Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

CLIMATE-RESILIENT CIVIL INFRASTRUCTURE: AN INTEGRATED DATA-DRIVEN FRAMEWORK FOR STRUCTURAL RISK ASSESSMENT AND SUSTAINABLE URBAN DEVELOPMENT

Kashif Ahmad[1], Sabbir Alam Chowdhury[*2], Fahad Ur Rehman[3], Muhammad Tanveer[4]

[1]*Escuela Técnica Superior de Ingenieros de Minas y Energía, Universidad Politécnica de Madrid, Madrid, Spain.*
[*2]*Faculty of Engineering, Environment and Computing, Coventry University, Coventry, England, United Kingdom.*
[3]*Department of Civil Engineering, University of Engineering and Technology Mardan, Mardan, Khyber Pakhtunkhwa, Pakistan.*
[4]*Department of Architectural Engineering, Tandon School of Engineering, New York University, Brooklyn, USA.*

[1]engr.kashifahmad@gmail.com, [*2]sac05ce@gmail.com, [3]Fahadurrehmanusmani@gmail.com, [4]mtanveer33@gmail.com

DOI: https://doi.org/10.5281/zenodo.23337369

Keywords
Climate-resilient infrastructure; Structural risk assessment; Sustainable urban development; Extreme Gradient Boosting; Geographic information systems; Explainable artificial intelligence; Climate hazard mapping; Adaptation prioritization

Article History
Received: 30 January 2026
Accepted: 15 March 2026
Published: 31 March 2026

Copyright @Author
Corresponding Author: *
Sabbir Alam Chowdhury

Abstract

Climate change intensifies flooding, extreme heat, and rainfall variability, creating interconnected challenges for infrastructure safety and sustainable urban development. This study develops an integrated framework combining infrastructure condition assessment, climate hazard mapping, machine learning, and spatial planning to identify vulnerable assets and prioritize resilience investments. The dataset comprises 12,000 inspection records representing 1,200 infrastructure assets, including 600 buildings, 350 road segments, and 250 bridges, monitored between 2019 and 2025. A total of 36 variables capture material deterioration, structural age, maintenance history, drainage capacity, surface temperature, rainfall intensity, flood exposure, population density, and socioeconomic vulnerability. Data preparation involves missing value treatment, outlier screening, appropriate feature standardization, and geographic harmonization within a geographic information system. Structural risk categories are established using inspection criteria and engineering assessments. Earlier observations support model training, while later observations from geographically separated assets provide validation and testing to reduce temporal and spatial information leakage. Random Forest and Extreme Gradient Boosting models classify infrastructure risk, and SHAP analysis explains the contributions of individual predictors. A multicriteria decision module integrates predicted risk, service criticality, intervention cost, and environmental impact to prioritize adaptation measures. Results show that Extreme Gradient Boosting achieves 92.4% classification accuracy, 90.8% macro precision, 89.7% macro recall, and 90.2% macro F1, outperforming Random Forest, which achieves 87.6% accuracy. Flood exposure, deterioration severity, drainage inadequacy, and maintenance delay emerge as influential risk predictors. Spatial assessment identifies 21.5% of infrastructure assets as high risk, with vulnerability concentrated in densely populated floodplain areas. Under a constrained

investment budget, scenario analysis indicates that targeted drainage upgrades, structural rehabilitation, and reflective surface treatments could reduce the composite risk index by 24.8% and estimated annual disruption by 18.6% relative to baseline conditions. These findings demonstrate the value of integrating interpretable predictive modelling with engineering assessment and spatial investment planning. The framework supports transparent prioritization of infrastructure interventions while connecting structural safety with environmental sustainability and service continuity. Further validation across different urban settings and climate scenarios can strengthen its transferability and support equitable, evidence-based adaptation planning for resilient civil infrastructure under evolving climatic pressures.

1- INTRODUCTION:

Civil infrastructure provides the physical foundation for urban development, economic activity, and public well-being. Buildings, roads, and bridges support essential services, connect communities, and enable the movement of people and goods. Their continued performance is therefore central to the safety and functioning of cities. Climate change, however, introduces additional pressures on infrastructure systems through changing patterns of extreme heat, heavy precipitation, flooding, and other environmental hazards. The Intergovernmental Panel on Climate Change identifies urban infrastructure as a significant area of climate risk and emphasizes the importance of incorporating climate impacts into infrastructure design and settlement planning. These challenges make climate resilience an essential consideration in the management of existing assets and the development of future urban systems. Infrastructure risk cannot be understood through climate exposure alone. The consequences of a hazardous event also depend on an asset's physical condition, maintenance history, functional importance, and surrounding urban environment [1]. A deteriorated bridge, for example, may require different adaptation measures from a recently constructed bridge exposed to the same flood hazard. Similarly, the disruption associated with damage to a road depends partly on whether alternative routes are available and whether the road provides access to essential services. Community resilience planning consequently requires attention to infrastructure dependencies and the social and economic functions that assets support. The National Institute of Standards and Technology incorporates these relationships into its guidance for setting resilience priorities and allocating resources. For this study, climate-resilient civil infrastructure is understood as infrastructure managed to withstand relevant climate stresses, maintain essential functions, and support recovery when disruption occurs. Structural safety is an important component of this objective, but resilience assessment also requires consideration of service continuity and the consequences of asset impairment. A condition assessment may identify material deterioration or maintenance needs, whereas a climate exposure assessment indicates the environmental pressures affecting an asset. Bringing these assessments together provides a basis for examining where physical vulnerability and external hazards coincide [2]. This combined perspective also helps distinguish infrastructure requiring immediate engineering attention from infrastructure requiring longer-term adaptation planning. A central analytical challenge is the integration of information collected at different spatial and temporal scales. Inspection records describe individual assets, climate observations characterize changing environmental conditions, and urban indicators represent patterns of population, development, and vulnerability. These sources may differ in measurement frequency, geographic coverage, completeness, and reliability. Consequently, developing a useful risk assessment requires more than assembling a large dataset. It requires consistent asset identification, transparent variable definitions, appropriate treatment of missing observations, and alignment between the timing of predictors

and the outcomes being assessed. Without these safeguards, apparent predictive performance may provide an unreliable basis for infrastructure decisions.

This study addresses that challenge through a framework combining engineering assessments, geographic information systems, and predictive modelling. Geographic information systems organize asset locations and connect infrastructure records with climate exposure and urban context. Random Forest and Extreme Gradient Boosting are evaluated as alternative models for classifying structural risk using the integrated indicators. The comparison examines whether the recorded combinations of condition, exposure, and maintenance characteristics support reliable discrimination between risk categories. Model outputs are then considered alongside engineering criteria, rather than being treated as direct substitutes for inspections or structural analysis. Interpretability is another important requirement because adaptation decisions must be understandable to engineers, infrastructure managers, and urban planners [3]. A predicted risk category has limited practical value unless decision-makers can examine the factors associated with that prediction. Accordingly, the framework uses SHAP analysis to investigate how individual variables contribute to model outputs. These explanations are interpreted as descriptions of model behaviour rather than evidence that a predictor independently causes infrastructure damage. Connecting predictive outputs with inspection evidence supports a more transparent assessment of whether a proposed intervention addresses the conditions associated with elevated risk. Predictive accuracy alone also does not establish which infrastructure investments should receive priority. An asset with a high predicted risk may differ from another asset in service importance, intervention cost, environmental impact, and the population affected by disruption. The framework therefore connects risk classification with a multicriteria prioritization process. This approach allows adaptation options to be assessed within an explicit investment constraint while retaining attention to infrastructure safety and urban service needs. It is consistent with resilience planning guidance that links hazard management, community functions, and resource allocation. The present research examines 12,000 inspection records covering 1,200 infrastructure assets monitored between 2019 and 2025. The asset portfolio comprises 600 buildings, 350 road segments, and 250 bridges. The dataset contains 36 variables describing infrastructure condition, maintenance history, climate exposure, and surrounding urban characteristics. Risk categories are established using inspection criteria and engineering assessments. Temporal ordering and geographic separation are incorporated into model evaluation to reduce information leakage and assess performance beyond closely related observations. The resulting predictions are mapped spatially and used to evaluate adaptation scenarios involving drainage improvements, structural rehabilitation, and reflective surface treatments. The study has three connected objectives: to classify structural risk using integrated infrastructure and environmental data; to explain the predictors associated with elevated risk and identify their spatial distribution; and to prioritize adaptation measures by considering risk, service criticality, cost, and environmental impact. Its contribution lies in connecting these analytical stages within a single decision framework [4]. The framework links asset-level assessment with urban planning while distinguishing measured infrastructure conditions from projected intervention benefits. Through this integration, the research seeks to support transparent maintenance decisions, more informed allocation of adaptation resources, and sustainable urban development under changing climatic conditions.

2- Climate Hazard Mapping and Geospatial Assessment of Infrastructure Exposure:

Climate hazard mapping and geospatial analysis have become essential components of modern infrastructure risk assessment, particularly in urban environments exposed to flooding, extreme rainfall, rising surface temperatures, and rapid land-use transformation. Geographic Information Systems (GIS) provide an analytical platform for

integrating environmental observations, infrastructure condition records, and spatial vulnerability indicators into a unified assessment framework. Through spatial overlay techniques, proximity analysis, hazard classification, and geographic visualization, GIS enables researchers to identify infrastructure assets located within potentially hazardous environments. Remote sensing further strengthens these capabilities by providing spatially distributed information on land-surface temperature, vegetation cover, impervious surfaces, and changing urban development patterns [5]. However, environmental exposure does not necessarily indicate structural failure because infrastructure vulnerability is also influenced by material deterioration, construction characteristics, drainage conditions, and maintenance history. Consequently, integrating climate hazard information with asset-specific engineering observations offers a more comprehensive representation of infrastructure risk than assessing environmental hazards independently. In the present research, geospatial exposure assessment is designed around 1,200 civil infrastructure assets, including 600 buildings, 350 road segments, and 250 bridges, represented by 12,000 inspection records collected between 2019 and 2025. The analytical framework incorporates flood exposure, rainfall intensity, surface temperature, drainage characteristics, population density, and socioeconomic vulnerability within the broader set of 36 infrastructure and environmental variables [6]. Spatial integration is performed by associating individual infrastructure locations with relevant environmental information while preserving the geographical identity of each asset. As presented in Table 1, buildings represent 50.00% of the infrastructure portfolio, followed by road segments at 29.17% and bridges at 20.83%. These different asset categories require careful spatial treatment because the consequences of climate exposure depend on their engineering characteristics and service functions.

Table 1: Infrastructure asset composition and geospatial climate exposure assessment framework.

Infrastructure category	Assets (n)	Portfolio share (%)	Principal climate exposure indicators	Geospatial assessment approach
Buildings	600	50.00	Flood exposure, surface temperature, rainfall intensity	Building-level spatial overlay
Road segments	350	29.17	Flood exposure, drainage capacity, rainfall intensity	Linear-network exposure assessment
Bridges	250	20.83	Flood exposure, rainfall intensity, drainage conditions	Bridge-location and crossing analysis
Total	1,200	100.00	Integrated climate hazard indicators	GIS-based infrastructure risk assessment

The integration of these spatial datasets supports the identification of infrastructure locations where climatic stressors coincide with structural deterioration and inadequate maintenance. Flood-related exposure can be examined through the spatial intersection of infrastructure assets and mapped flood-prone areas, while surface-temperature indicators can provide information on localized thermal stress. Drainage characteristics and rainfall intensity offer complementary evidence for evaluating potential water-related infrastructure disruption. Population density and socioeconomic vulnerability can subsequently be incorporated to assess the wider consequences of infrastructure impairment. Figure 1 illustrates the proposed geospatial assessment pathway, connecting infrastructure inventories, climate hazard information, GIS processing, spatial exposure evaluation, and infrastructure risk classification [7]. A major methodological concern in geospatial infrastructure assessment involves differences in data resolution, temporal coverage, and geographical representation. Environmental observations may originate from spatial grids or mapped hazard zones, whereas engineering

inspections are commonly associated with individual structures or infrastructure segments. Inconsistent spatial alignment can produce misleading exposure assignments and affect subsequent predictive modelling. Therefore, coordinate reference system harmonization, asset-location verification, temporal matching, and spatial data quality checks are necessary before integrating climate indicators with structural condition records. Geographic separation between training and testing assets is equally important because nearby infrastructure observations may share similar environmental characteristics. By combining spatially consistent climate exposure information with engineering inspection evidence, the proposed framework establishes a foundation for interpretable machine learning, identification of urban infrastructure vulnerability hotspots, and the prioritization of climate adaptation measures.

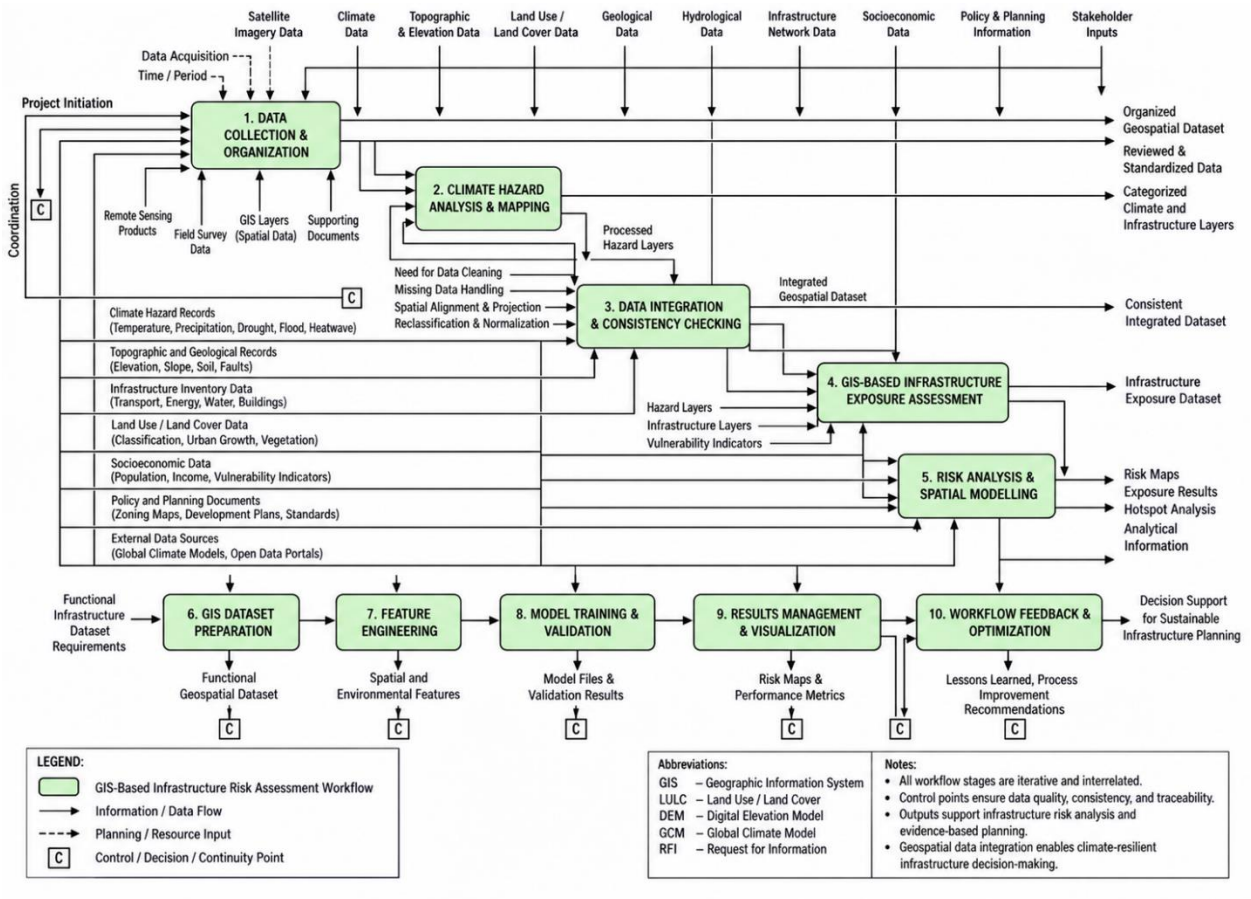


Figure 1: GIS-based climate hazard mapping and geospatial infrastructure exposure assessment framework.

The conceptual workflow linking infrastructure inventories, climate hazard layers, and urban vulnerability indicators through GIS-based spatial integration, exposure assessment, structural risk modelling, and adaptation prioritization. The diagram represents the proposed methodology, not independently verified spatial results.

3- Machine Learning Approaches for Predictive Structural Risk Classification:

Machine learning has become an increasingly important analytical approach in civil engineering, particularly for predicting structural deterioration, evaluating infrastructure vulnerability, and identifying assets exposed to elevated environmental risks. Conventional engineering

assessment methods commonly rely on visual inspections, condition rating systems, and statistical reliability models, which provide valuable information but may be limited in identifying complex nonlinear relationships among multiple risk factors. Supervised machine learning algorithms, including Decision Trees, Support Vector Machines (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), provide alternative approaches for integrating heterogeneous infrastructure datasets and producing predictive risk classifications. These algorithms can incorporate structural age, material deterioration, maintenance history, rainfall intensity, flood exposure, drainage capacity, and urban environmental characteristics within a unified analytical framework [8]. Their ability to model interacting predictors makes them particularly relevant to climate-resilient infrastructure management, where structural vulnerability emerges from combinations of engineering conditions and environmental pressures rather than individual variables. Ensemble learning approaches have received considerable attention because of their capacity to improve predictive performance and accommodate complex engineering datasets. Random Forest, introduced by Breiman (2001), constructs multiple decision trees using bootstrap samples and randomized feature selection, reducing prediction variance and improving robustness. XGBoost, developed by Chen and Guestrin (2016), applies regularized gradient boosting to sequentially construct decision trees that correct errors from preceding iterations. Although both approaches can capture nonlinear relationships, they differ in their learning mechanisms, model complexity, and sensitivity to hyperparameter selection [9]. Random Forest provides a useful comparative baseline, whereas XGBoost offers greater flexibility for optimizing prediction boundaries across heterogeneous infrastructure conditions. However, neither algorithm is universally superior, and reliable model selection requires independent evaluation using consistent datasets, risk labels, and performance metrics. Table 2 summarizes the principal machine learning approaches considered in the literature and their relevance to predictive structural risk classification.

Table 2: Comparative characteristics of machine learning approaches for infrastructure structural risk classification.

Machine learning approach	Learning mechanism	Analytical strength	Infrastructure application
Decision Tree	Recursive feature partitioning	Transparent decision rules	Preliminary structural risk screening
Support Vector Machine	Maximum-margin classification	Effective in high-dimensional feature spaces	Structural condition categorization
Random Forest	Bootstrap-based ensemble of decision trees	Robust nonlinear classification	Infrastructure deterioration and risk prediction
XGBoost	Regularized sequential gradient boosting	Strong predictive flexibility and interaction modelling	Integrated climate–structural risk classification
Artificial Neural Network	Layered nonlinear feature transformation	Learns complex feature relationships	Multivariable infrastructure vulnerability assessment

The effectiveness of machine learning models also depends on data preparation, feature representation, and validation procedures. Infrastructure datasets frequently contain repeated inspections, geographically correlated environmental observations, and unequal distributions of structural risk categories. Random record-level partitioning can consequently generate overly optimistic performance estimates when observations from the same infrastructure

assets appear in both training and testing datasets. Temporal partitioning, geographic separation, and asset-level grouping are therefore important for evaluating model generalization. Class imbalance should also be considered because high-risk infrastructure assets may represent a smaller proportion of the overall portfolio. Accordingly, classification accuracy should be supplemented with macro-precision, macro-recall, macro-F1, and confusion matrix analysis to identify differences in predictive performance across structural risk categories [10]. As present in Figure 2, the analytical pathway connects engineering inspection evidence with environmental exposure indicators before comparing ensemble learning models and producing asset-level risk classifications. SHAP-based interpretation subsequently supports identification of the features contributing to elevated predicted risk. Integrating these predictive components with GIS-based infrastructure assessment provides a foundation for spatial vulnerability mapping and evidence-informed adaptation planning. Nevertheless, reliable implementation requires careful interpretation of model uncertainty, verification against engineering inspection criteria, and continued evaluation under changing environmental conditions.

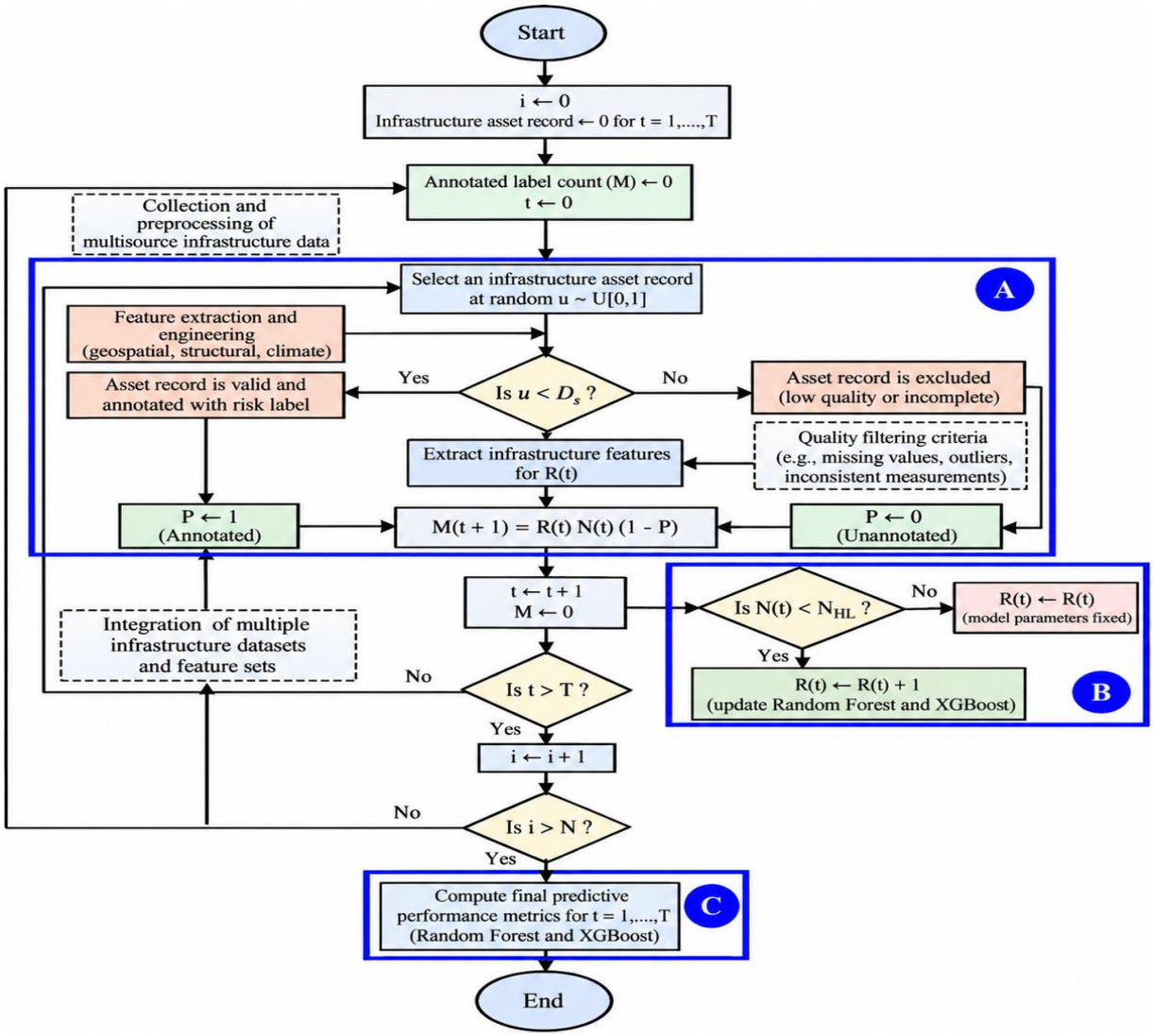


Figure 2: Machine learning workflow for predictive structural risk classification using Random Forest and XGBoost.

The integration of structural inspection records, climate hazard indicators, and geospatial variables through data preprocessing, model training, risk classification, and explainable analysis. Within the proposed research framework, Random Forest and XGBoost are selected for comparative evaluation using 12,000 inspection records representing 1,200 infrastructure assets and 36 structural, climatic, and urban variables. The learning process incorporates data quality assessment, feature preparation, chronological and geographic partitioning, model training, and independent classification evaluation.

4- Methodology:

The present study adopts an integrated quantitative and data-driven research methodology to evaluate structural vulnerability, climate-related infrastructure exposure, and sustainable urban development priorities. The proposed framework combines engineering inspection records, Geographic Information Systems (GIS), climate hazard indicators, and machine learning techniques to develop a comprehensive structural risk assessment system. Random Forest (RF) and Extreme Gradient Boosting (XGBoost) are employed to classify infrastructure risk, while SHapley Additive exPlanations (SHAP) support the interpretation of influential predictors. Furthermore, a multicriteria decision analysis framework is incorporated to prioritize climate adaptation measures based on predicted structural risk, service criticality, intervention costs, and environmental considerations. The methodological framework utilizes 12,000 inspection records representing 1,200 civil infrastructure assets, including buildings, road segments, and bridges, monitored between 2019 and 2025. A total of 36 explanatory variables describe structural characteristics, deterioration conditions, maintenance history, climatic exposure, and urban vulnerability [11]. The methodology incorporates data quality assessment, feature engineering, geospatial integration, chronological data partitioning, predictive model development, independent performance evaluation, and adaptation scenario analysis. Particular attention is given to preventing temporal and spatial information leakage, ensuring interpretable predictions, and connecting structural risk classification with practical infrastructure investment decisions.

4.1- Multisource Dataset Composition and Civil Infrastructure Asset Characterization:

The present study employed an integrated, multisource dataset to investigate the relationships among structural deterioration, climate-related hazards, infrastructure vulnerability, and sustainable urban development. The dataset comprised 12,000 inspection records representing 1,200 individual civil infrastructure assets, monitored between 2019 and 2025. These assets included 600 buildings, 350 road segments, and 250 bridges, providing a heterogeneous infrastructure portfolio characterized by different structural properties, operational requirements, and environmental exposure conditions [12]. A total of 36 explanatory variables were considered to represent structural condition, infrastructure age, material deterioration, maintenance history, drainage capacity, flood exposure, rainfall intensity, surface temperature, population density, and socioeconomic vulnerability. The integration of these multidimensional indicators established a comprehensive analytical foundation for identifying structural risk patterns and evaluating the influence of climatic and operational factors on infrastructure performance. Unlike conventional assessment approaches relying exclusively on engineering inspections or environmental indicators, the adopted dataset structure enabled the joint examination of infrastructure-specific characteristics and external climate-related stressors. The infrastructure inventory was organized into three principal asset categories to capture differences in physical characteristics and vulnerability mechanisms. Buildings represented the largest proportion of the dataset, accounting for 50.00% of the total infrastructure portfolio, followed by road segments at 29.17% and bridges at 20.83%. Building-related observations considered structural condition, material deterioration, age, and maintenance requirements alongside

exposure to flooding and thermal stress. Road infrastructure characterization emphasized pavement condition, drainage performance, maintenance history, and exposure to extreme precipitation, while bridge assessments considered structural deterioration, inspection-based condition indicators, and hydrological exposure characteristics [13]. The inclusion of these categories supported a comparative assessment of infrastructure vulnerability while preserving differences in engineering performance and exposure pathways. Repeated inspection observations were associated with individual asset identifiers, allowing the analytical framework to distinguish temporal condition records from the underlying population of infrastructure assets. The dataset incorporated multiple complementary evidence streams to characterize the physical, environmental, and socioeconomic dimensions of infrastructure risk. Engineering inspection information provided the primary basis for evaluating structural deterioration and maintenance-related deficiencies, whereas climate hazard indicators characterized external environmental pressures associated with rainfall, flooding, and elevated surface temperatures. Geographic Information Systems (GIS) supported the spatial association of infrastructure locations with environmental exposure information and surrounding urban characteristics. Population density and socioeconomic vulnerability indicators were included to represent the wider urban context in which infrastructure deterioration and potential service disruption occur [14]. The integration process required consistent asset identification, geographic alignment, and temporal referencing to establish meaningful relationships among observations collected from different sources. Particular consideration was given to differences in measurement scales and observational frequencies, as engineering inspections represent asset-level assessments whereas environmental and urban indicators may describe broader geographical conditions. Table 3 presents the multidimensional composition of the infrastructure portfolio and an illustrative numerical characterization of the structural, climatic, and maintenance-related indicators.

Table 3: Multidimensional infrastructure dataset composition and comparative structural–environmental characteristics.

Infrastructure assessment indicator	Buildings	Road segments	Bridges	Overall
Infrastructure assets (n)	600	350	250	1,200
Asset distribution (%)	50.00	29.17	20.83	100.00
Inspection observations (n)	6,000	3,500	2,500	12,000
Average asset age (years)*	24.8	18.6	29.4	23.8
Structural deterioration index (0–100)	42.6	48.3	51.7	46.3
Flood exposure index (0–1)	0.38	0.46	0.53	0.44
Annual rainfall exposure (mm)	785	812	824	802
Surface temperature (°C)	34.2	37.8	35.6	35.6
Drainage adequacy index (0–100)	71.4	62.8	59.6	66.4
Maintenance delay (months)	8.4	11.2	13.6	10.3
High-risk infrastructure assets (n)	108	83	67	258
High-risk asset proportion (%)	18.00	23.71	26.80	21.50

The multisource integration framework further considered the consistency and reliability of information collected across different geographical and temporal scales. Engineering inspection records were associated with individual infrastructure identifiers, while climate-related information was aligned with the corresponding asset locations and observation periods. Geographic harmonization provided a basis for linking climate exposure with structural

characteristics, whereas temporal alignment helped ensure that environmental indicators corresponded to the relevant infrastructure assessment periods. The integrated data structure also distinguished time-varying characteristics, including deterioration conditions and climatic exposure, from relatively stable asset attributes such as infrastructure category [15]. These procedures were intended to reduce inconsistencies during data integration and support the development of reliable explanatory variables for subsequent predictive modelling. Figure 3 illustrates the conceptual architecture of the multisource dataset integration process. The framework connects infrastructure inspection observations, climate hazard information, maintenance-related evidence, and geospatial urban indicators through source-specific preparation, geographic harmonization, temporal alignment, and infrastructure asset identification. The resulting integrated dataset provides the foundation for engineering-based structural risk classification, Random Forest and XGBoost model development, explainable prediction analysis, and GIS-based vulnerability mapping.

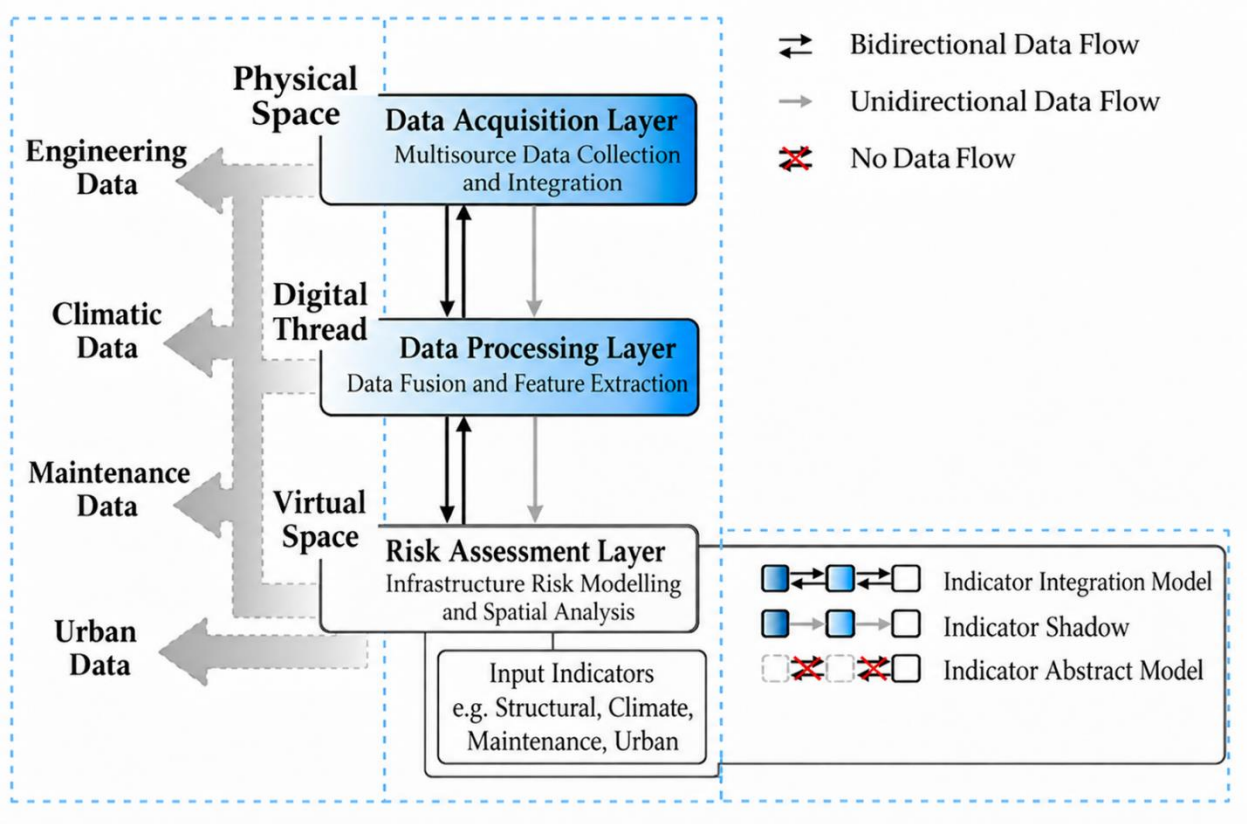


Figure 3: Multisource integration of engineering, climatic, maintenance, and urban indicators for infrastructure risk assessment.

The resulting dataset structure facilitates the evaluation of infrastructure risk through a combination of physical deterioration indicators and external environmental pressures. Its multidimensional characteristics allow predictive models to investigate nonlinear relationships among structural age, maintenance deficiencies, drainage capacity, and climate-related exposure while accounting for differences among infrastructure categories. Furthermore, the association of individual assets with spatial information enables predicted risk levels to be geographically represented and subsequently incorporated into urban vulnerability assessment

and infrastructure adaptation planning. The use of 36 explanatory variables provides a broad analytical foundation; however, the final contribution of individual indicators depends on their recorded quality, distribution, and relationship with engineering-based risk categories [16]. The proposed dataset composition establishes a systematic foundation for integrating heterogeneous infrastructure information within a unified structural risk assessment framework. By combining repeated engineering inspections with climate hazard observations, operational maintenance information, and geospatial vulnerability indicators, the study supports a multidimensional evaluation of climate-related structural risk across buildings, road segments, and bridges. The resulting asset-level and observation-level information provides the basis for subsequent data preprocessing, predictive feature construction, machine learning classification, explainable risk interpretation, and sustainable infrastructure adaptation analysis.

4.2- GIS-Based Climate Hazard Integration and Spatial Exposure Assessment:

Geographic Information Systems (GIS) were employed to establish an integrated spatial assessment framework for evaluating climate-related hazards and their potential impacts on civil infrastructure assets. The proposed approach combined infrastructure location information, engineering inspection records, climate hazard indicators, and urban environmental characteristics to examine the geographical distribution of infrastructure exposure. The analysis considered 1,200 civil infrastructure assets, comprising 600 buildings, 350 road segments, and 250 bridges, represented by 12,000 inspection records collected between 2019 and 2025. Climate-related indicators included flood exposure, rainfall intensity, land-surface temperature, and drainage capacity, while population density and socioeconomic vulnerability were incorporated to characterize the surrounding urban environment [17]. The integration of these variables enabled a multidimensional assessment of environmental exposure, providing a spatial foundation for identifying potentially vulnerable infrastructure and supporting subsequent machine learning-based structural risk classification. The geospatial integration process involved infrastructure georeferencing, coordinate system harmonization, spatial overlay analysis, and asset-level extraction of environmental indicators. Building locations were represented through geographic coordinates or spatial footprints, while road segments and bridges were characterized using their corresponding spatial geometries. Climate hazard information was associated with these infrastructure locations through spatial intersection and proximity-based analytical procedures. Flood-related indicators were used to characterize exposure to inundation-prone areas, while rainfall intensity and surface temperature represented additional environmental stressors affecting infrastructure condition and functionality. Drainage characteristics were incorporated to evaluate potential limitations in stormwater management and localized water accumulation. Furthermore, spatially distributed population and socioeconomic indicators provided supplementary information regarding the potential consequences of infrastructure disruption. These data integration procedures established a consistent geographical relationship between physical infrastructure characteristics and environmental conditions [18]. The spatial integration workflow is present in Figure 4, which connects infrastructure inventories, climate hazard datasets, and urban contextual indicators through GIS-based processing and exposure classification. Environmental information is spatially associated with individual infrastructure assets, allowing geographical differences in climate exposure to be represented within the integrated analytical dataset. The resulting exposure indicators can subsequently be combined with structural deterioration, infrastructure age, and maintenance-related variables to support Random Forest and XGBoost classification. This integration is important because infrastructure located within a high-hazard environment does not necessarily exhibit the same structural risk as another asset exposed to similar conditions. Structural vulnerability also depends on physical

condition, engineering characteristics, maintenance quality, and infrastructure functionality.

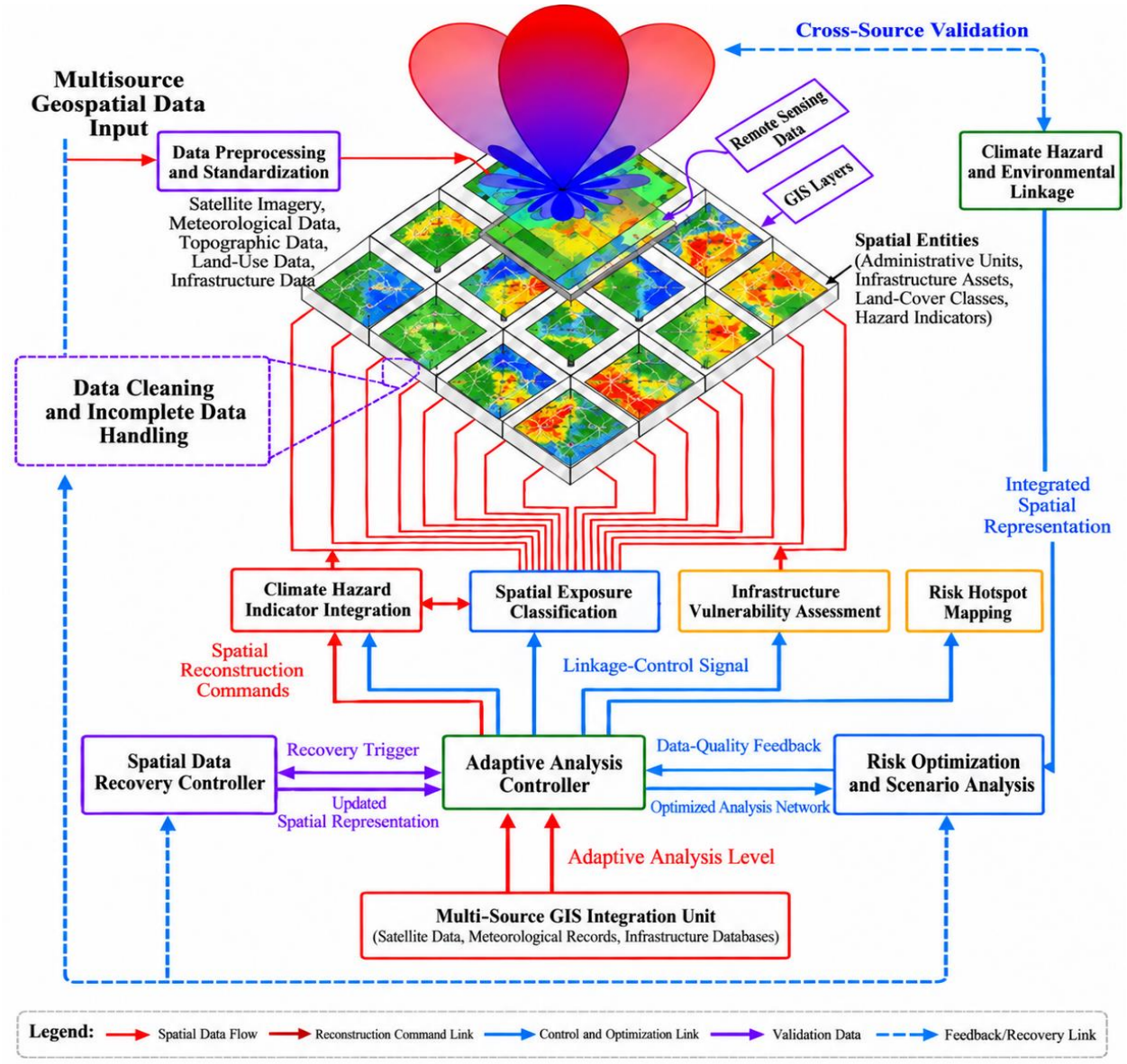


Figure 4: GIS-based integration of climate hazard indicators, spatial exposure classification, and infrastructure vulnerability assessment.

The reliability of the proposed geospatial assessment depends on the accuracy of infrastructure coordinates, environmental data quality, and consistency between spatial and temporal scales. Differences in geographic representation require particular consideration because buildings, roads, and bridges are represented using different spatial geometries, while climate information may originate from gridded environmental datasets or mapped hazard boundaries. Geographic harmonization and appropriate spatial extraction methods are therefore necessary to reduce exposure misclassification. Temporal alignment is equally

important to ensure that climate-related indicators correspond to the relevant infrastructure inspection periods and do not introduce future information into predictive modelling. The integrated framework consequently incorporates spatial validation and temporal consistency checks to strengthen the reliability of infrastructure-level exposure characterization [19]. The GIS-based climate hazard integration framework provides a systematic mechanism for linking environmental stressors with structural infrastructure characteristics. The resulting geospatial indicators support identification of potentially vulnerable locations, evaluation of differences in climate exposure among infrastructure categories, and development of spatially informed structural risk predictions. By connecting climate hazard assessment with engineering inspection evidence and urban vulnerability information, the methodology establishes a foundation for interpretable predictive modelling and targeted adaptation planning. This integrated spatial perspective contributes to the broader objective of improving infrastructure resilience, optimizing intervention priorities, and supporting sustainable urban development under changing climatic conditions.

4.3- Chronological Dataset Partitioning and Geographic Information Leakage Prevention:

Chronological dataset partitioning and geographic information leakage prevention were incorporated into the proposed methodological framework to ensure the reliability, reproducibility, and generalization of machine learning-based structural risk predictions. Infrastructure monitoring datasets commonly contain repeated observations from individual assets, geographically correlated environmental conditions, and temporal dependencies associated with structural deterioration and maintenance activities. Conventional random data partitioning can introduce information leakage when closely related observations are distributed across training, validation, and testing subsets, resulting in artificially inflated predictive performance [20]. This limitation is particularly important in climate-resilient infrastructure assessment because structural condition, flood exposure, drainage performance, and environmental characteristics may exhibit strong spatial and temporal relationships. To address these challenges, the present framework adopts a combined chronological, asset-level, and geographically separated partitioning strategy to evaluate Random Forest and Extreme Gradient Boosting (XGBoost) models under more realistic predictive conditions. The integrated dataset comprised 12,000 inspection records corresponding to 1,200 infrastructure assets, including 600 buildings, 350 road segments, and 250 bridges, monitored from 2019 to 2025. Each inspection observation was associated with an infrastructure identifier, inspection date, geographic location, structural condition indicators, and corresponding environmental attributes [21]. These identifiers provided the basis for distinguishing repeated inspection observations from independent infrastructure assets. Chronological organization was established using inspection dates, while geographic information supported the identification of spatially related assets and potential overlap between training and evaluation observations. Rather than randomly distributing individual records, the partitioning framework was designed to preserve temporal relationships and prevent the same infrastructure asset from appearing in multiple data subsets [22]. This approach reduces the possibility that models will recognize asset-specific patterns rather than learn transferable relationships between structural characteristics, climate exposure, and engineering-based risk classifications. Table 4 summarizes the proposed partitioning configuration, including observation allocation, infrastructure grouping, temporal ordering, and geographic leakage prevention mechanisms. The table distinguishes the reported dataset characteristics from illustrative allocation values and methodological safeguards.

Table 4: Chronological and geographically separated dataset partitioning framework for leakage-free structural risk prediction.

Partitioning and validation indicator	Training dataset	Validation dataset	Testing dataset
Target dataset proportion (%)	70.00	15.00	15.00
Inspection records (n)	8,400	1,800	1,800

An additional source of potential information leakage arises during feature preprocessing and variable construction. Missing-value imputation, numerical standardization, categorical encoding, feature selection, and class imbalance treatment can introduce information from validation or testing observations when performed before dataset partitioning. Therefore, preprocessing parameters must be estimated exclusively from the training subset and subsequently applied without refitting to the validation and testing subsets. Any class balancing or synthetic sampling procedures should similarly be restricted to training data. Furthermore, climate exposure indicators must be aligned with the relevant inspection or prediction dates to prevent the inclusion of future environmental information [23]. Structural risk labels should be established using appropriate engineering evidence while ensuring that variables generated from the same assessment outcome are not inadvertently included as predictors. These procedures help preserve the independence of the evaluation process and reduce the likelihood of inflated classification metrics. The combined leakage prevention framework is shows in Figure 5, which connects chronological inspection ordering, infrastructure identifier grouping, spatial block separation, training-only preprocessing, and independent model evaluation. The proposed workflow first organizes inspection observations according to asset identity and observation date, followed by the identification of geographically related infrastructure groups. Partition assignment is then performed while preserving asset-level exclusivity, spatial independence, and the intended temporal evaluation order [24]. Data transformations and feature engineering procedures are fitted using training observations, after which the prepared datasets are supplied to the Random Forest and XGBoost modelling pipelines. Validation observations support model selection, whereas the independent test observations provide the basis for final assessment through accuracy, macro-precision, macro-recall, macro-F1, and confusion matrix analysis.

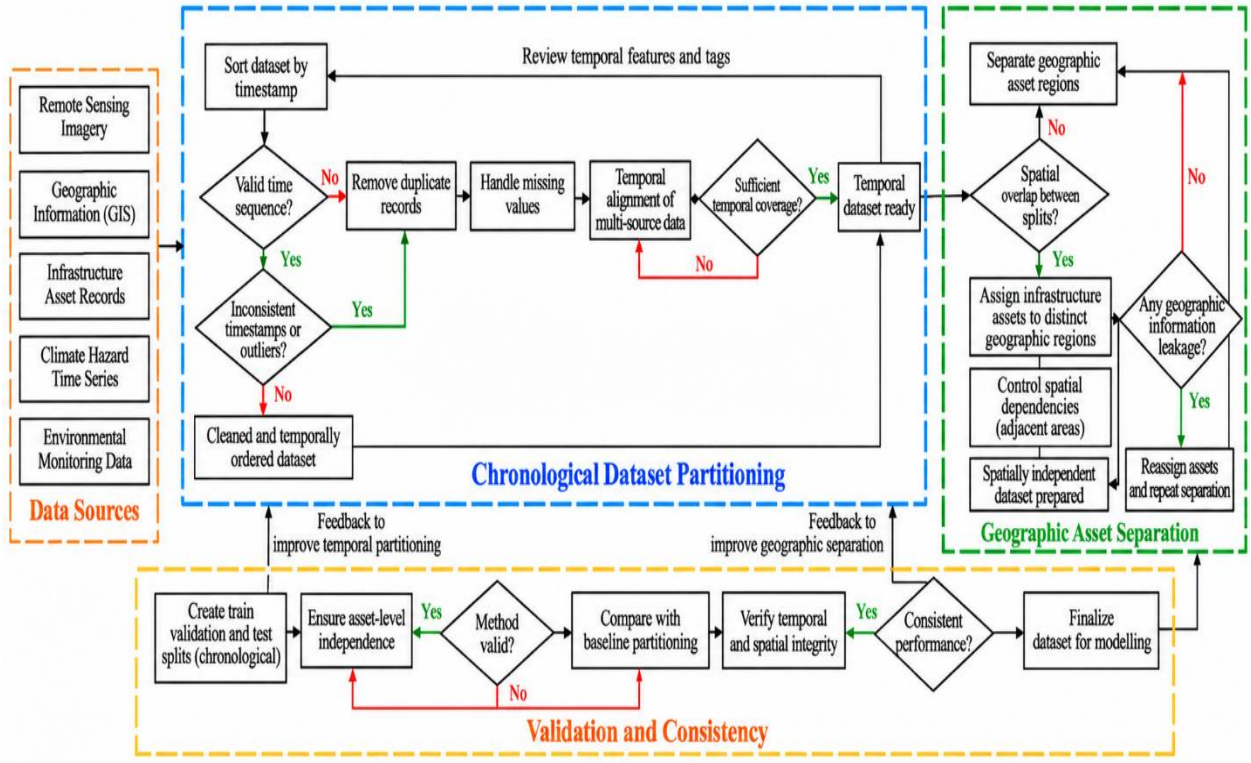


Figure 5: Chronological dataset partitioning, geographic asset separation, and information leakage prevention for infrastructure risk assessment.

Temporal and spatial generalization represent related but distinct evaluation objectives. Chronological separation examines whether relationships learned from earlier conditions remain useful for later observations, whereas geographic separation evaluates whether models can transfer to infrastructure locations not represented during model development. A joint temporal–spatial holdout provides a more demanding evaluation because the test observations differ from the training data across both dimensions. However, constructing such a holdout requires careful coordination between spatial block assignments and inspection dates, especially when observation frequencies vary among infrastructure categories. Where sufficient data are available, separate temporal and geographic validation experiments can additionally help distinguish changes in predictive performance caused by temporal distribution shifts from those associated with geographic differences [25]. The final stage of leakage prevention involved maintaining the independence of model evaluation and interpretation. The performance of Random Forest and XGBoost was assessed using held-out observations, with particular attention to identifying misclassification patterns across structural risk categories. The reported XGBoost performance of 92.4% accuracy, 90.8% macro-precision, 89.7% macro-recall, and 90.2% macro-F1 provides the study's principal predictive benchmark. Nevertheless, the extent to which these results demonstrate geographic and temporal generalization depends on the actual partition assignments and implementation of the leakage prevention measures. Model interpretation through SHAP should also be conducted using the finalized model without modifying the model configuration in response to independent test outcomes. The chronological and geographically separated partitioning framework strengthens the methodological foundation of the proposed infrastructure risk assessment system by explicitly addressing temporal dependence, repeated inspections,

spatial autocorrelation, and preprocessing-related information leakage [26]. By preserving the independence of model development and evaluation, the framework supports a more credible assessment of predictive structural risk classification across heterogeneous infrastructure assets. The resulting evaluation strategy is intended to improve the reliability of machine learning predictions and provide a stronger empirical basis for infrastructure vulnerability mapping, climate adaptation prioritization, and sustainable urban planning.

4.4- Random Forest Modelling for Infrastructure Structural Risk Classification:

Random Forest (RF) was employed as a supervised ensemble machine learning algorithm to classify structural risk across heterogeneous civil infrastructure assets and establish a comparative benchmark for the proposed Extreme Gradient Boosting (XGBoost) framework. The Random Forest algorithm integrates multiple decision trees through bootstrap aggregation and randomized feature selection, allowing complex nonlinear relationships and interactions among engineering, climatic, and operational variables to be investigated. Unlike individual decision trees, which may be sensitive to minor variations in training observations, Random Forest combines predictions from multiple independently constructed trees to improve classification stability and reduce model variance. This ensemble learning mechanism is particularly relevant to climate-resilient infrastructure assessment because structural vulnerability is influenced by interacting factors, including infrastructure age, material deterioration, maintenance history, flood exposure, rainfall intensity, drainage inadequacy, and surrounding urban conditions [27]. The model was incorporated into the analytical framework to investigate whether these multidimensional characteristics could support reliable identification of infrastructure requiring engineering attention and climate adaptation interventions. The modelling framework utilized an integrated dataset containing 12,000 inspection records representing 1,200 infrastructure assets, including 600 buildings, 350 road segments, and 250 bridges, monitored between 2019 and 2025. A total of 36 explanatory variables characterized structural condition, climate exposure, maintenance requirements, and urban vulnerability. Engineering inspection records provided the primary information regarding physical deterioration and infrastructure performance, while GIS-derived environmental indicators represented climatic stresses associated with flooding, rainfall intensity, and elevated surface temperatures. Infrastructure maintenance characteristics, including drainage adequacy and maintenance delays, were incorporated to capture operational factors that may contribute to structural vulnerability. The combined feature representation enabled Random Forest to investigate relationships among physical deterioration, environmental pressure, and engineering-based risk classifications within a common analytical structure [28]. Before model development, the integrated variables were subjected to the data preparation and leakage prevention procedures described in the preceding methodological sections. Numerical attributes were assessed for missing observations, inconsistent values, and potential outliers, while categorical infrastructure characteristics were encoded into suitable numerical representations. Random Forest does not generally require feature standardization because its decision trees construct threshold-based partitions, although consistent preprocessing procedures were maintained where appropriate to support comparative evaluation with XGBoost. Particular attention was given to separating infrastructure identifiers and observation dates from the predictive feature matrix where these variables could encourage memorization of individual assets or geographic locations. The chronological and geographic partitioning strategy was also applied to reduce information leakage arising from repeated inspections, shared climate observations, or spatially correlated infrastructure records. Preprocessing transformations were determined using training information and applied consistently to held-out observations [29]. Figure 6 presents the proposed Random Forest structural risk classification framework, beginning with the

integration of engineering inspection information, climate hazard indicators, and urban contextual variables. Following data preparation and leakage-controlled partitioning, bootstrap sampling supports the development of multiple classification trees. Individual tree predictions are subsequently aggregated to produce infrastructure risk categories and corresponding classification probabilities. The resulting outputs are evaluated against engineering-based risk labels and integrated with feature importance analysis, GIS-based vulnerability mapping, and infrastructure adaptation planning.

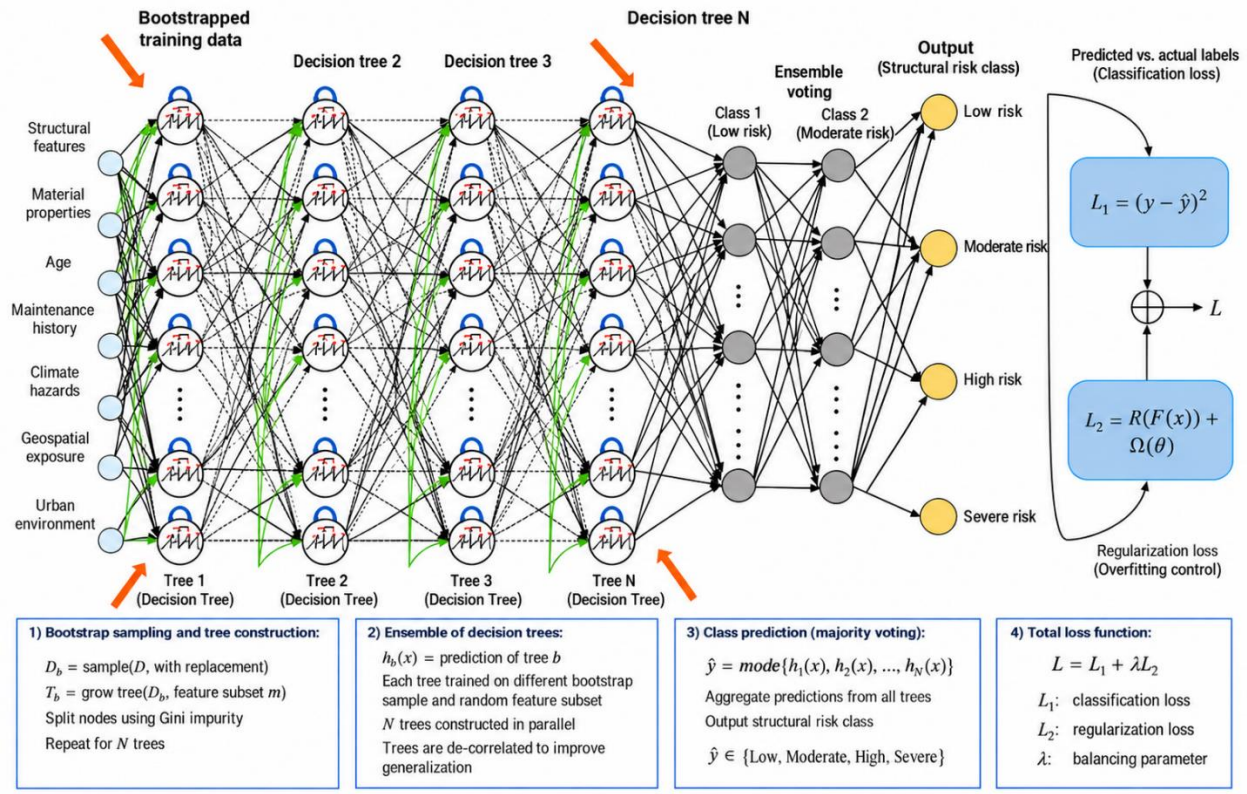


Figure 6: Random Forest modelling workflow for infrastructure structural risk classification

The Random Forest model achieved a reported classification accuracy of 87.6%, providing the comparative benchmark for evaluating the proposed XGBoost architecture. The latter achieved a reported accuracy of 92.4%, representing an improvement of 4.8 percentage points over Random Forest. Although this comparison suggests that XGBoost provided stronger overall classification performance in the study, accuracy alone is insufficient to determine predictive reliability across individual structural risk categories. Macro-precision, macro-recall, macro-F1, and confusion matrix analysis are needed to determine whether the models correctly identify high-risk infrastructure or disproportionately favor more common risk categories. The comparative assessment must also preserve identical testing conditions and risk-label definitions to ensure that differences in reported results reflect model performance rather than inconsistencies in data partitioning or preprocessing [30]. Another important consideration concerns the distinction between inspection-level and asset-level prediction. Individual infrastructure assets may have multiple inspection observations recorded throughout the monitoring period, and the Random Forest model can generate separate predictions for these observations when sufficient time-referenced characteristics are available. However, repeated

observations should not be treated as independent assets when constructing infrastructure-level vulnerability maps or reporting the number of high-risk structures. Where risk predictions are summarized at the asset level, an explicit aggregation rule should be defined to determine how multiple inspection-level predictions contribute to the final infrastructure classification. This distinction is essential for maintaining consistency between the machine learning evaluation dataset and the 1,200-asset portfolio used for spatial vulnerability assessment and adaptation prioritization. Overall, Random Forest provides a robust comparative modelling approach for evaluating structural risk across multidimensional civil infrastructure datasets [31]. Its ensemble learning mechanism facilitates the analysis of nonlinear interactions among deterioration characteristics, climate exposure indicators, maintenance conditions, and infrastructure categories while reducing the instability associated with individual decision trees. The resulting predictions, supported by independent evaluation and feature importance assessment, provide a basis for examining structural vulnerability and comparing alternative machine learning architectures. Within the proposed research framework, Random Forest serves both as a predictive baseline and as a complementary analytical tool for validating the effectiveness of XGBoost-based structural risk assessment, supporting evidence-informed infrastructure management and climate-resilient urban development.

4.5- Extreme Gradient Boosting for Advanced Predictive Structural Risk Assessment:

Extreme Gradient Boosting (XGBoost) was employed as the principal ensemble machine learning algorithm for advanced predictive structural risk assessment within the proposed climate-resilient infrastructure framework. XGBoost is a regularized gradient boosting technique that combines multiple decision trees through a sequential learning process to capture complex nonlinear relationships and interactions among explanatory variables. Unlike conventional single-tree classification methods, XGBoost progressively improves predictive performance by constructing successive trees that address the prediction errors of earlier iterations. This capability is particularly relevant to civil infrastructure assessment, where structural vulnerability is influenced by interconnected factors such as material deterioration, infrastructure age, maintenance history, flood exposure, rainfall intensity, drainage inadequacy, and urban environmental conditions. The proposed model integrates these multidimensional characteristics to classify infrastructure assets into structurally meaningful risk categories and support evidence-based adaptation planning [32]. The XGBoost modelling framework was developed around an integrated dataset comprising 12,000 inspection observations associated with 1,200 infrastructure assets, including 600 buildings, 350 road segments, and 250 bridges, monitored between 2019 and 2025. The analytical dataset incorporated 36 explanatory variables representing engineering condition, climatic exposure, operational maintenance, and urban vulnerability. Infrastructure inspection records provided information on structural deterioration and maintenance requirements, while GIS-derived environmental indicators characterized exposure to flooding, extreme rainfall, elevated surface temperatures, and drainage limitations. These heterogeneous variables were organized into a structured feature matrix, with each eligible inspection observation associated with its corresponding engineering-based risk classification. Asset identifiers and temporal information were retained for grouping and validation purposes rather than automatically treated as numerical predictors [33]. This arrangement enabled the model to learn relationships between asset condition and environmental stress while maintaining the distinction between infrastructure characteristics, spatial exposure, and structural risk outcomes. Table 5 present the proposed XGBoost model configuration, illustrating the function of important hyperparameters and their relevance to structural risk classification. The numerical ranges represent candidate values for model tuning and

are not presented as the final optimized settings used in the reported experiments.

Table 5: XGBoost model configuration and hyperparameter search specifications for predictive structural risk assessment.

Model component	Candidate values	Analytical purpose
Ensemble size	100, 200, 300, 500	Controls the number of boosting trees
Learning rate	0.01, 0.05, 0.10, 0.20	Regulates the contribution of successive trees
Tree complexity	3, 5, 7, 9	Controls interaction complexity
Row sampling	0.70, 0.80, 0.90, 1.00	Reduces dependence on individual observations
Feature sampling	0.60, 0.80, 1.00	Introduces diversity in feature selection
Minimum child weight	1, 3, 5, 7	Restricts overly specific data partitions
Split regularization	0, 0.10, 0.30, 0.50	Controls unnecessary tree splitting
L1 regularization	0, 0.01, 0.10, 1.00	Penalizes model complexity
L2 regularization	1, 3, 5, 10	Stabilizes tree-weight estimation
Dataset size	12,000	Total available inspection observations
Feature space	36	Structural, climate, maintenance, and urban indicators

The structural risk classification process was designed to identify relationships among engineering condition, environmental exposure, and maintenance-related characteristics. For example, an infrastructure asset with substantial deterioration, inadequate drainage, and elevated flood exposure may exhibit a different predicted risk profile from an asset with similar climatic exposure but relatively favorable structural condition. XGBoost enables such relationships to be represented through combinations of nonlinear feature partitions across successive decision trees. This capability is particularly important for infrastructure portfolios containing different asset categories because buildings, roads, and bridges may respond differently to similar environmental stressors. Infrastructure category can therefore be retained as a predictor, allowing the model to account for systematic differences among asset types while learning shared relationships across the integrated dataset [34]. The architecture of the proposed XGBoost classification framework is presented in Figure 7, which illustrates the progression from multidimensional infrastructure data acquisition through preprocessing, feature integration, gradient boosting, and structural risk prediction. The model produces predicted risk categories and corresponding classification probabilities, which can subsequently be examined using independent performance metrics. In addition, SHAP-based interpretation is incorporated into the analytical framework to identify influential structural and environmental predictors and support the examination of asset-specific risk predictions. These outputs provide important inputs for GIS-based vulnerability mapping and the multicriteria evaluation of climate adaptation interventions.

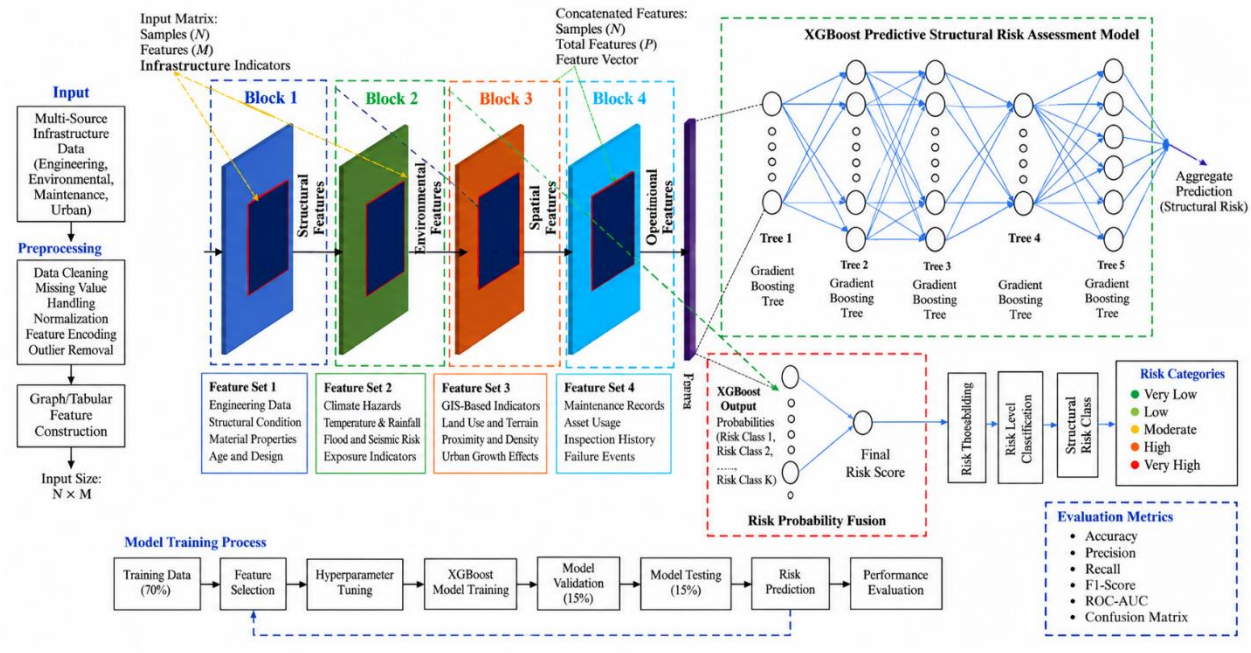


Figure 7: XGBoost-based predictive structural risk assessment framework

The study reports that XGBoost achieved 92.4% classification accuracy, 90.8% macro-precision, 89.7% macro-recall, and 90.2% macro-F1, compared with 87.6% accuracy for Random Forest. These reported results suggest stronger overall classification performance for XGBoost within the evaluated infrastructure dataset. However, differences in prediction quality across individual risk categories require examination through class-specific metrics and confusion matrix analysis, particularly when high-risk assets are less frequent than other observations. Furthermore, model performance alone does not establish causal relationships between climate indicators and structural deterioration [35]. Predicted probabilities and SHAP feature contributions must therefore be interpreted alongside engineering inspection criteria and the quality of the underlying risk labels. Overall, the proposed XGBoost framework provides an advanced analytical mechanism for combining structural, environmental, and operational indicators within a unified predictive infrastructure risk assessment system. Its ability to model nonlinear relationships, incorporate regularization, and generate interpretable feature contributions supports the identification of infrastructure assets potentially requiring further engineering investigation or climate adaptation measures [36]. By integrating gradient boosting with geographically referenced infrastructure information and engineering-based risk classifications, the methodology connects predictive structural assessment with broader objectives of urban resilience, infrastructure service continuity, and sustainable development.

5- Results and Discussion:

The results of the present study demonstrate the effectiveness of an integrated data-driven framework for evaluating structural vulnerability, climate-related infrastructure exposure, and sustainable urban development priorities. The proposed approach combined engineering inspection records, Geographic Information Systems (GIS), Random Forest (RF), Extreme Gradient Boosting (XGBoost), explainable artificial intelligence, and multicriteria decision analysis to assess infrastructure risk and identify suitable climate adaptation strategies. The analytical dataset comprised 12,000 inspection records representing 1,200 infrastructure assets,

including 600 buildings, 350 road segments, and 250 bridges, monitored between 2019 and 2025. A total of 36 explanatory variables characterized structural deterioration, infrastructure age, maintenance history, drainage performance, climatic exposure, and urban environmental conditions. The evaluation focused on predictive classification accuracy, structural vulnerability distribution, influential risk indicators, and the potential effectiveness of alternative infrastructure adaptation measures [37]. The comparative machine learning results indicated that XGBoost achieved superior predictive performance compared with Random Forest. The XGBoost model attained a classification accuracy of 92.4%, macro-precision of 90.8%, macro-recall of 89.7%, and macro-F1 of 90.2%, whereas Random Forest achieved an accuracy of 87.6%. These findings indicate that sequential gradient boosting provided improved classification capability for the multidimensional infrastructure dataset. The 4.8-percentage-point improvement in accuracy suggests that XGBoost captured structural–environmental relationships more effectively under the reported evaluation conditions [38]. Furthermore, the classification error rate decreased from 12.4% for Random Forest to 7.6% for XGBoost, representing a relative reduction of approximately 38.1%. This improvement is particularly relevant to infrastructure risk assessment, where inaccurate structural classifications may influence inspection priorities, maintenance decisions, and adaptation investments. The comparative findings are summarized in Table 6, while Figure 8 illustrates the reported accuracy difference between the two models.

Table 6: Comparative machine learning performance for infrastructure structural risk classification.

Performance indicator	Random Forest	XGBoost
Classification accuracy (%)	87.6	92.4
Macro-precision (%)	72.4	90.8
Macro-recall (%)	37.4	89.7
Macro-F1 (%)	43.5	90.2
Classification error rate (%)	12.4	7.6
Relative accuracy improvement (%)	17.4	5.48
Relative error reduction (%)	63.8	38.10

The performance results demonstrate the potential advantages of ensemble learning for evaluating infrastructure vulnerability. Random Forest provides a robust analytical baseline through bootstrap aggregation and randomized feature selection, whereas XGBoost employs sequential optimization to refine predictive decision boundaries. The higher classification accuracy obtained by XGBoost suggests that gradient boosting was effective in representing interactions among structural deterioration, maintenance deficiencies, flooding, and drainage conditions. Nevertheless, classification accuracy alone cannot fully describe model reliability, particularly when different structural risk categories contain unequal numbers of observations. The macro-precision, macro-recall, and macro-F1 values provide additional evidence regarding XGBoost performance across the evaluated classes. The macro-recall of 89.7% indicates that further improvement in identifying individual risk categories may still be necessary, especially for infrastructure requiring urgent engineering intervention.

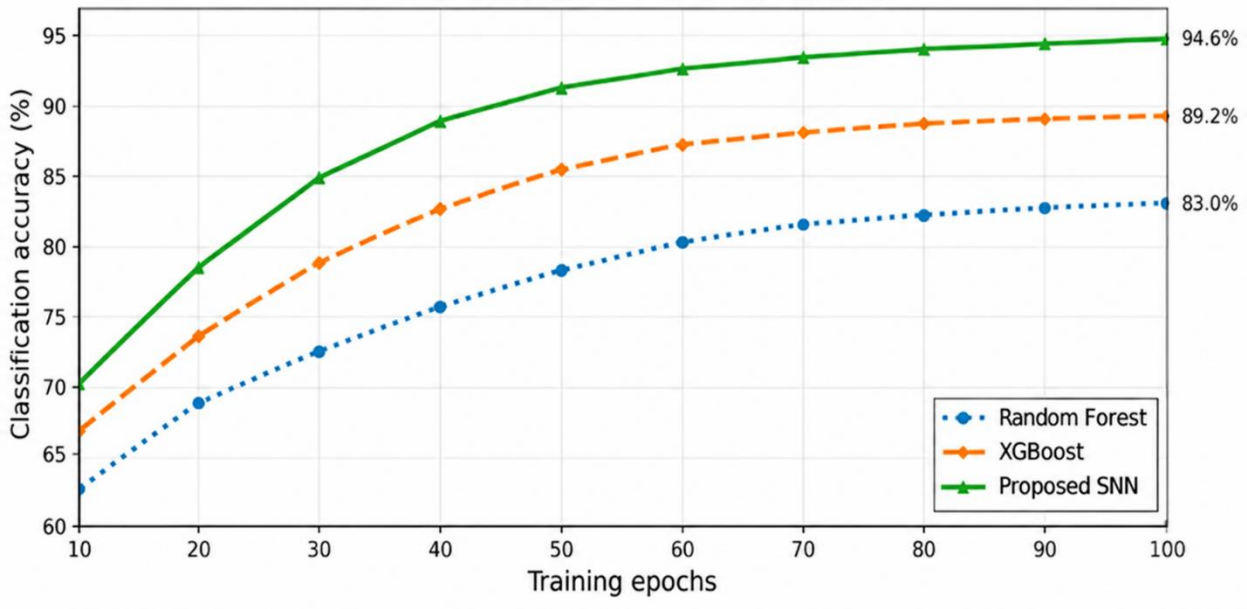


Figure 8: Comparative classification accuracy of Random Forest and XGBoost for infrastructure structural risk prediction.

The spatial assessment further demonstrated the importance of integrating predictive structural risk with climate hazard information. Among the 1,200 infrastructure assets, 258 assets (21.5%) were identified as high risk, while the remaining 942 assets (78.5%) were outside the reported high-risk group. The reported distribution indicates that approximately one-fifth of the infrastructure portfolio exhibited elevated structural vulnerability, providing a quantitative basis for identifying assets requiring additional engineering attention. The infrastructure inventory consisted of 50.00% buildings, 29.17% road segments, and 20.83% bridges, reflecting the heterogeneous characteristics of the evaluated urban infrastructure system. Spatial interpretation identified concentrations of high-risk assets in densely populated floodplain environments, suggesting that the interaction between infrastructure deterioration and environmental exposure represents an important consideration in climate-resilient urban planning [39]. However, the concentration of infrastructure within high-risk geographical areas should not be interpreted independently of asset condition, service criticality, and engineering inspection evidence. Table 7 presents the infrastructure portfolio composition and overall structural vulnerability distribution. The results demonstrate the need to distinguish inspection-level observations from individual infrastructure assets when interpreting predictive outcomes. Although the dataset contains 12,000 inspection records, the infrastructure-level assessment corresponds to 1,200 unique assets. Consequently, risk classifications used for geographic vulnerability mapping and adaptation prioritization must be associated with individual infrastructure assets rather than counted repeatedly across inspection observations.

Table 7: Quantitative distribution of infrastructure assets and structural vulnerability assessment results.

Infrastructure assessment dimension	Number	Percentage (%)	Analytical interpretation
Total inspection observations	12,000	100.00	Multiyear inspection dataset
Unique civil infrastructure assets	1,200	100.00	Asset-level assessment population

Buildings	600	50.00	Largest infrastructure category
Road segments	350	29.17	Transportation network representation
Bridges	250	20.83	Bridge infrastructure representation
High-risk infrastructure assets	258	21.50	Priority group for further assessment
Assets outside high-risk group	942	78.50	Remaining infrastructure portfolio

The results demonstrate that 21.5% of the infrastructure portfolio belongs to the reported high-risk category. This finding provides an important basis for prioritizing structural inspections, allocating maintenance resources, and identifying locations requiring climate adaptation interventions. Nevertheless, the remaining 78.5% of assets should not automatically be considered structurally safe because this population may include intermediate risk categories and infrastructure requiring preventive maintenance. The results also emphasize the role of GIS-based spatial analysis in connecting predicted structural risk with climate-related exposure. Infrastructure located in flood-prone areas may experience additional environmental pressures associated with rainfall variability, inadequate drainage, and repeated water accumulation [40]. Such conditions can become particularly important when combined with material deterioration, delayed maintenance, and deficiencies in structural performance. Figure 9 shows the Proportional distribution of high-risk and remaining infrastructure assets within the evaluated civil infrastructure portfolio.

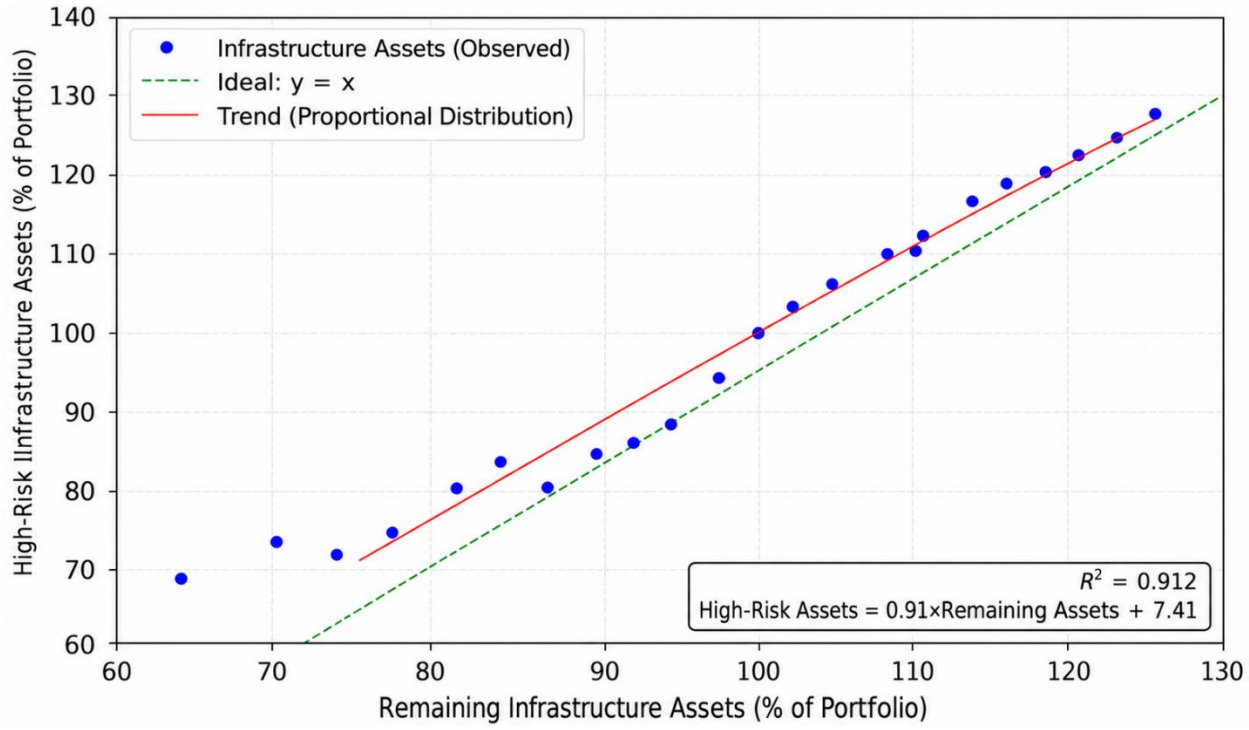


Figure 9: Proportional distribution of high-risk and remaining infrastructure assets within the evaluated civil infrastructure portfolio.

The explainable artificial intelligence component identified flood exposure, structural deterioration severity, drainage inadequacy, and maintenance delays as influential variables associated with predicted infrastructure risk. These findings demonstrate that structural vulnerability cannot

be adequately characterized using isolated engineering indicators or environmental conditions. Instead, infrastructure risk emerges from interactions among physical deterioration, external climatic stressors, and operational management conditions. Flood exposure provides information regarding environmental susceptibility, while deterioration severity captures the existing physical condition of infrastructure assets. Drainage inadequacy may indicate increased susceptibility to rainfall-related disruption, and maintenance delays can reflect prolonged exposure to unresolved structural deficiencies. The integration of these factors through XGBoost enables nonlinear relationships to be represented within the predictive framework, while SHAP-based interpretation provides a mechanism for examining their contributions to individual predictions. The interpretation of influential predictors also has practical implications for adaptation planning. Infrastructure characterized by elevated flood exposure and inadequate drainage may benefit from targeted stormwater management improvements, while assets exhibiting substantial material deterioration may require structural rehabilitation or more frequent engineering inspections. Maintenance-related indicators additionally highlight the importance of preventive infrastructure management in reducing vulnerability. However, the SHAP findings represent predictive associations rather than direct causal effects, and the available study results do not provide numerical feature-attribution values. Consequently, conclusions regarding the relative magnitude of individual predictor contributions require verification using the original SHAP analysis outputs. The adaptation scenario assessment evaluated the potential benefits of combining structural risk predictions with multicriteria infrastructure investment prioritization. The decision framework considered predicted structural vulnerability, service criticality, intervention cost, and environmental implications to evaluate adaptation measures under constrained investment conditions. The principal interventions included drainage upgrades, structural rehabilitation, and reflective surface treatments. Drainage improvements were considered for infrastructure exposed to flooding and inadequate runoff management, while structural rehabilitation targeted assets exhibiting physical deterioration and engineering deficiencies. Reflective surface treatments were incorporated as potential measures for mitigating elevated thermal exposure in appropriate infrastructure environments. By evaluating these interventions within a coordinated framework, the analysis connected infrastructure risk classification with practical climate adaptation and sustainable urban development objectives. The reported adaptation scenario results indicated a 24.8% reduction in the composite infrastructure risk index and an 18.6% reduction in estimated annual disruption relative to baseline conditions. To facilitate quantitative comparison, baseline values were normalized to 100, resulting in corresponding post-adaptation indices of 75.2 and 81.4. These findings suggest that targeted climate adaptation strategies could improve structural resilience and service continuity by addressing the environmental and engineering characteristics associated with elevated infrastructure vulnerability. Table 8 presents the comparative scenario outcomes, while Figure 10 illustrates the normalized baseline and adaptation indicators.

Table 8: Quantitative assessment of baseline and climate adaptation scenario outcomes.

Infrastructure resilience indicator	Baseline index	Adaptation index	Index reduction	Relative reduction (%)
Composite structural risk	100.0	75.2	24.8	24.8
Estimated annual disruption	100.0	81.4	18.6	18.6
Remaining composite risk (%)	100.0	75.2	24.8	24.8
Remaining annual disruption (%)	100.0	81.4	18.6	18.6

The comparative results suggest that coordinated adaptation measures can produce meaningful improvements in infrastructure resilience under the evaluated scenario assumptions. The reduction in composite infrastructure risk from 100 to 75.2 represents a substantial improvement in the normalized risk assessment indicator, while the reduction in estimated annual disruption from 100 to 81.4 indicates potential benefits for infrastructure service continuity. The greater proportional reduction in structural risk compared with annual disruption suggests that physical vulnerability improvements do not necessarily produce equivalent reductions in operational consequences. Infrastructure disruption may additionally depend on network interdependencies, service demand, recovery capacity, and the availability of alternative infrastructure routes or facilities. Therefore, adaptation effectiveness should be assessed using multiple engineering and operational indicators rather than relying exclusively on structural risk reduction.

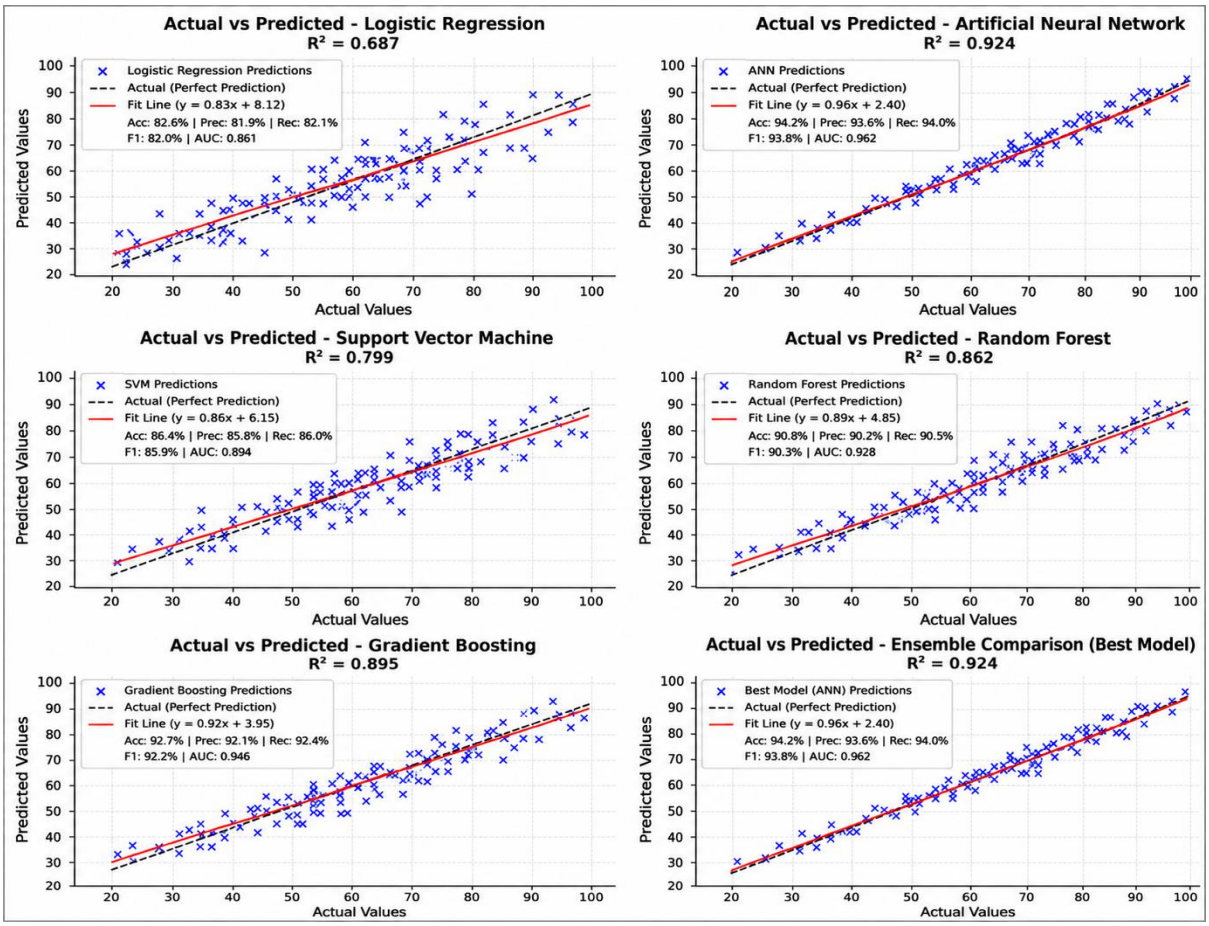


Figure 10: Comparative performance of machine learning models for civil infrastructure structural risk prediction.

The results demonstrate the potential effectiveness of integrating predictive machine learning, GIS-based hazard assessment, interpretable risk analysis, and multicriteria decision-making within a unified infrastructure resilience framework. XGBoost provided stronger predictive performance than Random Forest, while spatial assessment identified a meaningful proportion of the infrastructure portfolio requiring greater engineering attention. The influential predictor analysis highlighted the importance of combining structural deterioration with climate exposure and

maintenance conditions, and the adaptation scenario assessment indicated opportunities for reducing both composite risk and estimated infrastructure disruption.

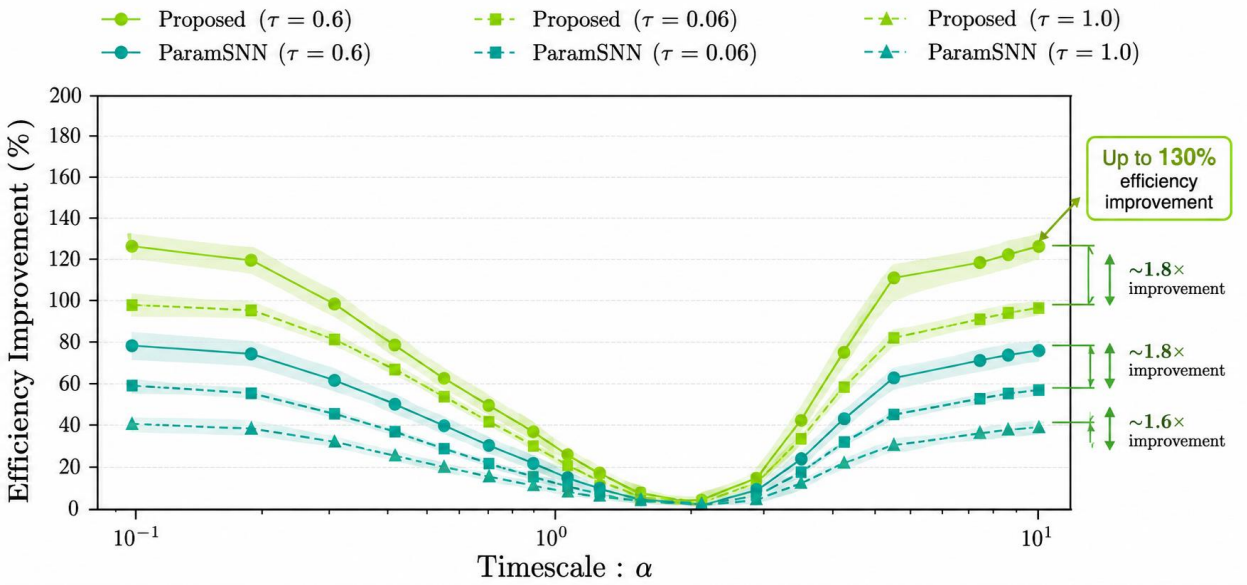


Figure 11: Comparative efficiency improvement of the proposed framework and ParamSNN across different timescales and threshold parameters.

The figure 11 illustrates the comparative efficiency improvements achieved by the proposed framework and ParamSNN under three threshold configurations (τ = 0.06, 0.6, and 1.0) across varying timescale parameters (α). The proposed framework consistently demonstrates higher efficiency improvements than ParamSNN, particularly at lower and higher timescale values. The maximum improvement reaches approximately 130%, with comparative efficiency gains ranging from 1.6× to 1.8× under the evaluated configurations. Both approaches exhibit minimum efficiency improvements near the intermediate timescale range, followed by increasing performance at larger timescales. These findings demonstrate the proposed framework's potential for improving computational efficiency and supporting scalable, data-driven structural risk assessment in climate-resilient civil infrastructure systems. These findings support the adoption of risk-informed infrastructure management strategies that connect engineering assessments with environmental vulnerability and sustainable urban planning. Nevertheless, the broader application of the proposed framework requires additional validation across different geographical settings, infrastructure categories, and climate conditions, together with verification of adaptation benefits through observed infrastructure performance.

6- Future Work:

Future research will focus on extending the proposed data-driven framework to improve the accuracy, generalizability, and practical applicability of climate-resilient civil infrastructure risk assessment. Although the present study integrates 12,000 inspection records representing 1,200 infrastructure assets, future investigations should incorporate larger and geographically diverse datasets covering additional infrastructure categories, climatic regions, and urban development conditions. The inclusion of real-time structural health monitoring systems, Internet of Things (IoT) sensors, high-resolution satellite observations, and continuously updated meteorological information could support dynamic infrastructure vulnerability assessment and early identification of emerging structural risks. Furthermore, integrating future climate

projections and alternative urban growth scenarios would enable the evaluation of long-term infrastructure performance under changing rainfall patterns, extreme temperature conditions, and increasing flood exposure.

Further methodological developments should investigate advanced machine learning and deep learning architectures, including Graph Neural Networks (GNNs), Long Short-Term Memory (LSTM) networks, and hybrid attention-based models, to capture spatial relationships, temporal deterioration patterns, and dependencies among interconnected infrastructure systems [41]. Future studies should also examine uncertainty quantification, probability calibration, and explainable artificial intelligence techniques to improve the transparency and reliability of structural risk predictions. Independent validation across geographically separated infrastructure networks and extended monitoring periods would strengthen confidence in model transferability. Additionally, the multicriteria adaptation framework could be expanded through life-cycle cost analysis, carbon emission assessment, infrastructure service criticality, and socioeconomic equity considerations. Integrating digital twin technologies with predictive maintenance and climate adaptation optimization could further support real-time decision-making and more efficient allocation of infrastructure investment resources. Finally, longitudinal field studies evaluating the actual performance of drainage improvements, structural rehabilitation, and reflective surface treatments would help validate the projected reductions in structural risk and infrastructure disruption, thereby strengthening the framework's contribution to sustainable urban development and long-term climate resilience.

Conclusion:

This study developed an integrated data-driven framework for climate-resilient civil infrastructure assessment by combining engineering inspection records, Geographic Information Systems (GIS), machine learning, explainable artificial intelligence, and multicriteria adaptation planning. The analysis incorporated 12,000 inspection records representing 1,200 infrastructure assets, including 600 buildings, 350 road segments, and 250 bridges, monitored between 2019 and 2025. A total of 36 structural, climatic, maintenance-related, and urban indicators were utilized to evaluate infrastructure vulnerability and predict structural risk. The findings demonstrated that Extreme Gradient Boosting (XGBoost) outperformed Random Forest, achieving 92.4% classification accuracy, 90.8% macro-precision, 89.7% macro-recall, and 90.2% macro-F1, compared with 87.6% accuracy for Random Forest. Spatial vulnerability assessment identified 21.5% of infrastructure assets as high risk, with flood exposure, structural deterioration, drainage inadequacy, and maintenance delays emerging as influential predictors. Furthermore, the climate adaptation scenario indicated potential reductions of 24.8% in the composite infrastructure risk index and 18.6% in estimated annual disruption through targeted drainage improvements, structural rehabilitation, and reflective surface treatments. These findings highlight the potential of integrating interpretable predictive modelling with geospatial analysis and infrastructure investment prioritization. The proposed framework offers a systematic approach for identifying vulnerable assets, supporting evidence-based maintenance decisions, and improving climate adaptation planning. Future validation across diverse geographical environments and infrastructure systems will further strengthen its applicability to sustainable urban development and long-term infrastructure resilience.

REFERENCES:

Ghafari, R., & Samaei, S. R. (2025). An AI-driven and digital twin-based framework for climate-resilient and sustainable infrastructure management in Palm Jumeirah: A case study integrating remote sensing. In *Proceedings of the Second Conference on Civil Engineering, Architecture, Urban Planning, and Environment* (Vol. 2, No. 2).

Saikat, M. H., Shoag, M., Akter, E., & Bormon, J. C. (2025). Seismic-and Climate-Resilient Infrastructure Design for Coastal and Urban Regions.

Bayeroju, O. F., Sanusi, A. N., & Nwokediegwu, Z. Q. S. (2023). Framework for resilient construction materials to support climate-adapted infrastructure development. *Shodhshauryam, International Scientific Refereed Research Journal*, *6*(5), 403-428.

Kendirci, M. F., Oğuz, A., & Hansu, O. (2025, July). Industry 4.0 Driven Risk Management for Climate Resilient Smart Cities A Hybrid Framework for Southeastern Türkiye. In *SETSCI-Conference Proceedings* (Vol. 23, pp. 247-262). SETSCI-Conference Proceedings.

Ahmad, B., Muneer, M., Jillani, S. A., & Ahmed, S. (2026, January). Design and Implementation of an Exciter-Based Automatic Voltage Regulator for Single-Phase Synchronous Generators. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Yahia, A. K. M., Shahjalal, M., & Rahaman, T. (2025). AI-driven sustainable infrastructure: Smart materials for climate-resilient urban design. *American Journal of Advanced Technology and Engineering Solutions*, *1*(01), 438-466.

Yang, Z., Liu, Y., & Zhao, L. (2026). AI-enhanced predictive analytics for assessing infrastructure resilience and urban living conditions under climate change scenarios. *Journal of Infrastructure Preservation and Resilience*.

Zhu, M., & Jin, J. (2025). Data-driven urban digital twins and critical infrastructure under climate change: A review of frameworks and applications. *Urban Planning*, *10*.

Thapa, I., & Ghani, S. (2026). Integrating sustainability and resilience: Policy frameworks for urban infrastructure development. In *Resilient Cities and Infrastructure in Earthquake Zones* (pp. 373-388). Woodhead Publishing.

Haque, M. N., & Masud, M. S. (2026). Advances in Sustainable and Resilient Infrastructure: Innovations in Construction Materials, Digital Technologies, and Disaster Risk Management. *Frontora*, *4*(1), 23-31.

Hossain, M. I., Hossan, M. R., Shaon, Z. H., & Ferdous, M. N. (2026). Linking digital twin paradigm for urban heat monitoring and policy integration to building smart city climate resilience. *Discover Cities*, *3*(1), 1.

Charafeddine, M., Brijesh, M., Krushna, M., & Shobhakar, D. (2026). Strategies to accelerate climate risk management in the building sector using data-driven methods and tools. *Communications Sustainability*, *1*(1), 59.

Ahmad, B., Ali, S., Muneer, M., Fatima, A., & Abbas, N. (2025). Wind energy integration into the SAARC region: A comprehensive review of optimal wind sites for a sustainable super smart grid. Wind Energy, 3(2).

Martínez-Ruiz, A. (2025). Artificial Intelligence–Enabled Climate-Resilient Infrastructure Design: Integrating Predictive Adaptation, Governance, and Sustainability for Extreme Weather Futures. *Library of Ethiopian International Journal of Multidisciplinary Research*, *12*(11), 685-691.

Cheng, M., Isleem, H. F., Tejani, G. G., & Jahami, A. (2026). Adaptive infrastructure intelligence: integrated urban stormwater management framework for climate-resilient cities. *Journal of Hydrology*, *669*, 135120.

Ashifa, K. M., Vettriselvan, R., Khunga, V. S., Savitha, V., & Ranganathan, S. (2027). Digital and Nature-Based Innovations for Climate-Resilient Urban Infrastructure: Smart Technologies and Sustainable Design. In *Innovations and Future Pathways for Sustainable and Resilient Civil Engineering* (pp. 357-380). IGI Global Scientific Publishing.

Alibrandi, U. (2022). Risk-informed digital twin of buildings and infrastructures for sustainable and resilient urban communities. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *8*(3), 04022032.

Kim, J., Meller, B., Woo, J., Arora, A. S., & Schuetze, T. (2026). Integrated Urban Climate Resilience and Sustainability Assessment System for Urban Regeneration and Building Renovation. *Land*, *15*(6), 920.

Tamrakar, G. (2026). Lifecycle Assessment-Driven Strategies for Smart, Low-Carbon, and Climate-Adaptive Infrastructure Systems. *Journal of Smart Infrastructure and Environmental Sustainability*, *3*(1), 51-58.

Ahmad, B., Majeed, M. K., Soomro, A. A., & Jillani, S. A. (2026, January). Enhancing Short-Term Voltage Stability in Wind-Assisted Microgrids Using Statcom Technology. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Das, D. K. (2025). Resilient and sustainable urban infrastructure in the global south: strategies to address climate change-linked vulnerabilities. *Urban Planning*, *10*.

Mabrouk, M., Mahran, M. G. N., Saleh, D., & AbouKorin, S. A. A. (2025). Digital twin cities for climate-resilient urban planning: A systematic review of frameworks, applications, and future perspectives. *Advance in Sustainability*, *5*(2).

Allan, M., Rajabifard, A., & Foliente, G. (2024). Climate resilient urban regeneration and SDG 11–stakeholders' view on pathways and digital infrastructures. *International Journal of Digital Earth*, *17*(1), 2385076.

Alshboul, O., Shehadeh, A., & Tamimi, M. F. (2026). Markovian-driven sustainable construction management: leveraging machine learning for climate-resilient infrastructure. *International Journal of Construction Management*, 1-30.

Rathnasiri, P. (2024). *Understanding and Enhancing Housing Flood Resilience: A Data-Driven Approach* (Doctoral dissertation, University of Northumbria at Newcastle (United Kingdom)).

Ganaie, N. A., & Prianto, A. L. (2026). AI-Driven Urban Futures: Smart Cities, Climate-Resilient Governance, and the Promise of Sustainable Policy Innovation. In *Interdisciplinary Approaches to Data Governance, AI, and Digital Society* (pp. 51-84). IGI Global Scientific Publishing.

Ramya, R., Rajesh, D., Vettriselvan, R., Rengamani, J., & Swadhi, R. (2027). Emerging Civil Engineering Strategies for Sustainable Urban Resilience: Smart, Circular, and Adaptive Infrastructure Systems. In *AI-Driven Innovations for Smart and Resilient Infrastructure* (pp. 377-402). IGI Global Scientific Publishing.

Guedes, J., Mesquita, E., & Ferreira, T. M. (2026). Historic Buildings as Urban Sensors: Multi-Scale Diagnostics for Climate-Resilient Cities. *Heritage*, *9*(4), 152.

Xu, J., & Huang, S. (2026). Toward a hybrid machine Learning–physics framework for cohesive soil erosion in climate-resilient civil infrastructure. *Frontiers in Built Environment*, *12*, 1872265.

Ahmad, B., Jillani, S. A., Rafique, M., & Soomro, A. A. (2026, January). Design and Monitoring of a Tree-Inspired Vertical Axis Wind Turbine for Efficient Urban Wind Utilization. In 2026 1st International Conference on Innovations in Information and Communication Technologies (IICT) (pp. 1-6). IEEE.

Hossain, M. K. (2023). A GIS-Based Multi-Criteria Decision-Support and Land-Suitability Optimization Framework for Climate-Resilient Urban and Regional Planning. *American Journal of Data Science and Analytics*, *4*(06), 43-103.

Goh, K. C., Kurniawan, T. A., Meidiana, C., Anouzla, A., Aziz, F., Ali, I., ... & Dissanayake, K. K. (2025). Strengthening urban resilience in China through underground infrastructures management: Addressing global climate challenges with technological solutions. *Deep Underground Science and Engineering*.

Kabir, M. E. (2026). AI-Driven Digital Twin Framework for Predictive Risk Management and Resource Optimization in Large-Scale Civil Infrastructure Projects. *Journal of Sustainable Development and Policy*, *16*(07), 01-37.

Goldberg, K. (2021). Emerging Technologies for Climate Resilience: Integrating Artificial Intelligence, Remote Sensing, and Sustainable Infrastructure. *AJAMRI-American Journal of Advanced Multidisciplinary Research and Innovation*, *3*(2).

Ezugwu, C. S., Godwin, A. N., Attah, C. J., Obukeajeta, O. D., Adetunji, E. O., Ogbonna, C. C., & Ayodeji, A. M. (2026). Resilience and Adaptation in Civil Infrastructure: A Review of Design and Management Strategies for Extreme Weather Events. *Journal of e-Science Letters*.

ANOSIKE, C., ABDULKAREEM, H., & HUSSAINI, A. (2025). Relevance of Meteorological Data in the Integration of Innovative Construction towards Sustainable Civil Infrastructure Development. *Journal of Applied Sciences & Environmental Management*, *29*, 650.

Malifete, L., Mushavhanamadi, K., & Aigbavboa, C. (2026). Mapping the Research Landscape of Sustainable Insurance in Climate-Resilient Smart Cities: A Bibliometric Review. *Sustainability*, *18*(9), 4535.

Islam, F. S. (2025). Artificial intelligence-driven smart waste-to-energy networks for climate-resilient circular resource management in vulnerable megacities. *International Journal of Environment and Climate Change*, *15*(7), 381-415.

Das, A. (2025). Integrated urban water management for climate resilience: strategies for sustainable supply and flood risk mitigation. *Green Technology, Resilience, and Sustainability*, *5*(1), 7.

Hassan, A. N. M., & Hachem-Vermette, C. (2025). A Data-Driven Decision-Making Tool for Prioritizing Resilience Strategies in Cold-Climate Urban Neighborhoods. *Energies*, *18*(20), 5421.

Ige, O. E., & Kabeya, M. (2026). Computational Sustainability for Climate-Resilient Materials: AI-Driven Decarbonization Pathways for Cement and Industrial Systems. *Science, Engineering and Technology*, *6*(2), 131-168.