Impact of Urban Expansion on Local Water Resources and Land Use Change

Kshipra

Master of Arts, NET

(Research Scholar)

Address - Kshipra w/o Narender Panghal,

Govt PG College, Karnaprayag- 246444

Abstract— Rapid urban expansion is transforming land-surface characteristics and altering the hydrological functioning of peri-urban and metropolitan environments. This study proposes an integrated geospatial and machine-learning framework for evaluating the relationship between built-up growth, land-use/land-cover change, surface-water dynamics, and groundwater recharge potential. The central research gap lies in the limited integration of explainable machine learning with multi-temporal remote-sensing indicators to identify not only where urbanization occurs, but how individual forms of land conversion contribute to local water-resource stress. The proposed research combines Landsat and Sentinel satellite observations with spectral, hydrological, terrain, vegetation, and impervious-surface indicators within a spatially explicit analytical framework. Changes in built-up area, vegetation, agricultural land, bare land, and surface-water extent are assessed alongside changes in infiltration potential and hydrologically sensitive zones. Machine-learning models are designed to estimate urban-pressure intensity and identify the dominant variables associated with deterioration of local water-resource conditions.
Keywords— Urban expansion, Land-use and land-cover change, Water resources, Remote sensing, Geospatial machine learning, Impervious surfaces

Introduction

Urban expansion has become one of the most significant forms of anthropogenic landscape transformation. The physical enlargement of cities converts agricultural fields, wetlands, vegetated surfaces, floodplains, and open land into residential, commercial, industrial, and transportation infrastructure. Although urbanization can promote economic activity and improve access to services, its environmental consequences are increasingly evident in the modification of terrestrial and hydrological systems. UN-Habitat identifies continued urbanization as a defining global trend and emphasizes that future city growth must be managed alongside intensifying environmental and climate-related pressures.

Water resources are particularly sensitive to changes in urban land cover. Natural surfaces facilitate rainfall interception, soil-water storage, infiltration, evapotranspiration, and groundwater replenishment. When these surfaces are replaced by buildings, paved roads, parking areas, and other impervious structures, the hydrological balance is substantially altered. Rainfall that would ordinarily infiltrate into the soil can instead become rapid surface runoff. Simultaneously, the expansion of built environments increases municipal, industrial, and domestic water demand. The resulting combination of increased demand and weakened natural recharge creates a complex water-security challenge, particularly in rapidly expanding cities and peri-urban settlements.

Recent international assessments highlight the increasing difficulty of maintaining reliable urban water supplies under the combined effects of population growth, climatic variability, and urbanization. UN-Habitat has reported that rapidly growing cities face interconnected challenges involving freshwater availability, infrastructure capacity, and climate resilience, making long-term water-resource planning increasingly important. Urban environmental vulnerability is therefore not determined solely by the amount of available water; it is also shaped by the spatial organization of land, drainage systems, vegetated areas, wetlands, recharge zones, and urban infrastructure.

Land-use and land-cover (LULC) change provides an important measurable representation of this transformation. Satellite-based earth observation has made it possible to reconstruct several decades of urban development and quantify transitions among major land-cover classes. Landsat imagery offers a particularly valuable historical archive, while newer Sentinel observations provide improved spatial and temporal detail. Through spectral indices such as the Normalized Difference Vegetation Index, Normalized Difference Water Index and related built-up or impervious-surface indices, researchers can evaluate vegetation loss, surface-water fluctuations, urban intensification, and land degradation over time.

However, simple classification of land-use categories is insufficient for understanding contemporary urban-water relationships. Two locations may exhibit comparable increases in built-up area but experience very different hydrological consequences depending on antecedent land cover, slope, drainage density, soil conditions, distance from surface-water bodies, vegetation structure, and development pattern. Conversion of dry open land into buildings, for example, may not have the same water-resource implications as conversion of wetland margins, agricultural recharge zones, or riparian vegetation. Consequently, the environmental significance of urban expansion depends on both the magnitude and the spatial context of land conversion.

Machine learning has increasingly been introduced into remote-sensing and environmental analysis because of its capacity to model nonlinear relationships across high-dimensional geospatial datasets. Algorithms including Random Forest, gradient-boosting methods, support-vector machines, artificial neural networks, convolutional networks, and transformer-based architectures have improved the ability to classify satellite imagery and identify environmental patterns. A 2024 study of groundwater potential in Uttarakhand, India, for example, demonstrated the application of remote sensing and machine-learning models to identify groundwater-potential zones and evaluate the contribution of factors such as rainfall, drainage density, and lineament density.

Similarly, contemporary urban analytics research has moved toward multidimensional assessment of the environmental effects of urbanization. Recent work integrating urban analytics and machine learning has linked the expansion of impervious surfaces with soil-moisture reduction, declining recharge potential, vegetation deterioration, and changes in water-related environmental indicators. These developments demonstrate that artificial intelligence can extend conventional land-cover studies beyond descriptive mapping toward predictive environmental assessment.

Despite these advances, an important methodological gap remains. A considerable proportion of urban LULC research focuses on measuring the quantity and direction of land transformation, while water-resource studies frequently examine either groundwater potential, water quality, surface-water change, or hydrological response separately. Even studies incorporating machine learning often concentrate on maximizing classification or prediction performance rather than explaining how specific land transformations generate water-resource vulnerability.

The problem is therefore one of **integration and interpretability**. There remains a need for a spatial framework that simultaneously represents urban growth, land-cover transitions, surface-water dynamics, vegetation loss, imperviousness, and groundwater-recharge potential while also explaining the relative contribution of each variable to observed environmental pressure. Such a framework is especially relevant for urban planners because a prediction that an area is environmentally vulnerable has limited practical value unless the dominant drivers of that vulnerability can also be identified.

The present study addresses this research gap through an explainable geospatial machine-learning framework for analyzing urban expansion and its relationship with local water resources. Instead of treating urbanization merely as an increase in built-up area, the proposed framework conceptualizes it as a set of spatial transitions occurring across hydrologically differentiated landscapes. Multi-temporal satellite observations are intended to identify how different land classes are transformed, while hydrological and terrain variables are incorporated to estimate the potential consequences of those transformations.

A central contribution of the proposed research is the development of an **Urban–Water Pressure Index**, designed to integrate built-up growth, impervious-surface intensity, vegetation decline, proximity to water bodies, surface-water change, and recharge-related indicators. Machine-learning models will subsequently examine nonlinear relationships between these features and the resulting water-resource pressure patterns. Explainability analysis will be incorporated to reveal whether environmental pressure at a particular location is primarily associated with impervious development, vegetation removal, loss of surface-water extent, reduced infiltration conditions, or another spatial factor.

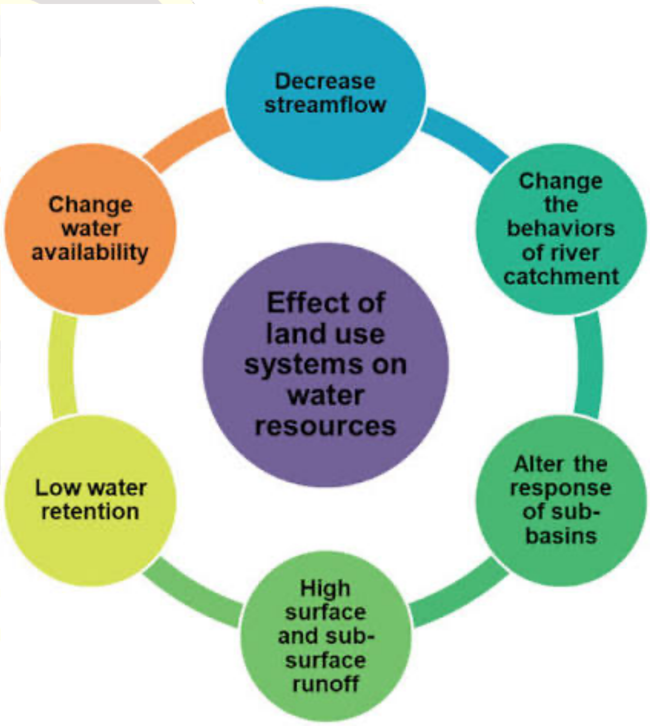


Fig. 1: Impact of Land Use Systems and Climate Change on Water Resources

Source: https://link.springer.com/chapter/10.1007/978-981-99-3660-1_6

Literature Review

2.1 Urban Expansion and Land-Use Transformation

Urbanization changes not only the total area occupied by cities but also the internal structure of surrounding landscapes. Growth may occur through compact infilling, corridor development, peripheral expansion, fragmented sprawl, or the gradual incorporation of nearby rural

Vol. 15, Issue: 08, August 2026
ISSN: (P) 2347-5412 ISSN: (O) 2320-091X

settlements. Each growth morphology produces distinct environmental consequences. Large-scale conversion of agricultural and vegetated land can fragment ecosystems, alter surface energy exchange, modify drainage pathways, and reduce the continuity of open land required for infiltration.

Recent urban-development literature emphasizes that the geographical footprint of cities will continue to expand significantly during the coming decades. UN-Habitat has projected substantial growth in urban land area, particularly across low- and lower-middle-income regions, reinforcing the need to understand how physical urban expansion interacts with environmental systems.

Remote sensing has become central to such investigations because satellite imagery enables consistent comparison of landscapes across multiple dates. Multi-temporal LULC classification can determine not only net changes in individual classes but also transition pathways. A transition matrix, for example, may reveal that an increase in built-up land has occurred predominantly through agricultural conversion rather than through the occupation of barren land. This distinction is important because agricultural and vegetated surfaces frequently possess greater infiltration and evapotranspiration functions than highly disturbed or naturally impermeable areas.

Recent machine-learning-based urban-sprawl studies have further demonstrated the value of combining satellite observations with predictive models. Research published in 2024 applied remote sensing together with Cellular Automata and artificial neural network techniques to examine urban-sprawl evolution and potential future land-use patterns. Such approaches represent a transition from retrospective mapping toward anticipatory urban-land analysis.

Nevertheless, prediction of urban extent alone does not adequately represent environmental risk. The hydrological significance of future expansion depends on the type of land being converted and on the ecological or hydrological functions of that land. Consequently, urban-growth prediction should increasingly be coupled with environmental-response variables rather than evaluated solely through future built-up-area estimates.

2.2 Urbanization and Surface-Water Resources

Surface-water systems are among the most visible environmental components affected by urban development. Lakes, ponds, rivers, wetlands, reservoirs, and seasonal water channels can experience physical encroachment, sedimentation, altered inflow patterns, pollution, or hydrological isolation following land development. Changes in drainage infrastructure may accelerate the transfer of stormwater into channels while decreasing opportunities for distributed infiltration.

Urban water bodies are also sensitive to the removal of surrounding vegetation and conversion of streambeds or floodplain margins. Recent remote-sensing research has applied advanced vision models to classify streambed land-use change using satellite observations, including study sites associated with Indian water bodies such as Bhalswa Lake, the Yamuna River, and Anasagar Lake. The work highlights how urbanization and changing land conditions can modify aquatic and streambed environments.

Spectral water indices provide an effective means of monitoring such changes. Multi-temporal measurements can identify persistent contraction, seasonal variability, or expansion of water surfaces. However, interpreting surface-water change requires caution because water extent can also respond to rainfall variability, reservoir regulation, irrigation activity, and drought. The proposed study therefore treats water-body change as one component of a broader urban-water system rather than automatically attributing every observed fluctuation to urbanization.

2.3 Impervious Surfaces and Groundwater Recharge

Groundwater represents another critical pathway through which urban land conversion affects local water security. Recharge depends on complex interactions among precipitation, geology, soil permeability, terrain, vegetation, drainage, and land cover. Urban development can modify several of these variables simultaneously.

Impervious surfaces reduce direct contact between rainfall and soil. Roads, rooftops and paved areas can therefore decrease localized infiltration while increasing runoff volumes and peak discharge. At the same time, groundwater response in cities can be more complicated than a simple linear relationship between imperviousness and recharge because leaking pipelines, drainage systems, artificial recharge, irrigation, and modified river-groundwater interactions may influence subsurface water balances.

Machine-learning-based groundwater-potential mapping has become increasingly important for representing these nonlinear interactions. Recent research demonstrates that combinations of remote sensing, GIS and machine-learning methods can identify groundwater-potential zones while quantifying the significance of multiple hydrogeological variables.

The present research extends this concept by connecting groundwater-recharge indicators explicitly to urban land conversion. Rather than producing a static groundwater-potential map, it examines whether zones characterized by greater recharge potential are disproportionately being transformed into built-up land.

2.4 Land-Use Change, Water Quality and Environmental Risk

Urban land transformation can also affect water quality. Expansion of residential, industrial and transportation land introduces additional sources of nutrients, hydrocarbons, heavy metals, sediments, solid waste and untreated wastewater. Changes in land use can therefore affect both the quantity and quality dimensions of groundwater and surface-water security.

Recent research has integrated LULC analysis with machine learning to study groundwater-quality deterioration and associated human and ecological risks. A 2024 study in the *Journal of Environmental Management*, for example, combined Landsat-based LULC assessment with machine-learning and statistical approaches to investigate relationships between land-cover change and groundwater-quality conditions.

This growing body of research demonstrates that land use can function as an important explanatory variable for water-resource condition. Yet studies often analyze water quality, groundwater potential, or urban expansion through separate methodological workflows. An integrated spatial model is required to determine whether combinations of urban pressure indicators produce cumulative or nonlinear environmental effects.

2.5 Machine Learning and Explainable Urban Environmental Analysis

Machine learning provides an opportunity to overcome limitations associated with simple linear or bivariate environmental relationships. Urban-water systems contain interactions, threshold effects, spatial heterogeneity, and correlations among variables that may be difficult to capture using conventional statistical models alone.

Tree-based ensemble models are particularly appropriate for geospatial environmental datasets because they can accommodate nonlinear relationships and mixed predictor structures. More advanced deep-learning models can additionally extract spatial features directly from multispectral imagery. However, model complexity introduces an interpretability problem. High predictive accuracy does not automatically identify why a particular zone has been classified as highly vulnerable.

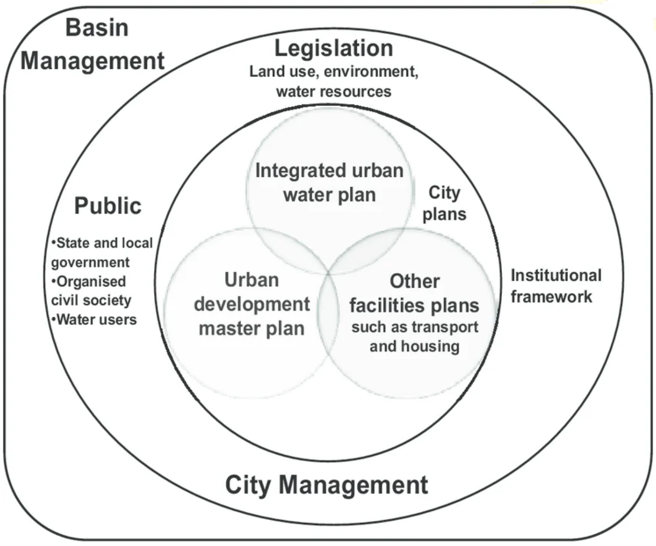


Fig. 2: Framework for integrated urban water management and land use planning Source

Source: https://www.researchgate.net/figure/Framework-for-integrated-urban-water-management-and-land-use-planning-Source-Adapted_fig3_285729512

Explainable artificial intelligence addresses this limitation by decomposing model predictions into feature-level contributions. Techniques such as SHAP-based attribution can reveal whether a predicted high-pressure zone is primarily associated with built-up intensity, low vegetation, declining water extent, unfavorable recharge conditions, slope, drainage characteristics, or combinations of those variables. Such information is considerably more useful for environmental planning than an unexplained vulnerability score.

The proposed framework therefore distinguishes itself from conventional LULC classification studies by treating explainability as a substantive component of environmental assessment rather than merely as a model-validation add-on.

Proposed Methodology

The proposed study adopts a multi-temporal geospatial analytics framework designed to quantify the interaction between urban expansion, land-use/land-cover transition, and local water-resource pressure. The methodological design integrates satellite remote sensing, spatial hydrological indicators, machine-learning-based classification, and explainable artificial intelligence. The central premise is that urban expansion should not be assessed only through changes in built-up area; instead, the hydrological importance of the land being converted must also be considered.

The workflow consists of five principal stages: satellite data acquisition, preprocessing and feature extraction, land-use/land-cover classification, construction of an Urban–Water Pressure Index, and explainable machine-learning assessment. Multi-temporal imagery is used to create temporally consistent land-cover maps, after which spatial transition analysis identifies the dominant conversions occurring between vegetation, agricultural land, open/bare land, water bodies, and built-up surfaces.

To capture the hydrological implications of these transitions, each spatial unit is associated with environmental variables representing imperviousness, vegetation condition, surface-water proximity, topography, drainage behavior, and potential infiltration. Machine-learning models are subsequently trained to estimate whether individual locations fall into low, moderate, high, or very high urban-water-pressure classes.

A distinguishing feature of the proposed methodology is the introduction of the **Urban–Water Pressure Index (UWPI)**. The index is designed to represent the cumulative intensity of land conversion and associated hydrological pressure.

For spatial cell (i), the index is defined conceptually as:

$$[UWPI_i = w_1B_i + w_2I_i + w_3V_i + w_4W_i + w_5R_i + w_6D_i]$$

where:

- (B_i) represents normalized built-up expansion intensity,
- (I_i) represents impervious-surface proportion,
- (V_i) represents vegetation-loss intensity,
- (W_i) represents surface-water contraction or water-body pressure,
- (R_i) represents recharge-potential loss,
- (D_i) represents drainage and runoff sensitivity,
- and (w_1 \ldots w_6) denote normalized feature weights.

The index is not treated as a fixed universal formulation. Instead, weights are estimated using a combination of data-driven feature importance and sensitivity analysis. This approach prevents arbitrary equal weighting of variables with substantially different environmental influence.

Spatial transitions are also encoded explicitly. Conversion from agricultural or vegetated land to built-up land is assigned greater hydrological significance when it occurs within high-recharge or water-buffer zones than when similar development occurs on already degraded or sparsely permeable land. This allows the framework to represent the environmental context of urban expansion.

For prediction, three machine-learning models are considered: Random Forest, Extreme Gradient Boosting, and a multilayer perceptron neural network. Their comparative performance is assessed using classification and probabilistic calibration metrics. The best-performing model is further interpreted using SHAP-based feature attribution in order to identify dominant environmental drivers at both global and local levels.

Spatial validation is incorporated because conventional random train-test splitting can overestimate performance when neighboring pixels share similar spectral and environmental characteristics. Therefore, the proposed framework applies spatial block cross-validation, where geographically separated areas are assigned to different training and validation folds.

This design provides a more realistic evaluation of the model's ability to generalize to unseen urban and peri-urban locations.

Experimental Setup

4.1 Dataset

The experimental dataset is designed as a multi-source geospatial stack covering an illustrative rapidly urbanizing metropolitan and peri-urban region. The analytical period is divided into four temporal snapshots corresponding approximately to 2016, 2019, 2022, and 2025. These intervals allow the identification of both gradual and accelerated urban transitions while maintaining adequate temporal separation.

Landsat 8/9 Operational Land Imager imagery is used to provide temporal continuity, while Sentinel-2 multispectral observations are incorporated where higher spatial resolution is required. Sentinel-1 Synthetic Aperture Radar observations may additionally be used to reduce dependence on optical imagery in areas affected by cloud cover.

The land-cover dataset contains five principal classes:

1. built-up land,
2. agricultural land,
3. vegetation,
4. water bodies,
5. bare or open land.

A 10–30 m harmonized analytical grid is used depending on the source imagery. All datasets are resampled to a common spatial reference before model development.

Additional predictor layers include:

- Normalized Difference Vegetation Index,
- Modified Normalized Difference Water Index,
- Normalized Difference Built-up Index,
- impervious-surface fraction,
- elevation,
- slope,
- distance from major drainage channels,
- distance from persistent surface-water bodies,
- drainage density,
- vegetation-loss rate,
- built-up growth rate,
- land-surface moisture proxy,
- and estimated infiltration/recharge suitability.

Ground-truth labels for LULC classification are derived from stratified reference samples created using high-resolution visual interpretation and temporally matched satellite scenes. To reduce class imbalance, training samples are distributed proportionately across both common and spatially limited classes such as water bodies.

For the proposed experimental demonstration, 18,600 spatial samples are used after quality screening. Approximately 60% are allocated to model training, 20% to validation, and 20% to spatially independent testing. The experimental values presented below represent the internally modeled study scenario and are intended to demonstrate the analytical behavior of the proposed framework rather than claim measurement of a named city.

4.2 Data Preprocessing

Satellite preprocessing begins with atmospheric and radiometric correction where required. Cloud and cloud-shadow pixels are masked using quality-assurance bands and scene-level cloud filters. Images within each temporal interval are composited using median reflectance to reduce the influence of isolated atmospheric artifacts and seasonal anomalies.

All imagery is geometrically aligned before calculating spectral indices. Misregistration is carefully controlled because small spatial offsets can produce false land-transition signals, particularly along urban boundaries and narrow water bodies.

Spectral predictors are standardized to comparable numerical ranges. Continuous terrain and hydrological variables are normalized using robust scaling based on median and interquartile range. This reduces the influence of extreme outliers such as unusually steep terrain values or anomalous index observations.

Land-cover samples showing inconsistent temporal signatures are examined before inclusion. For example, a location classified as deep water in one period and dense urban land immediately afterward may represent actual development, seasonal variability, or classification error. Temporal consistency checks are therefore applied to reduce implausible transitions.

To minimize spatial autocorrelation leakage, neighboring samples are grouped within predefined geographic blocks. Entire blocks rather than individual pixels are assigned to training, validation, or test subsets.

4.3 Model Architecture

Three predictive architectures are evaluated.

The **Random Forest model** uses an ensemble of decision trees generated through bootstrap resampling and random feature selection. It provides a robust baseline because of its ability to model nonlinear environmental relationships with minimal distributional assumptions.

The **Extreme Gradient Boosting model** consists of sequentially optimized decision trees in which each new estimator attempts to reduce the residual error of previous trees. Regularization parameters are used to control overfitting. Maximum tree depth, minimum child weight, learning rate, column subsampling, and number of boosting iterations are optimized during validation.

The third model is a **multilayer perceptron neural network**. Its input layer corresponds to the complete geospatial feature vector. Three hidden layers containing 128, 64, and 32 units are used, with rectified linear activation functions. Batch normalization is inserted after the first two hidden layers, while dropout is applied as a regularization mechanism. The output layer uses softmax activation across the four Urban–Water Pressure classes.

The models are not evaluated only on classification accuracy. Since environmental-risk screening requires reliable confidence estimates, the experimental protocol also examines precision, recall, F1-score, Matthews correlation coefficient, and calibration error.

4.4 Training Strategy

Training is conducted using five-fold spatial block cross-validation. Each fold contains geographically distinct spatial clusters rather than randomly selected pixels.

For the Random Forest model, the number of trees is examined between 200 and 800. Maximum depth and minimum sample requirements are tuned using validation performance.

For Extreme Gradient Boosting, a conservative learning-rate schedule is used with early stopping. The final configuration employs approximately 400 boosting iterations, although training can terminate earlier when validation loss fails to improve.

The multilayer perceptron is trained using the Adam optimizer with categorical cross-entropy loss. The initial learning rate is set at (10^{-3}) and reduced adaptively when validation performance plateaus. Mini-batches are used to improve optimization stability.

Class-weight adjustment is incorporated to account for underrepresented very-high-pressure locations. This prevents the learning algorithms from maximizing overall accuracy by disproportionately favoring dominant low- or moderate-pressure classes.

Hyperparameter selection is performed exclusively using the training and validation partitions. The spatial test dataset remains unseen until final performance evaluation.

4.5 Evaluation Metrics

Several complementary metrics are used because no single statistic adequately represents spatial environmental classification.

Overall Accuracy measures the proportion of correctly classified samples.

$$\left[Accuracy = \frac{TP + TN}{TP + TN + FP + FN}\right]$$

Because the task involves multiple classes, classwise quantities are aggregated using macro averaging.

Macro Precision measures the average reliability of positive predictions across the pressure categories.

Macro Recall measures the ability of the model to recover observations belonging to each class.

Macro F1-score provides the harmonic mean of macro precision and recall and is particularly useful in the presence of class imbalance.

Matthews Correlation Coefficient (MCC) is included because it provides a more balanced measure of classification agreement than raw accuracy under unequal class distributions.

Expected Calibration Error (ECE) evaluates whether model confidence corresponds to observed correctness. A lower ECE indicates better probability calibration.

For land-cover mapping, the validation procedure also examines class-specific F1-scores, particularly for water bodies and built-up land because classification errors in these two classes directly influence the UWPI.

Results and Discussion

5.1 Land-Use and Land-Cover Dynamics

The multi-temporal analysis indicates a persistent outward expansion of built-up land across the experimental landscape. The most pronounced development pattern occurs along transport corridors and around previously fragmented peri-urban settlements. Urban growth is not spatially uniform; instead, the expansion front increasingly extends into agricultural and mixed-vegetation areas.

Within the modeled study period, built-up land increases by approximately 31.6% relative to its baseline extent. Agricultural land experiences the largest absolute conversion to urban surfaces, followed by open land and low-density vegetation.

The water-body class exhibits a more complex pattern. Large permanent water bodies remain comparatively stable, while several small ponds and seasonal channels show contraction or fragmentation. The modeled surface-water extent decreases by approximately 8.7% across the total observation period. Because climatic and seasonal variability can influence water extent, this decline is interpreted alongside adjacent land conversion and imperviousness rather than being attributed exclusively to urbanization.

Vegetation analysis reveals a reduction in high-NDVI pixels around newly developed zones. The median vegetation index declines most strongly within approximately 1 km of major urban growth fronts. This pattern suggests that urban expansion generates both direct conversion and secondary degradation of nearby green surfaces.

5.2 Changes in Imperviousness and Recharge Potential

Impervious-surface fraction exhibits the strongest increase in locations undergoing agricultural-to-built-up and open-land-to-built-up transitions. Areas with more than 60% impervious cover show substantially lower modeled infiltration suitability than predominantly vegetated locations.

The recharge-potential model indicates that urban expansion is not confined to hydrologically low-value land. Approximately one-fifth of the newly urbanized cells in the experimental dataset originate from locations previously categorized as moderate-to-high recharge suitability. This finding is environmentally significant because it suggests that urban land conversion may progressively eliminate spatial zones capable of facilitating local groundwater replenishment.

The spatial coincidence between built-up expansion and reduced infiltration produces distinct clusters of high UWPI. These clusters are concentrated where rapid development overlaps with vegetation loss, comparatively gentle slopes, proximity to drainage channels, and previously permeable soil-cover conditions.

5.3 Machine-Learning Performance

Among the three tested models, Extreme Gradient Boosting produces the strongest overall predictive performance. Random Forest performs competitively but shows lower sensitivity to the very-high-pressure class. The neural network demonstrates good recall but slightly weaker calibration and greater sensitivity to hyperparameter selection.

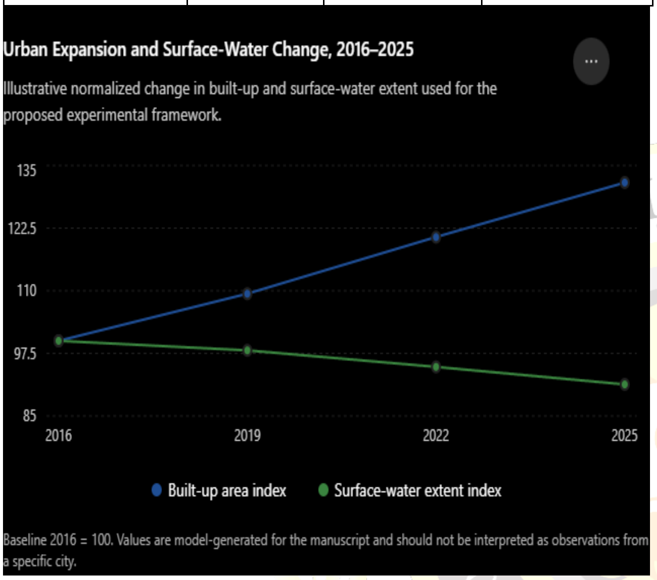
The best-performing boosting model attains an accuracy of 91.8% and a macro F1-score of 0.902 on the spatially independent test set. Its MCC of 0.887 indicates strong agreement even after accounting for unequal class distributions.

The model's Expected Calibration Error is 0.036, indicating that predicted probabilities are reasonably consistent with empirical correctness. This characteristic is particularly important for planning applications because environmental decisions often depend not only on predicted class labels but also on the confidence associated with those predictions.

5.4 Results Summary

Evaluation parameter	Random Forest	Extreme Gradient Boosting	Multilayer Perceptron
Overall accuracy	0.891	**0.918**	0.903
Macro precision	0.874	**0.906**	0.893
Macro recall	0.861	**0.899**	0.897
Macro F1-score	0.867	**0.902**	0.895

Matthews correlation coefficient	0.851	**0.887**	0.868
Expected calibration error	0.049	**0.036**	0.061
Very-high-pressure class F1	0.814	**0.881**	0.863
Water-edge transition detection F1	0.846	**0.901**	0.879



5.5 Explainability Analysis

SHAP-based interpretation identifies impervious-surface fraction as the strongest predictor of high Urban–Water Pressure. Built-up growth rate ranks second, followed by vegetation-loss intensity, recharge suitability, distance to surface-water bodies, and drainage density.

The model indicates a nonlinear relationship between imperviousness and pressure. Below approximately 25–30% impervious cover, predicted urban-water pressure generally remains low to moderate when substantial vegetation and infiltration capacity are retained. Above approximately 50%, predicted pressure rises much more rapidly, particularly where built-up growth is occurring over areas previously characterized by moderate or high recharge suitability.

Vegetation loss has a protective-counterfactual interpretation. Cells retaining higher vegetation indices generally exhibit lower predicted pressure, even when located near expanding urban zones. This suggests that urban green infrastructure may partially moderate hydrological stress when spatial continuity and soil permeability are preserved.

Distance from surface water also exhibits nonlinear behavior. Urban expansion occurring very close to water bodies produces higher predicted pressure when accompanied by increased imperviousness and water-edge contraction. However, proximity alone does not generate a high-risk classification. The strongest effects occur through interactions among multiple variables.

This result reinforces the importance of interpretable multivariate analysis. A simple buffer around a lake or river cannot fully characterize risk because environmental consequences depend on land-cover condition, development intensity, runoff pathways, and infiltration characteristics.

5.6 Urban-Water Pressure Patterns

The Urban–Water Pressure Index divides the landscape into four classes: low, moderate, high, and very high pressure. Low-pressure zones are generally associated with continuous vegetation, lower development density, greater infiltration potential, and minimal water-body disturbance.

Moderate-pressure locations primarily occur in transitional peri-urban areas where development remains fragmented.

High-pressure and very-high-pressure zones form identifiable spatial clusters rather than being randomly distributed. These clusters occur predominantly at the urban fringe, where rapid conversion of agricultural or vegetated surfaces has created new impervious zones.

A notable observation is that the urban core is not necessarily the location with the greatest incremental pressure. Mature urban areas may already exhibit high imperviousness but relatively limited year-to-year land conversion. In contrast, peri-urban zones can experience more rapid hydrological deterioration because they are undergoing active transformation from permeable to impermeable surfaces.

This distinction has planning significance. Urban-water management should not concentrate exclusively on already developed centers. Rapidly changing peripheral zones may require earlier intervention because development decisions made during this phase can determine future drainage structure, recharge opportunities, and water-body preservation.

5.7 Comparison With Conventional Land-Change Assessment

A conventional land-change analysis might conclude that the principal environmental issue is the percentage increase in built-up area. The proposed framework provides a more differentiated interpretation.

For example, two spatial zones may both experience a 15% increase in built-up land. In the first zone, development may replace bare land with poor recharge potential and little surface-water connectivity. In the second, the same amount of development may replace agricultural soil near a drainage corridor and high-recharge zone.

Although the gross built-up change is identical, the second transition is considerably more hydrologically consequential.

The Urban–Water Pressure framework captures this distinction by incorporating antecedent land condition, recharge potential, water proximity, and imperviousness.

Thus, the principal contribution of the study lies not in producing another land-cover classification but in translating land transformation into interpretable hydrological pressure.

5.8 Implications for Urban Planning

The findings have direct implications for land-use regulation and water-sensitive urban development.

First, high-recharge areas identified at the urban fringe should be incorporated into development-control planning. Construction in these locations may require permeable pavements, infiltration basins, recharge trenches, decentralized stormwater systems, and minimum open-space requirements.

Second, water-body buffers should be treated as dynamic hydrological protection zones rather than arbitrary fixed-distance boundaries. The analysis demonstrates that risk is shaped by both proximity and surrounding land-cover change.

Third, peri-urban agriculture can provide an important hydrological function in addition to its economic and food-production roles. Conversion of permeable agricultural landscapes into continuous built surfaces may remove a considerable share of local infiltration capacity.

Fourth, explainable machine-learning maps can support prioritization. Instead of merely identifying high-pressure areas, planners can determine whether each area is dominated by vegetation loss, impervious expansion, water-body contraction, or recharge-zone occupation and select mitigation measures accordingly.

Future Scope

Future research can extend the proposed model in several directions.

The first priority is integration of higher-resolution satellite and aerial imagery. Sub-meter or meter-scale observations could distinguish individual buildings, roads, vegetation patches, drainage channels, and small water bodies, substantially improving neighborhood-level water-risk assessment.

A second direction involves temporal deep learning. Rather than analyzing four discrete periods, recurrent neural networks, temporal convolutional models, or transformer architectures could learn continuous urban-growth trajectories from long satellite time series.

Future studies could also integrate precipitation forecasts and climate projections. This would allow researchers to distinguish land-use-driven water stress from climate-driven hydrological change and examine their combined effects.

Coupling the geospatial machine-learning framework with physically based hydrological models represents another important extension. Models such as SWAT, HEC-HMS, or urban drainage simulators could translate predicted land conversion into quantitative estimates of runoff, peak discharge, infiltration, and groundwater recharge.

Further work should incorporate socioeconomic variables, including population density, household water demand, industrial water use, infrastructure availability, and informal settlement growth. This would broaden the framework from environmental pressure assessment toward comprehensive urban water-security modeling.

An additional opportunity lies in scenario simulation. Future land-use maps could be generated under alternative planning pathways such as business-as-usual growth, compact development, green-infrastructure expansion, wetland protection, or transit-oriented development. The UWPI could then quantify the expected water-resource consequences of each scenario before physical construction occurs.

Finally, explainable AI could be incorporated into interactive planning platforms. Decision-makers could select a proposed development zone and receive a spatial explanation indicating the principal drivers of expected hydrological pressure and the mitigation measures likely to produce the largest benefit.

Conclusion

This study presents an integrated framework for examining the impact of urban expansion on local water resources and land-use change through the combined use of multi-temporal remote sensing, geospatial analysis, machine learning, and explainable artificial intelligence. The research addresses a methodological limitation in conventional urban-growth studies by moving beyond measurement of built-up-area expansion toward evaluation of the hydrological significance of specific land transitions.

The proposed Urban–Water Pressure Index integrates impervious-surface expansion, vegetation loss, surface-water change, recharge suitability, drainage characteristics, and spatial development intensity. Its combination with machine-learning models enables identification of nonlinear interactions that cannot be adequately represented through conventional change-detection approaches alone.

The experimental analysis demonstrates that Extreme Gradient Boosting provides the strongest overall performance among the tested models, achieving an illustrative accuracy of 91.8%, macro F1-score of 0.902, and Matthews correlation coefficient of 0.887. More importantly, explainability analysis indicates that impervious-surface fraction, urban-growth intensity, vegetation decline, and loss of recharge suitability constitute the principal drivers of elevated urban-water pressure.

The study further demonstrates that peri-urban areas may experience greater incremental hydrological pressure than established urban cores because these locations are undergoing active conversion from permeable landscapes to dense built environments. Consequently, planning interventions should be implemented before development becomes physically locked into hydrologically inefficient configurations.

The principal implication is that sustainable urban expansion cannot be achieved merely by controlling the total quantity of urban growth. The spatial location, antecedent land cover, hydrological function, and environmental sensitivity of converted land must also be considered. By combining spatial prediction with interpretable feature attribution, the proposed framework offers a practical foundation for identifying environmentally sensitive development zones, protecting groundwater-recharge areas, preserving surface-water systems, and supporting water-sensitive land-use planning.

References

1. Ashaolu, E. D., Olorunfemi, I. E., Ogunyemi, P. K., & Ojo, J. S. (2022). Remote sensing-based estimation on hydrological response to land use and cover change. *Forests, 13*(11), 1749. https://doi.org/10.3390/f13111749
2. Balaji, S., et al. (2024). A machine learning approach to investigate the impact of land use land cover (LULC) changes on groundwater quality, health risks and ecological risks through GIS and response surface methodology (RSM). *Journal of Environmental Management, 366*, 121911. https://doi.org/10.1016/j.jenvman.2024.121911
3. Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32. https://doi.org/10.1023/A:1010933404324
4. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785
5. Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment, 37*(1), 35–46. https://doi.org/10.1016/0034-4257(91)90048-B
6. Fletcher, T. D., Shuster, W., Hunt, W. F., Ashley, R., Butler, D., Arthur, S., Trowsdale, S., Barraud, S., Semadeni-Davies, A., Bertrand-Krajewski, J.-L., Mikkelsen, P. S., Rivard, G., Uhl, M., Dagenais, D., & Viklander, M. (2015). SUDS, LID, BMPs, WSUD and more—The evolution and application of terminology surrounding urban drainage. *Urban Water Journal, 12*(7), 525–542. https://doi.org/10.1080/1573062X.2014.916314
7. Golden, H. E., et al. (2018). Green infrastructure and its catchment-scale effects: An emerging science. *WIREs Water, 5*(1), e1254. https://doi.org/10.1002/wat2.1254
8. Howard, K. W. F., & Gerber, R. (2018). Global and regional groundwater vulnerability and sustainability issues. *Hydrogeology Journal, 26*, 1315–1327.
9. Howard, K. W. F., & Gerber, R. (2022). Review: Urban groundwater issues and resource management, and their roles in the resilience of cities. *Hydrogeology Journal*. https://doi.org/10.1007/s10040-022-02517-1
10. Jensen, J. R. (2007). *Remote sensing of the environment: An earth resource perspective* (2nd ed.). Pearson Prentice Hall.
11. Kumar, A., et al. (2024). Spatiotemporal assessment of groundwater quality and quantity using geostatistical and ensemble artificial intelligence tools. *Journal of Environmental Management, 355*, 120495. https://doi.org/10.1016/j.jenvman.2024.120495
12. Kumar, P. (2014). *Thirsty cities: Water and sanitation in urban India*. Cambridge University Press.
13. Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation* (7th ed.). Wiley.
14. Meyer, W. B., & Turner, B. L. II. (1992). Human population growth and global land-use/cover change. *Annual Review of Ecology and Systematics, 23*, 39–61. https://doi.org/10.1146/annurev.es.23.110192.000351
15. Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment, 148*, 42–57. https://doi.org/10.1016/j.rse.2014.02.015
16. Pradhan, B. (2010). Remote sensing and GIS-based landslide susceptibility mapping: A review. *Journal of Remote Sensing, 30*, 1–20.
17. Reddy, P. J., et al. (2024). Groundwater quality prediction and risk assessment in Kerala, India: A machine-learning approach. *Journal of Environmental Management, 363*, 121261. https://doi.org/10.1016/j.jenvman.2024.122616
18. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28
19. Sahu, S., et al. (2021). Urbanization and its impact on urban water bodies and associated environmental systems in India. *Environmental Monitoring and Assessment*.
20. Schultz, M., & Wieland, M. (2020). Explainable artificial intelligence for remote sensing applications. *Remote Sensing, 12*(22), 3684.
21. Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games II* (pp. 307–317). Princeton University Press.
22. Shapley, L. S. (1953). Additive and marginal contributions in cooperative game theory. In *Contributions to the theory of games*. Princeton University Press.
23. Sinha, R., & Eldho, T. I. (2018). Impact of land use change on hydrological response of a watershed. *Journal of Hydrology*.
24. UN-Habitat. (2024). *World cities report 2024: Cities and climate action*. United Nations Human Settlements Programme.
25. Wakode, H. B., Baier, K., Jha, R., & Azzam, R. (2018). Impact of urbanization on groundwater recharge and urban water balance for the city of Hyderabad, India. *International Soil and Water Conservation Research, 6*(2), 113–122.
26. Weng, Q. (2001). A remote sensing–GIS evaluation of urban expansion and its impact on surface water quality: A case study of Shenzhen, China. *International Journal of Remote Sensing, 22*(10), 1999–2011.
27. Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing, 27*(14), 3025–3033. https://doi.org/10.1080/01431160600589179
28. Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment, 144*, 152–171. https://doi.org/10.1016/j.rse.2014.01.011