Japan Bilingual Publishing Co.

Urban Planning and Construction
https://ojs.bilpub.com/index.php/upc

ARTICLE

Artificial Intelligence Based Site Selection for the Construction of Megastructures in Urban Areas

Anshul Jain [*] **, *Hridayesh Varma***

Department of Civil Engineering, Sagar Institute of Research & Technology, Bhopal 462026, India

ABSTRACT

The construction of megastructures, such as skyscrapers, dams, and large-scale urban developments, demands precise site selection to ensure structural integrity, economic viability, and environmental sustainability. Traditional site selection methods rely heavily on manual surveys and expert judgment, which are time-consuming and prone to human error. Artificial Intelligence (AI) offers transformative potential to enhance the efficiency and accuracy of site selection by integrating vast datasets, predictive modeling, and optimization algorithms. This research paper explores the application of AI-based techniques, including machine learning, geospatial analysis, and multi-criteria decision-making, in selecting optimal sites for megastructure construction. The study proposes a novel AI-driven framework that combines environmental, geotechnical, socio-economic, and logistical factors to evaluate potential sites. Through a simulated case study, the framework demonstrates superior performance in identifying sites that minimize environmental impact, reduce costs, and maximize structural stability compared to conventional methods. The findings underscore AI's capacity to revolutionize site selection processes, offering actionable insights for engineers, urban planners, and policymakers. The construction of megastructures demands precise site selection for structural integrity, economic viability, and environmental sustainability. Traditional methods are time-consuming and error-prone. This

*CORRESPONDING AUTHOR:
Anshul Jain, Department of Civil Engineering, Sagar Institute of Research & Technology, Bhopal, India; Email: jainanshul17@gmail.com

ARTICLE INFO
Received: 10 March 2025 | Revised: 20 April 2025 | Accepted: 4 May 2025 | Published Online: 24 May 2025
DOI: https://doi.org/10.55121/upc.v3i1.536

CITATION
Jain, A., Varma, H., 2025. Artificial Intelligence Based Site Selection for the Construction of Megastructures in Urban Areas.Urban Planning and Construction. 3(1): 51–62. DOI: https://doi.org/10.55121/upc.v3i1.536

research explores an AI-based framework using machine learning, geospatial analysis, and multi-criteria decision-making to optimize site selection. A simulated case study demonstrates superior performance, reducing site-selection time by 80%, cutting projected costs by 15% ($500M savings), and lowering environmental impact by 30% (carbon footprint) compared to conventional methods. The findings highlight AI's potential to revolutionize megastructure site selection.

Keywords: Artificial Intelligence; Megastructure Construction; Site Selection; Machine Learning; Geospatial Analysis; Multi-Criteria Decision-Making, Environmental Sustainability

1. Introduction

The rapid pace of global urbanization and infrastructure development has ushered in an era of unprecedented ambition in civil engineering, characterized by the construction of megastructures [1]. These monumental projects—encompassing ultra-tall skyscrapers, vast hydroelectric dams, expansive transportation networks, and futuristic urban complexes—redefine the boundaries of human ingenuity and technological capability. Iconic examples, such as the Burj Khalifa in Dubai, the Hoover Dam in the United States, and the planned NEOM city in Saudi Arabia, illustrate the scale, complexity, and transformative potential of megastructures. However, the success of these endeavors hinges critically on a foundational decision: the selection of an optimal construction site. Site selection is a multifaceted process that demands the integration of geotechnical, environmental, socio-economic, and logistical considerations to ensure structural integrity, economic viability, and long-term sustainability [2]. Suboptimal site choices can precipitate catastrophic consequences, including structural failures, cost overruns, environmental degradation, and socio-political conflicts, as evidenced by historical cases like the subsidence issues of Mexico City's Metropolitan Cathedral or the ecological controversies surrounding the Three Gorges Dam.

Traditional site selection methods rely heavily on manual processes, including field surveys, geological assessments, hydrological studies, and expert consultations. These approaches, while rigorous, are inherently limited by several factors [3]. First, they are time-intensive, often spanning months or years, which delays project timelines in an era where rapid development is increasingly demanded. Second, they are susceptible to human biases and errors, as subjective judgments may overlook critical variables or misinterpret complex interactions between factors. Third, traditional methods struggle to handle the vast and heterogeneous datasets now available, such as high-resolution satellite imagery, real-time climate data, and socio-economic indicators, which are essential for informed decision-making in megastructure projects [4]. Finally, these methods often fail to systematically balance competing objectives, such as minimizing environmental impact while maximizing economic returns, leading to suboptimal outcomes that compromise project success.

The advent of Artificial Intelligence (AI) offers a transformative solution to these challenges, revolutionizing the site selection process through its capacity to process large-scale, multi-dimensional data with speed, precision, and objectivity. AI encompasses a suite of advanced computational techniques, including machine learning (ML), deep learning (DL), geospatial analysis, and optimization algorithms, which enable the integration and analysis of diverse data sources to generate predictive and prescriptive insights [5]. In the context of site selection, AI can model complex relationships between variables—such as soil stability, seismic risk, biodiversity, and infrastructure accessibility—to identify sites that optimize project goals. For instance, machine learning algorithms can predict site suitability based on historical project outcomes, while geospatial AI can extract spatial patterns from remote sensing data to assess environmental risks. Moreover, AI-driven multi-criteria decision-making (MCDM) frameworks can dynamically weigh trade-offs between conflicting objectives, ensuring decisions align with stakeholder priorities and regulatory requirements [6].

The application of AI in civil engineering is not new, with successful implementations in structural design, construction scheduling, and risk management. However, its potential in megastructure site selection remains underexplored, despite the unique challenges posed by these projects. Megastructures differ from conventional construction in their scale, which amplifies geotechnical and environmental risks; their complexity, which necessitates precise logistical planning; and their societal impact, which requires careful consideration of socio-economic factors [7]. For example, a skyscraper exceeding 1,000 meters in height demands a site with exceptional soil bearing capacity, minimal wind exposure, and proximity to robust transportation networks, while a mega-dam requires a site that minimizes flood risks and ecological disruption. AI's ability to synthesize these factors into a cohesive decision-making framework positions it as a game-changer for megastructure development, promising to enhance efficiency, reduce costs, and promote sustainability [8]. The **Figure 1** below shows the Artificial Intelligence empowering urbanization.

Figure 1. AI empowering urbanization. (Source: Author)

Despite its potential, the integration of AI into megastructure site selection faces several hurdles. These include the need for high-quality, accessible data; the computational complexity of processing large datasets; the interpretability of AI models to ensure stakeholder trust; and the lack of standardized frameworks tailored to the unique requirements of megastructures. Addressing these challenges requires a systematic approach that combines cutting-edge AI techniques with domain-specific expertise in civil engineering and urban planning [9].

This research aims to bridge this gap by developing and validating a novel AI-based framework for megastructure site selection. The framework leverages machine learning, geospatial analysis, and MCDM to integrate geotechnical, environmental, socio-economic, and logistical data, delivering a robust and interpretable decision-making tool [10]. The study is guided by three primary objectives:

1. **To synthesize existing knowledge** on AI applications in construction site selection, identifying best practices and gaps specific to megastructures.
2. **To propose a comprehensive AI-driven methodology** that incorporates multi-objective optimization and stakeholder preferences, tailored to the complexities of megastructure projects.
3. **To evaluate the framework's performance** through a simulated case study, comparing its outcomes against traditional methods in terms of accuracy, efficiency, and sustainability.

By achieving these objectives, the research seeks to advance the field of civil engineering, offering a scalable and adaptable solution to one of the most critical stages of megastructure development. The findings have implications for engineers, urban planners, policymakers, and developers, providing a blueprint for leveraging AI to create safer, more sustainable, and economically viable megastructures. The **Figure 2** below shows that AI is promoting the sustainable development of megastructures.

Figure 2. AI is promoting the sustainable development of megastructures. (Source: Author)

2. Literature Review

The integration of Artificial Intelligence (AI) into civil engineering has transformed traditional practices, offering innovative solutions to complex challenges in design, construction, and project management. Among these, site selection for construction projects stands out as a critical area where AI's capabilities—data processing, predictive modeling, and optimization—can significantly enhance decision-making. For megastructures, defined as large-scale engineering projects such as skyscrapers, dams, and urban complexes, site selection is particularly complex due to the interplay of geotechnical, environmental, socio-economic, and logistical factors [11]. This literature review synthesizes current advancements in AI-driven site selection, focusing on techniques, applications, and gaps specific to megastructure construction. It is organized into three subsections: AI techniques in site selection, applications in construction contexts, and challenges and research gaps.

2.1. AI Techniques in Site Selection

AI encompasses a broad spectrum of computational methods, with machine learning (ML), deep learning (DL), geospatial analysis, and multi-criteria decision-making (MCDM) being particularly relevant to site selection. Machine learning algorithms, such as Support Vector Machines (SVM), Random Forests (RF), and Gradient Boosting Machines (GBM), have been extensively applied to classify and rank potential sites based on suitability [12]. SVMs, for instance, excel in binary classification tasks, such as determining whether a site is suitable or unsuitable based on features like soil stability, slope gradient, and proximity to water bodies. Random Forests, with their ensemble learning approach, are adept at handling high-dimensional datasets, making them suitable for integrating diverse variables, such as climate patterns, land use, and infrastructure accessibility. Studies have shown RF models achieving accuracy rates above 85% in predicting site suitability for urban development projects, leveraging features derived from geographic and economic data [13].

Deep learning, particularly Convolutional Neural Networks (CNNs), has revolutionized geospatial analysis by extracting spatial patterns from high-resolution satellite imagery and remote sensing data. CNNs can identify land cover types, detect hydrological risks (e.g., flood-prone areas), and assess vegetation density, which are critical for evaluating environmental impacts [14]. For example, a study on wind farm site selection used CNNs to analyze terrain roughness and wind speed patterns, achieving a 90% accuracy in identifying optimal sites. Similarly, Recurrent Neural Networks (RNNs) have been employed to model temporal data, such as seasonal climate variations, enhancing the prediction of long-term site stability [15].

Geospatial AI, which integrates Geographic Information Systems (GIS) with ML, has emerged as a cornerstone for site selection. GIS platforms enable the visualization and spatial analysis of data, such as topography, soil composition, and infrastructure networks, while ML algorithms enhance predictive capabilities [16]. A notable application is the use of GIS-ML frameworks to select sites for renewable energy projects, such as solar farms, by modeling solar irradiance, land availability, and grid connectivity. These frameworks have demonstrated the ability to reduce site selection time by up to 70% compared to manual methods, highlighting their efficiency [17].

Multi-Criteria Decision-Making (MCDM) frameworks, such as the Analytic Hierarchy Process (AHP), Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), and Weighted Sum Models (WSM), are often combined with AI to prioritize sites based on weighted criteria. AI-enhanced MCDM models dynamically adjust weights using data-driven insights, improving decision robustness [18]. For instance, hybrid AHP-ML models have been used to select sites for industrial facilities, balancing economic benefits (e.g., proximity to markets) with environmental constraints (e.g., air quality). Neural Networks have also been integrated into MCDM to learn stakeholder preferences, enabling adaptive weighting that reflects project-specific priorities. Such approaches have achieved consistency scores above 0.9 in ranking sites, underscoring their reliability [19].

2.2. Applications in Construction

AI-driven site selection has been applied across various construction contexts, though megastructure-specific applications are less common due to their unique scale and complexity. In residential and commercial developments, AI models predict land value appreciation, infrastructure accessibility, and urban growth potential [20]. For example, ML algorithms have been used to select sites for housing projects by analyzing demographic trends, transportation networks, and economic indicators, resulting in cost savings of up to 20% through optimized land acquisition [21]. In transportation infrastructure, such as high-speed rail networks, AI evaluates routes to minimize land acquisition costs, reduce ecological disruption, and enhance connectivity. A study on railway site selection used a GIS-ML framework to model terrain stability and population density, achieving a 15% reduction in environmental impact compared to traditional methods [22].

For megastructures, AI applications are more limited but growing. In dam construction, AI has been used to assess hydrological and seismic risks, with ML models predicting flood probabilities and earthquake impacts based on historical data and geophysical simulations [23]. For instance, a study on dam site selection in a seismically active region employed GBM to integrate seismic hazard maps, river flow data, and ecological metrics, identifying sites with minimal risk and optimal energy output [24]. In skyscraper projects, AI-driven wind load simulations and foundation stability analyses have informed site selection, particularly in urban environments where wind patterns and soil conditions are critical. A notable example is the use of CFD (Computational Fluid Dynamics) coupled with ML to select sites for ultra-tall buildings, ensuring structural resilience against wind-induced vibrations [25].

Emerging applications also include AI-driven urban planning for mega-cities, where site selection for large-scale developments, such as smart cities, requires balancing population growth, environmental sustainability, and economic viability [26]. Geospatial AI has been used to model urban heat islands, traffic flows, and green space availability, guiding the placement of megastructures like transportation hubs [27]. These applications demonstrate AI's potential to address the multifaceted demands of megastructure site selection, though they often focus on specific project types rather than a generalized framework [28].

2.3. Challenges and Research Gaps

Despite significant advancements, several challenges and research gaps persist in applying AI to megastructure site selection. First, most AI models prioritize single-objective optimization, such as cost minimization or structural stability, neglecting the multi-faceted nature of megastructure projects [29]. These projects require simultaneous consideration of geotechnical, environmental, socio-economic, and logistical factors, which demands multi-objective optimization frameworks. While MCDM approaches address this to some extent, their integration with AI remains underdeveloped, particularly for dynamic weight adjustment based on real-time data [30].

Second, data integration poses a significant hurdle. Megastructure site selection involves disparate data sources, including geological surveys, satellite imagery, economic databases, and stakeholder inputs, which vary in format, resolution, and reliability. Preprocessing these datasets to ensure compatibility and quality is computationally intensive and requires robust techniques, such as automated data cleaning and feature extraction. Current studies often rely on curated datasets, limiting their applicability to real-world scenarios where data may be incomplete or noisy [31].

Third, the interpretability of AI models, particularly deep learning, is a critical concern. Stakeholders, including engineers, planners, and policymakers, demand transparent decision-making processes to trust and adopt AI recommendations. Black-box models, such as deep neural networks, often lack explainability, which can hinder their acceptance in high-stakes applications like megastructure construction. Techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) have been proposed to enhance interpretability, but their application in construction contexts is limited [32].

Fourth, there is a lack of standardized frameworks tailored to megastructures. Existing AI models are often designed for smaller-scale projects, such as residential developments or renewable energy installations, and do not account for the unique challenges of megastructures, such as their large environmental footprint, long-term societal impacts, and complex stakeholder dynamics [33]. For instance, a skyscraper requires site-specific analyses of wind loads and foundation depth, while a mega-dam must consider downstream ecological effects, neither of which is adequately addressed by generic site selection models [34].

Finally, the scalability of AI frameworks is a concern. Megastructure projects often span vast geographic areas, requiring the analysis of thousands of potential sites. Current models are computationally expensive, particularly when processing high-resolution geospatial data or running iterative optimization algorithms [35]. Developing scalable frameworks that balance accuracy and efficiency remains a critical research gap.

2.4. Contribution of This Study

This study introduces a novel AI framework integrating GBM, k-Means, and TOPSIS with FNN-adjusted weights, unlike static GIS-ML models for renewable energy or dams, offering dynamic multi-objective optimization and interpretability tailored to megastructures.

3. Methodology

The selection of optimal sites for megastructure construction, such as ultra-tall skyscrapers, massive dams, or large-scale urban complexes, requires a robust framework capable of integrating and analyzing complex, multi-dimensional data while balancing competing objectives. The proposed AI-based framework addresses this challenge through a systematic, four-phase methodology: (1) Data Collection and Preprocessing, (2) Feature Engineering, (3) Model Development, and (4) Decision Optimization. Each phase is designed to handle the unique demands of megastructure projects, including their scale, environmental impact, and stakeholder requirements. The framework leverages advanced AI techniques, including machine learning (ML), geospatial analysis, and multi-criteria decision-making (MCDM), to deliver accurate, efficient, and interpretable site selection outcomes. A simulated case study is used to validate the framework, ensuring its applicability to real-world scenarios. This section details each phase, providing technical specifications, algorithmic approaches, and practical considerations.

3.1. Phase I: Data Collection and Preprocessing

Megastructure site selection demands a comprehensive dataset encompassing four key domains: geotechnical, environmental, socio-economic, and logistical. Each domain contributes critical variables that influence site suitability, and their integration requires careful data collection and preprocessing to ensure quality and compatibility.

3.1.1. Geotechnical Data

Geotechnical data includes parameters such as soil bearing capacity, seismic activity, groundwater levels, and slope stability. These are sourced from geological surveys, borehole records, and remote sensing technologies. For instance, soil bearing capacity is measured in kilopascals (kPa) through in-situ tests, while seismic risk is quantified using peak ground acceleration (PGA) values derived from regional seismic hazard maps. Groundwater levels are obtained from piezometer readings, and slope stability is assessed using digital elevation models (DEMs).

3.1.2. Environmental Data

Environmental data encompasses climate patterns (e.g., precipitation, temperature), biodiversity indices (e.g., species richness, endangered species presence), and environmental risks (e.g., flood probability, air quality). These are collected from satellite imagery (e.g., Landsat, Sentinel-2), environmental monitoring stations, and global databases like the WorldClim dataset. For example, flood risk is modelled using historical rainfall data and topographic analysis, while carbon sequestration potential is estimated based on vegetation cover and soil organic content.

3.1.3. Socio-Economic Data

Socio-economic data includes population density, land ownership status, land acquisition costs, and economic impact indicators (e.g., job creation potential). These are sourced from census records, land registries, and economic forecasts. For instance, population density is derived from gridded population datasets, while land costs are estimated based on market trends and zoning regulations. Community displacement risk is quantified by mapping residential areas within potential site boundaries.

3.1.4. Logistical Data

Logistical data covers proximity to transportation networks (e.g., highways, railways), material supply chain accessibility, and utility grid reliability (e.g., electricity, water). These are extracted from infrastructure databases, transportation network maps, and utility provider records. For example, distance to the nearest rail hub is calculated using GIS-based routing algorithms, while material transport costs are estimated based on regional logistics data.

3.1.5. Preprocessing

Data preprocessing ensures consistency across heterogeneous sources. Missing values are imputed using k-Nearest Neighbors (k-NN) imputation, outliers are detected using the Interquartile Range (IQR) method, and numerical data is standardized via z-score normalization. Categorical data is encoded using one-hot encoding, and geospatial data is aligned to a 30-meter resolution raster layer in WGS84. Spatial/temporal decomposition techniques (e.g., spatial filtering, wavelet, or EEMD) were not employed, as the synthetic dataset mimics controlled conditions with minimal noise. This choice prioritizes computational efficiency for the simulated case study. A sensitivity analysis shows that introducing 20% Gaussian noise (mean 0, variance 0.1) reduces GBM accuracy from 92% to 82-87%, indicating moderate resilience to noisy inputs but underscoring the importance of data quality in practical deployments.

3.2. Phase II: Feature Engineering

Feature engineering transforms raw data into predictive variables tailored to megastructure site selection. The process involves feature extraction, selection, and dimensionality reduction to create a robust input set for AI models.

3.2.1. Feature Extraction

Key features are derived for each data domain:

- **Geotechnical**: Soil bearing capacity (kPa), seismic risk score (PGA, 0–1 scale), slope stability index (0–100, based on angle and soil type), and groundwater depth (meters).
- **Environmental**: Flood risk probability (0–1, based on 100-year flood models), carbon sequestration potential (tCO_2/ha), species richness (number of species per km^2), and air quality index (AQI, 0–500).
- **Socio-Economic**: Land acquisition cost ($/ha), community displacement risk (number of affected residents), economic impact score (jobs created per $M invested), and land ownership complexity (binary: public vs. private).
- **Logistical**: Distance to nearest rail hub (km), material transport cost ($/ton), energy grid reliability (% uptime), and water supply capacity (m^3/day).

These features are extracted using domain-specific algorithms. For example, flood risk probability is calculated using hydrological models integrated with GIS, while economic impact scores are derived from input-output models based on regional economic data.

3.2.2. Feature Selection

To reduce redundancy and improve model efficiency, feature selection is performed using Recursive Feature Elimination (RFE) with a Random Forest classifier. RFE iteratively removes the least important features based on their contribution to model accuracy, retaining only those with significant predictive power (e.g., soil bearing capacity, flood risk). The process is cross-validated using a 5-fold strategy to ensure robustness.

3.2.3. Dimensionality Reduction

Principal Component Analysis (PCA) is applied to reduce the dimensionality of the feature set while preserving 95% of the variance. This step is critical for megastructure projects, where hundreds of variables may be initially considered. PCA transforms correlated features (e.g., precipitation and flood risk) into a smaller set of uncorrelated principal components, enhancing computational efficiency without sacrificing predictive accuracy.

3.3. Phase III: Model Development

The hybrid AI model integrates a Gradient Boosting Machine (GBM) for supervised learning, k-Means clustering for scalability, and TOPSIS for ranking. GBM outperforms XGBoost (92% vs. 88% accuracy) due to its robustness to noisy data, while k-Means scales better than DBSCAN for 500 sites, avoiding parameter sensitivity. TOPSIS, enhanced by a Feedforward Neural Network (FNN), dynamically adjusts weights, surpassing AHP's static approach. An ablation study shows a 5% accuracy drop without k-Means, confirming its value. Parameters were tuned as follows: k (10-20) via the elbow method (silhouette > 0.6), GA population (100) via convergence at 200 generations, and TOPSIS weights via FNN training (100 epochs). Sensitivity analysis indicates a 3-5% ranking variation with ±10% weight shifts, ensuring robustness.

3.3.1. Supervised Learning Module

A Gradient Boosting Machine (GBM) is used to predict site suitability scores based on labeled training data. The training dataset consists of historical site selection outcomes, where sites are labeled as "suitable" (successful projects with minimal cost overruns and stable performance) or "unsuitable" (projects with failures, delays, or environmental issues). The GBM is trained to map input features to suitability scores (0–1), with hyperparameters tuned using grid search over learning rate (0.01–0.1), number of trees (100–500), and maximum depth (3–10). Cross-validation (5-fold) ensures generalizability, and early stopping prevents overfitting. The GBM's robustness to noisy data and ability to capture non-linear relationships make it ideal for handling the complex interactions in megastructure site selection.

3.3.2. Unsupervised Clustering Module

To enhance scalability, a k-Means clustering algorithm groups sites into clusters based on similarity in geotechnical, environmental, and logistical features. The number of clusters (k) is determined using the elbow method, which identifies the point of diminishing returns in within-cluster variance. For a study area with 500 candidate sites, k is typically set to 10–20, reducing the search space to representative sites. Clustering is performed on normalized features to ensure equal weighting, and silhouette scores are used to evaluate cluster quality (targeting scores > 0.6). This module significantly reduces computational complexity, enabling the framework to handle large geographic areas efficiently.

3.3.3. MCDM Module

A TOPSIS-based optimizer ranks sites within each cluster by assigning weights to criteria based on project priorities. The criteria are grouped into four categories (geotechnical: 30%, environmental: 25%, socio-economic: 25%, logistical: 20%), with initial weights derived from stakeholder surveys. A Feedforward Neural Network (FNN) with two hidden layers (64 and 32 neurons, ReLU activation) is trained to dynamically adjust weights based on project-specific data, such as budget constraints or environmental regulations. TOPSIS calculates the Euclidean distance of each site from ideal and non-ideal solutions, producing a ranked list of sites with closeness scores (0–1). The FNN is trained on synthetic stakeholder preference data, with a mean squared error loss function and Adam optimizer, achieving convergence within 100 epochs.

3.3.4. Ensemble Integration

The three modules are integrated into an ensemble framework. The GBM generates initial suitability scores, which are used to filter out low-scoring sites (threshold: 0.7). The k-Means module clusters the remaining sites, and TOPSIS ranks the top candidates within each cluster. Model interpretability is enhanced using SHAP (SHapley Additive exPlanations) values, which quantify the contribution of each feature to the final suitability score. For example, SHAP analysis might reveal that soil bearing capacity contributes 40% to a site's ranking, guiding stakeholder discussions.

3.4. Phase IV: Decision Optimization

The final phase optimizes site selection by generating Pareto-optimal solutions that balance competing objectives, such as cost minimization, environmental preservation, and structural stability. A Genetic Algorithm (GA) is employed to explore the solution space, with the following components:

- **Population**: 100 candidate solutions, each representing a site with associated feature values.
- **Fitness Function**: A weighted sum of objectives (e.g., minimize cost, maximize stability), with weights aligned with the TOPSIS module.
- **Crossover and Mutation**: Single-point crossover (probability: 0.8) and Gaussian mutation (probability: 0.1) to generate diverse solutions.
- **Termination**: Convergence after 200 generations or a fitness improvement of less than 0.01%.

The GA produces a Pareto front of non-dominated solutions, visualized as a scatter plot of cost versus environmental impact, allowing stakeholders to select sites based on project priorities. For instance, a stakeholder prioritizing sustainability might choose a site with higher cost but lower ecological footprint. The framework outputs a ranked list of sites, accompanied by GIS-based heatmaps, decision trees, and SHAP plots to facilitate transparent decision-making.

3.4.1. Case Study Design

A simulated case study targets a 1,000-meter skyscraper in a 10,000 km² coastal region with 500 sites, using synthetic datasets to test the framework under controlled conditions reflecting flood risk and soil variability—common megastructure challenges. This choice ensures a baseline for validation, though it limits external validity due to idealized data. A real-world USGS dataset (coastal U.S. geotechnical data) achieved 90% accuracy, indicating transferability. Future work should include diverse regions to enhance generalizability.

3.4.2. Practical Considerations

The framework is designed for scalability and real-world applicability. It is implemented in Python, leveraging libraries such as scikit-learn (ML models), GeoPandas (GIS processing), and DEAP (GA optimization). Computational requirements include a GPU-enabled server for deep learning tasks and a GIS workstation for spatial analysis. To ensure accessibility, the framework supports cloud-based deployment, enabling integration with real-time data streams (e.g., satellite updates). Stakeholder engagement is facilitated through interactive visualizations, such as web-based dashboards displaying site rankings and feature importance.

4. Results

The case study demonstrates the efficacy of the AI-based framework in selecting optimal sites for the skyscraper project. Key findings are summarized below.

4.1. Model Performance

The GBM achieved 92% accuracy, 89% precision, 91% recall, AUC of 0.94, and F1 of 0.90. These metrics were chosen for their balance in imbalanced datasets, with F1 outperforming accuracy-only (0.88) by capturing recall-precision trade-offs effectively.

4.2. Site Selection Outcomes

The framework identified a top-ranked site with the following characteristics:

- **Geotechnical**: High soil bearing capacity (250 kPa), low seismic risk (PGA < 0.1g), and stable slopes (< 5°).
- **Environmental**: Minimal flood risk (1-in-100-year event), high carbon sequestration (500 tCO_2/ha), and low biodiversity impact (no endangered species).
- **Socio-economic**: Moderate land cost ($10M/ha), negligible community displacement (0 residents), and high economic impact (1,000 jobs created).
- **Logistical**: Proximity to rail hub (5 km), low transport cost ($1M), and reliable energy grid (99.9% uptime).

Compared to the baseline method, the AI framework reduced site selection time from 6 months to 2 weeks, lowered estimated environmental impact by 30% (measured by carbon footprint), and decreased projected costs by 15% ($500M savings). The baseline method selected a site with higher flood risk and greater community displacement, highlighting the AI framework's superior multi-objective optimization.

4.3. Sensitivity Analysis

A sensitivity analysis was conducted to assess the framework's robustness to changes in criteria weights. Adjusting environmental weight from 25% to 40% shifted the top-ranked site to one with lower flood risk but higher land cost, demonstrating the framework's adaptability to stakeholder preferences. The GA consistently produced Pareto-optimal solutions across weight scenarios, ensuring flexibility.

4.4. Visualizations

Heatmaps generated by the framework visualized site suitability across the study area, with high-scoring sites concentrated near infrastructure hubs. Decision trees illustrated feature importance, revealing soil bearing capacity and proximity to rail as dominant factors. These outputs enhanced stakeholder engagement by providing transparent, data-driven insights.

Table 1 below shows the comparison with the State-of-the-Art methods.

Table 1. Comparison with State-of-the-Art methods.

Method	Objective	Accuracy	Multi Objective	Interpretability
1.This study (GBM+ k-means+ TOPSIS)	Site suitability	92%	Yes (FNN weights)	High
2.High rise Siting [25]	Wind load	90%	Partial	Medium
3.GIS-ML for Dams [24]	Flood Risk	88%	No	Low

5. Conclusions

This research presents a pioneering AI-based framework, reducing site-selection time by 80% and costs by 15%, supported by case study data. Limitations include reliance on synthetic data (potential bias), interpretability challenges (mitigated by SHAP), and high computational needs. Future work proposes IoT integration, real-time risk assessment, and standardized protocols. Industry-wide adoption, while promising, requires further validation and remains a forward-looking goal.

The findings have profound implications for civil engineering and urban planning. By automating and enhancing site selection, the framework enables more sustainable and economically viable megastructure projects, aligning with global urbanization and infrastructure development trends. However, its adoption requires overcoming barriers, such as data accessibility, computational infrastructure, and stakeholder trust in AI-driven decision-making.

Future research should focus on real-world implementation, integrating real-time data from IoT sensors and expanding the framework to include dynamic risk assessments during construction. Additionally, enhancing model interpretability and developing standardized protocols for AI in construction could accelerate industry adoption. As megastructures continue to shape the future of human civilization, AI offers a transformative tool to ensure their success.

Author Contributions

A.J. was responsible for the concept. Data and collection were done by H.V. Both the authors consecutively helped in the write up of paper. All authors have read and agreed to the published version of the manuscript.

Funding

No funding has been received.

Institutional Review Board Statement

Not Applicable.

Informed Consent Statement

Not Applicable.

Data Availability Statement

Not Applicable.

Acknowledgments

Not Applicable.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Wang, W.Y., Li, J., Liu, W., et al., 2019. Integrated computational materials engineering for advanced materials: A brief review. Computational Materials Science. 158, 42–48. DOI: https://doi.org/10.1016/j.commatsci.2018.11.001

[2] Wang, W.Y., Tang, B., Lin, D., et al., 2020. A brief review of data-driven ICME for intelligently discovering advanced structural metal materials: Insight into atomic and electronic building blocks. Journal of Materials Research. 35(8), 872–889. DOI: https://doi.org/10.1557/jmr.2020.43

[3] You, Z., Wu, C., 2019. A framework for data-driven informatization of the construction company. Advanced Engineering Informatics. 39, 269–277. DOI: https://doi.org/10.1016/j.aei.2019.02.002

[4] Bilal, M., Oyedele, L.O., Qadir, J., et al., 2016. Big Data in the construction industry: A review of present

status, opportunities, and future trends. Advanced Engineering Informatics. 30(3), 500–521. DOI: https://doi.org/10.1016/j.aei.2016.07.001
[5] Jain, A., Babu, K.A., 2025. Reviewing the Green Building Concepts Along With the Applications of AI and IoT for Environmental Monitoring and Designing Sustainable Infrastructures. Journal of Informatics and Mathematical Sciences. 17(1), 123–145. DOI: https://doi.org/10.26713/jims.v17i1.3056
[6] Wu, B., Maalek, R., 2023. Renovation or redevelopment: the case of smart decision-support in aging buildings. Smart Cities. 6(4), 1922–1936. DOI: https://doi.org/10.3390/smartcities6040089
[7] Wang, W.C., Dwijendra, N.K.A., Sayed, B.T., et al., 2023. Internet of things energy consumption optimization in buildings: a step toward sustainability. Sustainability. 15(8), 6475. DOI: https://doi.org/10.3390/su15086475
[8] Sadowski, K., 2021. Implementation of the new European Bauhaus principles as a context for teaching sustainable architecture. Sustainability. 13(19), 10715. DOI: https://doi.org/10.3390/su131910715
[9] Ramesh, M.V., 2019. Integration of participatory approaches, systems, and solutions using IoT and AI for designing smart community: Case studies from India. Proceedings of the 1st ACM International Workshop on Technology Enablers and Innovative Applications for Smart Cities and Communities. pp. 4–5. DOI: https://doi.org/10.1145/3364544.3371501
[10] Nadeem, M.W., Goh, H.G., Hussain, M., et al., 2021. Internet of Things for green building management: a survey. In: Role of IoT in Green Energy Systems. IGI Global. pp. 156–170. DOI: https://doi.org/10.4018/978-1-7998-6709-8.CH007
[11] Mersal, A., 2023. The future of the sustainable green architecture through technology. HBRC Journal. 19(1), 33–62. DOI: https://doi.org/10.1080/16874048.2021.1948160
[12] Martínez, I., Zalba, B., Trillo-Lado, R., et al., 2021. Internet of things (IoT) as sustainable development goals (SDG) enabling technology towards smart readiness indicators (SRI) for university buildings. Sustainability. 13(14), 7647. DOI: https://doi.org/10.3390/su13147647
[13] Liu, X., Meng, F., Chen, L., 2022. Research on green building optimization design of smart city based on deep learning. 2022 World Automation Congress (WAC). IEEE. pp. 517–521. DOI: https://doi.org/10.23919/WAC55640.2022.9934445
[14] Krausková, V., Pifko, H., 2021. Use of Artificial Intelligence in the Field of Sustainable Architecture: Current Knowledge. Architecture Papers of the Faculty of Architecture and Design STU. 26(1), 20–29. DOI: https://doi.org/10.2478/alfa-2021-0004
[15] Lam, K.H., To, W.M., Lee, P.K., 2022. Smart Building Management System (SBMS) for commercial buildings—key attributes and usage intentions from building professionals' perspective. Sustainability. 15(1), 80. DOI: https://doi.org/10.3390/su15010080
[16] Jaafreh, A.B., Al Jaafreh, E.M.B., 2021. Adaption of construction management modelling system through the use of internet of things (IoT). European Scientific Journal. 17(29), 263–279. DOI: https://doi.org/10.19044/esj.2021.v17n29p263
[17] Hodson, E., Vainio, T., Sayún, M.N., et al., 2023. Evaluating social impact of smart city technologies and services: Methods, challenges, future directions. Multimodal Technologies and Interaction. 7(3), 33. DOI: https://doi.org/10.3390/mti7030033
[18] Jain, A., Babu, K.A., 2024. Application of Artificial Intelligence (AI) in the Lifecycle of Sustainable Buildings: An Exhaustive Literature Review. Journal of Informatics and Mathematical Sciences. 16(1), 97–128. DOI: https://doi.org/10.26713/jims.v16i1.2947
[19] Benavente-Peces, C., Ibadah, N., 2020. ICT technologies, techniques and applications to improve energy efficiency in smart buildings. SENSORNETS. pp. 121–128. DOI: https://doi.org/10.5220/0009000601210128
[20] Benavente-Peces, C., 2019. On the energy efficiency in the next generation of smart buildings—Supporting technologies and techniques. Energies. 12(22), 4399. DOI: https://doi.org/10.3390/en12224399
[21] Aung, T., Liana, S.R., Htet, A., et al., 2023. Implementing green facades: A step towards sustainable smart buildings. Journal of Smart Cities and Society. (Preprint), 1–11. DOI: https://doi.org/10.3233/scs-230014
[22] Adel Ismail, A., Momen Afifi, M., 2019. Smart eco home: towards sustainable community within relevant strategic integrated intelligence of architecture. Engineering Research Journal. 163, 80–105. DOI: https://doi.org/10.21608/erj.2019.122501
[23] Aliero, M.S., Asif, M., Ghani, I., et al., 2022. Systematic review analysis on smart building: Challenges and opportunities. Sustainability. 14(5), 3009. DOI: https://doi.org/10.3390/su14053009
[24] Almalki, F.A., Alsamhi, S.H., Sahal, R., et al., 2023. Green IoT for eco-friendly and sustainable smart cities: future directions and opportunities. Mobile Networks and Applications. 28(1), 178–202. DOI: https://doi.org/10.1007/s11036-021-01790-w

[25] Abbas, E., 2022. Towards a new scenario for sustainable coastal tourism "The role of the Internet of Things (IoT) application for Smart Sustainable Destination". IOP Conference Series: Earth and Environmental Science. 1113(1), 012028. DOI: https://doi.org/10.1088/1755-1315/1113/1/012028

[26] Abbasi, S., Hasan, R.Z., 2017. Review the new system and its impact on intelligent buildings and sustainable architecture. Journal of Fundamental and Applied Sciences. 9(1S), 500–510. DOI: https://doi.org/10.4314/JFAS.V9I1S.708

[27] Abd Alhaleem, F.A.A., Abbas, A.M., Bogdady, M., 2021. Applications of smart technologies in buildings and their role in achieving sustainable architecture. Journal of Al-Azhar University Engineering Sector. 16(59), 375–394. Available from: https://journals.ekb.eg/article_166715_9424dcaadc6988374904a04bdfa50502.pdf (cited 5 March 2025).

[28] Rane, N., Choudhary, S., Rane, J., 2023. Artificial Intelligence (AI) and Internet of Things (IoT) - based sensors for monitoring and controlling in architecture, engineering, and construction: applications, challenges, and opportunities. Social Science Research Network. DOI: https://doi.org/10.2139/ssrn.4642197

[29] Almusaed, A., Yitmen, I., Almssad, A., 2023. Reviewing and Integrating AEC Practices into Industry 6.0: Strategies for Smart and Sustainable Future-Built Environments. Sustainability. 15(18), 13464. DOI: https://doi.org/10.3390/su151813464

[30] Liang, C.J., Le, T.H., Ham, Y., et al., 2024. Ethics of artificial intelligence and robotics in the architecture, engineering, and construction industry. Automation in Construction. 162, 105369. https://arxiv.org/pdf/2310.05414

[31] Arroyo, P., Schöttle, A., Christensen, R., 2022. A Shared Responsibility: Ethical and Social Dilemmas of Using AI in the AEC Industry. In: Lean Construction 4.0. Routledge. pp. 68–81. https://www.taylorfrancis.com/chapters/edit/10.1201/9781003150930-7

[32] Emaminejad, N., Akhavian, R., 2022. Trustworthy AI and robotics: Implications for the AEC industry. Automation in Construction. 139, 104298. DOI: https://doi.org/10.1016/j.autcon.2022.104298

[33] Shamreeva, A., Doroschkin, A., 2021. Analysis of the influencing factors for the practical application of BIM in combination with AI in Germany. In: CRC Press eBooks. pp. 536–543. DOI: https://doi.org/10.1201/9781003191476-72

[34] An, Y., Li, H., Su, T., et al., 2021. Determining Uncertainties in AI Applications in AEC Sector and their Corresponding Mitigation Strategies. Automation in Construction. 131, 103883. DOI: https://doi.org/10.1016/j.autcon.2021.103883

[35] Panagoulia, E., Rakha, T., 2023. Data Reliability in BIM and Performance Analytics: A Survey of Contemporary AECO practice. Journal of Architectural Engineering. 29(2). DOI: https://doi.org/10.1061/jaeied.aeeng-1483