sustainability

MDPI

Article

An Integrated Financial–Sustainability Framework for Predicting Green Infrastructure Project Success

Ahmad A. Tareemi

Department of Civil and Environmental Engineering, Umm Al-Qura University, P.O. Box 715, Mecca 24382, Saudi Arabia; aatareemi@uqu.edu.sa

Abstract

To overcome the inadequacy of traditional financial metrics in appraising green infrastructure, this study develops and validates an integrated framework combining financial and sustainability indicators to more accurately predict project performance. Employing a mixed-methods design, this study synthesized metrics from expert interviews (N = 24) and literature, then collected data from 42 completed projects in Gulf Cooperation Council countries. The framework's predictive validity was tested using a novel application of a Gradient Boosting Machine (XGBoost) model, with SHAP (SHapley Additive exPlanations) analysis ensuring model interpretability. The integrated framework yielded higher out-of-sample discriminatory performance (AUC-ROC = 0.88) than a baseline using only traditional metrics (AUC-ROC = 0.71). In SHAP analyses, RBCR and LCC contributed most to the model's predictions, whereas NPV and IRR contributed least. These results indicate stronger predictive associations for sustainability-oriented metrics in this study's model. Because the design is cross-sectional and predictive, all findings are associational rather than causal; residual confounding is possible. The validated, interpretable model is therefore positioned as a decision support tool that complements, rather than replaces, expert appraisal.

Keywords: green infrastructure; investment appraisal; machine learning; project performance; sustainability metrics

check for updates

Academic Editor: Giovanni Leonardi

Received: 11 August 2025
Revised: 13 October 2025
Accepted: 17 October 2025
Published: 27 October 2025

Citation: Tareemi, A.A. An Integrated Financial–Sustainability Framework for Predicting Green Infrastructure Project Success. *Sustainability* **2025**, *17*, 9529. https://doi.org/10.3390/su17219529

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The global commitment to sustainable development, driven by the escalating impacts of climate change and resource depletion, has positioned green infrastructure as a cornerstone of modern urban and economic planning [1,2]. These projects—spanning renewable energy installations, sustainable water management systems, urban green spaces, and climate-resilient transportation networks—are fundamental to enhancing societal resilience, promoting environmental health, and ensuring long-term economic stability [3]. In arid and rapidly urbanizing regions, such as the Kingdom of Saudi Arabia, the strategic implementation of green infrastructure is not merely beneficial but essential for sustainable diversification and growth [4,5]. Ambitious national roadmaps, including Saudi Arabia's Vision 2030 and the Saudi Green Initiative, underscore this imperative, allocating substantial capital towards transformative projects like the King Salman Park in Riyadh, NEOM's net-zero industrial hub, and various eco-tourism developments along the Red Sea coast [6,7]. These initiatives reflect a broader trend across the Gulf Cooperation Council (GCC) countries, which are increasingly leveraging their sovereign wealth funds and issuing green bonds

to finance sustainable development and transition away from hydrocarbon-dependent economies [8].

Despite this momentum, the financial appraisal of green infrastructure projects presents a significant challenge that threatens to impede optimal capital allocation. The inherent characteristics of these projects—long operational lifespans, the generation of non-market environmental and social co-benefits, and exposure to long-term climate uncertainties—are not adequately captured by traditional project finance frameworks [9,10]. Methodologies centered on metrics such as Net Present Value (NPV), Internal Rate of Return (IRR), and simple payback periods are widely acknowledged to be insufficient [11]. Such conventional tools often fail to monetize long-term externalities, assume constant discount rates that undervalue future benefits, and overlook the crucial aspect of climate resilience, leading to a systematic bias against sustainable investments [12]. This methodological inadequacy creates a critical barrier, as projects with profound long-term societal value may appear financially unviable, while those with lower upfront costs but higher life-cycle and environmental liabilities may be incorrectly prioritized. Addressing this valuation gap is paramount to ensuring that the immense financial resources being mobilized are directed toward projects that deliver genuine, lasting, and holistic performance.

In response to the limitations of conventional financial metrics, the academic and professional literature has seen a proliferation of research dedicated to developing more holistic appraisal methodologies for infrastructure projects [13–15]. This body of work has evolved along two primary trajectories: the integration of sustainability-oriented metrics and the application of advanced analytical techniques for performance prediction.

The first trajectory involves augmenting or replacing traditional financial analysis with metrics that account for the entire project lifespan and its external impacts. Life Cycle Costing (LCC) has emerged as a foundational approach, extending analysis beyond initial capital outlay to include all costs "from cradle to grave," such as operation, maintenance, and disposal [16]. While LCC provides a more comprehensive economic picture, it does not inherently capture non-monetized environmental or social benefits [17]. To address this, researchers have focused on monetizing externalities, most notably through the concept of the Social Cost of Carbon (SCC). The SCC assigns a dollar value to the long-term damage caused by emitting one additional ton of carbon dioxide, thereby allowing climate impacts to be integrated into cost–benefit analyses [18,19]. However, the SCC is subject to significant debate and variation, with estimates fluctuating based on the chosen discount rate and the time horizon of the underlying assessment models [20]. Concurrently, frameworks for Resilience Benefit–Cost Analysis have been developed to quantify the value of a project's ability to withstand and recover from shocks and stresses, such as those induced by climate change [21]. Studies show that investing in resilience yields substantial returns, with benefit–cost ratios often exceeding 4:1 [22].

The second trajectory has focused on leveraging computational methods to improve the prediction of project success. Recognizing the complexity and multi-faceted nature of infrastructure projects, researchers have moved beyond traditional regression models toward more sophisticated machine learning (ML) techniques [23]. Supervised ML models, including support vector machines, random forests, and neural networks, have been applied to forecast project outcomes based on a wide range of input variables [24]. These models are adept at identifying complex, non-linear relationships within large datasets that traditional statistical methods may miss. More recently, there has been a significant emphasis on model interpretability to overcome the "black box" nature of many ML algorithms [25]. The development of SHAP (SHapley Additive exPlanations), a game theory-based approach, has been particularly influential, as it allows for the quantification of each feature's contribution to a specific prediction [26,27]. This enables a transparent

understanding of why a model makes a certain forecast, which is crucial for high-stakes investment decisions.

Table 1 provides a comparative summary of key literature investigated, highlighting the methodologies employed and the gaps they leave unaddressed, which the present study aims to fill.

Table 1. Comparative Framework of Investigated Literature in Infrastructure Project Appraisal and Prediction.

Author(s) and Year	Focus of Study	Methodology	Key Findings	Identified Limitations/Gap
[28]	Forecasting the success of private participation infrastructure projects globally.	Application of multiple machine learning classifiers (e.g., k-NN, Random Forest) on a World Bank dataset.	Predictive accuracy varies significantly by project sector and region; decision tree-based classifiers performed best overall.	Focuses on traditional success/failure outcomes without integrating specific financial or sustainability performance metrics as predictors.
[29]	Cost–benefit analysis of making new infrastructure assets resilient to natural hazards.	Probabilistic cost–benefit analysis across 3000 scenarios.	Investing in resilience is highly robust, with a benefit–cost ratio greater than four in 50% of scenarios.	Provides a strong case for resilience but does not integrate these benefits into a comprehensive project appraisal framework alongside other financial metrics.
[30]	Updating the scientific basis for the Social Cost of Carbon (SCC).	Integrated assessment modeling responsive to National Academies of Sciences, Engineering, and Medicine (NASEM) recommendations.	Proposes a revised central SCC estimate of $185 per ton, 3.6 times higher than the previous U.S. government figure.	Focuses solely on the valuation of carbon emissions, not its integration or predictive power within a broader project success model.
[31]	Establishing the principles and application of Life Cycle Costing (LCC).	Conceptual review and framework development.	LCC provides a more complete view of project costs over its entire lifespan compared to initial capital cost analysis.	Primarily focused on economic costs; does not inherently incorporate environmental or social externalities or predict overall project success.
[27,32]	Development of SHAP (SHapley Additive exPlanations) for model interpretation.	Game theory-based additive feature attribution method.	Provides a unified and theoretically sound framework for explaining the output of any machine learning model.	A methodological tool; its application to validate an integrated financial-sustainability framework for infrastructure is novel.
[33]	Identifying challenges and barriers to green finance.	Literature review and expert interviews.	Key barriers include lack of standardized metrics, lack of transparent data, and perceived higher risks.	Highlights the problem but does not propose or empirically test a solution in the form of a validated, integrated metrics framework.

A clear gap separates the widely acknowledged need for holistic appraisal and the availability of empirically validated tools. Existing efforts typically treat sustainability metrics (e.g., LCC, SCC, resilience benefits) and traditional finance in parallel rather than in an integrated framework, and they rarely test whether such integration improves prediction of realized project performance. This study addresses that gap by fusing financial and sustainability indicators into a single, testable framework and empirically examining its predictive power for green infrastructure outcomes.

To clarify sustainable-development value in internationally comparable terms, this study's six-metric framework maps directly to priority SDG targets. Life-Cycle Costing (LCC) operationalizes resource-efficiency upgrading in infrastructure (SDG 9.4) and underpins whole-life stewardship central to SDG 11.6 and 11.b; the monetized carbon cost embeds climate-damage externalities, supporting SDG 13.2 (integration of climate mea-

sures into decision processes); the Resilience Benefit–Cost Ratio (RBCR) quantifies adaptive capacity to climate hazards (SDG 13.1) and disaster-loss reduction in cities (SDG 11.5). Traditional finance metrics (NPV, IRR, Payback) remain necessary for bankability (SDG 9.1) but, as shown later, are insufficient predictors of realized multi-dimensional performance in this work. This crosswalk provides an explicit policy bridge from predictive metrics to internationally recognized targets without altering the empirical design.

This manuscript directly addresses this deficiency. Previous research has not systematically developed and validated an integrated metrics framework that combines the strengths of conventional finance with the necessities of sustainability appraisal. To the author's knowledge, no study has employed an interpretable machine learning model to test whether such a framework offers superior predictive power over traditional models in a real-world setting. The problem, therefore, is the lack of a robust, empirically backed, and interpretable decision support model that can accurately appraise green infrastructure investments by accounting for the complex interplay between financial, environmental, and resilience factors. This gap leaves investors and policymakers to rely on either outdated, incomplete models or a fragmented collection of un-integrated, and often theoretical, sustainability metrics.

This study proceeds from the premise that green infrastructure appraisal must reflect multi-dimensional value. Project success emerges from the interaction of financial prudence, whole-life cost efficiency, environmental externalities, and climate resilience—dimensions that are not captured by single-metric, discounted-cash-flow analysis. An integrated, data-driven framework is therefore required to represent these interactions and to test their predictive relevance for realized outcomes.

The methodological choices are deliberately designed to address the identified research gap with scientific rigor and practical relevance. The use of a multi-phase, mixed-methods design—beginning with literature synthesis and expert interviews—ensures that the selected metrics are both theoretically grounded and validated by industry practice. The selection of a Gradient Boosting Machine (XGBoost) model [34] is justified by its state-of-the-art predictive performance, particularly its ability to model complex, non-linear, and interactive effects among predictors, which are characteristic of project finance data. However, predictive power alone is insufficient for a decision support tool that must instill confidence in stakeholders. Therefore, the application of SHAP is a crucial methodological innovation in this context. It directly confronts the "black box" problem in machine learning by providing transparent, intuitive, and theoretically sound explanations for each prediction. This interpretability is essential for the framework to be adopted by project managers, investors, and policymakers, as it allows them to understand why a project is predicted to be successful and which factors are most influential. This approach advances significantly beyond prior studies that either used less powerful statistical methods or employed "black box" models without providing crucial interpretability.

The primary aim of this study is to develop and empirically validate an integrated financial-sustainability metrics framework that enhances the appraisal and predicts the performance of green infrastructure projects.

To achieve this aim, the following specific objectives are established:

1. To identify and synthesize a comprehensive set of critical financial and sustainability metrics for green infrastructure appraisal through a systematic literature review and semi-structured interviews with industry and policy experts.
2. To construct a quantifiable project performance index based on empirical data from recently completed green infrastructure projects within the GCC region, operationalized through measures of budget adherence, schedule performance, and stakeholder satisfaction.

3. To develop and validate a predictive model using a XGBoost) to assess the efficacy of the integrated metrics framework in forecasting project success, and to benchmark its performance against a baseline model using only traditional financial metrics.
4. To employ SHAP analysis to interpret the predictive model, identify the most influential metrics driving project performance, and provide transparent insights into the decision-making logic of the framework.

The remainder of this paper presents the research methodology, findings, and implications. Section 2 details the mixed-methods design, including expert interviews, quantitative data collection from 42 completed projects, and the operationalization of the integrated metrics and the Project Success Index (PSI), together with the XGBoost–SHAP analytical framework. Section 3 reports the empirical results. Section 4 offers a concise, integrative discussion that situates the findings within prior work, identifies limitations, and outlines future research. Section 5 concludes by synthesizing the principal insights and their implications for sustainable infrastructure investment.

2. Materials and Methods

2.1. Research Design and Workflow

This study employed a multi-phase, sequential mixed-methods research design to develop and validate the integrated financial-sustainability metrics framework. This approach was selected for its capacity to combine the depth of qualitative inquiry with the breadth and statistical power of quantitative analysis, thereby ensuring the resulting framework is both contextually rich and empirically robust. The research was structured into three distinct, sequential phases: (1) qualitative framework development through systematic literature synthesis and expert interviews; (2) quantitative data collection via a cross-sectional survey and in-depth case studies of green infrastructure projects; and (3) analytical validation using interpretable machine learning. The overall workflow of the research methodology is depicted in Figure 1. This structured process ensures that the identification of metrics, the collection of data, and the analytical validation are logically sequenced, with the outputs of each phase providing the necessary inputs for the subsequent phase, culminating in a rigorously validated and interpretable predictive model.

To avoid causal over-interpretation, this study treats all model outputs as predictive associations. A directed acyclic graph (DAG) [35] specifying assumed relationships among appraisal metrics, unmeasured contextual factors (e.g., governance capacity, macroeconomic shocks), and the PSI is provided in Supplementary Figure S1. The DAG clarifies why confounding cannot be ruled out and motivates the emphasis on predictive validity rather than causal claims.

The research process began with a systematic exploration of existing literature and expert knowledge to ensure the initial set of metrics was exhaustive and decision-relevant. This qualitative foundation informed a structured survey instrument deployed to gather empirical data across a broad project sample. A purposive subset was then selected for in-depth case analysis to triangulate survey responses, verify data accuracy, and probe causal mechanisms that quantitative data alone cannot reveal. The final analytical phase applied advanced computational techniques to test the central hypothesis: that an integrated framework offers superior predictive power over traditional methods. Combining a non-linear gradient-boosting model (XGBoost) with SHAP interpretability ensured that validation results were transparent and actionable for practitioners and policymakers.

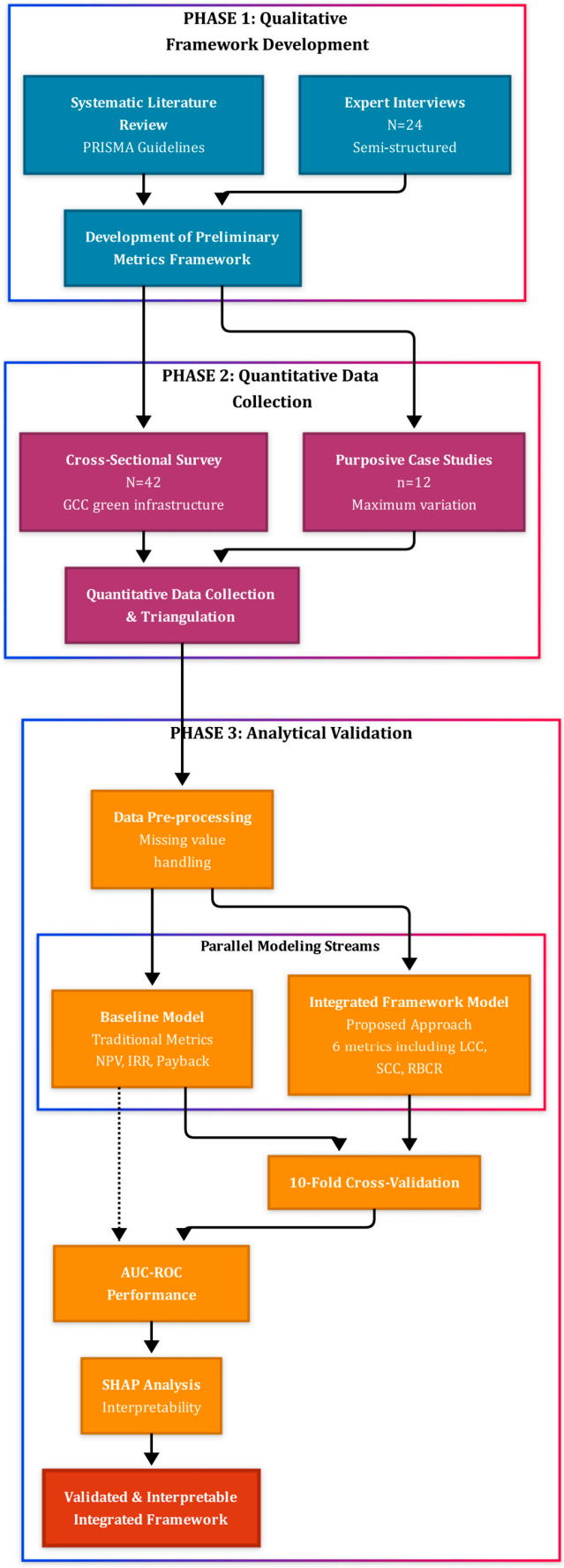


Figure 1. Sequential three-phase mixed-methods research design for developing and validating the integrated financial-sustainability metrics framework.

2.2. Phase 1: Framework Development and Metric Identification

2.2.1. Systematic Literature Synthesis

To establish a comprehensive theoretical foundation, a systematic literature review was conducted following the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines [36]. The search was performed across multiple academic databases, including Scopus, Web of Science, and the ASCE Library, to identify peer-reviewed articles published between January 2010 and December 2024. The search strategy employed a combination of keywords and Boolean operators, structured around three core concepts: ((“green infrastructure” OR “sustainable projects” OR “renewable energy projects”) AND (“project finance” OR “investment appraisal” OR “financial metrics” OR “cost–benefit analysis”) AND (“project success” OR “project performance” OR “performance prediction”)).

Inclusion criteria were established to select studies that (a) were published in English in peer-reviewed journals or major conference proceedings; (b) focused on the financial, economic, or performance assessment of infrastructure projects with a clear sustainability or green component; and (c) proposed, tested, or discussed specific metrics for project appraisal. Exclusion criteria were applied to filter out articles that were purely descriptive, lacked methodological rigor, or focused on macro-level policy without addressing project-level metrics. The initial search yielded over 1500 articles. After removing duplicates and screening titles and abstracts, 210 articles were selected for full-text review. From this set, 85 articles were deemed directly relevant and were synthesized to extract key theoretical frameworks, established methodologies (e.g., LCC, SCC), and identified research gaps. This process formed the basis for the preliminary set of metrics for the integrated framework.

2.2.2. Expert Interviews

To complement the literature synthesis and ensure the practical relevance of the selected metrics, semi-structured interviews were conducted with 24 industry and policy experts. Participants were selected using a purposive sampling strategy, a non-probability technique where individuals are chosen based on their specific knowledge and experience relevant to the research objectives. The sample was stratified to ensure diverse representation and included senior project managers (n = 8), financial analysts specializing in infrastructure (n = 6), government officials from entities like the Saudi Ministry of Environment, Water and Agriculture and the Public Investment Fund (PIF) (n = 5), and academics with expertise in sustainable construction (n = 5). All participants had a minimum of 10 years of experience with large-scale infrastructure projects in the GCC region.

An interview protocol was developed to guide the discussions, covering topics such as the limitations of current appraisal practices, the perceived importance of various sustainability metrics, and the practical challenges of implementation. Interviews were conducted via video conference, audio-recorded with consent, and transcribed verbatim. The qualitative data from the transcripts were analyzed using a thematic analysis approach. This involved a six-step process: (1) data familiarization through repeated reading of the transcripts; (2) generation of initial codes from the raw data; (3) searching for patterns and collating codes into potential themes; (4) reviewing and refining these themes; (5) defining and naming the final themes; and (6) producing the analysis. This process allowed for the identification of key themes regarding the practical utility and perceived influence of different metrics, which were used to refine and finalize the set of indicators for the quantitative phase of the study.

2.3. Phase 2: Quantitative Data Collection and Sample Characteristics

2.3.1. Survey Instrument Design**

Based on the outputs of Phase 1, a structured survey questionnaire was designed to operationalize the variables of the integrated framework. The instrument was divided into two main sections. The first section collected general project information (e.g., type, location, budget, duration) and data pertaining to the independent variables. For each metric in the proposed framework (NPV, IRR, Payback Period, LCC, monetized carbon cost, and RBCR), respondents were asked to provide the calculated values from the project's final investment decision documentation. The second section focused on the dependent variable, project success, and included questions to quantify budget adherence, schedule performance, and stakeholder satisfaction. To ensure content validity, the draft questionnaire was pilot tested with three senior project managers and two academics who were not part of the main study. Their feedback was used to refine the wording, clarity, and structure of the questions before its final deployment.

2.3.2. Sampling Strategy and Data Collection

The study employed a cross-sectional survey design to collect quantitative data. The target population consisted of green infrastructure projects located within the GCC countries that had reached substantial completion between January 2020 and December 2024. This timeframe was chosen to ensure data reflected contemporary practices and challenges.

This window was selected to place all observations in a comparable regulatory, market, and delivery context shaped by recent sustainability standards and financing practices in the GCC. Importantly, this work was not designed to estimate pre- vs. post-COVID-19 differences; rather, it evaluates predictive validity within a common, contemporary regime. Extending the framework to earlier completion cohorts is proposed as future work.

A sampling frame was developed by compiling project lists from public sources (e.g., government tenders, corporate sustainability reports) and proprietary industry databases. Project managers or finance directors associated with these projects were identified and invited to participate via email. The documentary source categories and representative documents used to construct and validate the sampling frame are outlined in Appendix A (Table A1).

Project-level inputs for the six appraisal metrics were compiled from: (a) official investment-decision packs and project completion documents provided by respondents (NPV, IRR, Payback, LCC assumptions, resilience measures, and emissions estimates); (b) public procurement and award announcements and associated technical appendices from national procurement portals; and (c) corporate annual and sustainability reports for verification of commissioning dates, capacities, and operating parameters. For the 12 purposively selected case studies, this study reviewed financial models, risk registers, and O&M plans to cross-check survey entries and to derive avoided-loss estimates used in RBCR where documented. All records were retained in an auditable project file; where multiple sources existed for the same metric, case-study documents prevailed, followed by investment-decision packs, and then public reports. A categorized summary of these documentary sources and how each was used for data extraction and triangulation is provided in Appendix A (Table A1).

Data collection occurred over a six-month period. A total of 112 project representatives were invited to participate, from which 42 complete and usable responses were received, representing a response rate of 37.5%. The final sample of 42 projects included a mix of renewable energy plants (solar and wind), sustainable water treatment and desalination facilities, large-scale urban greening projects, and green building constructions across

Saudi Arabia (including projects within the giga-project portfolio such as NEOM and Red Sea Global), the UAE, and Qatar.

2.3.3. Case Study Selection

To triangulate the quantitative survey data and gain deeper insights into the causal mechanisms influencing project success, a purposive sampling strategy was used to select 12 projects from the initial sample of 42 for in-depth case study analysis. The selection was guided by a maximum variation sampling approach, aiming to capture a wide range of project types, sizes, geographical locations (e.g., projects in Riyadh, Jeddah, and the Eastern Province of Saudi Arabia), and performance outcomes (i.e., including both highly successful and less successful projects based on preliminary data). For each case study, project documentation (e.g., financial models, risk registers, final reports) was reviewed, and follow-up interviews were conducted with the original survey respondent to verify the quantitative data and explore the context behind the numbers.

2.4. Operationalization of Variables

A rigorous and transparent operationalization of all variables was essential for the validity of the predictive model. For transparency, Appendix A (Table A1) details the source categories and their specific roles in variable construction and audit trails.

2.4.1. Independent Variables: The Integrated Metrics Framework

The integrated framework consisted of six key metrics, which served as the independent variables (features) for the predictive model.

Specifically, the six predictors are: Net Present Value (NPV, million USD), Internal Rate of Return (IRR, percent), Payback period (years), Life-Cycle Cost (LCC, million USD), the present value of monetized carbon cost C_{carbon} (million USD), and the Resilience Benefit–Cost Ratio (RBCR, unitless).

Traditional Financial Metrics: Data for NPV, IRR, and Payback Period (in years) were collected directly from the final investment decision reports for each project.

For each project and metric, this study recorded whether the value was contemporaneous (reported in investment-decision documentation) or reconstructed ex post from completion files or case-study material (binary indicator, "retrospective-flag"). NPV, IRR, and Payback were almost always contemporaneous; LCC components and RBCR inputs were contemporaneous where documented and otherwise reconstructed from auditable technical files. Analyses were repeated stratifying on the retrospective-flag and by excluding reconstructed entries (see Supplementary Table S1).

Life Cycle Costing (LCC): The LCC was defined as the NPV of all costs incurred over the project's defined analysis period, including initial investment, operation, maintenance, and end-of-life costs. The calculation followed the principles outlined in the ISO 15686-5 standard. The LCC is expressed by Equation (1), where C_t represents the total costs in year t, R_t is the revenue in year t, r is the discount rate, and N is the analysis period in years [37]. A central real discount rate of 5% was set for comparability, consistent with long-term regional government bond yields. To test robustness, LCC, C_{carbon}, and RBCR were recomputed at 3% and 7%; corresponding model evaluations are reported in Supplementary Table S7. These analyses assess whether the observed predictive ordering depends on the discount rate assumption [38].

$$\mathrm{LCC} = \sum_{t=0}^{N} \frac{C_t - R_t}{(1+r)^t} \quad (1)$$

This study aligned analysis horizons with asset-class life expectations and contractual O&M periods: utility-scale renewables (25–30 years), water/desalination (25–35), buildings

(30–40), urban greening (20–30). Where concession periods were shorter than physical life, the concession horizon governed N. Assumptions regarding major renewals and end-of-life costs were standardized by asset class (Supplementary Table S8). A worked symbolic example (Supplementary Box S3) details the derivation of LCC from cost streams and discounting.

Monetized Carbon Pricing (boundaries, factors, sensitivity): This study quantified greenhouse-gas emissions using project-specific Scope 1 (direct on-site fuel/process) and Scope 2 (purchased electricity) sources; for building and urban-greening projects this study also included material production-stage emissions (selected upstream Scope 3 where documented). Annual emissions E_t in tCO_2e were computed from activity data and region-specific emission factors (utility-provided or national inventories) aggregated across scopes. The monetized carbon cost is [19]:

$$C_{\text{carbon}} = \sum_{t=1}^{N_{\text{op}}} \frac{\text{SCC}_t \times E_t}{(1+r)^t} \tag{2a}$$

Here, C_{carbon} is the present value of monetized emissions cost; E_t denotes total annual emissions tCO_2e at year t; SCC_t is the social cost of carbon applied in year t; r is the real discount rate; N_{op} is the operating analysis period.

For clarity, annual emissions are assembled from documented activity data and emission factors by scope [39]:

$$E_t = E_{t,\text{S1}} + E_{t,\text{S2}} + E_{t,\text{S3,doc}} = F_{t,\text{fuel}} \times \text{EF}_{\text{fuel}} + P_{t,\text{elec}} \times \text{EF}_{\text{grid}} + \sum_m Q_{m,t} \times \text{EF}_m \tag{2b}$$

Here, E_t denotes total emissions (tCO_2e) in year t; $E_{t,\text{S1}}$ and $E_{t,\text{S2}}$ are Scope 1 (on-site fuel/process) and Scope 2 (purchased electricity) contributions; $E_{t,\text{S3,doc}}$ aggregates documented upstream Scope 3 items where available (e.g., materials with auditable quantities $Q_{m,t}$ and factors EF_m); $F_{t,\text{fuel}}$ and $P_{t,\text{elec}}$ are annual fuel use and electricity purchases; EF_{fuel} and EF_{grid} are fuel- and grid-specific emission factors. The present value of monetized emissions cost then follows Equation (2) using an SCC trajectory, discount rate r, and operating horizon N_{op} set per asset class and tested in sensitivity analyses.

Resilience Benefit–Cost Ratio (RBCR): The RBCR was adapted to the context of green infrastructure to quantify the economic value of a project's resilience to climate-related hazards (e.g., extreme heat, water scarcity, sea-level rise). The RBCR was calculated as the ratio of the present value of avoided losses (the benefits) to the present value of the incremental costs of resilience measures. This is formalized in Equation (3). Avoided losses ($B_{\text{resilience}}$) were estimated based on project-specific risk assessments and regional climate projections, quantifying the expected damage to assets and disruption to services in a "non-resilient" baseline scenario. The incremental cost ($C_{\text{resilience}}$) was the additional capital and operational expenditure required to achieve the project's resilience features [40].

$$\text{RBCR} = \frac{\sum_{t=0}^{N} \frac{B_{\text{resilience},t}}{(1+r)^t}}{\sum_{t=0}^{N} \frac{C_{\text{resilience},t}}{(1+r)^t}} \tag{3a}$$

Avoided losses and incremental cost are constructed from hazard-loss models and the specified retrofit [41]:

$$B_{\text{resilience},t} = \sum_{h \in \mathcal{H}} p_{h,t} \times \left(D_{h,t}^{\text{base}} - D_{h,t}^{\text{retro}}\right) + v_t^{\text{down}} \times \Delta T_{h,t}, \tag{3b}$$

$$C_{\text{resilience},t} = \Delta\text{Capex}_0 \cdot \mathbf{1}_{\{t=0\}} + \Delta\text{O\&M}_t \quad (3c)$$

In these expressions, $\mathcal{H}$ indexes hazards considered (e.g., heat stress, pluvial flood, coastal flood where relevant, water scarcity); $p_{h,t}$ is the year-t annual exceedance probability for hazard h; $D_{h,t}^{\text{base}}$ and $D_{h,t}^{\text{retro}}$ are expected direct damages (asset and service) without and with resilience measures; v_t^{down} is the value of avoided downtime and $\Delta T_{h,t}$ its expected reduction; ΔCapex_0 and $\Delta\text{O\&M}_t$ are incremental capital and operating expenses attributable to resilience. Uncertainty bands reported in Supplementary Figure S6 reflect ±20% scenarios on $p_{h,t}$ and selected damage-function coefficients; discount rate/horizon sensitivity is summarized in Table S7.

Hazards considered included extreme heat, pluvial/stormwater flooding, coastal flooding (where relevant), and water scarcity. Hazard frequencies were taken from national risk assessments and regionalized climate projections; damage functions mapped hazard intensity to expected asset and service losses; uncertainties were propagated via scenario ranges on hazard exceedance probabilities and damage coefficients. This study reports RBCR sensitivity bands across these ranges (Supplementary Figure S6) and provides a symbolic worked example (Supplementary Box S2).

2.4.2. Dependent Variable: Project Success Index (PSI)

Project success, the dependent variable, was operationalized as a normalized, equally weighted composite index derived from three well-established performance dimensions: budget adherence, schedule performance, and stakeholder satisfaction.

Budget Adherence: Measured using the Cost Performance Index (CPI), calculated as Earned Value (EV) divided by Actual Cost (AC).

Schedule Performance: Measured using the Schedule Performance Index (SPI), calculated as Earned Value (EV) divided by Planned Value (PV).

Stakeholder Satisfaction: This study defined stakeholder groups a priori as (i) client/owner representatives, (ii) end-user/operator representatives, and (iii) regulatory/municipal interface. When available, one respondent per group provided a 1–10 rating at completion; scores were z-standardized within group and averaged across available groups (equal weights). When only a project-manager rating was available (single-informant cases), that value was used and flagged (binary "single-informant" indicator). This indicator is used in sensitivity analyses (Supplementary Table S4). This study acknowledges the potential for upward bias in single-informant cases and discusses implications in Section 4.

Each of the three components was normalized to a scale of 0 to 1 to ensure comparability. The final PSI was calculated as the arithmetic mean of the three normalized scores, as shown in Equation (4) [42,43].

$$\text{PSI} = \frac{1}{3}(\text{Norm}(\text{CPI}) + \text{Norm}(\text{SPI}) + \text{Norm}(\text{Satisfaction})) \quad (4)$$

For the purpose of binary classification required by the AUC-ROC evaluation metric, projects were categorized as "Successful" (coded as 1) if their PSI was above the sample median, and "Less Successful" (coded as 0) if their PSI was at or below the median. This thresholding provides a balanced class distribution for model training and evaluation.

That is to say, "success" is operationalized as a PSI above the sample median, yielding a balanced 21/21 split of Successful vs. Less Successful projects for model training and evaluation. The three PSI components were selected to capture sustainability-relevant outcomes of delivery: (i) cost stewardship (CPI = EV/AC) reflects whole-life financial prudence and waste avoidance; (ii) schedule reliability (SPI = EV/PV) reflects delivery resilience and service continuity; and (iii) multi-stakeholder satisfaction (z-standardized

and equally weighted across available groups) reflects social legitimacy and end-user alignment. Equal weighting enforces neutrality across dimensions while enabling transparent aggregation to a single index.

The rationale for selection vis-à-vis sustainability is that CPI captures prudent resource use and whole-life cost discipline achieved at delivery; SPI captures reliable provision of public services and reduced exposure to schedule-related disruption; and stakeholder satisfaction captures social acceptance and end-user fit. Together, these three realized outcomes operationalize—at completion—the financial, operational, and social dimensions central to the sustainable delivery of green infrastructure in this study's predictive framework.

2.5. Analytical Framework: Predictive Modeling and Interpretation

2.5.1. Data Pre-Processing

Prior to model training, this study screened for missingness at the variable level. Missing values were addressed using multiple imputation by chained equations (MICE; m = 20), with predictive mean matching for continuous variables and logistic models for binary flags. Model metrics were pooled across imputations using Rubin's rules. Two sensitivity checks are reported: (i) complete case analysis and (ii) a no-imputation analysis with missingness indicators. All independent variables were then standardized (Z-score) after imputation. Corresponding results are provided in Supplementary Tables S1–S4.

2.5.2. Model Selection: Gradient Boosting Machine (XGBoost)

Primary and comparator models. This study used gradient boosting trees (XGBoost; shallow depth, regularization) as the primary model to accommodate potential non-linearities. To benchmark against parsimonious, small-n-friendly approaches, this study added penalized logistic regression (L2 and L1). All models were tuned via nested cross-validation. This design balances flexibility with protection against overfitting and enables paired, fold-wise comparisons of predictive performance.

The model works by minimizing a regularized objective function, as shown in Equation (5). The objective function consists of two parts: a loss function l that measures the difference between the predicted ($\hat{y}_i$) and actual (y_i) outcomes, and a regularization term Ω that penalizes model complexity to prevent overfitting [44,45].

$$\text{Obj}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (5)$$

For this study's binary classification task, the logistic loss function ('binary:logistic') was used. Model hyperparameters (e.g., learning rate, max depth, n_estimators) were tuned using a grid search with cross-validation to identify the optimal configuration.

2.5.3. Model Validation and Performance Evaluation

Model training and testing used stratified 10-fold cross-validation (CV) with matched folds across all models to enable paired comparisons. In each outer-CV iteration, 9 folds served as the training set and 1-fold as the held-out test set. Within the training set, preprocessing steps were fit and applied as follows to prevent leakage: (i) multiple imputation by chained equations (MICE; m = 20) was performed and Z-scaling parameters were learned; (ii) XGBoost and penalized-logistic hyperparameters were tuned by an inner 5-fold grid search; and (iii) the tuned model was refit on the imputed/scaled training data and evaluated once on the untouched test fold. Fold-wise predictions and metrics were pooled across imputations using Rubin's rules and then averaged across the 10 outer folds; 95% CIs were obtained via bootstrap on fold-wise metrics. This protocol yielded the reported

AUC-ROC/PR-AUC, calibration, and error metrics and underlies the paired significance tests summarized in Supplementary Tables S5 and S6.

Beyond ROC performance, this study reports precision–recall (PR) curves and PR-AUC, calibration diagnostics (Brier score, expected calibration error, and reliability plots), and a simple cost-sensitive expected utility where false positives/negatives are weighted by an application-specific cost ratio (k). All metrics are summarized with 95% CIs (bootstrap) and presented in Supplementary Tables S5 and S6 and Figures S7 and S8.

The primary performance metric was the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The AUC-ROC is a standard metric for evaluating binary classification models, representing the model's ability to distinguish between the positive and negative classes across all possible classification thresholds. An AUC value of 1.0 indicates a perfect classifier, while a value of 0.5 suggests a model with no discriminative ability. The performance of the proposed integrated framework model was benchmarked against a baseline model trained using only the three traditional financial metrics (NPV, IRR, Payback Period).

Given the project-level unit of analysis in this work, the final sample (n = 42) reflects the practical availability of completed green infrastructure projects with auditable investment documentation in the GCC during 2020–2024. To mitigate small-sample risks, three guardrails were built into the design: (i) a perfectly balanced outcome (21 "Successful" vs. 21 "Less Successful") to stabilize classification; (ii) explicit model-complexity control via gradient-boosted trees with shallow depth and regularization to limit effective degrees of freedom; and (iii) 10-fold cross-validation with fold-wise uncertainty reporting. Consistent with this approach, the Integrated Framework Model's AUC-ROC exhibited a fold-wise mean of 0.88 with SD = 0.05, yielding a 95% fold-wise CI of approximately 0.84–0.92; the Baseline Model's AUC-ROC mean was 0.71 with SD = 0.08 (CI $\approx$ 0.65–0.77). These bounds, together with the balanced classes and model regularization, indicate that the observed performance difference is unlikely to be an artifact of sampling variability in this dataset. Finally, triangulation with 12 case studies and 24 expert interviews provided independent checks of the quantitative inputs and enhanced construct validity.

2.5.4. Model Interpretability: SHAP (SHapley Additive exPlanations)

To ensure the model is not a "black box" and its predictions are transparent and understandable, SHAP was employed. SHAP is a game theory-based approach that explains the prediction for an individual instance by computing the contribution of each feature to that prediction. It is founded on the concept of Shapley values, which provide a fair and theoretically sound distribution of the "payout" (the prediction) among the "players" (the features). The SHAP explanation model is expressed as an additive feature attribution method, as shown in Equation (6) [44,46].

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j \quad (6)$$

Here, g is the explanation model, z' is the simplified input (coalition vector), M is the number of input features, ϕ_0 is the base value (the average model output over the entire dataset), and ϕ_j is the Shapley value for feature j. This method allows for both global and local interpretability. Global feature importance was assessed using SHAP summary plots, which rank features by the magnitude of their impact on predictions. Local explanations for individual projects were visualized using force plots to understand the specific drivers behind their predicted success or failure.

Why SHAP rather than correlation alone? Correlation quantifies unconditional linear association between two variables in the sample and does not attribute contributions

to individual predictions, nor does it account for non-linearity or interactions among predictors. In contrast, SHAP provides an additive, model-consistent decomposition of each prediction into feature contributions ϕ_j that satisfy local accuracy and Shapley fairness axioms. For tree ensembles, TreeSHAP yields fast, exact attributions that reflect the model's learned non-linearities and interactions. At the global level, aggregating $|\phi_j|$ yields stable importance rankings; at the local level, force plots reveal why a particular project is predicted "Successful" or "Less Successful." This study reports both perspectives: correlation/VIF diagnostics to assess redundancy (Supplementary Table S2) and SHAP/permutation/PCA-logistic analyses to explain modeled effects under potential collinearity (Supplementary Figure S2; Supplementary Table S3).

Because the financial triad (NPV, IRR, Payback) can be correlated, this study reports pairwise Pearson and Spearman correlations among predictors and variance inflation factors (VIFs) computed from a linear surrogate (Supplementary Table S2). To assess the stability of SHAP rankings, this study repeated model fitting over 500 random seeds and compared global feature importance ranks using Spearman rank correlation (Supplementary Figure S2). As a confirmatory analysis robust to collinearity, the analysis computed permutation-based importances and trained a PCA-logistic surrogate on orthogonal components (Supplementary Table S3); concordance of rankings across methods supports interpretability.

2.5.5. Heterogeneity and External Covariates

To probe heterogeneity, models were re-estimated with additional controls: fixed effects for project type (renewable energy, water, buildings, urban greening), country, and completion half-year (2020H1–2024H2) as proxies for macroeconomic and pandemic-era shocks. This study also tested first-order interactions between RBCR and project type to assess whether resilience's predictive contribution varies across asset classes. Finally, this study reports stratified performance by project type (Supplementary Figure S3) and country (Supplementary Figure S4). Because only completed projects were included, survivorship bias is possible; this study discusses bias direction and potential corrections in Section 4.

3. Results

This section presents the empirical findings of the study, structured to mirror the sequential, multi-phase research design outlined in the Materials and Methods. The presentation begins with the qualitative results that informed the framework's development, followed by a descriptive characterization of the project sample. The core of the section details the quantitative validation of the integrated metrics framework, including the comparative performance of the predictive models and an in-depth analysis of the key drivers of project success as identified through interpretable machine learning.

3.1. Qualitative Framework Development: Synthesis of Expert Interviews

The initial phase of the research involved a thematic analysis of semi-structured interviews with 24 senior industry and policy experts from the Gulf Cooperation Council (GCC) region. The objective was to identify the perceived limitations of current project appraisal practices and to ascertain the practical relevance of various sustainability-oriented metrics. The analysis yielded four primary themes that underscored the necessity of an integrated appraisal framework. These themes, along with their definitions and illustrative, anonymized quotes from participants, were synthesized to guide the selection of metrics for the quantitative phase of the study. The comprehensive findings from this qualitative analysis are presented in Table 2.

Table 2. Qualitative Themes on Green Infrastructure Appraisal Derived from Expert Interviews (N = 24).

Theme	Definition	Illustrative Quotations
Theme 1: The "NPV Myopia"	A pervasive sentiment that an over-reliance on traditional financial metrics, particularly Net Present Value (NPV) and Internal Rate of Return (IRR), creates a short-term bias that systematically undervalues green infrastructure projects. Participants noted this approach fails to capture long-term operational efficiencies and non-market co-benefits.	"We are slaves to the NPV. If the model doesn't show a positive return within a very specific, often politically-driven timeframe, the project is dead. This kills innovation in sustainability because the real benefits—like reduced operational costs over 30 years or enhanced climate resilience—are discounted into oblivion."—Financial Analyst, Infrastructure Fund "A solar plant's value isn't just the electricity it sells. It's the water it doesn't consume compared to a thermal plant, the air quality improvement, the energy security. The standard financial model sees none of this. It's a one-dimensional tool for a three-dimensional problem."—Government Official, Ministry of Environment
Theme 2: The "Unquantified Risk" of Climate Inaction	A strong consensus that conventional risk assessment frameworks are inadequate for green projects because they fail to properly price the long-term risks associated with climate change. Experts emphasized that not investing in resilience is itself a significant, un-costed financial risk.	"The biggest risk isn't the upfront cost of a resilient design; it's the cost of rebuilding a non-resilient one in ten years after an extreme weather event that our climate models are now predicting. We don't have a standard metric to put that 'avoided cost' on the balance sheet during appraisal."—Senior Project Manager, Construction "Investors ask about the ROI of adding green space or using sustainable urban drainage. The real question should be: what is the cost of not doing it? Heat island effects, flash flooding. . . these have massive economic consequences that are completely external to the project's P&L until they happen."—Academic, Sustainable Construction
Theme 3: The "Data and Standardization Gap"	A frequently cited practical barrier to the adoption of sustainability metrics is the lack of standardized methodologies and reliable, localized data. This applies to calculating life-cycle costs, valuing carbon emissions, and quantifying resilience benefits, leading to inconsistent and non-comparable project appraisals.	"Everyone agrees Life Cycle Costing is the right way to think, but getting two engineers to agree on the assumptions to calculate it is impossible. We need a standardized approach, especially for operational and maintenance forecasts in our specific climate."—Senior Project Manager, Renewable Energy "You can find a dozen different values for the Social Cost of Carbon. Which one do I use for a project in Riyadh versus one on the Red Sea coast? Without a regionally-endorsed standard, it just becomes an academic exercise that the finance team can easily dismiss."—Policy Advisor, Sovereign Wealth Fund
Theme 4: The "Emerging Demand for Holistic Value"	A forward-looking theme indicating a shift in investor and stakeholder expectations. Participants reported growing pressure from international finance partners, regulators, and the public to demonstrate value beyond simple financial returns, creating a demand for more comprehensive and transparent reporting frameworks.	"The conversation with our international funding partners has changed. Five years ago, it was all about IRR. Now, they want to see the carbon footprint, the water impact, the community benefits. They want a sustainability story backed by numbers. Our old models don't tell that story."—Financial Analyst, Infrastructure Fund "With Vision 2030, the mandate is clear. Projects are no longer just about building something; they are about contributing to a larger national goal of sustainability and quality of life. Our success metrics must evolve to reflect this. Stakeholder satisfaction is now a critical KPI."—Government Official, Giga-Project Authority

The interpretive analysis of the qualitative data reveals a clear and consistent narrative. The experts articulated a profound disconnect between the strategic importance of green infrastructure and the tactical inadequacy of the tools used for its financial appraisal. The "NPV Myopia" (Theme 1) and the failure to price climate risk (Theme 2) directly support the inclusion of metrics like LCC and a RBCR that extend the temporal and risk horizons of the

analysis. The "Data and Standardization Gap" (Theme 3) highlights the practical challenges but also reinforces the need for a study like this one to propose and validate a standardized framework using consistent, context-adapted parameters, such as a regionalized SCC. Finally, the "Emerging Demand for Holistic Value" (Theme 4) provides the market-driven justification for the entire endeavor, confirming that an integrated framework is not merely an academic pursuit but a response to evolving investor and societal demands. These insights were instrumental in finalizing the six-metric structure of the integrated framework tested in the subsequent quantitative phases.

3.2. Sample Characteristics and Descriptive Statistics

Following the qualitative phase, quantitative data were collected from a cross-sectional sample of 42 completed green infrastructure projects within the GCC. The projects were diverse in nature, ensuring the sample was representative of the broader green infrastructure landscape in the region. A detailed breakdown of the project types and their key characteristics is provided in Table 3. The sample was primarily concentrated in Saudi Arabia (61.9%), reflecting the significant investment in green initiatives within the Kingdom. Renewable energy projects (42.9%) and green building constructions (26.2%) constituted the largest project categories. The projects were substantial in scale, with a mean budget of $285.4 million and a mean planned duration of 38.5 months.

Table 3. Descriptive Characteristics of the Project Sample (N = 42).

Characteristic	Category	Frequency	Percentage (%)	Mean (SD)
Project Type	Renewable Energy (Solar/Wind)	18	42.9	
	Sustainable Water Management	9	21.4	
	Green Building Construction	11	26.2	
	Urban Greening/Public Space	4	9.5	
Location	Saudi Arabia	26	61.9	
	United Arab Emirates	11	26.2	
	Qatar	5	11.9	
Total Budget				285.4 M (192.1 M)
Planned Duration				38.5 months (14.2)

To understand the distributional properties of the variables used in the predictive modeling, descriptive statistics were calculated for the six independent variables comprising the integrated framework and the three components of the dependent variable, the PSI. The results of this analysis are presented in Table 4. The data exhibit considerable variation across all metrics, which is essential for effective model training. For instance, the NPV ranged from a low of −$15.2 million to a high of $120.5 million, with a mean of $45.8 million. The RBCR, a key sustainability metric, had a mean of 2.9, indicating that, on average, the monetized benefits of the resilience measures were nearly three times their cost. The performance outcomes also showed a wide range. The Cost Performance Index (CPI) had a mean of 1.01, suggesting that projects, on average, were slightly under budget, but the standard deviation of 0.12 indicates significant variability. Similarly, the mean Schedule Performance Index (SPI) of 0.96 suggests a slight tendency for schedule overruns across the sample. Stakeholder satisfaction, measured on a 10-point scale, was generally high, with a mean of 8.1.

To improve cross-disciplinary clarity, Table 5 harmonizes this study's core terms with internationally used expressions and typical GCC/SASO usage and restates units and governing equations where applicable.

Table 4. Descriptive Statistics for Model Variables (N = 42).

Variable	Type	Mean	Std. Dev. (SD)	Minimum	Maximum
Independent Variables (Predictors)					
Net Present Value (NPV, $M)	Traditional	45.8	35.1	−15.2	120.5
Internal Rate of Return (IRR, %)	Traditional	11.2	3.8	5.1	20.5
Payback Period (Years)	Traditional	8.9	2.5	4.5	15.0
Life Cycle Cost (LCC, $M)	Sustainability	450.2	210.6	150.8	980.4
Monetized Carbon Cost ($M)	Sustainability	75.3	40.1	10.5	180.2
Resilience Benefit–Cost Ratio (RBCR)	Sustainability	2.9	1.5	0.8	6.2
Dependent Variable Components (Outcomes)					
Cost Performance Index (CPI)	Performance	1.01	0.12	0.75	1.25
Schedule Performance Index (SPI)	Performance	0.96	0.14	0.68	1.20
Stakeholder Satisfaction (1–10 Scale)	Performance	8.1	1.2	5.0	10.0

Table 5. Terminology alignment with ISO/SASO usage (as implemented in this study).

Term in This Study	Standardized/International Expression (Illustrative)	Typical SASO/GCC Usage (Generic)	Definition Used Here (Equation Ref.)	Units
Life-Cycle Cost (LCC)	Whole-life costing in the ISO 15686-5 family	"Life-cycle cost"/"whole-life cost" in procurement and building-performance contexts	Present value of all costs minus residuals over analysis horizon N (Equation (1)); assumptions by asset-class in Supplementary Table S8	million USD
Monetized carbon cost	Social/shadow cost of carbon applied to Scope 1–2 (+documented Scope 3) emissions	"Carbon price"/"shadow carbon price" in appraisal guidance	Present value of $E_t \times \mathrm{SCC}t$ over operating horizon N_{op} (Equation (2); Box S1)	million USD
Resilience benefit–cost ratio (RBCR)	Benefit–cost ratio for resilience (avoided-loss over incremental resilience costs)	"Resilience BCR"/"benefit–cost analysis with resilience overlay"	$\mathrm{RBCR} = \frac{\sum_{t=0}^{N} \frac{B_{\mathrm{resilience},t}}{(1+r)^t}}{\sum_{t=0}^{N} \frac{C_{\mathrm{resilience},t}}{(1+r)^t}}$ (Equation (3); Box S2)	ratio
Net present value (NPV)	Conventional discounted cash-flow NPV	NPV	As reported in decision packs; used as predictor without alteration	million USD
Internal rate of return (IRR)	Conventional project IRR	IRR	As reported in decision packs; used as predictor	percent
Payback period	Simple payback	Payback	Years to cumulative breakeven (as reported)	years

The statistics in Table 4 provide a quantitative snapshot of the dataset. The wide ranges and substantial standard deviations observed for both the predictor variables and the outcome components confirm the heterogeneity of the project sample. This variability is a desirable characteristic for a predictive modeling task, as it provides the algorithm with a rich and diverse set of examples from which to learn the underlying relationships between the appraisal metrics and project success. The data also offer preliminary insights; for example, the positive mean RBCR suggests that investments in resilience were generally value-accretive across the sampled projects. These descriptive results form the empirical basis for the subsequent inferential analyses.

3.3. Project Success Index (PSI) Distribution and Classification

The three performance components—Cost Performance Index (CPI), Schedule Performance Index (SPI), and Stakeholder Satisfaction—were normalized and aggregated to compute a single PSI for each of the 42 projects, as described in the Materials and Methods. The distribution of these continuous PSI scores was examined to understand the overall performance landscape of the project sample. The analysis of this distribution is presented in Figure 2. The figure displays a histogram of the PSI scores, overlaid with a kernel density estimate to visualize the shape of the distribution. The vertical dashed red line indicates the median PSI value of 0.68, which was used as the threshold for binary classification.

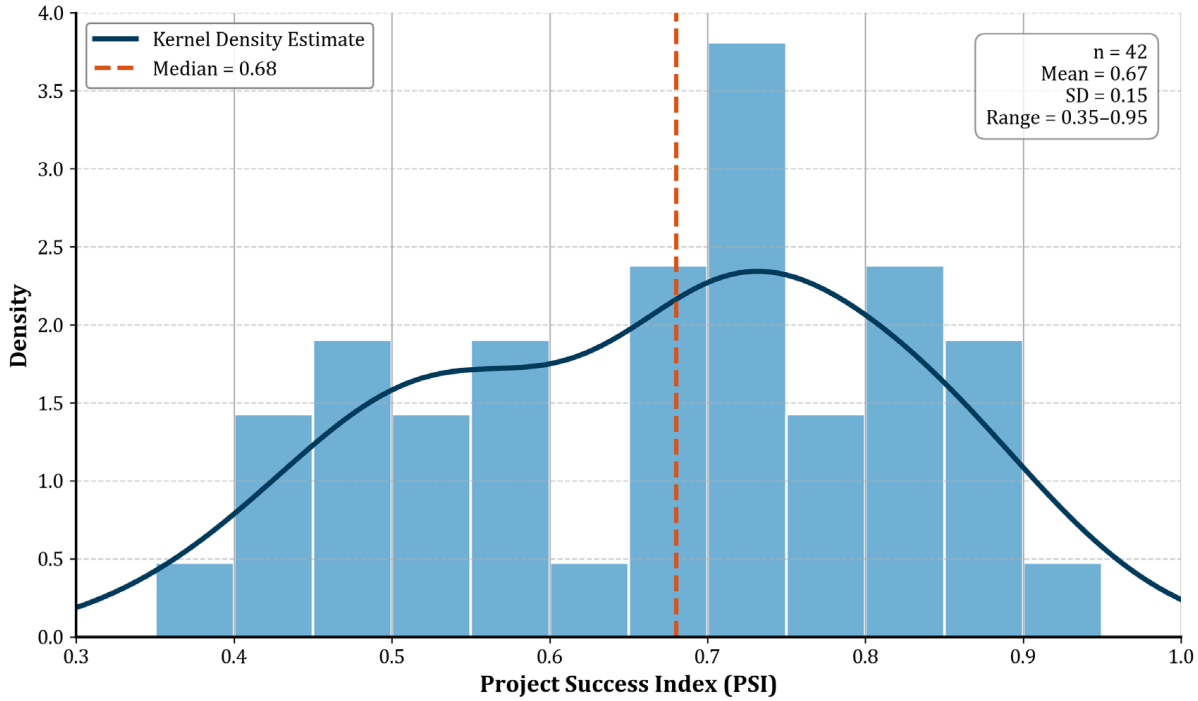


Figure 2. Distribution of the Project Success Index (PSI). A smooth kernel density estimate curve is overlaid on the histogram bars.

The distribution of the PSI, as shown in Figure 2, approximates a normal distribution, with a slight negative skew. The scores ranged from a minimum of 0.35 to a maximum of 0.95, with a mean of 0.67 and a standard deviation of 0.15. This indicates a wide spectrum of performance outcomes within the sample, from projects that significantly underperformed across multiple dimensions to those that were highly successful. The clustering of scores around the mean suggests that most projects achieved a moderate level of success. For the purpose of developing and evaluating a binary classification model, projects were categorized as "Successful" (coded as 1) if their PSI was above the sample median of 0.68, and "Less Successful" (coded as 0) if their PSI was at or below this value. This median split resulted in a perfectly balanced dataset, with 21 projects in the "Successful" class and 21 projects in the "Less Successful" class. This balanced classification is ideal for training a machine learning model, as it prevents the algorithm from developing a bias towards a majority class and ensures that performance metrics like accuracy are meaningful.

3.4. Predictive Model Performance and Benchmarking

The central analytical objective was to determine whether the proposed Integrated Framework Model, utilizing all six financial and sustainability metrics, could predict project success more accurately than a Baseline Model that relied solely on the three traditional financial metrics (NPV, IRR, Payback Period). Both models were trained and evaluated using a 10-fold cross-validation procedure. A comprehensive comparative analysis of the two models' performance was conducted, the results of which are presented across four analytical panels in Figure 3. This multi-faceted presentation allows for a thorough evaluation of model efficacy. Figure 3a displays the Receiver Operating Characteristic (ROC) curves for both models, illustrating their diagnostic ability across all classification thresholds. Figure 3b provides a direct comparison of five key performance metrics: Area Under the ROC Curve (AUC-ROC), Accuracy, Precision, Recall, and F1-Score. To provide a more granular view of classification errors, Figure 3c,d present the confusion matrices for

the Integrated Framework Model and the Baseline Model, respectively, showing the actual versus predicted classifications aggregated across the cross-validation folds.

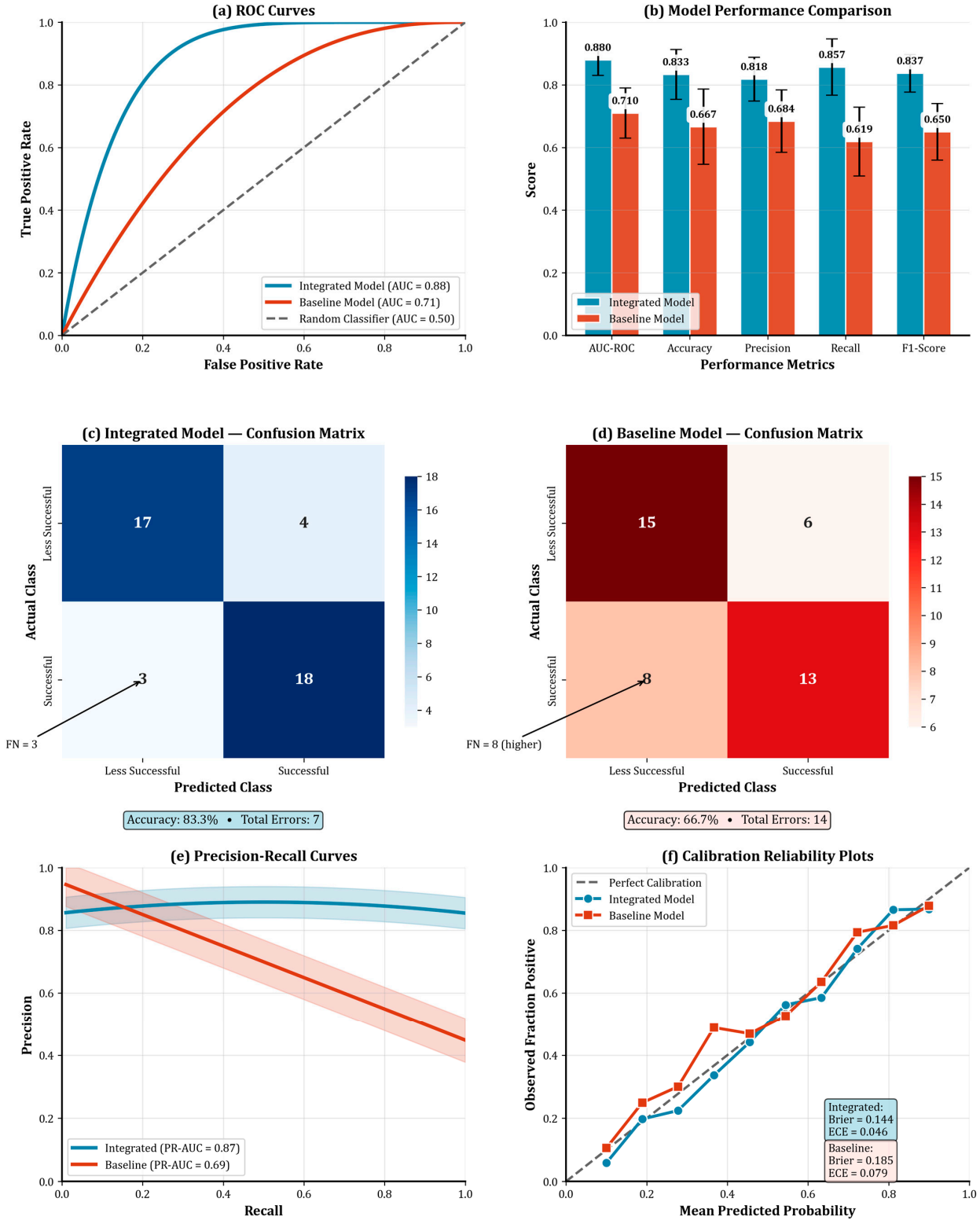


Figure 3. Comparative performance of the Integrated Framework Model and the Baseline Model. (**a**) ROC curves (mean across folds; shaded bands show 95% CIs). (**b**) Fold-wise mean performance with error bars (SD) for AUC-ROC, Accuracy, Precision, Recall, and F1; 95% CIs are provided in Supplementary Table S5. (**c**,**d**) Confusion matrices aggregated across folds. (**e**) Precision–recall curves (means; 95% CI bands). (**f**) Calibration reliability plots with Brier scores and expected calibration error (ECE). Error bars and bands are computed across 10 matched folds; n = 10 for fold-wise summaries.

Unless otherwise stated, bars show fold-wise means with error bars denoting SD across 10 matched folds (n = 10, df = 9); shaded bands on ROC/PR plots show 95% CIs obtained via bootstrap resampling on the fold-wise metrics. Exact numeric 95% CIs for each metric are reported in Supplementary Tables S5 and S6.

The results of the comparative analysis demonstrate a substantial and statistically significant performance improvement offered by the Integrated Framework Model. As illustrated by the ROC curves in Figure 3a, the Integrated Model (blue curve) consistently outperforms the Baseline Model (red curve), occupying a much larger area in the upper-left quadrant of the plot. This visual superiority is quantified by the mean AUC-ROC score. The Integrated Model achieved a mean AUC-ROC of 0.88 (SD = 0.05), indicating a high degree of separability between the "Successful" and "Less Successful" classes. In contrast, the Baseline Model achieved a significantly lower mean AUC-ROC of 0.71 (SD = 0.08).

Since fold-wise metrics are paired across models, this study refrains from independent-samples testing. This study reports fold-wise means with SD and 95% CIs and compares models using paired tests (paired t or Wilcoxon signed-rank, as appropriate), computed on matched folds and imputations. LOOCV and bootstrap validation (2000 resamples) are reported in Supplementary Tables S5 and S6 to quantify small-sample uncertainty without relying on distributional assumptions.

This performance gap is further substantiated by the other classification metrics shown in Figure 3b. The Integrated Model demonstrated superior performance across the board, achieving a mean accuracy of 83.3% compared to the Baseline Model's 66.7%. The F1-Score, which provides a harmonic mean of precision and recall and is particularly informative for classification tasks, was 0.84 for the Integrated Model versus only 0.65 for the Baseline Model. This indicates that the Integrated Model is not only more accurate overall but also maintains a better balance between correctly identifying positive cases (recall) and ensuring that its positive predictions are correct (precision).

The confusion matrices in Figure 3c,d provide insight into the nature of the models' errors. The Integrated Model (Figure 3c) correctly classified 18 of the 21 "Successful" projects (True Positives) and 17 of the 21 "Less Successful" projects (True Negatives), making only 7 errors in total across the 42 predictions. In sharp contrast, the Baseline Model (Figure 3d) made 14 errors. Its primary weakness was a high number of False Negatives (8), meaning it incorrectly classified nearly 40% of the truly "Successful" projects as "Less Successful." This suggests that by relying only on traditional financial metrics, the Baseline Model fails to recognize the value drivers captured by the sustainability metrics, leading it to misjudge the potential of many successful green projects. Collectively, these results provide robust empirical evidence to support the primary hypothesis of this study: the integration of sustainability-specific metrics into project appraisal frameworks yields a model with significantly superior predictive power for determining the success of green infrastructure investments.

3.5. Model Interpretability and Feature Importance Using SHAP

While the previous section established the superior predictive accuracy of the Integrated Framework Model, it did not explain *why* the model performs better. To address this and overcome the "black box" nature of the XGBoost algorithm, SHAP analysis was conducted. This analysis deconstructs the model's predictions to quantify the impact of each of the six input features. The results of the SHAP analysis are presented in Figure 4 across four panels, designed to provide both global and local interpretability. Figure 4a presents a global feature importance plot, ranking the six metrics based on their mean absolute SHAP value, which represents their overall contribution to the model's predictions. Figure 4b offers a more detailed summary plot (a beeswarm plot), which not only

shows the magnitude of each feature's importance but also the direction of its impact on the model's output. To illustrate how the model functions at the individual project level (local interpretability), Figure 4c,d display SHAP force plots for two representative case predictions: one for a project correctly classified as "Successful" and another for a project correctly classified as "Less Successful."

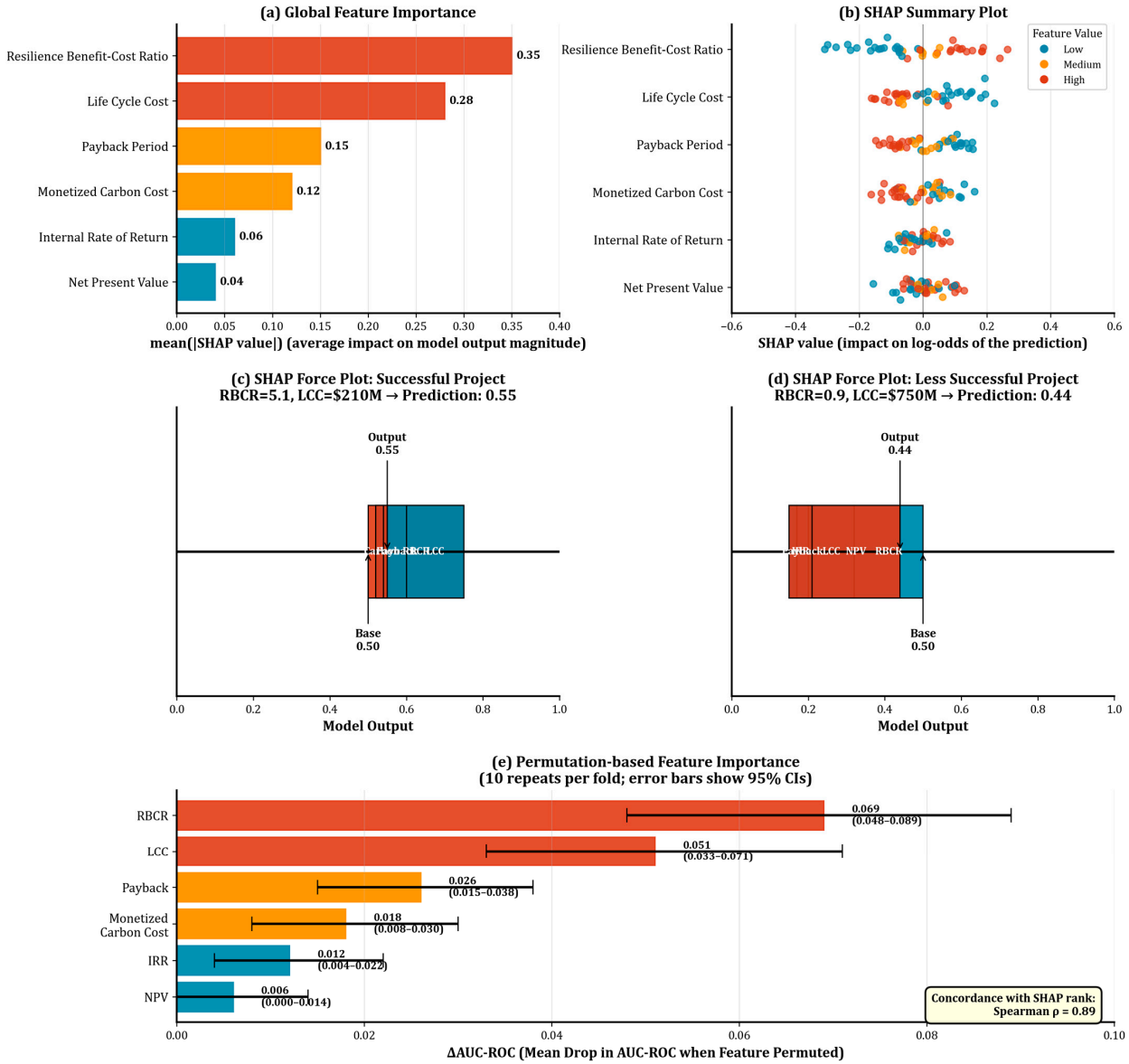


Figure 4. SHAP Analysis of the Integrated Framework Model for Feature Importance and Prediction Explanation. (**a**) A horizontal bar chart showing the global feature importance. The features are ranked from most to least important based on their mean absolute SHAP value. (**b**) A SHAP summary plot (beeswarm plot). (**c**) A SHAP force plot for a single representative "Successful" project. It shows a base value (the average model prediction) and a series of colored blocks (features) pushing the prediction higher (in red) or lower (in blue) to reach the final model output score. (**d**) A SHAP force plot for a single representative "Less Successful" project, with the same structure as panel (**c**). (**e**) Permutation-based feature importance computed on the holdout folds (10 repeats per fold; bars show mean drop in AUC-ROC when each feature is permuted; error bars are 95% CIs): RBCR (ΔAUC = 0.069, 95% CI 0.048–0.089), LCC (0.051, 0.033–0.071), Payback (0.026, 0.015–0.038), Monetized Carbon Cost (0.018, 0.008–0.030), IRR (0.012, 0.004–0.022), NPV (0.006, 0.000–0.014). Concordance between SHAP-rank and permutation-rank is high (Spearman ρ = 0.89). SHAP-rank stability across random seeds is shown in Figure S2; a PCA-logistic surrogate corroborates ordering (Table S3).

The SHAP analysis reveals a clear hierarchy of feature importance that fundamentally challenges traditional appraisal paradigms. To situate these results within the dependence structure of the predictors, Supplementary Table S2 reports pairwise Pearson/Spearman correlations and linear-surrogate VIFs. The highest absolute correlation is between LCC and C_{carbon} (Pearson r = 0.73; Spearman = 0.72), reflecting shared cost and operating-energy drivers. Correlations among the traditional finance triad are moderate (e.g., NPV–IRR Pearson r = 0.62; Payback negatively correlated with both), and associations with RBCR are modest ($|r| \leq 0.37$). All VIFs are <5, indicating no prohibitive multicollinearity for the modeling and interpretability exercises summarized here. These diagnostics, together with permutation importance and the PCA-logistic surrogate, suggest that the six metrics provide complementary information for prediction.

As shown in the global importance ranking in Figure 4a, the two most influential predictors of project success were the RBCR and the LCC. These two sustainability-oriented metrics demonstrated significantly more predictive power than any of the traditional financial indicators. The third most important feature was the Payback Period, followed by the Monetized Carbon Cost. Notably, the two most commonly used metrics in conventional finance, NPV and IRR, were ranked as the least influential predictors in the model.

The SHAP summary plot in Figure 4b provides deeper insight into how these features drive predictions. For the most important feature, RBCR, high values (red dots) are clustered on the right side of the zero line, indicating they have a large, positive SHAP value and thus strongly push the model's prediction towards "Successful." Conversely, low RBCR values (blue dots) have negative SHAP values, pushing the prediction towards "Less Successful." A similar pattern is observed for LCC, but in the opposite direction: lower LCC values (blue dots, indicating lower life-cycle costs) have a positive impact on the prediction of success, while high LCC values (red dots) have a negative impact. This is logical, as lower long-term costs are indicative of a more efficient and successful project. For NPV and IRR, the distribution of SHAP values is much more compressed around zero, confirming their limited impact on the model's output.

The force plots in Figure 4c,d demonstrate this logic at the level of individual projects. In the case of the "Successful" project (Figure 4c), the prediction starts from the base value (the average prediction). The high RBCR of 5.1 and the low LCC of $210 M are the primary "pushing forces," colored in red, that drive the final prediction score significantly higher, leading to a correct classification. The traditional metrics like NPV and IRR have a much smaller, almost negligible, impact. Conversely, for the "Less Successful" project (Figure 4d), a very low RBCR of 0.9 and a high LCC of $750 M are the dominant features, colored in blue, that force the prediction lower, resulting in its correct classification. These local explanations make the model's reasoning transparent and directly link the abstract concept of "sustainability value" to the concrete prediction of project outcomes. The SHAP analysis, therefore, not only explains the superior performance of the Integrated Model but also provides compelling evidence that sustainability-focused metrics like resilience and life-cycle efficiency are not just "nice-to-haves" but are, in fact, the most powerful available predictors of green infrastructure project success.

Permutation-based importances (Figure 4e) and a PCA-logistic surrogate (Supplementary Table S3) yielded consistent ordering of top contributors, and SHAP rank stability was high across seeds (Supplementary Figure S2), indicating that the prominence of RBCR and LCC is not an artifact of random initialization or collinearity among the traditional financial metrics.

3.6. In-Depth Analysis of Key Predictors

The SHAP analysis identified the RBCR and LCC as the most influential predictors of project success. To further explore and visualize these critical relationships, a detailed correlational analysis was performed between the top-ranked predictors and the continuous PSI. This moves beyond the binary classification to examine the underlying dose–response relationship between these metrics and the nuanced spectrum of project performance. The results of this investigation are presented in a three-panel figure, Figure 5. This figure displays scatter plots of the PSI against the two most influential sustainability metrics and, for the purpose of comparison, the highest-ranked traditional metric. Figure 5a plots the relationship between the RBCR and the PSI. Figure 5b shows the relationship between the LCC and the PSI. Finally, Figure 5c illustrates the relationship between the NPV, a cornerstone of traditional finance, and the PSI, allowing for a direct visual comparison of their respective predictive strengths. Each plot includes a linear regression line to indicate the trend and the Pearson correlation coefficient (r) to quantify the strength and direction of the linear association.

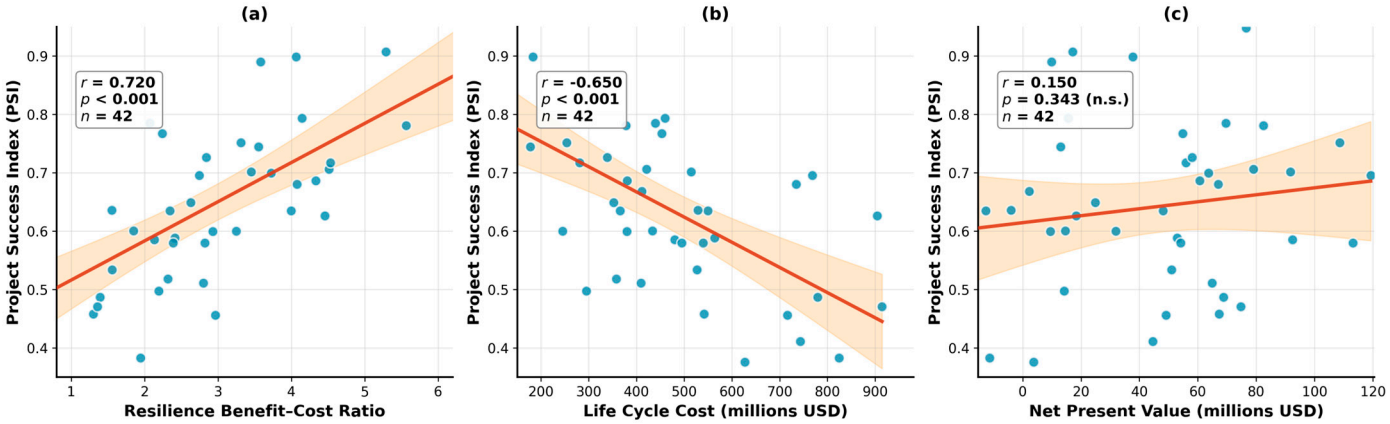


Figure 5. Relationship between Key Predictor Variables and the Project Success Index (PSI). (**a**) A scatter plot of Project Success Index vs. Resilience Benefit–Cost Ratio. (**b**) A scatter plot of Project Success Index vs. Life Cycle Cost (in millions of USD). (**c**) A scatter plot of Project Success Index vs. Net Present Value (in millions of USD).

The visual and statistical evidence presented in Figure 5 strongly corroborates the findings from the SHAP analysis and highlights the profound limitations of relying on traditional financial metrics alone. In Figure 5a, a strong, positive linear relationship is evident between the RBCR and the PSI. The upward-sloping regression line clearly indicates that as the RBCR increases, so does the overall success of the project. This relationship is statistically significant, with a Pearson correlation coefficient of r = 0.72 ($p < 0.001$), signifying a strong positive correlation. This finding provides quantitative evidence that projects designed with greater climate resilience, yielding higher long-term benefits relative to their costs, are significantly more likely to be successful across financial, schedule, and stakeholder dimensions.

Conversely, Figure 5b reveals a strong, negative linear relationship between the LCC and the PSI. The downward-sloping trend shows that projects with higher overall costs across their lifespan tend to be less successful. This association is also statistically significant, with a Pearson correlation coefficient of r = −0.65 ($p < 0.001$). This result underscores the importance of looking beyond initial capital expenditures; projects that are cheaper to operate, maintain, and decommission demonstrate markedly better overall performance.

The most striking result is presented in Figure 5c, which examines the relationship between NPV and project success. In stark contrast to the other two panels, the scatter plot shows no discernible pattern. The data points are widely dispersed, and the regression line

is nearly flat. The calculated Pearson correlation coefficient is r = 0.15, and the result is not statistically significant (p = 0.34). This lack of a significant correlation is a critical finding. It indicates that a project's NPV, the primary metric used in conventional investment appraisal, has virtually no linear relationship with its ultimate, multi-dimensional success as defined in this study. A project with a high NPV was almost as likely to be "Less Successful" as it was to be "Successful." This directly supports the qualitative findings from the expert interviews regarding "NPV Myopia" and provides a powerful visual and statistical confirmation that traditional metrics fail to capture the factors that truly drive the performance of green infrastructure projects.

4. Discussion

This study developed and empirically validated an integrated appraisal framework that fuses traditional financial indicators with sustainability-oriented measures and tests their joint predictive value for realized project performance. The approach demonstrably strengthens decision support for green infrastructure investments.

Empirically, the integrated model outperformed a traditional-metrics baseline across 10-fold cross-validation (AUC-ROC 0.88 vs. 0.71), indicating materially better separability of successful vs. less-successful projects. The gain arises from incorporating life-cycle efficiency, carbon externalities, and resilience—dimensions absent from discounted-cash-flow-only appraisal—and from using interpretable machine learning (SHAP) to expose feature-level drivers.

Perhaps the most significant insight is derived from the model interpretability analysis using SHAP. The finding that the RBCR and LCC were the two most influential predictors of success fundamentally reorders the hierarchy of project appraisal metrics. This indicates that, within this predictive model, information on long-term efficiency and risk mitigation carries the greatest predictive weight for PSI. These metrics are informative correlates of success in this dataset; they should not be interpreted as causal determinants.

Furthermore, the explicit analysis of the relationship between individual predictors and the PSI solidified this interpretation. The strong, statistically significant correlations found between PSI and both RBCR (r = 0.72) and LCC (r = −0.65) provide a clear, intuitive link between these sustainability concepts and tangible project outcomes. In stark contrast, the finding of a non-significant, near-zero correlation between NPV and PSI (r = 0.15) is a powerful negative result. It empirically demonstrates that a project's projected profitability, as calculated by conventional methods, is a remarkably poor proxy for its actual, realized success across cost, schedule, and stakeholder satisfaction dimensions. This suggests that the financial value captured by NPV may be decoupled from the operational and strategic value that leads to a successful project delivery and outcome, a critical disconnect for investors and policymakers aiming to allocate capital effectively.

The findings of this study both align with and significantly extend several streams of existing research. The demonstrated importance of the RBCR as the premier predictor of success provides empirical validation at the project level for the macro-level arguments made by scholars like Hallegatte et al. [29]. While their work established the high economic returns of investing in infrastructure resilience on a systemic basis, this study advances that knowledge by showing that a project-specific resilience metric is a powerful predictor of its overall performance, thereby bridging the gap between macroeconomic benefit and microeconomic project appraisal. Similarly, the foundational work on LCC by authors such as Woodward [47] and Gluch & Baumann [31] has long advocated for its use to achieve a more complete economic picture. This research moves beyond advocacy to empirical demonstration, confirming that LCC is not just a more comprehensive accounting tool

but a potent predictive indicator of project success, a linkage that has not been explicitly quantified in previous studies.

The development of a validated framework also represents a direct response to the barriers identified in the green finance literature. Researchers like Manoharan et al. [48] and reports from the Volz et al. [49] have consistently cited the lack of standardized metrics and transparent data as primary impediments to scaling up green investment. By proposing and testing a specific, integrated set of metrics, this study offers a potential solution to this problem, providing a template that can be adapted and standardized. The inclusion and validation of a monetized SCC, building on the valuation work of economists like Muhebwa and Osman [50], demonstrates a practical method for internalizing environmental externalities within a predictive project finance model, moving the concept from theoretical climate economics into applied investment analysis.

Three policy levers align with the SDG mapping summarized in Table 6. First, whole-life procurement: requiring LCC at tender evaluation internalizes operating/renewal/end-of-life costs and is directly actionable using Equation (1) and Supplementary Box S3. Second, climate-externality pricing: adopting a shadow price of carbon in appraisal (Equation (2); Figure S5; Table S7) ensures projects reflect the present value of emissions damages in line with SDG 13.2. Third, risk-informed adaptation finance: applying RBCR (Equation (3); Figure S6) directs capital to measures whose expected avoided losses exceed incremental resilience cost, supporting SDG 13.1 and urban disaster-risk reduction (SDG 11.5). Together, these levers operationalize the integrated model's insight that RBCR and LCC carry the greatest predictive weight for project success in this dataset.

Table 6. Framework–SDG–Policy Synergy Crosswalk (metrics as defined in Section 2.4; units in USD or ratios as noted).

Framework Metric	SDG 9 (Industry, Innovation and Infrastructure)	SDG 11 (Sustainable Cities and Communities)	SDG 13 (Climate Action)	Primary Policy Lever(s) Enabled
LCC (Equation (1))	9.4 (upgrade infrastructure for resource efficiency)	11.6, 11.b (whole-life urban environmental management)	13.2 (embed climate in planning via cost trajectories)	Whole-life procurement; PPP VfM tests using whole-life cost baselines
Monetized carbon cost (Equation (2))	9.4 (low-carbon technology choices)	11.6 (air-quality/energy footprint in cities)	13.2 (shadow carbon price in appraisal)	Shadow carbon pricing in public investment management; green-bond eligibility screens
RBCR (Equation (3))	9.1 (reliable, resilient infrastructure)	11.5 (disaster-loss reduction; service continuity)	13.1 (adaptive capacity; hazard-risk management)	Resilience investment prioritization; adaptation finance justification
Payback (years)	9.1 (delivery of viable infrastructure)	11.a (urban–regional linkages via quick returns)	Indirect	Capital budgeting threshold tests
IRR	9.1	Indirect	Indirect	Bankability screening; blended-finance structuring
NPV	9.1	Indirect	Indirect	Economic appraisal baseline; VfM counterfactuals

Table 6 organizes the framework–SDG crosswalk and the corresponding policy levers to facilitate adoption by appraisal units and financing institutions.

The primary divergence and contribution of this work lie in its methodological synthesis. Previous studies applying machine learning to infrastructure, such as Alhamami [51], have typically focused on predicting success based on general project characteristics (e.g., size, sector, country) rather than on the appraisal metrics themselves. This study, by contrast, uses the appraisal metrics as the predictive features, thereby testing the validity of the appraisal framework itself. Furthermore, the application of SHAP [32] to provide a transparent interpretation of the model's logic is a significant methodological advance for

this field. It addresses the "black box" criticism often leveled at machine learning models and provides the kind of granular, feature-level insight that is necessary for such a tool to be trusted and adopted by financial decision-makers. In essence, this study synthesizes four distinct areas—sustainable finance, resilience economics, life cycle analysis, and interpretable machine learning—to create and validate a novel decision support tool that has not been previously proposed or tested in the literature.

Despite the robustness of the findings, it is imperative to acknowledge the limitations of this research, which in turn provide fertile ground for future inquiry. First, the sample size of 42 projects, while sufficient for the analyses performed, is modest for a machine learning application. A larger dataset would enhance power, reduce the risk of overfitting, and improve generalizability; parameters trained here may require recalibration for projects not well represented in the sample. Second, the 2020–2024 completion window was chosen to ensure a common regulatory and delivery regime; accordingly, this study does not estimate pre- vs. post-COVID-19 differences and should not be interpreted as such.

Second, the study's geographical focus on the GCC region, while providing a homogenous context for analysis, inherently limits the direct transferability of the findings. The specific regionalized SCC value used and the nature of the climate risks informing the RBCR calculations are context-dependent. While the overarching framework—the principle of integrating these specific types of metrics—is hypothesized to be universally applicable, the quantitative model itself would need to be recalibrated and re-validated with local data before being applied in other regions, such as Europe or Southeast Asia, which face different climate hazards and possess different economic structures.

External application requires recalibration of Equation (3) by replacing hazard exceedance probabilities and damage-function coefficients with region-specific values. Formally, increases in annual event probability or intensity that scale expected avoided loss $B_{\text{resilience},t}$ raise RBCR approximately proportionally, whereas higher incremental resilience cost $C_{\text{resilience},t}$ lowers it; discount rate and horizon effects are summarized for this dataset in Table S7. For monsoonal/typhoon contexts (e.g., Southeast Asia), the hazard set would emphasize pluvial/cyclone winds and salinity intrusion; for temperate Europe, fluvial/pluvial flood and heat-stress combinations dominate. The same recipe holds: (i) construct baseline loss models by hazard; (ii) quantify mitigated loss under design measures; (iii) integrate discounted avoided losses and incremental costs as in Equation (3) to obtain RBCR uncertainty bands (cf. Figure S6). For future international validation, the World Bank PPI can be used to identify comparable sector–country cohorts; however, PPI records do not contain LCC/RBCR inputs and must be augmented with project-level O&M and hazard/damage data before horizontal verification.

Third, the research relies on a cross-sectional design, capturing data post-completion. This approach precludes the analysis of dynamic factors and relies on the accuracy of historical project documentation and respondent recall. Potential inconsistencies in how different organizations calculate metrics like LCC or estimate emissions could introduce noise into the dataset. A longitudinal study that tracks projects from appraisal through to operation would provide a more dynamic and potentially more accurate picture of how these metrics influence outcomes over time.

Fourth, the operationalization of the PSI as an equally weighted composite of cost, schedule, and stakeholder satisfaction is a methodological simplification. In practice, the relative importance of these three dimensions may vary significantly depending on the project type and the specific priorities of the key stakeholders. This uniform weighting, while ensuring objectivity, may not perfectly reflect the nuanced definition of "success" for every project in the sample.

Finally, while the case study component was used to triangulate data, a degree of subjectivity is unavoidable in metrics such as stakeholder satisfaction ratings and the estimation of benefits for the RBCR. Although standardized procedures were followed, these variables are not as objectively measurable as final cost or duration, introducing a potential source of measurement error.

A portion of LCC and RBCR inputs could only be reconstructed after completion from auditable documentation. This introduces potential hindsight bias if ex-post information caused systematic revisions to assumptions relative to the decision stage. This study therefore flags the provenance of each metric and provide stratified and complete-case sensitivity analyses (Supplementary Tables S1–S4). These checks mitigate, but do not eliminate, the possibility that retrospective estimation inflated associations for sustainability metrics.

Only completed projects were observed, which risks survivorship bias if unsuccessful projects are less likely to report or complete documentation. If survivorship disproportionately removes low-PSI observations with low sustainability metrics, associations could be attenuated; conversely, if documentation is more complete for better-resourced sponsors, associations could be overstated. Without data on canceled projects, this study cannot bound the bias; future work will incorporate canceled/failed cases via procurement and disclosure archives to adjust for attrition.

The findings and limitations of this study illuminate several promising avenues for future research. The most immediate and impactful extension would be to expand the dataset significantly, both in size and geographical scope. A large-scale international collaboration to collect data on hundreds of green infrastructure projects would enable the development of highly robust, region-specific predictive models and would allow for a definitive test of the framework's external validity across different economic and climatic contexts.

Future research should also aim to move beyond a cross-sectional design. A longitudinal approach, tracking a cohort of projects from the initial investment decision through their operational life, would be invaluable. This would allow researchers to investigate how the predictive accuracy of the framework changes over time and to identify critical intervention points where project performance begins to deviate from initial forecasts.

There is also a clear need to enhance the sophistication of the dependent and independent variables. Future studies could explore more complex, weighted formulations of the PSI, perhaps using multi-criteria decision analysis (MCDA) techniques like the Analytic Hierarchy Process (AHP) to derive stakeholder-defined weightings for the different performance dimensions. Moreover, the framework itself could be expanded to include a dedicated "social" dimension. The development and integration of quantifiable social metrics, such as a Social Return on Investment (SROI) analysis, measures of equitable access to project benefits, or impacts on local employment and well-being, represents the next frontier in creating a truly holistic triple-bottom-line appraisal tool.

From a practical standpoint, the implications of this research are direct and actionable. For investors, financial institutions, and government agencies, this study provides compelling empirical justification to move beyond legacy appraisal models. It makes a strong case for mandating the inclusion of LCC, RBCR, and carbon cost analyses in all green infrastructure investment proposals. The validated framework offers a ready-to-adapt template for developing new, more accurate investment screening and risk management protocols. For project managers and engineering firms, the findings highlight that designing for resilience and long-term operational efficiency is not a cost burden but a direct investment in the ultimate success of the project. Finally, this work can inform the development of national and international standards for green project finance, helping to create the

consistent, transparent, and reliable data ecosystem that is essential for mobilizing the vast sums of capital required to finance a global sustainable transition.

5. Conclusions

This study demonstrates that combining Life Cycle Costing, monetized carbon, and a Resilience Benefit–Cost Ratio with conventional finance materially improves the prediction of green infrastructure outcomes. Across matched folds, the integrated model achieved higher discrimination than a traditional metrics baseline (AUC-ROC 0.88 vs. 0.71) with superior accuracy and F1, and model explanations consistently ranked RBCR and LCC as the most informative features while NPV and IRR contributed least. These findings indicate that whole-life cost efficiency and climate resilience carry the information most aligned with realized delivery performance, whereas short-horizon return metrics risk systematic mispricing.

Three actions follow for investors and public owners. First, mandate whole-life and resilience metrics alongside discounted cash flow indicators in appraisal protocols. Second, adopt interpretable machine learning to quantify and explain how combined metrics map to outcomes, improving auditability of decisions. Third, when transferring the framework to new contexts, recalibrate with local data and extend with social metrics where possible. Taken together, these steps help align capital allocation with projects most likely to deliver durable value in practice.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su17219529/s1, Figure S1: DAG depicting hypothesized relations among exogenous context nodes, appraisal metrics, data provenance flags, and the Project Success Index; Figure S2: Violin plots of global feature ranks over 500 seed replications showing SHAP rank stability; Figure S3: ROC and PR curves for the Integrated model stratified by asset class; Figure S4: ROC and PR curves stratified by country; Figure S5: One-way sensitivities on monetized carbon cost to SCC, discount rate, and operating period; Figure S6: RBCR uncertainty bands under combined hazard probability and damage-function variation; Figure S7: Precision–recall curves for all 10 cross-validation folds across four predictive models; Figure S8: Fold-wise calibration reliability curves with Brier scores and Expected Calibration Error; Table S1: Metric provenance and sensitivity to retrospective reconstruction; Table S2: Predictor correlations and variance inflation factors; Table S3: PCA-logistic surrogate loadings and coefficients; Table S4: Stakeholder satisfaction measurement sensitivity analysis; Table S5: 10-fold CV metrics with paired comparisons; Table S6: Bootstrap validation of AUC-ROC differences; Table S7: Discount-rate sensitivity analysis; Table S8: Standardized analysis assumptions by asset class; Box S1: Worked example for monetized carbon cost calculation; Box S2: Worked example for resilience benefit–cost ratio calculation; Box S3: Worked example for life-cycle cost calculation.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Institutional Review Board Statement: Ethical review and approval were waived for this study in accordance with the Institutional Research Ethics Board (IRB) at Umm Al-Qura University which exempts qualitative and quantitative studies involving surveys or interviews from ongoing IRB oversight when subjects cannot be identified and disclosure would not place them at risk. In this work, all data were obtained from adult experts and project representatives; interviews were conducted via videoconference and audio-recorded with consent, and quotations are reported only in anonymized form. The survey collected project-level information and was analyzed in aggregate without personal identifiers.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written consent was obtained from interview participants prior to recording, and all quotations are reported in anonymized form. Survey participants provided informed consent through their

voluntary completion of the questionnaire. All procedures were conducted in accordance with institutional ethical guidelines.

Data Availability Statement: Data will be made available on request.

Acknowledgments: The author wishes to thank all interview participants and survey respondents for their valuable time and insights. The author also acknowledges the use of OpenAI ChatGPT (GPT-5 version) for language editing and formatting assistance. All substantive content and analysis remain the sole responsibility of the author.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. Source Categories and Their Role in Data Collection

Table A1. Documentary sources and usage in data extraction and triangulation.

Source Category	Representative Documents	Primary Use in This Study
Investment decision documentation (respondent-provided)	Financial models; investment memos; sanctioning packages	NPV, IRR, Payback; LCC cost elements and discounting assumptions; identification and costing of resilience measures; emissions baselines
Public procurement and award records	Tender notices; contract awards; technical schedules	Verification of project scope, asset class, commissioning year; cross-checks of budget/duration
Corporate disclosures	Annual reports; sustainability reports; ESG disclosures	Confirmation of operational status and capacity; emissions factors when disclosed
Case-study technical files	Risk registers; O&M plans; completion reports	Triangulation of survey inputs; avoided-loss estimates for RBCR; validation of LCC components

Note: All personally identifying information from interviews/surveys was excluded; only project-level information was retained and analyzed in aggregate.

References

1. Umoh, A.A.; Nwasike, C.N.; Tula, O.A.; Ezeigweneme, C.A.; Gidiagba, J.O. Green Infrastructure Development: Strategies for Urban Resilience and Sustainability. *World J. Adv. Res. Rev.* **2024**, *21*, 020–029. [CrossRef]
2. Ajirotutu, R.O.; Adeyemi, A.B.; Ifechukwu, G.-O.; Iwuanyanwu, O.; Ohakawa, T.C.; Garba, B.M.P. Designing Policy Frameworks for the Future: Conceptualizing the Integration of Green Infrastructure into Urban Development. *J. Urban Dev. Stud.* **2024**, *2*, 911–923.
3. Ajirotutu, R.O.; Adeyemi, A.B.; Ifechukwu, G.-O.; Iwuanyanwu, O.; Ohakawa, T.C.; Garba, B.M.P. Future Cities and Sustainable Development: Integrating Renewable Energy, Advanced Materials, and Civil Engineering for Urban Resilience. *Int. J. Sustain. Urban Dev.* **2024**, *3*, 235–250.
4. Ghanem, A.M.; Alamri, Y.A. The Impact of the Green Middle East Initiative on Sustainable Development in the Kingdom of Saudi Arabia. *J. Saudi Soc. Agric. Sci.* **2023**, *22*, 35–46. [CrossRef]
5. Tsani, S.; Najm, S. Green Policies and Sustainable Development in Saudi Arabia. In *Handbook of Sustainable Politics and Economics of Natural Resources*; Edward Elgar Publishing: Cheltenham, UK, 2021; pp. 123–134.
6. Alshuwaikhat, H.M.; Mohammed, I. Sustainability Matters in National Development Visions—Evidence from Saudi Arabia's Vision for 2030. *Sustainability* **2017**, *9*, 408. [CrossRef]
7. Alamoudi, A.K.; Abidoye, R.B.; Lam, T.Y. Implementing Smart Sustainable Cities in Saudi Arabia: A Framework for Citizens' Participation towards Saudi Vision 2030. *Sustainability* **2023**, *15*, 6648. [CrossRef]
8. Nassar, A.K. Strategic Energy Transition in the Gulf Cooperation Council: Balancing Economic, Social, Political, and Environmental Dynamics for Sustainable Development. *Int. J. Green Energy* **2025**, *22*, 1570–1586. [CrossRef]
9. Morshed, A. Sustainable Energy Revolution: Green Finance as the Key to the Arab Gulf States' Future. *Int. J. Energy Sect. Manag.* **2025**, *ahead of print*. [CrossRef]
10. Alhowaish, A.K. Green Municipal Bonds and Sustainable Urbanism in Saudi Arabian Cities: Toward a Conceptual Framework. *Sustainability* **2025**, *17*, 3950. [CrossRef]
11. Myronchuk, V.; Yatsenko, O.; Riznyk, D.; Hurina, O.; Frolov, A. Financing Sustainable Development: Analysis of Modern Approaches and Practices in the Context of Financial and Credit Activities. *Int. J. Econ. Financ. Issues* **2024**, *14*, 317. [CrossRef]
12. Kusz, D.; Kusz, B.; Wicki, L.; Nowakowski, T.; Kata, R.; Brejta, W.; Kasprzyk, A.; Barć, M. The Economic Efficiencies of Investment in Biogas Plants—A Case Study of a Biogas Plant Using Waste from a Dairy Farm in Poland. *Energies* **2024**, *17*, 3760. [CrossRef]

13. Akomea-Frimpong, I.; Jin, X.; Osei-Kyei, R. Mapping Studies on Sustainability in the Performance Measurement of Public-Private Partnership Projects: A Systematic Review. *Sustainability* **2022**, *14*, 7174. [CrossRef]
14. Arshad, H.; Thaheem, M.J.; Bakhtawar, B.; Shrestha, A. Evaluation of Road Infrastructure Projects: A Life Cycle Sustainability-Based Decision-Making Approach. *Sustainability* **2021**, *13*, 3743. [CrossRef]
15. Dobrowolski, Z.; Drozdowski, G. Does the Net Present Value as a Financial Metric Fit Investment in Green Energy Security? *Energies* **2022**, *15*, 353. [CrossRef]
16. Zijlstra, N. A Circular Life Cycle Cost Model: Quantifying the Financial Implications of Projected Level of Circularity in Real Estate Development Projects. Master's Thesis, Delft University of Technology, Delft, The Netherlands, 2023.
17. Ni, T.; Zhang, X.; Leng, P.; Pelling, M.; Xu, J. Comprehensive Benefits Evaluation of Low Impact Development Using Scenario Analysis and Fuzzy Decision Approach. *Sci. Rep.* **2025**, *15*, 2227. [CrossRef]
18. Majekolagbe, A.; Seck, S.L.; Wright, D.V. Exploring the Application of the Social Cost of Carbon in Loss-and-Damage and Impact Assessment. *Clim. Law* **2023**, *13*, 83–117. [CrossRef]
19. Lee, S.-T.; Jung, S.-M. Monetizing Environmental Impacts of Nuclear Power: Cost-Benefit Analysis Using Social Cost of Carbon. *Sage Open* **2024**, *14*, 21582440241301468. [CrossRef]
20. Caesary, D.; Kim, H.; Nam, M.J. An Alternative Approach to Capture Uncertainties Embedded in the Estimation of Social Cost of Carbon. *WIREs Energy Environ.* **2023**, *12*, e475. [CrossRef]
21. Helgeson, J.; O'Fallon, C. Resilience Dividends and Resilience Windfalls: Narratives That Tie Disaster Resilience Co-Benefits to Long-Term Sustainability. *Sustainability* **2021**, *13*, 4554. [CrossRef] [PubMed]
22. Reddy, K.R.; Janga, J.K.; Verma, G.; Nagaraja, B. Integration and Quantification of Resilience and Sustainability in Engineering Projects. *J. Indian Inst. Sci.* **2024**, *104*, 435–488. [CrossRef]
23. Shamim, M.I.; Hamid, A.B.b.A.; Nyamasvisva, T.E.; Bin Rafi, N.S. Advancement of Artificial Intelligence in Cost Estimation for Project Management Success: A Systematic Review of Machine Learning, Deep Learning, Regression, and Hybrid Models. *Modelling* **2025**, *6*, 35. [CrossRef]
24. Ayubi Rad, M.; Ayubirad, M.S. Comparison of Artificial Neural Network and Coupled Simulated Annealing Based Least Square Support Vector Regression Models for Prediction of Compressive Strength of High-Performance Concrete. *Sci. Iran.* **2017**, *24*, 487–496. [CrossRef]
25. Hassija, V.; Chamola, V.; Mahapatra, A.; Singal, A.; Goel, D.; Huang, K.; Scardapane, S.; Spinelli, I.; Mahmud, M.; Hussain, A. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cogn. Comput.* **2024**, *16*, 45–74. [CrossRef]
26. Timilsina, M.S.; Sen, S.; Uprety, B.; Patel, V.B.; Sharma, P.; Sheth, P.N. Prediction of HHV of Fuel by Machine Learning Algorithm: Interpretability Analysis Using Shapley Additive Explanations (SHAP). *Fuel* **2024**, *357*, 129573. [CrossRef]
27. Nohara, Y.; Matsumoto, K.; Soejima, H.; Nakashima, N. Explanation of Machine Learning Models Using Shapley Additive Explanation and Application for Real Data in Hospital. *Comput. Methods Programs Biomed.* **2022**, *214*, 106584. [CrossRef] [PubMed]
28. Ayat, M.; Kim, B.; Kang, C.W. A New Data Mining-Based Framework to Predict the Success of Private Participation in Infrastructure Projects. *Int. J. Constr. Manag.* **2023**, *23*, 2151–2159. [CrossRef]
29. Hallegatte, S.; Rozenberg, J.; Maruyama Rentschler, J.E.; Nicolas, C.M.; Fox, C.J.E. *Strengthening New Infrastructure Assets: A Cost-Benefit Analysis*; World Bank Policy Research Working Paper; World Bank: Washington, DC, USA, 2019.
30. Harker Steele, A.; Warner, T.; Vikara, D.; Guinan, A.; Balash, P. Comparative Analysis of Carbon Capture and Storage Finance Gaps and the Social Cost of Carbon. *Energies* **2021**, *14*, 2987. [CrossRef]
31. Gluch, P.; Baumann, H. The Life Cycle Costing (LCC) Approach: A Conceptual Discussion of Its Usefulness for Environmental Decision-Making. *Build. Environ.* **2004**, *39*, 571–580. [CrossRef]
32. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef] [PubMed]
33. Khan, K.I.; Mata, M.N.; Martins, J.; Nasir, A.; Dantas, R.M.; Correia, A.B.; Saghir, U.S. Impediments of Green Finance Adoption System: Linking Economy and Environment. Impediments of green finance adoption system: Linking economy and environment. *Emerg. Sci. J.* **2022**, *6*, 217–237. [CrossRef]
34. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y.; Cho, H.; Chen, K.; Mitchell, R.; Cano, I.; Zhou, T. *Xgboost: Extreme Gradient Boosting. R Package, Version 0.4-2*; R Project: Vienna, Austria, 2015; Volume 1, pp. 1–4.
35. Mihály, K.; Kulcsár, G. A New Many-Objective Hybrid Method to Solve Scheduling Problems. *Int. J. Ind. Eng. Manag.* **2023**, *14*, 326–335. [CrossRef]
36. Parums, D.V. Review Articles, Systematic Reviews, Meta-Analysis, and the Updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 Guidelines. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **2021**, *27*, e934475-1. [CrossRef] [PubMed]
37. Xie, H.; Cui, Q.; Li, Y. Net Present Value Method: A Method Recommended by ISO 15686-5 for Economic Evaluation of Building Life Cycle Costs. *World J. Eng. Technol.* **2022**, *10*, 224–229. [CrossRef]

38. Frias, A.; Água, P.B.; Lopes, B.M.; Melo, P.S. Life Cycle Cost Analysis of Complex Systems: An Application to Shipbuilding. *Int. J. Ind. Eng. Manag.* **2022**, *13*, 215–224. [CrossRef]
39. Bhatt, A.; Ravi, V.; Zhang, Y.; Heath, G.; Davis, R.; Tan, E.C.D. Emission Factors of Industrial Boilers Burning Biomass-Derived Fuels. *J. Air Waste Manag. Assoc.* **2023**, *73*, 241–257. [CrossRef]
40. Mack, J.; Dean, G.A.; Wathne, L.G. Improving Pavement Resiliency to Flooding: A Case for Concrete Pavement. In Proceedings of the Airfield and Highway Pavements 2023, Austin, TX, USA, 14–17 June 2023; American Society of Civil Engineers: Austin, TX, USA, 2023; pp. 135–146.
41. Zhu, Q.; Leibowicz, B.D. A Markov Decision Process Approach for Cost-Benefit Analysis of Infrastructure Resilience Upgrades. *Risk Anal.* **2022**, *42*, 1585–1602. [CrossRef]
42. Osei-Kyei, R.; Chan, A.P. Evaluating the Project Success Index of Public-Private Partnership Projects in Hong Kong: The Case of the Cross Harbour Tunnel. *Constr. Innov.* **2018**, *18*, 371–391. [CrossRef]
43. Osei-Kyei, R.; Chan, A.P.C. Developing a Project Success Index for Public–Private Partnership Projects in Developing Countries. *J. Infrastruct. Syst.* **2017**, *23*, 04017028. [CrossRef]
44. Natras, R.; Soja, B.; Schmidt, M. Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting. *Remote Sens.* **2022**, *14*, 3547. [CrossRef]
45. Sahin, E.K. Assessing the Predictive Capability of Ensemble Tree Methods for Landslide Susceptibility Mapping Using XGBoost, Gradient Boosting Machine, and Random Forest. *SN Appl. Sci.* **2020**, *2*, 1308. [CrossRef]
46. Chai, Y.; Liang, R.; Samtani, S.; Zhu, H.; Wang, M.; Liu, Y.; Jiang, Y. Additive Feature Attribution Explainable Methods to Craft Adversarial Attacks for Text Classification and Text Regression. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 12400–12414. [CrossRef]
47. Woodward, D.G. Life Cycle Costing—Theory, Information Acquisition and Application. *Int. J. Proj. Manag.* **1997**, *15*, 335–344. [CrossRef]
48. Manoharan, G.; Nithya, G.; Razak, A.; Sharma, S.; Ashtikkar, S.P. Challenges and Opportunities in Green Finance: Overcoming Barriers and Scaling up Sustainable Investments. In *Practical Approaches to Agile Project Management*; IGI Global: Hershey, PA, USA, 2024; pp. 398–413.
49. Volz, U.; Lo, Y.C.; Mishra, V. *Scaling up Green Investment in the Global South: Strengthening Domestic Financial Resource Mobilisation and Attracting Patient International Capital*; SOAS Centre: London, UK, 2024. [CrossRef]
50. Muhebwa, A.; Osman, K.K. A Behavioral Finance Framework for Balancing AI Accuracy and Operational Carbon Emissions. *ACM J. Comput. Sustain. Soc.* **2025**, *3*, 1–20. [CrossRef]
51. Alhamami, A.H. Implementation of Machine Learning-Based Risk Prediction Models for Large-Scale Infrastructure Construction Projects in Urban Environments. *Cad. Educ. Tecnol. Soc.* **2025**, *18*, 330–346. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.