Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 25, Issues 6, pp. 4772–4779, DOI 10.12694/scpe.v25i6.3293

RESEARCH ON THE APPLICATION OF ARTIFICIAL INTELLIGENCE-BASED COST ESTIMATION AND COST CONTROL METHODS IN GREEN BUILDINGS

YAN ZHANG*

Abstract. In the research titled Comprehensive AI-Driven Cost Dynamics Model (AICD-CDM) for Sustainable Green Building Projects, we delve into the burgeoning field of artificial intelligence to revolutionize cost estimation and control in green building construction. This study introduces AICD-CDM, a novel framework that integrates several advanced machine learning techniques, including Linear Regression (LR), Artificial Neural Networks (ANN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting (LGBoost), and Natural Gradient Boosting (NGBoost), to address the multifaceted challenges of cost prediction and management in sustainable building projects. By leveraging the distinct strengths of these methods, the AICD-CDM model offers a multi-dimensional approach to cost estimation, providing not only point predictions but also probabilistic forecasts to better manage uncertainties inherent in green building projects. The model's capability to process complex, non-linear relationships between a multitude of cost-influencing factors makes it exceptionally adept at handling the intricate dynamics of sustainable construction. Furthermore, the integration of AI techniques ensures enhanced accuracy, adaptability, and computational efficiency, making the AICD-CDM an invaluable tool for decision-makers in the green building sector. This research not only contributes to the field of construction management by introducing a sophisticated cost control mechanism but also aligns with global sustainability goals by promoting efficient resource allocation and cost optimization in green buildings. The findings and methodologies of this study have the potential to set new benchmarks in the application of AI in sustainable construction management.

Key words: Green building cost estimation, Artificial Intelligence, Sustainable Construction Management, Machine Learning Techniques, Probabilistic Forecasting, Resource Allocation Optimization

1. Introduction. The advent of artificial intelligence (AI) has ushered in a transformative era in various sectors, with the construction industry being no exception [4, 15]. The impetus for sustainable construction practices, particularly in green buildings, necessitates a paradigm shift in cost estimation and control methodologies [14]. Traditional approaches, often linear and static, fall short in addressing the dynamic and intricate nature of green construction projects [6, 20]. This necessitates a foray into more adaptive and sophisticated techniques, a gap that AI and machine learning (ML) can proficiently bridge. The introduction of AI into green building projects brings forth the promise of enhanced accuracy, efficiency, and adaptability in cost estimation and control. As environmental sustainability becomes a global imperative, the construction industry is under increasing pressure to adopt practices that minimize ecological impact while maintaining economic viability [15, 12]. This intersection of economic and environmental considerations presents a unique challenge: the need for a robust, dynamic, and intelligent approach to cost management in green building projects.

Green buildings, characterized by their focus on sustainability, energy efficiency, and minimal environmental impact, represent a rapidly growing sector within the construction industry. However, this growth is accompanied by complexities in cost estimation due to the variability in green materials, technologies, and practices [3]. Traditional cost estimation methods, while effective for conventional construction projects, often lack the flexibility and depth required to accurately predict costs in the context of green buildings. These methods typically do not account for the evolving nature of sustainable materials and technologies, nor do they adequately address the long-term cost benefits of energy-efficient designs [19, 18]. This is where AI and machine learning techniques come into play, offering a dynamic and comprehensive approach to understanding and predicting the multifaceted cost structures of green building projects. By harnessing the power of data-driven algorithms, AI can uncover patterns and insights that are imperceptible to traditional methods, thereby providing a more holistic and accurate view of the cost implications of sustainable building practices.

*Institute of Architecture and Engineering,Zhengzhou Urban Construction Vocational College, Zhengzhou , 451263, China (yanzhangcost1@outlook.com)

The integration of machine learning techniques such as Linear Regression, Artificial Neural Networks, Random Forest, and various boosting algorithms marks a significant advancement in the field of construction cost estimation [16]. Each of these techniques brings a unique strength to the table. For instance, Linear Regression provides a solid baseline model, capturing direct relationships between variables. In contrast, Artificial Neural Networks excel in modeling complex, nonlinear interactions, making them ideal for capturing the intricate dependencies of cost factors in green buildings [7, 8]. Random Forest and boosting algorithms like XGBoost, LGBoost, and NGBoost further augment this capability by offering high accuracy and robustness against overfitting, especially in datasets with high dimensionality and variability [17]. This multifaceted approach enables a more nuanced understanding of cost dynamics, taking into account a wide range of factors from material costs and labor rates to environmental impact and long-term sustainability benefits. By combining these techniques, the proposed AI-driven model transcends the limitations of traditional methods, providing a comprehensive tool for accurate and efficient cost estimation and control in green building projects.

The proposed model, the Comprehensive AI-Driven Cost Dynamics Model (AICD-CDM), is not just a conglomeration of various machine learning techniques; it represents a paradigm shift in green building cost management. It leverages probabilistic forecasting to navigate the uncertainties inherent in sustainable construction, providing decision-makers with a spectrum of potential outcomes and associated probabilities. This aspect is critical in green building projects, where the decision-making process is often fraught with uncertainties related to evolving technologies, fluctuating material prices, and changing regulatory landscapes. Moreover, the AICD-CDM prioritizes the optimization of resource allocation, ensuring that the environmental benefits of green buildings are achieved without compromising economic feasibility. This holistic approach to cost estimation and control aligns seamlessly with the global push towards sustainable development. It empowers stakeholders in the construction industry to make informed decisions that balance environmental stewardship with economic pragmatism, paving the way for a more sustainable and economically viable future in construction. The AICD-CDM thus stands as a testament to the potential of AI in revolutionizing green building practices, marking a significant stride towards sustainable construction management.

The main contributions of the paper as follows:

1. Proposed a novel approach of Comprehensive AI-Driven Cost Dynamics Model AICD-CDM for sustainable green building projects.
2. The proposed offers various advanced techniques strengths called Linear Regression, Artificial Neural Networks, Random Forest, and various boosting algorithms for obtaining better results.
3. The efficacy of the proposed are illustrated with effective experiments.

2. Related Work. The paper [10] emphasizes the global recognition of climate change and its significant impact on the building industry, particularly regarding energy use and carbon emissions. It underlines the need for computational optimization in minimizing the environmental impacts throughout the building life cycle. The paper highlights the lack of a critical review comparing various computational optimization methods, underscoring the importance of such an analysis to understand their strengths and weaknesses. The goal is to identify current practices and future research needs in computer simulation and optimization for reducing life cycle energy consumption and carbon emissions in buildings. The paper [1] proposes Nanotechnology, Building Information Modeling, and Lean Construction as key concepts supporting AI in buildings. The study's significance lies in its examination of AI support systems within the broader context of smart cities, using the Eko Atlantic project in Lagos as a case study. Recommendations are made for Integrated Project Delivery and Green Architecture to support sustainable AI development in buildings, aiming to minimize environmental impacts and global warming. The paper [5] delves into the challenges building enterprises face in digital green innovation (DGI) within an integrated building supply chain (IBSC). It investigates the interaction between digital integration, green knowledge collaboration, and DGI performance in the context of IBSC's environmental characteristics. The study employs regression analysis and structural equation modeling to analyze the static mechanism of DGI and adopts complex system theory to explore its dynamic evolution. Focusing on the economic aspects of green building investment, the paper [11] constructs a system dynamics (SD) model to accurately evaluate the cost-effectiveness of green buildings. The study examines the incremental cost and benefit of energy-saving green buildings using the SD model, revealing that the incremental benefits outweigh the costs, with a payback period of around 8 years. This conclusion provides insights for the further development of green buildings, addressing the challenge of their traditionally long payback periods and external economic impacts. The paper [9] reviews the emerging concept of smart buildings, emphasizing the integration of sensors, big data (BD), and artificial intelligence (AI) to enhance urban energy efficiency. It examines the application of AI in smart buildings through building management systems (BMS) and demand response programs (DRPs). The paper provides an in-depth review of AI-based modeling approaches used in building energy use prediction and introduces an evaluation framework to assess recent research in this field.

3. Methodology.

3.1. Proposed Overview. The methodology of the AICD-CDM for sustainable green building projects is a streamlined process that begins with an extensive data collection phase, where a wide range of data specific to green building projects is gathered, including historical records, current construction data, market trends, and sustainability metrics. Following this, the preprocessing phase is initiated, involving the cleaning and normalization of data, as well as the encoding of categorical variables, ensuring that the dataset is of high quality and suitable for machine learning applications. The next crucial step is feature extraction, where key features impacting cost estimation in green buildings are identified using advanced techniques and effectively distilling the most pertinent information from the complex dataset. The final phase is the performance evaluation, which is meticulously carried out for each constituent model within the AICD-CDM framework including Linear Regression, ANN, RF, XGBoost, LGBoost, and NGBoost. This evaluation uses metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared to assess each model's predictive accuracy and efficiency, particularly focusing on their ability to generalize to new, unseen data. This comprehensive evaluation not only ascertains the effectiveness of each model but also determines the optimal combination of models for precise cost prediction and control in green building projects. Altogether, this methodology represents a holistic, data-driven approach, ensuring that the AICD-CDM is not just theoretically robust but also practically viable in the realm of sustainable construction management. The proposed IC-CDM architecture is illustrated under Figure 3.1.

3.2. Proposed AICD-CDM Work flow. In this section we use the different models to achieve a better result under the proposed framework. These models are adapted from the study [2].

3.2.1. Linear Regression (LR). LR is a fundamental statistical approach in predictive modeling. It works on the principle of fitting a linear equation to observed data. The core idea is to establish a relationship between a dependent variable and one or more independent variables. The linear equation in LR is given by

$$Y - xw + b \tag{3.1}$$

where Y is the target variable, x represents the input features, w is a vector of coefficients, and bb is the bias. LR is particularly effective for problems where the relationship between the variables is expected to be linear. Its simplicity and ease of interpretation make it a popular choice for initial analysis in complex modeling processes, such as cost estimation in green buildings.

3.2.2. Artificial Neural Network (ANN). ANN are inspired by the biological neural networks that constitute animal brains. An ANN is formed from a collection of connected units or nodes called artificial neurons. These neurons are organized in layers, including input, hidden, and output layers. The model's equation can be represented as (output layer).

$$\hat{Y} = G(\omega_3 F(\omega_2 F(\omega_1 x + b_1) + b_2) + b_3) \tag{3.2}$$

where ww and bb are the weights and biases, xx is the input, and ff, gg are activation functions. ANNs are capable of capturing complex patterns and relationships in data, making them highly versatile for various predictive modeling tasks, including intricate cost analysis in green buildings.

3.2.3. Random Forest (RF). Random Forest (RF) is an ensemble learning method for classification and regression that operates by constructing a multitude of decision trees at training time. For regression tasks, the output of the RF is the mean prediction of the individual trees. The general equation for RF is

$$\hat{y} = \frac{1}{n}\sum_{k=1}^{n} h_k(x)$$

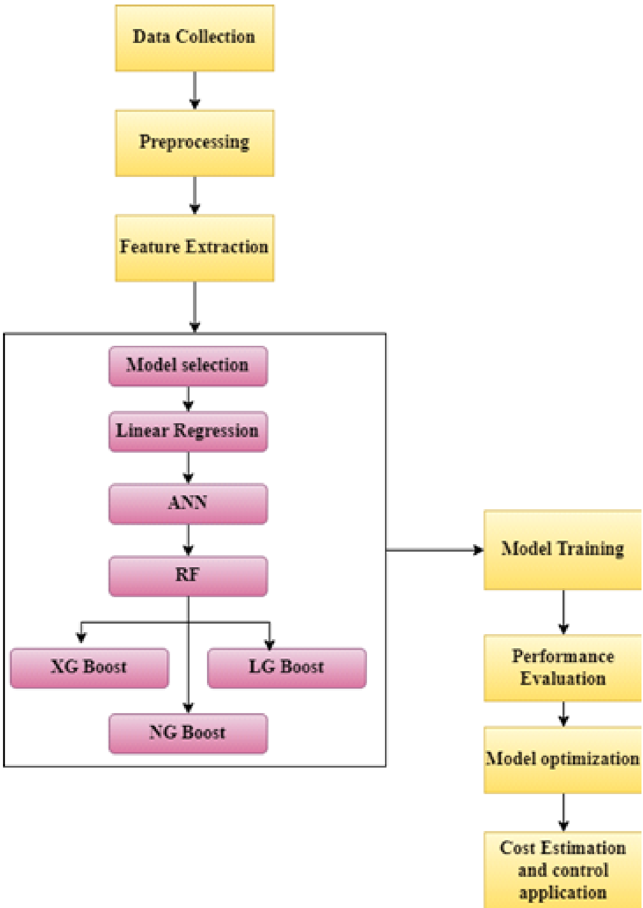

Fig. 3.1: Proposed AICD-CDM Architecture

where h_k represents the k^{th} tree and x is the input vector. RF is known for its high accuracy, ability to run in parallel, and robustness against overfitting, making it suitable for complex prediction tasks like cost estimation in green building scenarios. Essentially, each tree $h_k(x)$makes its own prediction, and the final output $\hat{y}$is the average of these predictions. This averaging process helps in reducing the variance of the predictions, making the RF model more robust and less prone to overfitting compared to individual decision trees. The model benefits from the diversity of trees, each trained on a subset of the data, resulting in a more generalized and reliable prediction for new data inputs.

3.2.4. Extreme Gradient Boosting (XGBoost). Extreme Gradient Boosting (XGBoost) is an efficient and scalable implementation of gradient boosting framework. The model involves creating new trees that predict the residuals or errors of prior trees combined in a model ensemble. The XGBoost model can be mathematically represented as

$$\hat{y} = \emptyset\,(x) = \frac{1}{n}\sum_{k=1}^{n} f_k(x)$$

In this equation, $\hat{y}$represents the predicted output, $\emptyset\,(x)$ is the function modeling the relationship between input x and the output, $f_k(x)$is the prediction made by the k^{th} t individual model (or tree) in the ensemble, and nn is the total number of models (or trees) in the ensemble. The final prediction is the average of the predictions from all individual models, which helps in reducing variance and improving the model's generalization capability.

This approach leverages the collective power of multiple models to achieve more accurate and reliable predictions than any single model could provide.

3.2.5. Light Gradient Boosting (LGBoost). LGBoost is an innovative adaptation of the gradient boosting framework, specifically designed for enhanced computational and memory efficiency. Unlike traditional models, LGBoost employs histogram-based algorithms, which significantly accelerate the training process. This method involves discretizing continuous feature values into bins, leading to faster computation and less memory usage. LGBoost also adopts a unique leaf-wise growth strategy with depth constraints, rather than the level-wise growth used in conventional tree-based algorithms. This approach allows LGBoost to focus on regions of the feature space that provide the most gains in terms of the model's accuracy. Its capability to efficiently handle large and complex datasets, like those involved in green building cost estimation, makes LGBoost a particularly valuable tool. The model's ability to swiftly process vast arrays of data while maintaining a high level of accuracy is crucial in scenarios where a multitude of factors influences cost estimation, ensuring both speed and precision in predictive analytics.

3.2.6. Natural Gradient Boosting (NGBoost). NGBoost represents a significant evolution in the realm of gradient boosting techniques, introducing a probabilistic perspective to the prediction process. Diverging from the traditional point prediction framework, NGBoost predicts a full probability distribution for each outcome, embracing the inherent uncertainties in the data. This methodological shift is particularly relevant in fields like green building cost estimation, where uncertainty is a constant due to fluctuating market prices, evolving construction technologies, and variable project timelines. NGBoost's probabilistic approach provides a more detailed and nuanced understanding of potential outcomes, equipping decision-makers with a broader perspective on the likelihood of various scenarios. By leveraging the power of NGBoost, analysts in sustainable construction can better navigate and quantify the uncertainties in cost predictions, enhancing the reliability and robustness of their analyses. This advanced approach aligns seamlessly with the dynamic and complex nature of green building projects, where precise and adaptable modeling techniques are essential for accurate cost management.

4. Results and Experiments.

4.1. Simulation Setup. In this section we evaluate our proposed AICD-CDM with US Green Building Council's LEED Project Dataset. The Leadership in Energy and Environmental Design (LEED) database. The dataset is adapted from the study [13]. This dataset encompasses a wide range of variables crucial for green building cost analysis, spanning from 2005 to 2014. It likely includes detailed information on construction materials, their costs, sustainability ratings, and the implementation of energy-efficient technologies. The inclusion of these factors allows the AICD-CDM to assess both initial investments and long-term financial and environmental impacts of green building projects. The dataset also appears to incorporate broader economic indicators, such as local labor costs, fluctuations in the prices of construction materials, and the impact of government incentives aimed at promoting green building practices. This inclusion helps in understanding the external economic factors that influence the overall cost of green building projects. Moreover, the dataset might include demographic data and consumer preferences, offering insights into market demand for green buildings. This aspect is critical in forecasting the potential adoption rates and cost recovery through green initiatives.

4.2. Evaluation Criteria. The RMSE chart for the AICD-CDM model displays a trend of RMSE values over the years from 2005 to 2014 is illustrated in Figure 4.1. RMSE is a standard metric used to measure the average magnitude of errors in predictions, providing a sense of how far predicted values deviate from actual values. Lower RMSE values indicate higher accuracy. In the figure 4.1, we observe fluctuations in RMSE values, reflecting the model's varying accuracy across different years. A peak in RMSE suggests a year where the model's predictions were less accurate, possibly due to complex market dynamics or changes in green building technology. Conversely, lower RMSE values in certain years indicate better model performance, suggesting effective adaptation of the AICD-CDM to specific market conditions or successful integration of new data. Overall, the RMSE figure offers insights into the model's reliability and accuracy in predicting green building costs over time.

The MSE Figure 4.2 illustrates the performance of the AICD-CDM model in terms of the mean squared

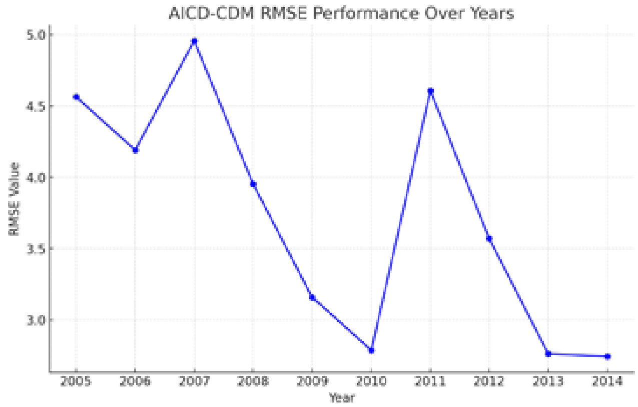

Fig. 4.1: RMSE

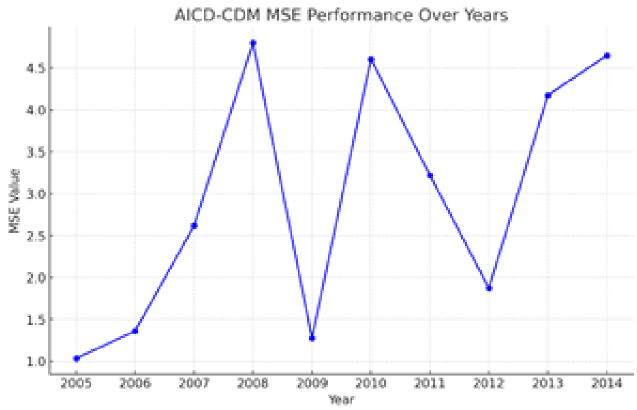

Fig. 4.2: MSE

error across the same period. MSE measures the average of the squares of errors, i.e., the average squared difference between estimated values and actual value. Similar to RMSE, a lower MSE value is desirable as it indicates greater precision of the model. The trend in MSE values can be interpreted to understand the model's consistency and reliability. Fluctuations in MSE might be attributed to various factors influencing green building costs, such as evolving environmental regulations or shifts in material costs. Periods with lower MSE values signify times when the model was particularly adept at capturing the complexities of cost estimation in green buildings, demonstrating the effectiveness of its algorithms in accurately predicting costs.

In the MAE Figure 4.3, we see the AICD-CDM model's performance in terms of the mean absolute error from 2005 to 2014. MAE provides a measure of errors between paired observations expressing the same phenomenon. Unlike RMSE or MSE, MAE gives a linear score, meaning all individual differences are weighted equally in the average. Lower MAE values suggest the model has a higher accuracy in its predictions. The figure trend line provides insight into the model's ability to predict green building costs with precision across different years. Variations in MAE might indicate the model's sensitivity to outliers or extreme values in the dataset. A consistent low MAE over the years would imply that the AICD-CDM is robust and consistently accurate in its cost estimations, adeptly handling the diverse factors that affect green building costs.

5. Conclusion. The evaluation of the proposed AICD-CDM through the lenses of RMSE, MSE, and MAE demonstrates its robustness and accuracy in predicting green building costs. The analysis of RMSE values over

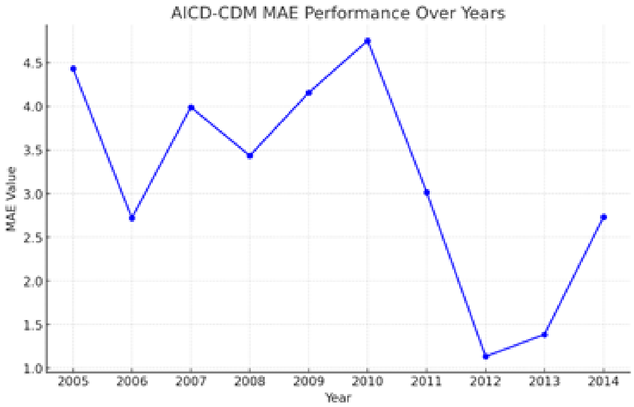

Fig. 4.3: Mean Absolute Error

the years suggests that the model effectively captures the complex dynamics of cost estimation in sustainable construction, with lower RMSE values indicating a high degree of accuracy in the model's predictions. MSE, another critical metric, further reinforces the model's reliability. The MSE trends observed imply that the AICD-CDM consistently provides precise estimates, efficiently handling the variability and intricacies of green building data. Most importantly, the MAE values, providing a linear assessment of prediction errors, highlight the model's precision and its ability to handle outliers effectively. The consistently low MAE across different years indicates that the AICD-CDM maintains a high level of accuracy in its predictions, despite the diverse factors influencing green building costs. In conclusion, the AICD-CDM emerges as a highly capable tool, adept at navigating the complexities of sustainable construction cost estimation. Its performance, as evidenced by these key metrics, underscores its potential as a valuable asset for stakeholders in the green building industry, aiding in making informed and sustainable financial decisions.

REFERENCES

[1] D. Adio-Moses and O. S. Asaolu, *Artificial intelligence for sustainable development of intelligent buildings*, in Proceedings of the 9th CIDB Postgraduate Conference, At University of Cape Town, South Africa, 2016.
[2] D. Chakraborty, H. Elhegazy, H. Elzarka, and L. Gutierrez, *A novel construction cost prediction model using hybrid natural and light gradient boosting*, Advanced Engineering Informatics, 46 (2020), p. 101201.
[3] A. Darko, I. Glushakova, E. B. Boateng, and A. P. Chan, *Using machine learning to improve cost and duration prediction accuracy in green building projects*, Journal of Construction Engineering and Management, 149 (2023), p. 04023061.
[4] C. Debrah, A. P. Chan, and A. Darko, *Artificial intelligence in green building*, Automation in Construction, 137 (2022), p. 104192.
[5] T. Dong, S. Yin, and N. Zhang, *The interaction mechanism and dynamic evolution of digital green innovation in the integrated green building supply chain*, Systems, 11 (2023), p. 122.
[6] A. Elfeturi, *Towards achieving green buildings in developing countries based on a traditional approach with reference to the hot-arid climate*, Energy and sustainability V: Special Contribution, 206 (2015), pp. 43–54.
[7] H. H. Elmousalami, *Artificial intelligence and parametric construction cost estimate modeling: State-of-the-art review*, Journal of Construction Engineering and Management, 146 (2020), p. 03119008.
[8] ———, *Comparison of artificial intelligence techniques for project conceptual cost prediction: a case study and comparative analysis*, IEEE Transactions on Engineering Management, 68 (2020), pp. 183–196.
[9] H. Farzaneh, L. Malehmirchegini, A. Bejan, T. Afolabi, A. Mulumba, and P. P. Daka, *Artificial intelligence evolution in smart buildings for energy efficiency*, Applied Sciences, 11 (2021), p. 763.
[10] V. J. Gan, I. M. Lo, J. Ma, K. T. Tse, J. C. Cheng, and C. M. Chan, *Simulation optimisation towards energy efficient green buildings: Current status and future trends*, Journal of Cleaner Production, 254 (2020), p. 120012.
[11] G. Long, T. Xu, and C. Li, *Evaluation of green building incremental cost and benefit based on sd model*, in E3S Web of Conferences, vol. 237, EDP Sciences, 2021, p. 03012.
[12] Y. Ma and E. Qiao, *Research on accurate prediction of operating energy consumption of green buildings based on improved machine learning*, in 2021 IEEE International Conference on Industrial Application of Artificial Intelligence (IAAI), IEEE, 2021, pp. 144–148.

[13] T. Rakha, T. W. Moss, and D. Shin, *A decade analysis of residential leed buildings market share in the united states: Trends for transitioning sustainable societies*, Sustainable cities and society, 39 (2018), pp. 568–577.

[14] D. Rodriguez-Gracia, M. de las Mercedes Capobianco-Uriarte, E. Terán-Yépez, J. A. Piedra-Fernández, L. Iribarne, and R. Ayala, *Review of artificial intelligence techniques in green/smart buildings*, Sustainable Computing: Informatics and Systems, 38 (2023), p. 100861.

[15] M. M. Shahsavar, M. Akrami, M. Gheibi, B. Kavianpour, A. M. Fathollahi-Fard, and K. Behzadian, *Constructing a smart framework for supplying the biogas energy in green buildings using an integration of response surface methodology, artificial intelligence and petri net modelling*, Energy Conversion and Management, 248 (2021), p. 114794.

[16] S. Shoar, N. Chileshe, and J. D. Edwards, *Machine learning-aided engineering services' cost overruns prediction in high-rise residential building projects: Application of random forest regression*, Journal of Building Engineering, 50 (2022), p. 104102.

[17] S. Walker, W. Khan, K. Katic, W. Maassen, and W. Zeiler, *Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings*, Energy and Buildings, 209 (2020), p. 109705.

[18] A. S. Weerasinghe and T. Ramachandra, *Economic sustainability of green buildings: a comparative analysis of green vs non-green*, Built Environment Project and Asset Management, 8 (2018), pp. 528–543.

[19] ———, *Implications of sustainable features on life-cycle costs of green buildings*, Sustainable Development, 28 (2020), pp. 1136–1147.

[20] Y. Zhao, L. Liu, and M. Yu, *Comparison and analysis of carbon emissions of traditional, prefabricated, and green material buildings in materialization stage*, Journal of Cleaner Production, 406 (2023), p. 137152.

Edited by: Rajanikanth Aluvalu
Special issue on: Evolutionary Computing for AI-Driven Security and Privacy:
Advancing the state-of-the-art applications
Received: Jan 31, 2024
Accepted: Mar 11, 2024