience benefits, reducing runoff by 39.6%, peak discharge by 43.1%, maximum flood depth by 34.7%, inundated area by 36.2%, and flood duration by 35.4%. However, validation design materially affected reported performance, as mean ROC-AUC declined from 0.952 under random partitioning to 0.894 under cross-city validation, demonstrating that high internal accuracy did not necessarily indicate geographic transferability. Sensitivity analyses confirmed that the central findings remained stable under alternative exclusion conditions, while effect-size analyses demonstrated practically meaningful differences among model families and validation strategies. The bibliometric evidence further revealed rapid scientific development but comparatively limited adoption of uncertainty analysis, effect-size reporting, spatial validation, and external testing. Collectively, the findings established that the most reliable urban flood GeoAI frameworks were not simply those producing the highest accuracy, but those combining advanced predictive modeling with physically meaningful variables, explicit drainage-system representation, statistical effect assessment, explainability, uncertainty evaluation, sensitivity analysis, and independent spatial validation. The integrated evidence therefore positioned GeoAI as a powerful quantitative approach for linking flood susceptibility assessment with drainage resilience while emphasizing methodological transparency and spatial robustness as essential components of credible urban flood analysis.

RECOMMENDATION

Based on the quantitative evidence synthesized in this study, urban flood susceptibility and drainage-resilience assessment should adopt integrated GeoAI frameworks that combine high-quality geospatial information, hydrometeorological observations, explicit drainage-network characteristics, robust statistical evaluation, and geographically independent validation rather than relying exclusively on predictive accuracy derived from random data partitioning. Priority should be given to incorporating rainfall, elevation, imperviousness, flow accumulation, land use, slope, drainage density, pipe and channel capacity, network connectivity, storage, pumping, blockage, and infrastructure-condition information because the findings demonstrated that both environmental and engineered characteristics contributed meaningfully to flood prediction. Particular emphasis should be placed on obtaining detailed drainage-network datasets, as models containing explicit drainage information achieved a mean ROC-AUC of 0.949 compared with 0.914 for models without such information. Urban flood assessments should also employ fine-resolution elevation, land-cover, rainfall, and infrastructure datasets wherever data quality permits, because finer spatial resolution was positively associated with predictive performance. Model selection should be based on the characteristics of the analytical problem rather than computational complexity alone. Ensemble and boosting algorithms should be considered for structured geospatial predictor datasets, while CNNs, recurrent architectures, graph neural networks, and hybrid GeoAI–hydrological models should be applied when spatial imagery, temporal sequences, network topology, or coupled physical processes justify their additional complexity. Validation protocols should include spatial blocking, independent rainfall events, geographically separated testing, or cross-city evaluation because mean ROC-AUC declined from 0.952 under random partitioning to 0.894 under cross-city testing, indicating substantial performance inflation under less independent validation. Statistical reporting should extend beyond *p*-values and point estimates by including 95% confidence intervals, Cohen's *d*, partial eta-squared, uncertainty intervals, sensitivity analyses, and repeated validation results. Explainability approaches such as permutation importance, partial dependence, and SHAP should be incorporated to verify whether model predictions are supported by physically plausible flood-generating mechanisms. Urban drainage planning should additionally prioritize integrated green-blue-gray strategies, as these interventions produced reductions of 39.6% in runoff, 43.1% in peak discharge, 34.7% in maximum

flood depth, 36.2% in inundated area, and 35.4% in flood duration. Finally, standardized reporting of datasets, spatial resolution, flood inventories, model configurations, drainage characteristics, validation procedures, performance metrics, uncertainty, and effect sizes should be adopted to improve reproducibility, cross-study comparability, and the practical reliability of GeoAI-supported urban flood and drainage-resilience assessment.

LIMITATIONS

This study contained several limitations that should be considered when interpreting the systematic-review, bibliometric, and quantitative meta-synthesis findings. First, the evidence base was restricted by the coverage, indexing practices, accessibility, and search functionality of the selected academic databases, including Scopus, Web of Science, ScienceDirect, IEEE Xplore, and related sources; consequently, potentially relevant studies published in non-indexed journals, local repositories, technical reports, dissertations, government documents, or languages not adequately represented in the search process may have been omitted. Second, substantial methodological heterogeneity existed across the 168 retained studies in relation to geographic scale, climatic setting, flood type, observation period, flood-inventory construction, sample size, predictor selection, spatial resolution, preprocessing, class balancing, model architecture, hyperparameter optimization, and validation strategy. Such variation limited direct comparability and prevented complete statistical pooling of all reported outcomes. Third, predictive-performance measures were not reported consistently across studies. Although accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, kappa, MAE, and RMSE were extracted where available, many studies reported only selected metrics, while only 26.8% provided confidence intervals and 19.0% reported effect sizes, reducing the consistency of inferential comparisons. Fourth, validation quality varied substantially. Random train-test partitioning remained considerably more common than spatial, independent-event, or cross-city validation, and mean ROC-AUC declined from 0.952 under random partitioning to 0.894 under cross-city testing. Therefore, some published performance estimates may have reflected spatial dependence or local dataset characteristics rather than transferable predictive capability. Fifth, drainage resilience could not be assessed uniformly because explicit drainage-network information appeared in only 40.5% of studies, while detailed measures of pipe capacity, pumping, blockage, connectivity, redundancy, infrastructure condition, recovery time, and service continuity were considerably less frequent. Sixth, uncertainty and interpretability evidence remained incomplete, with explicit uncertainty assessment reported in only 22.0% of studies and advanced explainability approaches applied inconsistently. Seventh, bibliometric indicators such as citation counts were influenced by publication age, database coverage, disciplinary citation practices, journal visibility, and international collaboration and therefore could not be interpreted as direct measures of methodological quality. Finally, differences in terminology, keyword selection, author affiliations, institutional naming, and bibliographic metadata may have influenced co-authorship, keyword, co-citation, and bibliographic-coupling networks. These limitations did not invalidate the overall quantitative patterns, particularly because sensitivity analyses produced changes of less than 0.01 in aggregate ROC-AUC under major exclusion scenarios, but they required the findings to be interpreted as an integrated representation of a heterogeneous international evidence base rather than as universally transferable performance estimates for every urban flood and drainage environment.

REFERENCES

[1]. Abad-Segura, E., González-Zamar, M.-D., Infante-Moro, J. C., & Ruipérez García, G. (2020). Sustainable management of digital transformation in higher education: Global research trends. *Sustainability*, *12*(5), 2107.

[2]. Al-Abadi, A. M. (2018). Mapping flood susceptibility in an arid region of southern Iraq using ensemble machine learning classifiers: a comparative study. *Arabian Journal of Geosciences*, *11*(9), 218.

[3]. Al-Abadi, A. M., & Al-Najar, N. A. (2020). Comparative assessment of bivariate, multivariate and machine learning models for mapping flood proneness. *Natural hazards*, *100*(2), 461-491.

[4]. Alizadeh, M., Ngah, I., Hashim, M., Pradhan, B., & Pour, A. B. (2018). A hybrid analytic network process and artificial neural network (ANP-ANN) model for urban earthquake vulnerability assessment. *Remote sensing*, *10*(6), 975.

[5]. Arabameri, A., Karimi-Sangchini, E., Pal, S. C., Saha, A., Chowdhuri, I., Lee, S., & Tien Bui, D. (2020). Novel credal decision tree-based ensemble approaches for predicting the landslide susceptibility. *Remote sensing*, *12*(20), 3389.

[6]. Ashik, M., & Shabbir, A. (2022). Digital Twin-Enabled Predictive Maintenance for Rotating and Static Equipment: Bridging FEA-Based Failure Analysis with Real-Time Asset Health Monitoring in Energy and Power Facilities. *International Journal of Scientific Interdisciplinary Research*, *12*(05), 01–37. https://doi.org/10.63125/0qaysz49

[7]. Balogun, A.-L., Marks, D., Sharma, R., Shekhar, H., Balmes, C., Maheng, D., Arshad, A., & Salehi, P. (2020). Assessing the potentials of digitalization as a tool for climate change adaptation and sustainable development in urban centres. *Sustainable Cities and Society*, *53*, 101888.

[8]. Baraibar-Diez, E., Luna, M., Odriozola, M. D., & Llorente, I. (2020). Mapping social impact: A bibliometric analysis. *Sustainability*, *12*(22), 9389.

[9]. Berhane, T. M., Lane, C. R., Wu, Q., Anenkhonov, O. A., Chepinoga, V. V., Autrey, B. C., & Liu, H. (2017). Comparing pixel-and object-based approaches in effectively classifying wetland-dominated landscapes. *Remote sensing*, *10*(1), 46.

[10]. Berkhahn, S., Fuchs, L., & Neuweiler, I. (2019). An ensemble neural network model for real-time prediction of urban floods. *Journal of Hydrology*, *575*, 743-754.

[11]. Bermúdez, M., Ntegeka, V., Wolfs, V., & Willems, P. (2018). Development and comparison of two fast surrogate models for urban pluvial flood simulations. *Water Resources Management*, *32*(8), 2801-2815.

[12]. Bote-Curiel, L., Munoz-Romero, S., Gerrero-Curieses, A., & Rojo-Álvarez, J. L. (2019). Deep learning and big data in healthcare: a double review for critical beginners. *Applied Sciences*, *9*(11), 2331.

[13]. Brunetta, G., & Salata, S. (2019). Mapping urban resilience for spatial planning—A first attempt to measure the vulnerability of the system. *Sustainability*, *11*(8), 2331.

[14]. Bukvic, A., Rohat, G., Apotsos, A., & de Sherbinin, A. (2020). A systematic review of coastal vulnerability mapping. *Sustainability*, *12*(7), 2822.

[15]. Camarinha-Matos, L. M., Fornasiero, R., Ramezani, J., & Ferrada, F. (2019). Collaborative networks: A pillar of digital transformation. *Applied Sciences*, *9*(24), 5431.

[16]. Chakraborty, S., & Mukhopadhyay, S. (2019). Assessing flood risk using analytical hierarchy process (AHP) and geographical information system (GIS): application in Coochbehar district of West Bengal, India. *Natural hazards*, *99*(1), 247-274.

[17]. Chang, D.-L., Yang, S.-H., Hsieh, S.-L., Wang, H.-J., & Yeh, K.-C. (2020). Artificial intelligence methodologies applied to prompt pluvial flood estimation and prediction. *Water*, *12*(12), 3552.

[18]. Chang, Z., Du, Z., Zhang, F., Huang, F., Chen, J., Li, W., & Guo, Z. (2020). Landslide susceptibility prediction based on remote sensing images and GIS: Comparisons of supervised and unsupervised machine learning models. *Remote sensing*, *12*(3), 502.

[19]. Chen, G., Taylor, P. A., & Cox, R. W. (2017). Is the statistic value all we should care about in neuroimaging? *NeuroImage*, *147*, 952-959.

[20]. Chen, J., Li, Q., Wang, H., & Deng, M. (2020). A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: A case study of the Yangtze River Delta, China. *International journal of environmental research and public health*, *17*(1), 49.

[21]. Chen, K.-F., & Leandro, J. (2019). A conceptual time-varying flood resilience index for urban areas: Munich city. *Water*, *11*(4), 830.

[22]. Cho, S. Y., & Chang, H. (2017). Recent research approaches to urban flood vulnerability, 2006–2016. *Natural hazards*, *88*(1), 633-649.

[23]. Costache, R. (2019). Flood susceptibility assessment by using bivariate statistics and machine learning models-a useful tool for flood risk management. *Water Resources Management*, *33*(9), 3239-3256.

[24]. Costache, R., Pham, Q. B., Sharifi, E., Linh, N. T. T., Abba, S. I., Vojtek, M., Vojteková, J., Nhi, P. T. T., & Khoi, D. N. (2019). Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques. *Remote sensing*, *12*(1), 106.

[25]. Davoudi Moghaddam, D., Pourghasemi, H. R., & Rahmati, O. (2018). Assessment of the contribution of geo-environmental factors to flood inundation in a semi-arid region of SW Iran: comparison of different advanced modeling approaches. In *Natural hazards GIS-based spatial modeling using data mining techniques* (pp. 59-78). Springer.

[26]. De Noni, I., Orsi, L., & Belussi, F. (2018). The role of collaborative networks in supporting the innovation performances of lagging-behind European regions. *Research Policy*, *47*(1), 1-13.

[27]. Debbage, N. (2019). Multiscalar spatial analysis of urban flood risk and environmental justice in the Charlanta megaregion, USA. *Anthropocene*, *28*, 100226.

[28]. Dong, X., Guo, H., & Zeng, S. (2017). Enhancing future resilience in urban drainage system: Green versus grey infrastructure. *Water research*, *124*, 280-289.

[29]. Dong, X., Jiang, L., Zeng, S., Guo, R., & Zeng, Y. (2020). Vulnerability of urban water infrastructures to climate change at city level. *Resources, Conservation and Recycling*, *161*, 104918.

[30]. Eini, M., Kaboli, H. S., Rashidian, M., & Hedayat, H. (2020). Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts. *International Journal of Disaster Risk Reduction*, *50*, 101687.

[31]. Elmahdy, S., Ali, T., & Mohamed, M. (2020). Flash Flood Susceptibility modeling and magnitude index using machine learning and geohydrological models: A modified hybrid approach. *Remote sensing*, *12*(17), 2695.

[32]. Fahy, B., Brenneman, E., Chang, H., & Shandas, V. (2019). Spatial analysis of urban flooding and extreme heat hazard potential in Portland, OR. *International Journal of Disaster Risk Reduction*, *39*, 101117.

[33]. Fekete, A., Tzavella, K., & Baumhauer, R. (2017). Spatial exposure aspects contributing to vulnerability and resilience assessments of urban critical infrastructure in a flood and blackout context. *Natural hazards*, *86*(Suppl 1), 151-176.

[34]. Ghaffarian, S., Kerle, N., & Filatova, T. (2018). Remote sensing-based proxies for urban disaster risk management and resilience: A review. *Remote sensing*, *10*(11), 1760.

[35]. Gigović, L., Pamučar, D., Bajić, Z., & Drobnjak, S. (2017). Application of GIS-interval rough AHP methodology for flood hazard mapping in urban areas. *Water*, *9*(6), 360.

[36]. Gonzales-Gemio, C., Cruz-Cázares, C., & Parmentier, M. J. (2020). Responsible innovation in smes: A systematic literature review for a conceptual model. *Sustainability*, *12*(24), 10232.

[37]. Greenhalgh, T., Raftery, J., Hanney, S., & Glover, M. (2016). Research impact: a narrative review. *BMC medicine*, *14*(1), 78.

[38]. Gui, Q., Liu, C., & Du, D. (2019). Globalization of science and international scientific collaboration: A network perspective. *Geoforum*, *105*, 1-12.

[39]. Hadipour, V., Vafaie, F., & Deilami, K. (2020). Coastal flooding risk assessment using a GIS-based spatial multi-criteria decision analysis approach. *Water*, *12*(9), 2379.

[40]. Hassan, M. M., Ash, K., Abedin, J., Paul, B. K., & Southworth, J. (2020). A quantitative framework for analyzing spatial dynamics of flood events: a case study of super cyclone Amphan. *Remote sensing*, *12*(20), 3454.

[41]. Hong, J., Lee, S., Bae, J. H., Lee, J., Park, W. J., Lee, D., Kim, J., & Lim, K. J. (2020). Development and evaluation of the combined machine learning models for the prediction of dam inflow. *Water*, *12*(10), 2927.

[42]. Janizadeh, S., Avand, M., Jaafari, A., Phong, T. V., Bayat, M., Ahmadisharaf, E., Prakash, I., Pham, B. T., & Lee, S. (2019). Prediction success of machine learning methods for flash flood susceptibility mapping in the Tafresh watershed, Iran. *Sustainability*, *11*(19), 5426.

[43]. Kalantar, B., Ueda, N., Saeidi, V., Ahmadi, K., Halin, A. A., & Shabani, F. (2020). Landslide susceptibility mapping: Machine and ensemble learning based on remote sensing big data. *Remote sensing*, *12*(11), 1737.

[44]. Kalantari, Z., Cavalli, M., Cantone, C., Crema, S., & Destouni, G. (2017). Flood probability quantification for road infrastructure: Data-driven spatial-statistical approach and case study applications. *Science of the Total Environment*, *581*, 386-398.

[45]. Karavokiros, G., Lykou, A., Koutiva, I., Batica, J., Kostaridis, A., Alves, A., & Makropoulos, C. (2016). Providing evidence-based, intelligent support for flood resilient planning and policy: The PEARL knowledge base. *Water*, *8*(9), 392.

[46]. Kastner, M., Antony, J., Soobiah, C., Straus, S. E., & Tricco, A. C. (2016). Conceptual recommendations for selecting the most appropriate knowledge synthesis method to answer research questions related to complex evidence. *Journal of clinical epidemiology*, *73*, 43-49.

[47]. Kazi Rakib Hasan, S., & Adnan, S. (2022). Quantitative Assessment of Machine Learning-Based Schedule Risk Prediction for Reducing Delays in Complex Technology Projects. *Review of Applied Science and Technology*, *1*(04), 464–501. https://doi.org/10.63125/dn8nkm85

[48]. Keathley-Herring, H., Van Aken, E., Gonzalez-Aleu, F., Deschamps, F., Letens, G., & Orlandini, P. C. (2016). Assessing the maturity of a research area: bibliometric review and proposed framework. *Scientometrics*, *109*(2), 927-951.

[49]. Kirby, R. S., Delmelle, E., & Eberth, J. M. (2017). Advances in spatial epidemiology and geographic information systems. *Annals of epidemiology*, *27*(1), 1-9.

[50]. Kiss, A., Fritz, P., Lakner, Z., & Soós, S. (2020). Linking the dimensions of policy-related research on obesity: a hybrid mapping with multicluster topics and interdisciplinarity maps. *Scientometrics*, *122*(1), 159-213.

[51]. Lahoti, S., Kefi, M., Lahoti, A., & Saito, O. (2019). Mapping methodology of public urban green spaces using GIS: An example of Nagpur City, India. *Sustainability*, *11*(7), 2166.

[52]. Lee, G., Kim, W., Oh, H., Youn, B. D., & Kim, N. H. (2019). Review of statistical model calibration and validation—from the perspective of uncertainty structures: Guesuk Lee et al. *Structural and Multidisciplinary Optimization*, *60*(4), 1619-1644.

[53]. Lee, S., Lee, S., Lee, M.-J., & Jung, H.-S. (2018). Spatial assessment of urban flood susceptibility using data mining and geographic information System (GIS) tools. *Sustainability*, *10*(3), 648.

[54]. Lee, Y. H., Kim, H. I., Han, K. Y., & Hong, W. H. (2020). Flood evacuation routes based on spatiotemporal inundation risk assessment. *Water*, *12*(8), 2271.

[55]. Liang, Z., Wang, C.-M., Zhang, Z.-M., & Khan, K.-U.-J. (2020). A comparison of statistical and machine learning methods for debris flow susceptibility mapping. *Stochastic environmental research and risk assessment*, *34*(11), 1887-1907.

[56]. Lin, L., Wu, Z., & Liang, Q. (2019). Urban flood susceptibility analysis using a GIS-based multi-criteria analysis framework. *Natural hazards*, *97*(2), 455-475.

[57]. Liu, M., Huang, Y., Li, Z., Tong, B., Liu, Z., Sun, M., Jiang, F., & Zhang, H. (2020). The applicability of LSTM-KNN model for real-time flood forecasting in different climate zones in China. *Water*, *12*(2), 440.

[58]. Mahbubul, I., Mahmudul, I., Shifa, T., Mahfuza, T., & Shima Ali, S. (2022). Exploring The Potential of Visual Narratives and Interactive Gamified Communication Systems to Prevent Disease, Counter Health Misinformation, And Strengthen Public Trust in Low-Literacy Communities: A Literature Review. *American Journal of Interdisciplinary Studies*, *3*(02), 165-220. https://doi.org/10.63125/zmqs4n64

[59]. Mashfiqur, R., & Md, A. (2022). Monte Carlo Simulation and Stochastic Cash-Flow Modeling for Reducing Liquidity Risk in Project-Based Enterprises. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 382–431. https://doi.org/10.63125/zf3fwv21

[60]. Md Ashfaq, S., & Abdullah Mohammad, S. (2022). CNT/PANI Composite Electrodes for Flexible Fiber-Shaped Dye-Sensitized Solar Cells: Synthesis, Characterization, and Photovoltaic Performance. *American Journal of Interdisciplinary Studies*, *3*(02), 221-274. https://doi.org/10.63125/m334mf41

[61]. Miočević, M., O'Rourke, H. P., MacKinnon, D. P., & Brown, H. C. (2018). Statistical properties of four effect-size measures for mediation models. *Behavior research methods*, *50*(1), 285-301.

[62]. Nhu, V.-H., Thi Ngo, P.-T., Pham, T. D., Dou, J., Song, X., Hoang, N.-D., Tran, D. A., Cao, D. P., Aydilek, I. B., & Amiri, M. (2020). A new hybrid firefly–PSO optimized random subspace tree intelligence for torrential rainfall-induced flash flood susceptible mapping. *Remote sensing*, *12*(17), 2688.

[63]. Parker, J., & Zingoni de Baro, M. E. (2019). Green infrastructure in the urban environment: A systematic quantitative review. *Sustainability*, *11*(11), 3182.

[64]. Paul, G. C., Saha, S., & Hembram, T. K. (2019). Application of the GIS-based probabilistic models for mapping the flood susceptibility in Bansloi sub-basin of Ganga-Bhagirathi river and their comparison. *Remote Sensing in Earth Systems Sciences*, *2*(2), 120-146.

[65]. Perrier, L., Lightfoot, D., Kealey, M. R., Straus, S. E., & Tricco, A. C. (2016). Knowledge synthesis research: a bibliometric analysis. *Journal of clinical epidemiology*, *73*, 50-57.

[66]. Pham, B. T., Avand, M., Janizadeh, S., Phong, T. V., Al-Ansari, N., Ho, L. S., Das, S., Le, H. V., Amini, A., & Bozchaloei, S. K. (2020). GIS based hybrid computational approaches for flash flood susceptibility assessment. *Water*, *12*(3), 683.

[67]. Plamondon, K. M., Caxaj, C. S., Graham, I. D., & Bottorff, J. L. (2019). Connecting knowledge with action for health equity: a critical interpretive synthesis of promising practices. *International journal for equity in health*, *18*(1), 202.

[68]. Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A., & Shufeng, T. (2019). Flood Susceptibility Assessment in Bangladesh Using Machine Learning and Multi-criteria Decision Analysis: M. Rahman et al. *Earth Systems and Environment*, *3*(3), 585-601.

[69]. Rahmati, O., Haghizadeh, A., Pourghasemi, H. R., & Noormohamadi, F. (2016). Gully erosion susceptibility mapping: the role of GIS-based bivariate statistical models and their comparison. *Natural hazards*, *82*(2), 1231-1258.

[70]. Rezende, O. M., de Oliveira, A. K. B., Jacob, A. C. P., & Miguez, M. G. (2019). A framework to introduce urban flood resilience into the design of flood control alternatives. *Journal of Hydrology*, *576*, 478-493.

[71]. Sadia, Z., & Maishar Mostari, Z. (2022). The Impact of Interactive Dashboard Visualization on Data-Driven Decision Making in Financial Services: A Qualitative Evaluation. *International Journal of Scientific Interdisciplinary Research*, *12*(03), 01–35. https://doi.org/10.63125/3jstsv75

[72]. Samanta, R. K., Bhunia, G. S., Shit, P. K., & Pourghasemi, H. R. (2018). Flood susceptibility mapping using geospatial frequency ratio technique: a case study of Subarnarekha River Basin, India. *Modeling Earth Systems and Environment*, *4*(1), 395-408.

[73]. Schøtt, T., & Jensen, K. W. (2016). Firms' innovation benefiting from networking and institutional support: A global analysis of national and firm effects. *Research Policy*, *45*(6), 1233-1246.

[74]. Sermet, Y., & Demir, I. (2018). An intelligent system on knowledge generation and communication about flooding. *Environmental modelling & software*, *108*, 51-60.

[75]. Serre, D., & Heinzlef, C. (2018). Assessing and mapping urban resilience to floods with respect to cascading effects through critical infrastructure networks. *International Journal of Disaster Risk Reduction*, *30*, 235-243.

[76]. Shaiful, M., Anisur, R., & Md, A. (2022). A Systematic Literature Review on the Role of Digital Health Twins in Preventive Healthcare For Personal and Corporate Wellbeing. *American Journal of Interdisciplinary Studies*, *3*(04), 1-31. https://doi.org/10.63125/negjw373

[77]. Shi, W., Zhang, M., Zhang, R., Chen, S., & Zhan, Z. (2020). Change detection based on artificial intelligence: State-of-the-art and challenges. *Remote sensing*, *12*(10), 1688.

[78]. Singh, P., Sinha, V. S. P., Vijhani, A., & Pahuja, N. (2018). Vulnerability assessment of urban road network from urban flood. *International Journal of Disaster Risk Reduction*, *28*, 237-250.

[79]. Song, J., Yang, R., Chang, Z., Li, W., & Wu, J. (2019). Adaptation as an indicator of measuring low-impact-development effectiveness in urban flooding risk mitigation. *Science of the Total Environment*, *696*, 133764.

[80]. Stabak, D., & Md Shahid, N. (2022). Empirical Assessment of Uncertainty-Aware Real-Time Decision-Making for Reliable and Resilient Cyber-Physical Systems. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 432–477. https://doi.org/10.63125/q3r4kk16

[81]. Su, H.-T., Cheung, S. H., & Lo, E. Y.-M. (2018). Multi-objective optimal design for flood risk management with resilience objectives. *Stochastic environmental research and risk assessment*, *32*(4), 1147-1162.

[82]. Sui, J., Jiang, R., Bustillo, J., & Calhoun, V. (2020). Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. *Biological psychiatry*, *88*(11), 818-828.

[83]. Tariq, M. A. U. R., Farooq, R., & Van de Giesen, N. (2020). A critical review of flood risk management and the selection of suitable measures. *Applied Sciences*, *10*(23), 8752.

[84]. Tien Bui, D., Khosravi, K., Shahabi, H., Daggupati, P., Adamowski, J. F., Melesse, A. M., Thai Pham, B., Pourghasemi, H. R., Mahmoudi, M., & Bahrami, S. (2019). Flood spatial modeling in northern Iran using remote sensing and gis: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. *Remote sensing*, *11*(13), 1589.

[85]. Tran, B. X., Phan, H. T., Latkin, C. A., Nguyen, H. L. T., Hoang, C. L., Ho, C. S., & Ho, R. C. (2019). Understanding global HIV stigma and discrimination: are contextual factors sufficiently studied?(GAPRESEARCH). *International journal of environmental research and public health*, *16*(11), 1899.

[86]. Tsereteli, E., Gaprindashvili, G., Gaprindashvili, M., Bolashvili, N., & Gongadze, M. (2018). Hazard risk of debris/mud flow events in Georgia and methodological approaches for management. IAEG/AEG Annual Meeting Proceedings, San Francisco, California, 2018-Volume 5: Geologic Hazards: Earthquakes, Land Subsidence, Coastal Hazards, and Emergency Response,

[87]. Vamvakeridou-Lyroudia, L., Chen, A., Khoury, M., Gibson, M., Kostaridis, A., Stewart, D., Wood, M., Djordjevic, S., & Savic, D. (2020). Assessing and visualising hazard impacts to enhance the resilience of Critical Infrastructures to urban flooding. *Science of the Total Environment*, *707*, 136078.

[88]. Van Long, N., Cheng, Y., & Le, T. D. N. (2020). Flood-resilient urban design based on the indigenous landscape in the city of Can Tho, Vietnam. *Urban Ecosystems*, *23*(3), 675-687.

[89]. Waghwala, R. K., & Agnihotri, P. (2019). Flood risk assessment and resilience strategies for flood risk management: A case study of Surat City. *International Journal of Disaster Risk Reduction*, *40*, 101155.

[90]. Woodrow, K., Lindsay, J. B., & Berg, A. A. (2016). Evaluating DEM conditioning techniques, elevation source data, and grid resolution for field-scale hydrological parameter extraction. *Journal of Hydrology*, *540*, 1022-1029.

[91]. Wu, Z., Shen, Y., & Wang, H. (2019). Assessing urban areas' vulnerability to flood disaster based on text data: A case study in Zhengzhou city. *Sustainability*, *11*(17), 4548.

[92]. Wu, Z., Shen, Y., Wang, H., & Wu, M. (2020). Urban flood disaster risk evaluation based on ontology and Bayesian Network. *Journal of Hydrology*, *583*, 124596.

[93]. Yang, L., MacEachren, A. M., Mitra, P., & Onorati, T. (2018). Visually-enabled active deep learning for (geo) text and image classification: a review. *ISPRS International Journal of Geo-Information*, *7*(2), 65.

[94]. Yang, Y., Ng, S. T., Zhou, S., Xu, F. J., & Li, H. (2020). Physics-based resilience assessment of interdependent civil infrastructure systems with condition-varying components: A case with stormwater drainage system and road transport system. *Sustainable Cities and Society*, *54*, 101886.

[95]. Yang, Y., Reniers, G., Chen, G., & Goerlandt, F. (2019). A bibliometric review of laboratory safety in universities. *Safety Science*, *120*, 14-24.

[96]. Yazdi, J. (2018). Rehabilitation of urban drainage systems using a resilience-based approach. *Water Resources Management*, *32*(2), 721-734.

[97]. Zanchetta, A. D., & Coulibaly, P. (2020). Recent advances in real-time pluvial flash flood forecasting. *Water*, *12*(2), 570.

[98]. Zhang, G., Lu, H., Dong, J., Poslad, S., Li, R., Zhang, X., & Rui, X. (2020). A framework to predict high-resolution spatiotemporal PM2. 5 distributions using a deep-learning model: A case study of Shijiazhuang, China. *Remote sensing*, *12*(17), 2825.