

Integrating Gradient Boosting and Parametric Architecture for Optimizing Energy Use Intensity in Net-Zero Energy Buildings

Maqbul Kamaruddin^{1,2}, Martin C. T. Manullang^{3,4}, Jurng-Jae Yee^{1*}

¹ Department of ICT Integrated Ocean Smart City Engineering, Dong-A University, Busan 49315, Korea.

² Department of Architecture, Institut Teknologi Sumatera, South Lampung Regency 35365, Indonesia.

³ Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei 10607, Taiwan.

⁴ Department of Informatics, Institut Teknologi Sumatera, South Lampung Regency 35365, Indonesia.

Received 04 November 2024; Revised 06 February 2025; Accepted 11 February 2025; Published 01 March 2025

Abstract

Achieving net-zero energy building (NZEB) status requires accurate energy use intensity (EUI) calculations, as conventional methods often fail to capture the complexity of design and climatic conditions. In this research, a parametric energy modeling approach was used to conduct 1,350 simulations and analyze the impact of design parameters on building EUI. These simulations covered six building types—an existing building and I-, L-, T-, U-, and H-shaped buildings—across eight locations in different climate zones. A case study was conducted in Busan, Korea, where on-site measurements were obtained using portable devices to validate the simulation results. I-shaped buildings exhibited the lowest EUI, reaching 109 kWh/m²/yr at 0° and 180° orientations. The simulation results indicated that building orientations of 140°, 90°, 135°, and 270° tended to produce higher EUI values, whereas 0° and 180° showed lower EUI values of 122 and 123 kWh/m²/yr, respectively. The use of triple-pane insulated glass effectively reduced the I-shaped building's EUI to 103 kWh/m²/yr. Implementing photovoltaic (PV) systems further reduced the EUI significantly, with the I-shaped building achieving an EUI of -14 kWh/m²/yr at a 20% PV efficiency. Analysis using an extreme gradient boosting (XGBoost) model revealed that the climate zone, PV area, and type of heating, ventilation, and air-conditioning system significantly affected the EUI. This model, processed using Colab, was highly effective, with an R-squared value of 0.99, a root mean square error of 4.57, and a mean absolute error of 1.99. These findings demonstrate that the XGBoost model can effectively capture complex data patterns and can be used for high-accuracy EUI estimation.

Keywords: Gradient Boosting; Parametric Architecture; Energy Use Intensity; Net-Zero Energy Buildings.

1. Introduction

The building sector continues to dominate global energy consumption, accounting for over 40% of energy use and a significant share of greenhouse gas emissions [1]. This trend underscores the urgent need for energy-efficient solutions, particularly net-zero energy buildings (NZEBs), which are designed to offset their annual energy use through renewable energy sources [2]. However, despite the implementation of global initiatives, such as the Paris Agreement and national energy policies, the adoption of NZEBs remains inconsistent, particularly in developing regions, where energy consumption is rising significantly [3, 4].

The use of NZEBs is a pivotal strategy in addressing energy and environmental challenges by combining energy efficiency with renewable energy systems. These buildings aim to minimize energy demand while maximizing on-site

* Corresponding author: jjyee@dau.ac.kr

<http://dx.doi.org/10.28991/CEJ-2025-011-03-06>



© 2025 by the authors. Licensee C.E.J, Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

renewable energy generation, significantly reducing greenhouse gas emissions and reliance on fossil fuels [5]. The importance of NZEBs can be summarized as follows:

- **Energy Efficiency:** By integrating energy-saving tools and passive design strategies, NZEBs achieve high energy efficiency [6], thereby helping meet global climate objectives and reducing carbon footprints [7].
- **Cost Savings:** Although the initial investment in NZEB technology may be higher than average, the long-term reductions in energy bills can be substantial, making NZEBs financially attractive over their lifetime [8].
- **Regulatory Compliance:** Many governments have policies that require or encourage the construction of NZEBs as part of broader sustainability initiatives [9]. This aligns with international agreements aiming to combat climate change [10].

Achieving NZEB status involves numerous parameters and challenges. One of the main challenges is the complexity of design and planning, as each building element must be designed to minimize energy consumption [11]. Energy simulation, thermal analysis, and the modeling of various scenarios are needed to ensure that building designs meet the NZEB criteria [12]. Several parameters must be considered, such as energy efficiency, air quality, natural lighting, heating and cooling, and the integration of renewable energy.

In the context of NZEBs, energy use intensity (EUI) is a critical indicator that quantifies annual energy consumption per square meter of building floor area. A lower EUI indicates a more environmentally beneficial and efficient building. [13]. The EUI is particularly useful in the design of buildings with minimal energy consumption because it enables the comparison of building performance based on dimension, function, and location [14, 15]. Table 1 provides a comparison of international green building certifications that emphasize energy efficiency, energy conservation, and renewable energy integration.

Numerous international certification systems, including Leadership in Energy and Environmental Design (LEED), Green Star, and the Building Research Establishment Environmental Assessment Methodology (BREEAM), have precise criteria for renewable energy production, conservation, and efficiency. For instance, as illustrated in Table 1, LEED promotes the utilization of renewable energy credits and sets four specific benchmarks for passive and active strategies. Green Star emphasizes energy conservation through operational and design practices. South Korea's Green Standard for Energy and Environmental Design (G-SEED) also prioritizes the utilization of on-site renewable energy and promotes sustainable, energy-efficient design. Simson et al. [16] observed that the focus of NZEBs varies significantly across countries and climate zones. This position was supported by Garcia & Kranzl [17], who argued that an adaptive approach is required to establish NZEB standards suitable for various regions. Therefore, parametric architecture is essential for the design of buildings that meet NZEB criteria.

Table 1. Comparison of Green Building Certifications and Their Criteria

Certification	Level	Energy Efficiency	Energy Conservation	Energy Production	Country/Region
LEED [18, 19]	Certified, Silver, Gold, and Platinum	High	Emphasizes both passive and active strategies	Encourages renewable energy credits	USA, International
Green Star [20, 21]	1 Star to 6 Stars	High	Focus on reducing energy consumption through design and operational practices	Encourages on-site renewable energy	Australia, New Zealand
BREEAM [22, 23]	Pass, Good, Very Good, Excellent, Outstanding	High	Encourages reduction in energy demand and efficient use of energy	Promotes integration of renewable energy sources	United Kingdom, International
Comprehensive Assessment System for Built Environment Efficiency [24, 25]	C, B-, B+, A, S	Moderate to high	Promotes energy-saving design and technologies	Encourages on-site renewable energy production	Japan
Green Globes [26, 27]	One Green Globes to Four Green Globes	Moderate to high	Evaluates energy performance and encourages improvements	Supports renewable energy integration	USA, Canada
Excellence in Design for Greater Efficiencies [22, 28]	Certification	Moderate to high	Focus on energy-saving measures and technologies	Encourages renewable energy sources	International
G-SEED [29-31]	Green 1 to Green 4	High	Emphasizes sustainable, energy-efficient design	Encourages on-site renewable energy	South Korea
Architecture 2030 [28, 32]	Target	Very high	Aims for zero fossil fuel energy use	Emphasizes on-site renewable energy production	International

Parametric architecture, where algorithms are used to evaluate multiple design scenarios simultaneously, provides a powerful methodology for designing energy-efficient buildings [33]. Building on recent research by Stevanović et al. [34], the current study highlights the integration of parametric techniques with machine learning models, such as extreme gradient boosting (XGBoost), to improve the predictive accuracy and performance of energy modeling.

Regarding methodological frameworks, Lu et al. [35] and Labib [36] demonstrated the advantages of combining parametric design and machine learning for building performance optimization. For instance, Veiga et al. [37], Seyedzadeh et al. [38], Amasyali and El-Gohary [39], and Solmaz [40] conducted parametric analyses using the building

energy modeling program EnergyPlus to obtain optimal results. Data on building shapes, external materials, internal loads, occupancy behavior, and other features were collected as input files (.idf) for further training and validation in Python. The authors investigated which parameters significantly affect a building's energy consumption. The interaction between parametric architecture and machine learning can be crucial for building design, construction, and operation.

Machine learning techniques, such as artificial neural networks (ANNs), and gradient boosting models, such as XGBoost, are highly effective for simulating building energy performance. In the study of Guo et al. [41], ANN models outperformed other approaches in predicting the energy consumption of detached residential buildings. Similarly, Barbaresi et al. [42] highlighted the capacity of machine learning models, including support vector machines (SVMs), random forests (RFs), and XGBoost, to replace computationally intensive simulation tools, such as EnergyPlus, with faster, more accurate prediction methods. These findings emphasize the growing importance of machine learning in energy optimization workflows.

XGBoost is a powerful tool for managing large datasets with complex interactions, offering high accuracy and resilience to overfitting. Kumar et al. [43] effectively utilized XGBoost to forecast energy consumption patterns across diverse building typologies considering dynamic environmental conditions. Ni et al. [44] further showcased XGBoost's adaptability by integrating it into a digital twin framework for real-time energy performance optimization. Gan & Gao [45] expanded on this by combining XGBoost with parametric modeling tools, enabling multi-objective optimization that balances energy with occupant comfort.

Recent research continues to affirm XGBoost's effectiveness in energy management applications. Abdelsattar et al. [46] demonstrated XGBoost's superiority to RF, multilayer perceptron (MLP), SVM, and similar models in improving solar power forecasting accuracy. Although the MLP and SVM models were promising, their limitations in generalizing to unfamiliar data underscored the robustness of gradient-boosting models in the parametric architectural context. These advancements highlight the critical role of machine learning in simplifying decision-making during early design stages and achieving reliable energy performance predictions.

Previous research has explored parametric architecture and machine learning separately to optimize architectural parameters and building energy consumption. However, the integration of these two approaches, especially to optimize the EUI while maintaining occupant comfort, is rarely discussed in the literature. This gap is an untapped opportunity to combine the benefits of parametric architecture in enhancing building design efficiency with the adaptive and predictive capabilities of machine learning in managing energy consumption in real time.

This gap is critical because integrating parametric architecture with machine learning can create a more holistic system for building energy optimization. This methodology may balance occupant comfort and energy efficiency more effectively. Without integration, energy savings may be achieved at the expense of occupant comfort, or vice versa. This study aims to address the abovementioned gaps through the following:

- Investigation of Inefficiencies in Traditional NZEB Renovation

The weaknesses and inefficiencies of conventional renovation methods for NZEBs are identified. The focus is on prolonged, repeated iterative processes required to achieve EUI values that comply with standards.

- Redesign of Case Study Building Using Parametric Architecture

Parametric architecture techniques are used to redesign a case study building in Busan. This approach enables the flexible, efficient customization of various design elements, such as building orientations, window types, heating, cooling, electrical systems, and photovoltaic (PV) systems.

- Development of Machine Learning Models for EUI Reduction

Machine learning models (XGBoost, LightGBM, and CatBoost) are implemented to predict and optimize the EUI. The models are trained using simulated and real-world data to ensure accuracy and reliability.

- Significant EUI Reductions to Meet NZEB Standards

Machine learning models are integrated into parametric architecture workflows and iterative optimization is performed to reduce the EUI significantly. The ultimate target is to meet or exceed the applicable NZEB benchmarks, such as the G-SEED or Architecture 2030 standards.

This study will contribute significantly to building energy optimization, particularly in achieving NZEB standards efficiently and effectively.

2. Research Methodology

As shown in Figure 1, this study covers the key aspects of EUI optimization in NZEBs through parametric architecture and machine learning. These phases are an investigation of inefficiencies in conventional renovation, redesign of a case study building, development of machine learning models, and fulfilment of NZEB criteria.

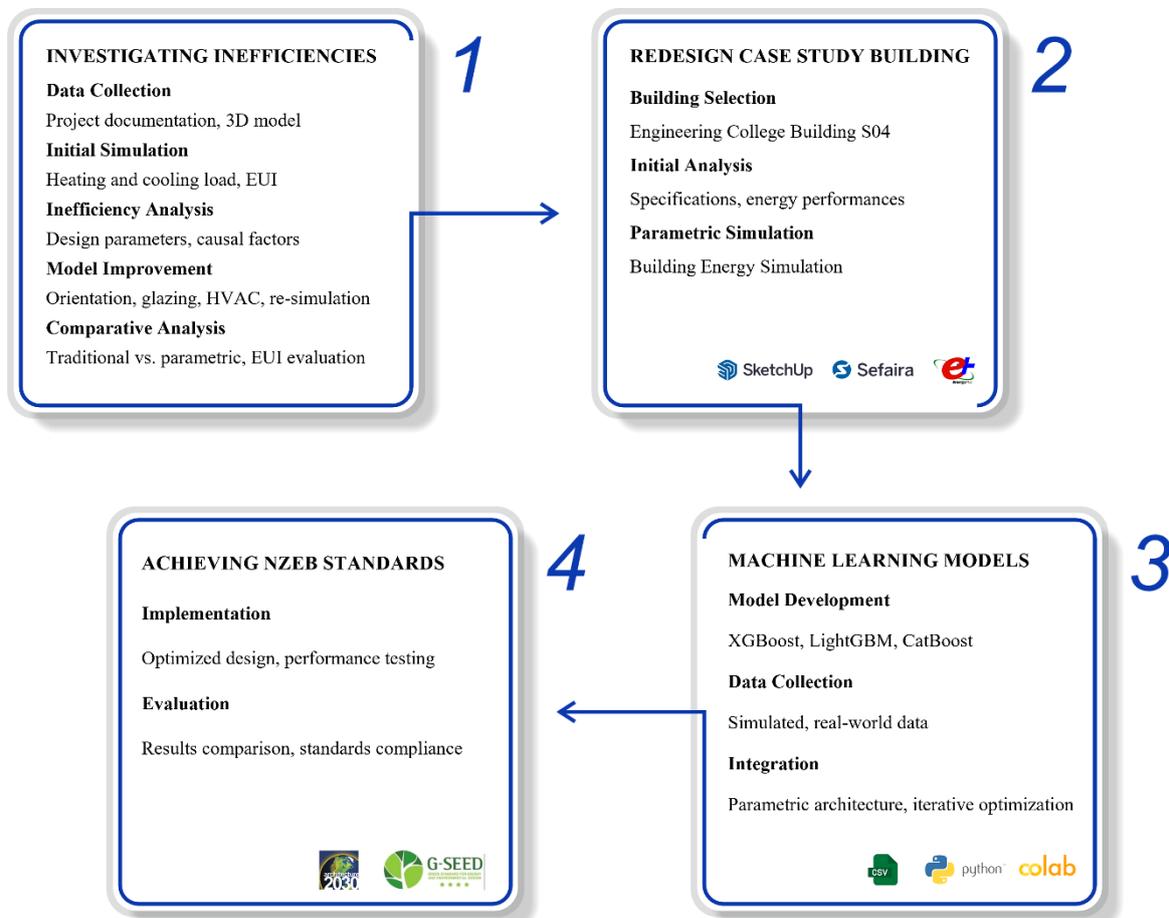


Figure 1. Overall Methodology Flowchart for Optimizing EUI in NZEBs

2.1. Inefficiencies in Traditional NZEB Renovation

This research aims to evaluate and understand the inefficiencies in traditional renovation methods for achieving NZEB standards. These inefficiencies were identified, analyzed, and addressed through several systematic methodological steps, as depicted in Figure 1.

1. Initial Data Collection

- **Project Documentation:** All documentation related to a renovation project was collected, including building plans, technical specifications, historical energy consumption data, and audit reports.
- **Building Remodeling:** The software SketchUp and Sefaira were used to create a three-dimensional (3D) model of the building to be renovated. The model included all critical components, such as the walls, roofs, windows, and heating, ventilation, and air-conditioning (HVAC) systems.

2. Initial Energy Simulation

- **Initial Simulation:** The created 3D model was used to perform an initial energy simulation to obtain an initial EUI value and identify areas of high energy consumption.
- **Simulation Tools:** The initial energy simulation was conducted using the energy simulation software Sefaira and EnergyPlus. The resulting data were analyzed to understand energy consumption patterns and identify inefficiencies.

3. Inefficiency Analysis

- **Evaluation of Design Parameters:** The design parameters that influence the energy performance of the building were analyzed, such as the building orientation, window type, HVAC system, and thermal insulation.

- Identification of Causal Factors: The main causal factors of energy inefficiency were determined. This involved an in-depth analysis of building design and operational parameters.

4. Iterative Process of Model Improvement

- 3D Model Improvement: The building's 3D model was enhanced based on the analysis of inefficiencies. These improvements included adjusting the building orientation, selecting more efficient types of glass, and optimizing the HVAC system.
- Renewable Energy System Integration: After design optimization, analysis was performed to determine whether the building still required additional energy to achieve NZEB status.
- Re-Simulation: After each improvement, the energy simulation was repeated to evaluate the effects of the design changes on the EUI. This process was conducted iteratively until an EUI value that met NZEB standards was achieved.

5. Comparative Analysis

- Method Comparison: The energy simulation results obtained using the traditional renovation method were compared with those obtained using the proposed parametric architecture approach to obtain insights into the efficiency and effectiveness of each method.
- Energy Performance Evaluation: The building's energy performance after the design improvements was assessed to ensure that the EUI values met or exceeded NZEB standards.

With this methodology, this study illustrates the inefficiencies in traditional renovation methods for achieving NZEB standards. The results will provide practical guidance and recommendations for enhancing building energy efficiency through a systematic, pragmatic approach.

2.2. Location and Parametric Redesign of Case Study Building

The Engineering College building S04 of Dong-A University, Seunghak Campus, Busan, Korea, was the focus of this NZEB study. The building location is shown in Figure 2-a, with a broader view of South Korea and the building location in Busan illustrated in Figure 2-b. The building is situated at 35.12°N, 129.03°E. Busan, included in Zone 3A in the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) 90.1 classification [45], experiences hot, humid summers and cool winters. As shown in Figure 2b, building S04 is oriented at 140°N and consists of five floors, with the ground floor being used primarily for access to the first floor.

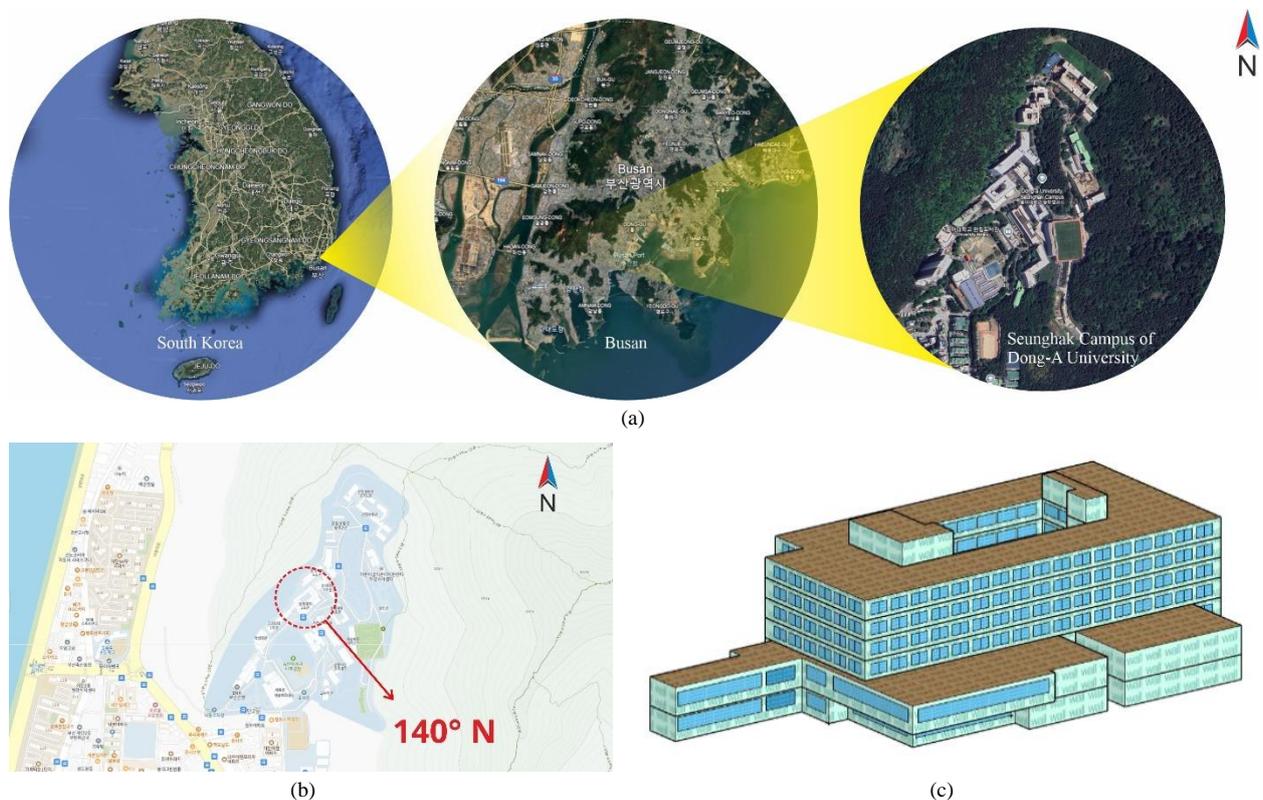


Figure 2. Case Study Building: (a) Study Location, (b) Building Orientation (140°N), (c) Sefaira Model

This building functions as an educational facility, comprising various classrooms, lecture rooms, laboratories of multiple technical fields, and supporting amenities (such as a student canteen, printing area, lavatory, and mechanical, electrical, and plumbing service room). The building, which is rectangular, has a central courtyard and a single-corridor layout, ensuring that all rooms receive adequate lighting and air circulation, as illustrated in Figure 2-c.

Several key building parameters are simulated and analyzed to achieve NZEB status. In this simulation, building energy performance is assessed under various scenarios, including different building shapes, numbers of floors, total areas, and climate zones. The key parameters to be tested include the type of floor finish, wall and roof U-values, HVAC system type, and PV panel efficiency and orientation. For example, the U-value of an existing fence is 1.5 W/m²K and compared with those of concrete blocks (1.8 W/m²K) and brick plaster (2.62 W/m²K). Concrete roofs with a U-value of 0.9 W/m²K are tested against other types of roofs, such as metal decks with a U-value of 7.1 W/m²K. Additional parameters, such as the air infiltration rate, equipment power density (EPD), and lighting power density (LPD), are also varied to determine their impact on building energy performance.

Building S04 was compared with building scenarios in various other climate locations, such as Singapore (0A), Ho Chi Minh City (1A), Durban (2A), Seoul (4A), Berlin (5A), Helsinki (6A), and Murmansk (7), as shown in Figure 3. The eight locations were selected on the basis of the ASHRAE 90.1 climate zones, a globally recognized standard for energy performance modeling. These zones represented a diverse range of climatic conditions, including variations in temperature, humidity, and seasonal patterns, ensuring that the study covered key environmental factors that significantly impact the EUI.

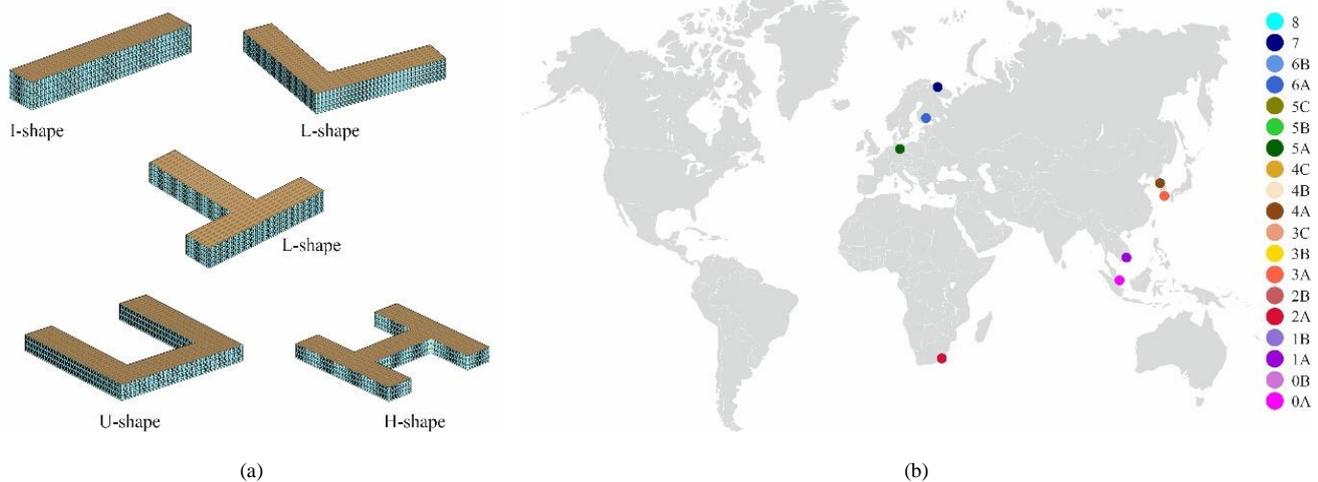


Figure 3. Building Types and Climate Zones: (a) Different Building Shapes for Case Study, (b) ASHRAE Climate Zone Map

This selection was made for a broad coverage of global climate diversity, encompassing hot, temperate, and cold climates. This ensures that the findings are relevant for green building standards and NZEB benchmarks, which are often tailored to regional climate conditions. By aligning with ASHRAE standards, this study helps elucidate energy performance in diverse climates while supporting regions that are adopting the NZEB framework or its counterparts.

Each scenario was evaluated using Sefaira, which was used to record the energy performance input and output data for different configurations. This software provided the climate data used in this simulation, sourced from EnergyPlus Weather (EPW). This enabled the automated use of appropriate climate data based on location.

The building comprises various parameters, as shown in Table 2, such as façade glass U-value, façade glass solar heat gain coefficient (SHGC) value, wall type, wall U-value, floor finish, floor U-value, roof type, roof U-value, infiltration rate in Air Changes per Hour (ACH), EPD, LPD, HVAC system type, PV panel efficiency, PV panel orientation, PV panel tilt, and PV panel area. Output data, including the heating and cooling equipment design capacity and EUI, were recorded. This simulation was performed to evaluate how changes in design parameters could optimize the building's energy performance toward the NZEB target.

Table 2. Baseline and Simulation Variants for Building Energy Performance Analysis

Parameter Category	Baseline	Simulation Variants
Building orientation	140°N	0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°
ASHRAE climate zone	3A (Busan, Korea)	0A (Singapore), 1A (Ho Chi Minh City, Vietnam), 2A (Durban, South Africa), 3A (Busan, Korea), 4A (Seoul, Korea), 5A (Berlin, Germany), 6A (Helsinki, Finland), 7 (Murmansk, Russia)
Building shape	Rectangular with courtyard	Rectangular (I, L, U, T, H)
Number of floors	5	3, 4, 5
Wall type	Precast concrete	Precast concrete, brick plaster, concrete block
Wall U-value	1.5 W/m ² K	1.5, 2.62, 1.8 W/m ² K
Floor finish	Tile	Tile, carpet, hardwood
Floor U-value	0.8 W/m ² K	0.8, 2.5, 2.1 W/m ² K
Facade glazing solar heat gain coefficient	0.7	0.86 (single pane), 0.7 (double pane), 0.5 (triple pane), 0.35 (low E), 0.65 (argon filled), 0.4 (vacuum insulated)
Facade glazing U-value	3 W/m ² K	6 (single pane), 3 (double pane), 1.2 (triple pane), 1.6 (low E), 2 (argon filled), 0.7 (vacuum insulated)
Infiltration rate	Well-sealed	Airtight (0.25 ACH), well-sealed (0.5 ACH), average (1.0 ACH)
EPD	15 W/m ²	10, 15, 20 W/m ²
LPD	9 W/m ²	6, 9, 12 W/m ²
HVAC system type	Fan coil units and central plant	Variable-refrigerant-flow fan coils, fan coil units, central plant
PV efficiency	n/a	15%, 20%
PV panel orientation	n/a	South facing (180°), southeast facing (135°), southwest facing (-45°)
PV panel tilt	n/a	0°, 30°, 35°

2.3. Performance Benchmark of XGBoost for EUI Prediction

The EUI, a key benchmark for assessing building energy performance, is crucial in the context of NZEBs. It is the annual energy consumption of a building relative to its total area, expressed in kWh/m²/yr or kBtu/sf/yr, depending on regional preferences or standards. Lower EUI values indicate higher energy efficiency. This section describes the implementation of gradient boosting models, namely, linear regression (LR), XGBoost, RF regression (RFR), gradient boosting regression (GBR), and SVM, to predict the EUI. These models were chosen for their ability to handle complex data and predict accurately. The training process and parameters used to develop these models are detailed in the following subsections.

EUI prediction is pivotal in optimizing building energy performance, especially for NZEB standards. Leveraging machine learning models, such as XGBoost, enhances accuracy and scalability in addressing the complexities of energy modeling. These models are not only adept at processing large datasets but also excel in identifying nonlinear relationships and interactions between variables. The use of XGBoost alongside other models in this study provides a comprehensive benchmark for evaluating their predictive capabilities and contributions toward energy-efficient building design. This evaluation ensures that the chosen models are suitable for making precise, data-driven decisions critical to reducing energy consumption and fulfilling NZEB standards.

2.3.1. Dataset and Splitting

The machine learning models were developed using a dataset sourced from the Sefaira simulation. This dataset comprised six buildings from eight locations, which were modified to generate 32 scenarios, resulting in 1,350 samples. These samples were saved in .csv format and loaded into Python for processing.

We used fivefold cross-validation to minimize bias and increase confidence in model performance. This standard machine learning method provides a robust framework for model training and validation. By splitting the dataset into five distinct folds, we created five unique data distributions, which helped in obtaining an objective measure of model performance. The use of a single fold in model development would have resulted in a significant risk of high bias and an unrepresentative performance value. Fivefold cross-validation mitigates this risk by ensuring that every dataset sample is used for both training and validation [46].

In practice, a dataset is randomly divided into five equal parts: four parts are utilized as training data in each fold, and the remaining part serves as validation data. This technique is conducted five times, with each fold taking a turn as the validation set. Eventually, all dataset parts are used as training and validation data. The final performance metrics of a model are obtained by averaging the results from each folds. This averaging process provides a comprehensive, dependable estimate of model performance. The fivefold cross-validation process is visualized in Figure 4.

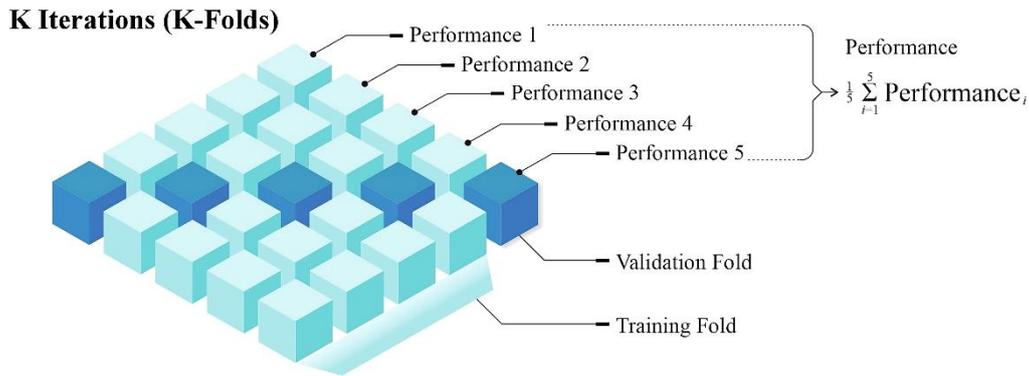


Figure 4. Illustration of Fivefold Cross-Validation

2.3.2. Defining Features, Non-Numerical Features, and Target

Twenty-one features were used to develop the models: the ASHRAE 90.1 climate zone, building orientation, total floor area, window façade U-value, window façade SHGC, wall type, wall U-value, floor finish, floor U-value, roof type, roof U-value, infiltration rate, EPD, LPD, HVAC system type, PV panel efficiency, panel orientation, panel tilt, panel area, heating equipment design capacity, and cooling equipment design capacity.

In the training process in machine learning, all features should have numerical values. Therefore, one-hot encoding was used to convert categorical features into numerical ones. One-hot encoding transforms categorical variables into a numerical format compatible with machine learning models. It creates a binary column for each class within a categorical variable, where each column indicates the presence or absence of that category. This transformation enables the model to effectively learn the correlation between these categories and the target variable.

The model target was the EUI result (RES EUI), or the annual building energy usage per square meter of floor area. In this dataset, the EUI value was obtained from the Sefaira modeling results.

2.3.3. Model Benchmark

The use of XGBoost for EUI prediction provides a compelling advantage due to its ability to handle heterogeneous data and capture intricate patterns. Its tree-based ensemble approach efficiently manages feature interactions and nonlinear dependencies, which are common in energy modeling datasets. This characteristic makes XGBoost particularly suitable for EUI prediction, where input features, such as building orientation, façade properties, and HVAC systems, exhibit complex relationships. Compared with traditional models, such as LR, and even other ensemble methods, such as RFs, XGBoost consistently demonstrates higher accuracy, making it a valuable tool for energy efficiency evaluation and design optimization in achieving NZEB goals.

Machine learning models operate based on theoretical principles and mathematical equations. In this study, five types of traditional machine learning models were compared: LR, XGBoost, RFR, GBR, and SVM. The parameter values used for each model were determined through parameter optimization according to existing literature. XGBoost, RFR, and GBR were optimized using grid search and random search to identify the optimal combination of parameters for the data. The foundations and mathematical representations of the models are summarized as follows:

- **LR**

LR assumes a linear relationship between the input features (X_1, X_2, \dots, X_n) and the target variable Y . The equation is [47]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon, \quad (1)$$

where β_0 is the intercept; $\beta_1, \beta_2, \dots, \beta_n$ are coefficients for the predictors; and ϵ is the residual error.

- **XGBoost**

XGBoost minimizes a loss function $L(\theta)$ using gradient boosting, thus improving predictions iteratively. The objective function is [48]:

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^k \Omega(fk), \quad (2)$$

where $l(y_i, \hat{y}_i)$ is the loss between the predicted and actual values; $\Omega(fk)$ is the regularization term, which prevents overfitting; Θ includes the model parameters; and fk represents decision trees.

- **RFR**

An RF averages multiple decision tree predictions to reduce variance. The prediction y is computed as [49]:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(X), \tag{3}$$

where T is the total number of decision trees and $h_t(X)$ is the prediction from the t th tree.

- **GBR**

Gradient boosting iteratively builds models to minimize a loss function. For regression, the gradient descent algorithm updates predictions as [50]:

$$F_{m+1}(x) = F_m(x) + \gamma \cdot h_m(x), \tag{4}$$

where $F_m(x)$ is the model at iteration m , $h_m(x)$ is the weak learner (decision tree) at iteration m , and γ is the learning rate.

- **SVM**

An SVM maps data into higher dimensions to find an optimal hyperplane for classification or regression. The optimization objective is [51]:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \varepsilon_i \tag{5}$$

subject to

$$y_i(\omega \cdot x_i + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, \tag{6}$$

where ω is the weight vector, C is the penalty parameter, and ε_i represents slack variables.

Additionally, parameters such as the number of estimators (`n_estimators`) in the ensemble models (XGBoost, RFR, and GBR) were selected based on preliminary experiments, with insights from previous studies indicating that these parameters effectively improve model accuracy.

For LR, which was more straightforward than the other models, the parameters were selected following standard implementations commonly found in the literature, as outlined in Table 3. The radial basis function kernel was selected for the SVM, with the parameters C and `epsilon` optimized to balance bias and variance.

Table 3. Summary of Analyzed Models

Model	Library/Module	Parameters	Description
LR	<code>Sklearn.linear_model</code>	n/a	LR models the correlation between a dependent variable and one or more independent variables by assuming a linear correlation and learns the best-fitting straight line through the data points.
XGBoost	<code>xgboost</code>	<code>objective='squared error', n_estimators=100</code>	It is efficient and accurate, especially with large datasets and complex relationships. It uses boosting loops to improve performance and handle complex data.
RFR	<code>sklearn.ensemble</code>	<code>n_estimators=100</code>	This ensemble learning method creates several decision trees and generates an average prediction from each. It reduces overfitting and improves model accuracy and robustness.
GBR	<code>sklearn.ensemble</code>	<code>n_estimator=100</code>	Models are built sequentially to correct errors from previous models. It effectively improves model accuracy but requires large amounts of computational resources.
SVM	<code>sklearn.svm</code>	<code>'kernel='rbf', C=1.0, epsilon=0.1</code>	It uses a radial basis function kernel. The model balances the trade-off between matching the training data and maintaining smoothness. It focuses on significant deviations rather than minor errors.

2.3.4. Evaluation Metrics

We used three metrics to evaluate the model performance in this benchmarking process: the R-squared, root mean square error (RMSE), and mean absolute error (MAE) values.

- **R-squared**, also called the coefficient of determination, is a numerical measure that indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. It shows how well the independent variables explain the variability of the dependent variable. An R-squared value ranges from 0 to 1, where 0 means the model does not explain any variability and 1 means the model explains all the variability in the data around its mean. A higher R-squared value indicates a better fit of the model to the data.

- The **RMSE** is a numerical measure used to assess the accuracy of a predictive model. It is the square root of the average squared difference between the predicted and actual values. The RMSE shows how well the model's predictions match the observed data, with smaller values indicating better model performance. It is useful for understanding the magnitude of prediction errors and for comparing the accuracy of different models.
- The **MAE** is a statistical measure used to estimate the accuracy of a predictive model. It represents the average absolute difference between the predicted and actual values. The MAE provides a straightforward calculation of prediction errors, with smaller values indicating improved model performance. It helps understand the average magnitude of prediction errors without considering their direction.

Using these metrics, we comprehensively assessed the performance of the developed machine learning models considering their goodness of fit and the magnitude of their prediction errors. We also included the standard deviation (SD) of each metric to show the range of variation between the five folds.

The use of evaluation metrics, such as R-squared, RMSE, and MAE, highlights the importance of balancing predictive accuracy with interpretability in the context of building energy performance. XGBoost, with its robust optimization capabilities, often yields high R-squared values while maintaining low RMSE and MAE values, indicating prediction precision and reliability. Beyond being academically valuable, these results have practical implications in guiding design decisions. For instance, accurate EUI prediction enables architects and engineers to prioritize modifications in building features, such as glazing properties or HVAC efficiency, that yield the greatest energy savings. This supports an informed, targeted approach to achieving NZEB targets.

By leveraging XGBoost for EUI prediction, we bridge the gap between advanced machine learning and practical energy performance benchmarks. The insights derived from accurate EUI prediction can directly inform the design and retrofit of energy-efficient buildings, ensuring that interventions align with NZEB standards. This demonstrates not only the technical superiority of XGBoost but also its critical role in advancing sustainable architectural practices.

2.4. Validation on Real-World Data

Model validation was previously conducted to compare the simulation results with real-world data collected from a Davis Vantage Pro2 weather station and a TR-72nw temperature and humidity logger. The weather station was placed on the rooftop of building S04, and the temperature and humidity logger was in room 0310 in the same building, as shown in Figure 5.



Figure 5. Data Collection Setup: (a) Outdoor Weather Station; (b) Room 0310, Building S04; (c) Indoor Temperature Sensors

The validation began with the collection and processing of experimental data from these primary sources. The Vantage Pro2 and TR-72nw data were imported and cleaned by removing empty or irrelevant values. Next, these data were harmonized to ensure consistency in the subsequent analyses. The timestamps were converted into a standard format and the data were resampled to the same time interval, such as hourly, to ensure uniformity in the analysis.

The processed experimental data were compared with the simulation findings. The simulation results were imported and adjusted to match the time series of the experimental data. The two datasets were then combined for a direct comparison of the experimental and simulated data. The results were analyzed and visualized to measure the variations between the experimental and simulation data and identify potential discrepancies and areas requiring model adjustment. The differences between the experimental and simulated temperatures and humidities were calculated, and the abovementioned error metrics were used to estimate model accuracy. Graphically visualizing experimental and simulation data helps highlight relationships and deviations that may not be noticeable in numerical data analysis alone. By mapping these data, we observed the alignment between the real-world and simulation results and identified areas needing adjustments. Model adjustments were made based on the comparison results to improve the accuracy of EUI prediction and other building performance metrics.

This comprehensive validation ensured that the simulation model could align predictions with real-world conditions, thereby reliably and accurately predicting building energy performance. By iterating and adjusting based on this validation, the model becomes increasingly precise in reflecting building energy use.

3. Results and Discussion

3.1. Analysis of Simulation Results

A total of 1,350 simulations were performed using Sefaira to analyze the impact of various design parameters on building EUI. The simulations were diverse, covering six building types—existing building and I-, L-, T-, U-, and H-shaped buildings—and eight locations according to the ASHRAE 90.1 climate zones—Singapore (0A), Ho Chi Minh City (1A), Durban (2A), Busan (3A), Seoul (4A), Berlin (5A), Helsinki (6A), and Murmansk (7). Due to the extensive quantity of simulation results, only a subset of the results is displayed here.

The simulation outcomes show considerable EUI discrepancies depending on the orientation of the existing structure in Busan. As presented in Figure 6, the 0° and 180° orientations demonstrate reduced EUI values of 122 and 123 kWh/m²/yr, respectively. These orientations optimize solar heat uptake, ensuring uniform sunlight exposure throughout the day while reducing the cooling demand. Conversely, the 90°, 135°, and 270° orientations have elevated EUI values, often reaching 125 kWh/m²/yr. This rise is mostly attributable to the heightened sun exposure in the morning and afternoon, which amplifies cooling demands.

Existing Building: Orientation		AIA 2030 COMMITMENT		Share Project	
Busan, KR		Upload Project			
Use different weather location					
		Heating Equipment ...	Cooling Equipment ...	EUI	Annual Net CO ₂ e Em...
Baseline Concept 15,588 m ²	HVAC System Type Fan Coil Units and Central Plant	624.5 kW	1,085.8 kW	125 kWh/m²/yr	826,406 kgCO₂
Orientation 0°	Fan Coil Units and Central Plant	624.2 ↓0%	1,075.0 ↓0%	122 ↓2%	818,197 ↓0%
Orientation 45°	Fan Coil Units and Central Plant	624.2 ↓0%	1,121.7 ↑3%	124 ↓0%	829,199 ↑0%
Orientation 90°	Fan Coil Units and Central Plant	624.8 ↑0%	1,140.2 ↑5%	125 ↔0%	832,778 ↑0%
Orientation 135°	Fan Coil Units and Central Plant	623.8 ↓0%	1,098.4 ↑1%	125 ↔0%	827,474 ↑0%
Orientation 180°	Fan Coil Units and Central Plant	624.1 ↓0%	1,055.7 ↓2%	123 ↓1%	818,343 ↓0%
Orientation 225°	Fan Coil Units and Central Plant	555.1 ↓11%	880.3 ↓18%	123 ↓1%	793,839 ↓3%
Orientation 270°	Fan Coil Units and Central Plant	624.9 ↑0%	1,114.8 ↑2%	125 ↔0%	834,396 ↑0%
Orientation 315°	Fan Coil Units and Central Plant	624.6 ↑0%	1,072.9 ↓1%	124 ↓0%	828,565 ↑0%

Figure 6. Simulation Results for Existing Building Orientation in Busan

Energy performance shows modest but significant variations across different angles. The 45° and 225° orientations exhibit marginally enhanced EUI values (124 kWh/m²/yr) relative to the peak recorded values, indicating a minor decrease in direct solar influence during essential heat-gain intervals. These distinctions highlight the intricate link between building orientation and energy performance: some angles may be more advantageous than others while not attaining optimal efficiency.

The measured EUI changes demonstrate a clear association between building orientation and cooling demand in Busan's subtropical environment. The detected patterns in the data establish a comprehensive basis for comprehending the impact of orientation on energy performance and pave the way for further exploration of other factors influencing the EUI, including building shape, material selection, and system efficiency.

EUI analysis based on building orientation in Busan shows significant variations between building types. Buildings with the existing form and courtyard structure often record higher EUIs than other building shapes, according to the simulation data in Figure 7. For example, at the 90°N, 135°N, and 140°N orientations, the existing building achieves a peak EUI of 125 kWh/m²/yr. This can be explained by the building's structure, which enables increased exposure to sunlight, particularly in summer. This high exposure, which increases the exposed surface area, affects the building's heat absorption and release, thereby increasing energy consumption.

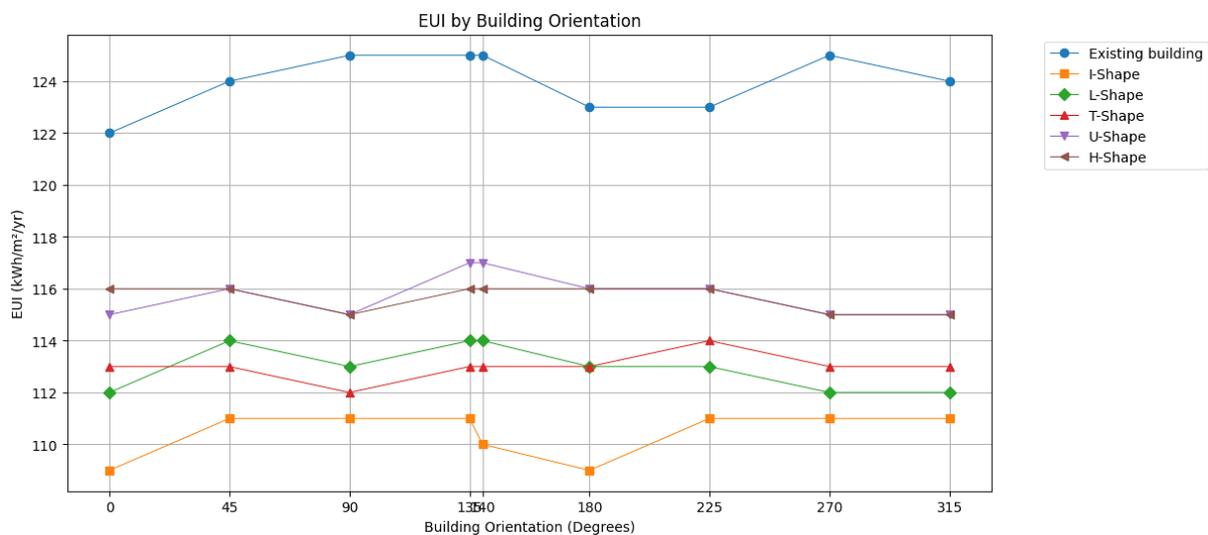


Figure 7. EUI vs. Building Orientation

By contrast, I-shaped buildings show lower EUI values, with the lowest values reaching 109 kWh/m²/yr at the 0° and 180° orientations. The linear, open nature of the I-shaped design allows for more efficient light distribution and minimizes the effect of excess heat, thus lowering the cooling and heating energy load. The effectiveness of this design in reducing the EUI demonstrates how architecture that strategically considers building orientation and shape can enhance energy efficiency.

The average EUI observed in Busan is approximately 114 kWh/m²/yr, indicating that building orientation and design are crucial for determining energy efficiency. A higher EUI is typically associated with orientations that maximize exposure to direct sunlight, resulting in greater cooling loads. By contrast, orientations that optimize natural light reception without overheating tend to result in lower EUIs. This condition highlights the importance of considering orientation in building planning to maximize energy efficiency and minimize buildings' environmental impact.

The EUI values at the 140°N orientation notably vary across climatic zones and architectural forms, as shown in Figure 8. Murmansk, situated in the coldest climate zone (ASHRAE Zone 7), exhibits the greatest energy consumption among all sites, with the existing building registering a peak EUI of 305 kWh/m²/yr. The optimized designs, including the I- and U-shaped structures, have notably elevated EUI values of 227 and 246 kWh/m²/yr, respectively. This shows the primary impact of heating needs in frigid areas, where sustaining interior thermal comfort necessitates considerable energy. The persistently elevated EUI values across all configurations in this region highlights the difficulty of attaining energy efficiency in extreme cold.

In warmer climates, such as Durban (ASHRAE Zone 2A), buildings have significantly lower EUI values. The I-shaped design attains an EUI of 103 kWh/m²/yr, whereas the T- and L-shaped structures exhibit comparable efficiencies, registering 101 and 102 kWh/m²/yr, respectively, as seen in Figure 8. The notable decrease in energy consumption is mostly ascribed to the advantageous climatic conditions, which diminish dependence on active heating and cooling systems. The more intricate designs, such as the U- and H-shaped designs, exhibit superior performance in warmer regions, achieving EUI values of approximately 101 kWh/m²/yr. These patterns demonstrate the efficacy of passive cooling and natural ventilation in tropical and subtropical areas.

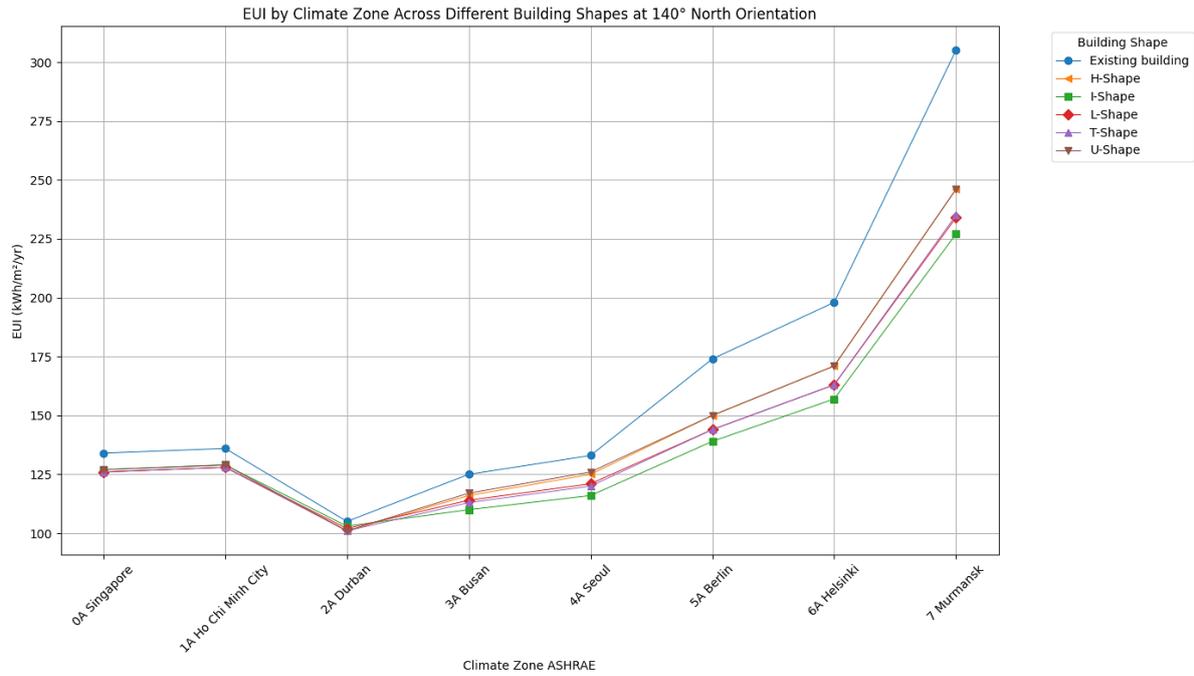


Figure 8. EUI vs. Climate Zone for Different Building Shapes at 140°N Orientation

Moderate climatic zones, such as Busan (ASHRAE Zone 3A), exhibit intermediate EUI values, with the existing structure having an EUI of 125 kWh/m²/yr. The I-shaped design is the most efficient among the compared forms, exhibiting an EUI of 110 kWh/m²/yr, followed by the L- and T-shaped designs, which have EUI values of 114 and 113 kWh/m²/yr, respectively. Figure 8 illustrates that the more intricate forms, specifically the U- and H-shaped designs, exhibit marginally elevated energy consumption, having EUI values of 117 and 116 kWh/m²/yr, respectively. These results underscore the relationship between climatic conditions and architectural geometries: simpler, more linear forms generally exhibit superior energy efficiency, even under mild environmental conditions.

Figure 9 illustrates the EUI variations between different building types and façade materials in Busan. The metal deck of the current structure exhibits the greatest EUI: 143 kWh/m²/yr. This signifies inadequate thermal efficiency, resulting in heightened cooling requirements during the warmer months. The brick plaster on the existing structure likewise yields a significant EUI of 133 kWh/m²/yr. This material, albeit esthetically appealing, has inadequate energy efficiency.

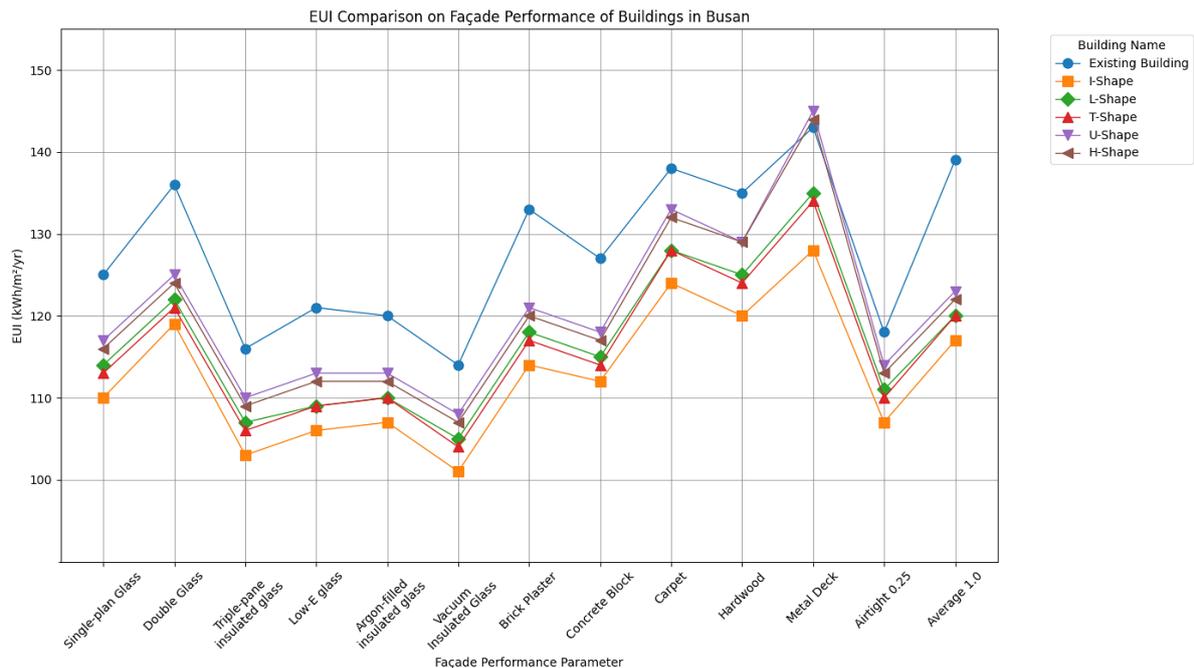


Figure 9. EUI vs. Façade Material for Buildings in Busan

Advanced materials, such as triple-pane insulated glass and vacuum-insulated glass, exhibit superior performance. Triple-pane insulated glass attains an EUI of 103 kWh/m²/yr in the I-shaped structure and 116 kWh/m²/yr in the existing edifice. Vacuum-insulated glass exhibits low EUI values of 108 kWh/m²/yr in the U-shaped structure and 101 kWh/m²/yr in the I-shaped structure. These materials regularly diminish heat transfer, thus enhancing energy efficiency.

Materials such as single-pane glass and metal decking typically yield elevated EUI values of over 125 kWh/m²/yr. Materials with moderate performance, including low-E glass and argon-filled insulated glass, have an energy consumption range of 105–121 kWh/m²/yr. Building geometry significantly affects these results: simpler designs, such as the I-shaped one, perform excellently across all façade materials. Figure 9 illustrates the significance of integrating suitable materials with effective design configurations to enhance building energy performance.

The results in Figure 10 reveal significant EUI differences across various internal load parameters and HVAC system types for different building shapes in Busan. Buildings with high internal loads, namely, a high EPD (20 W/m²) and a high LPD (12 W/m²), exhibit consistently higher EUI values. For example, the H-shaped design reaches a maximum EUI of 127 kWh/m²/yr at a high EPD, whereas the U-shaped design has an EUI of 128 kWh/m²/yr at the same EPD. These values highlight the effect of elevated EPDs and LPDs on energy demand, particularly in buildings with complex shapes, such as H- and U-shaped ones.

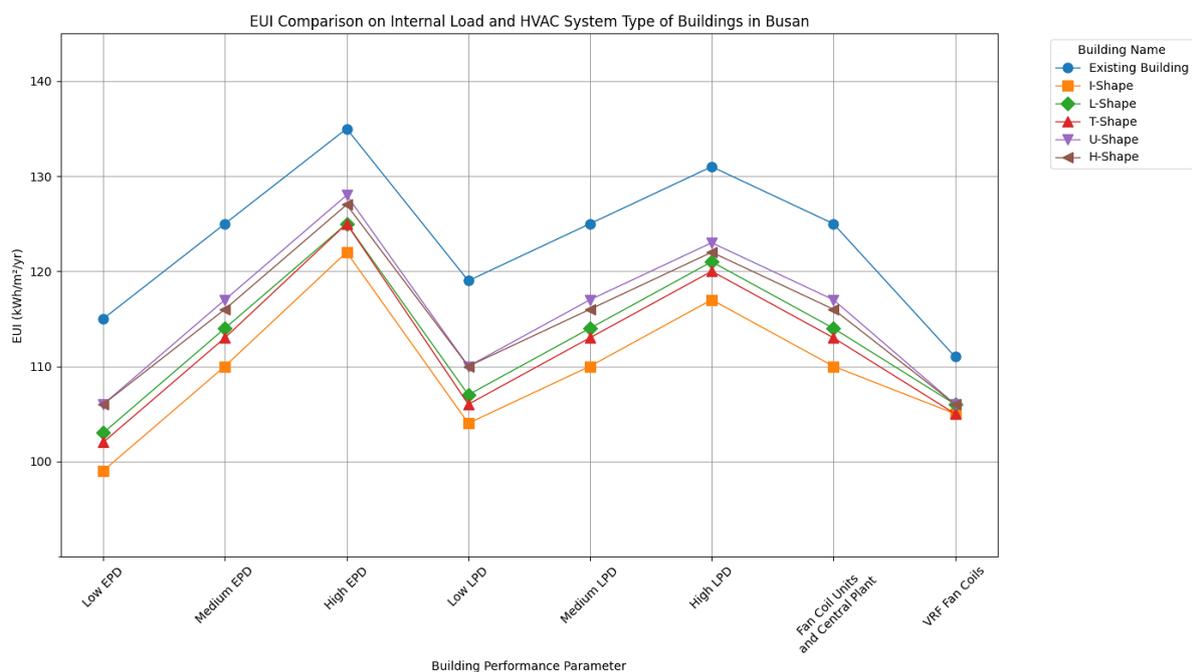


Figure 10. EUI vs. Internal Load and HVAC System Type

The I-shaped building demonstrates superior energy performance at lower internal loads. At a low EPD (10 W/m²), the I-shaped building achieves the lowest recorded EUI of 99 kWh/m²/yr. Even under medium EPD and medium LPD conditions, its EUI values remain low, ranging between 104 and 110 kWh/m²/yr. These results indicate the efficiency of linear, compact designs in reducing energy consumption, particularly at minimal internal loads. As shown in Figure 10, buildings with simpler geometries, such as the I- and L-shaped ones, consistently outperform their more complex counterparts under similar conditions.

The HVAC system type also plays a critical role in determining the EUI. Buildings equipped with variable-refrigerant-flow (VRF) fan coils achieve significantly lower EUI values than those using fan coil units and central plants. For instance, the existing building's EUI with fan coil units, 125 kWh/m²/yr, decreases to 111 kWh/m²/yr with VRF fan coils. This trend is consistent across all building shapes, with VRF systems reducing the EUI by 5%–10%. Therefore, the combination of efficient HVAC systems and optimized internal loads can substantially improve building energy performance.

Figure 11 highlights the substantial influence of PV systems on the EUI of buildings in Busan. The existing building, lacking any PV system, has an EUI of 125 kWh/m²/yr. This value diminishes markedly to 91 kWh/m²/yr at a 15% PV efficiency and further declines to 80 kWh/m²/yr at a 20% PV efficiency. These results underscore the significant decrease in energy usage attainable through the integration of PV systems into architectural design.

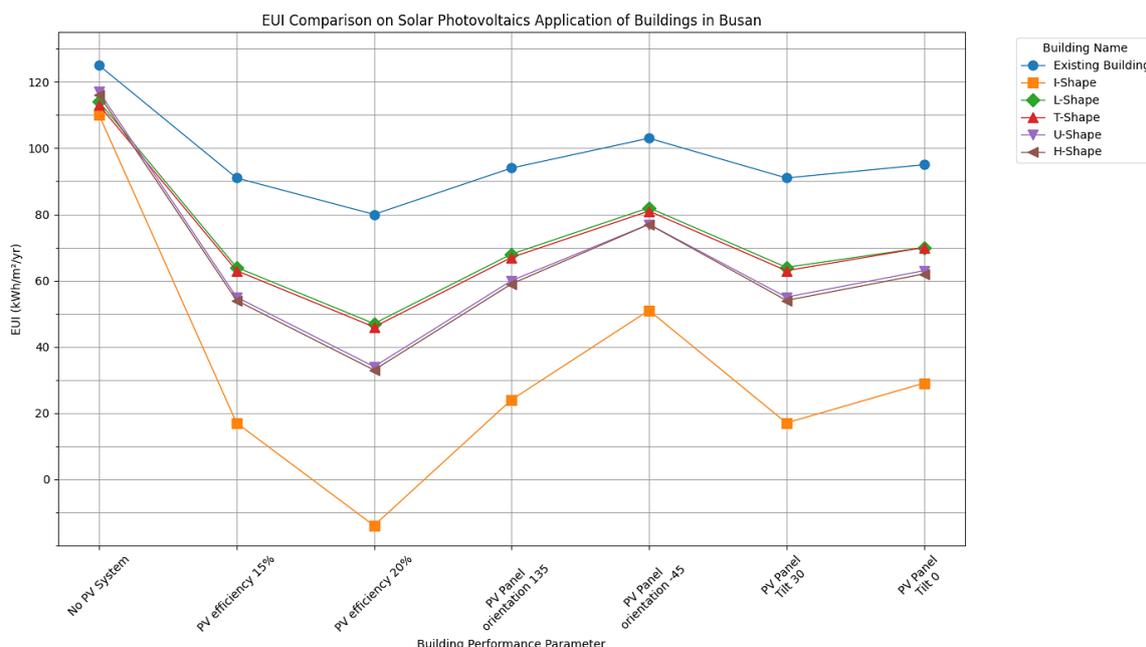


Figure 11. EUI vs. Solar PV Application in Buildings in Busan

The I-shaped structure exhibits exceptional energy efficiency when integrated with PV systems. In the absence of PV systems, it exhibits an EUI of 110 kWh/m²/yr. At a 20% PV efficiency, its EUI decreases considerably to -14 kWh/m²/yr, signifying that the building produces more energy than it utilizes. At a 15% PV efficiency, the EUI diminishes to just 17 kWh/m²/yr. These results demonstrate the efficacy of PV systems in attaining energy-positive results, especially in efficient building geometries, such as the I-shaped design.

The direction and inclination of PV panels also significantly influence the EUI. A southeast-facing panel orientation of 135° yields an EUI of 94 kWh/m²/yr in the existing building, but a southwest-facing orientation of -45° results in an EUI of 103 kWh/m²/yr. A 30° tilt performs excellently compared with a 0° tilt. The I-shaped building has an EUI of 17 kWh/m²/yr at a 30° tilt but 29 kWh/m²/yr at a 0° tilt. This underscores the need to optimize panel positioning to enhance solar energy capture.

EUI reduction increases with PV efficiency across all building types. In the absence of PV systems, the average building EUI values are 116–125 kWh/m²/yr. At a 15% PV efficiency, the average EUI lowers to 64–91 kWh/m²/yr, and at a 20% PV efficiency, it further diminishes to 33–80 kWh/m²/yr. Thus, PV systems efficiently reduce energy consumption while yielding energy-positive results for specific building types.

3.2. Model Performance and Deployment

We examined the benchmark model's EUI predictions and implemented it with a graphical interface for easier use.

3.2.1. Model Benchmarking Results

According to the model benchmarking results in Table 4, XGBoost achieves an exceptionally high R-squared value of 0.99 (±0.01), indicating that this model can explain 99% of the variability in the target variable. This high R-squared value is primarily attributed to the characteristics of the dataset, which was created using Sefaira, a parametric architectural tool that uses a robust mathematical model to forecast RES EUI by considering various parametric conditions. The machine learning models were essentially developed to replicate the output of the Sefaira base model. Sefaira's output remains consistent with its internal model when anomalies are introduced into the input. This consistent output is subsequently used as the truth data during training.

Table 4. Benchmarking Results of Five Models on Three Metrics

	LR	XGBoost	RFR	GBR	SVM
R-Squared (SD)	0.94 (0.01)	0.99 (0.01)	0.98 (0.01)	0.98 (0.01)	0.06 (0.04)
RMSE (SD)	12.09 (1.50)	4.57 (0.84)	5.73 (0.83)	6.98 (1.03)	48.04 (3.01)
MAE (SD)	7.62 (0.60)	1.99 (0.20)	3.01 (0.28)	3.29 (0.21)	31.67 (2.43)

Consequently, the machine learning model continues to exhibit exceptionally high accuracy. RFR and GBR demonstrate significant predictive capability, both achieving an R-squared value of 0.98 (±0.01). RFR marginally outperforms GBR in terms of the RMSE and MAE. LR likewise has a good R-squared value of 0.94 (±0.01), but its error measures are higher than those of the top three models.

Interestingly, the top-performing models (XGBoost, RFR, and GBR) are all ensemble methods, highlighting their effectiveness in capturing complex patterns in data and showing the nonlinear relationship between the input features and RES EUI. By contrast, the SVM model performs poorly, with an R-squared of only 0.06 (± 0.04) and a much higher error value, indicating its unsuitability for this task without extensive tuning. These results underscore the importance of model selection in energy consumption prediction and suggest that ensemble methods, particularly XGBoost, may be particularly suitable for EUI estimation. However, these models should be further tested using more varied data to assess their robustness under less controlled conditions.

3.2.2. Feature Importance

We assessed the importance of each feature to determine the most significant weights in EUI prediction based on each machine learning model. We then ranked them from largest to smallest, according to the best-performing model (XGBoost). However, in tree-based models, including XGBoost, feature importance does not directly indicate a causal relationship or correlation between features and target variables. This reflects the relative importance of features for model prediction.

Climate zones are the most influential factor in analyzing the feature importance data for predicting RES EUI using the XGBoost model. The model assigns the highest importance to climate zone 7 (0.6753), followed by zones 6A (0.0836) and 2A (0.0165), indicating that geographic and weather-related variables play a crucial role in defining energy consumption patterns. Interestingly, the second most important feature is related to PV systems (PV panel area, 0.1561), suggesting that on-site renewable energy generation significantly impacts the overall EUI. The HVAC system type, particularly that with fan coil units and a central plant (0.0189), is also a top predictor, highlighting the importance of heating and cooling loads in building energy consumption. These findings are presented in Table 5.

Table 5. Feature Importance of Compared Models (Sorted Using XGBoost)

	LR	XGBoost	RFR	GBR	SVM
Climate zone: 7	4528797652	0.6753106	0.57895963	0.5191214	5.93E-10
PV panel area	-0.0100613	0.15610759	0.17401452	0.17547178	n/a
Climate zone: 6A	4528797591	0.0835563	0.06479539	0.04178501	2.52E-11
HVAC type: fan coil units and central plan	6.23664769	0.01894984	0.00436442	0.00351075	4.79E-06
Climate zone: 2A	4528797576	0.01653481	0.02195132	0.02026711	0.00853889
Climate zone: 5A	4528797585	0.01320698	0.02153283	0.01535783	8.62E-11
Roof type: concrete	-0.5872072	0.0046927	0.00059881	0.00180147	1.02E-10
Climate zone: 4A	4528797546	0.0046745	0.00220305	0.02706119	3.22E-11
Heating equipment design capacity	0.10618511	0.00431184	0.06011499	0.14479743	n/a
PV panel efficiency	-3.6500301	0.003307	0.00365433	0.00500414	n/a
Floor finish: tiles	-3.1312618	0.00327052	0.00043843	0.00066108	6.54E-10
Cooling equipment design capacity	-0.0006291	0.00303468	0.05025901	0.03294517	n/a
EPD	2.00703666	0.00291028	0.00143681	0.00206033	n/a
Climate zone: 3A	4528797563	0.00241579	0.00162409	0.00159645	2.90E-09
LPD	2.1575934	0.00157245	0.00062005	0.00069449	n/a
PV panel orientation	-0.0341373	0.00121757	0.00185087	0.00160071	n/a
Infiltration rate	13.1414843	0.00120967	0.00060339	0.00046044	n/a
Climate zone: 0A	4528797630	0.00105655	0.00026757	0.00022571	2.57E-08
Façade glazing U-value	2.17273578	0.00101224	0.00081678	0.00056295	0.00040307
PV panel tilt	0.04200055	0.00033519	0.00030575	0.00048796	n/a
Total floor area	-0.0022201	0.00031315	0.00077557	4.87E-05	0.00941174
Wall type: precast concrete	-0.6289629	0.0002461	7.99E-05	0	3.09E-09
Façade glazing SHGC	-0.0530887	0.00020767	0.00069462	0.00074941	n/a
Wall type: concrete block	-0.0128642	0.00016681	3.59E-05	0	3.68E-10
Wall type: brick plaster	0.64179694	0.0001651	2.78E-05	0	9.52E-10
Climate zone: 1A	4528797626	0.00012642	0.00010864	0.00025551	1.93E-07
Floor finish: hardwood	2.75393596	5.15E-05	8.16E-05	0	3.78E-09
Building orientation	0.00172187	2.86E-05	0.00011267	0	0.11951629
Floor finish: carpet	0.37733367	7.56E-06	5.52E-05	0	1.08E-11
Floor U-value	4.22156619	0	0.00038788	0.00103373	n/a
Roof U-value	2.40748417	0	0.00050358	0.00063877	n/a
Walls U-value	0.7149665	0	0.00010879	0	n/a
Roof type: metal deck	0.58718838	0	0.00077127	0.00038755	5.07E-09
HVAC type: VRF fan coils	-6.2366422	0	0.00584456	0.00141293	1.31E-07

This hierarchy of feature importance reveals that macro level factors, such as climate conditions and major building systems, substantially influence RES EUI more than specific building envelope characteristics or materials. Notably, attributes such as wall type, roof type, and glazing properties exhibit relatively low importance in the model. Thus, although detailed building characteristics are relevant, their effect on the EUI may be overshadowed by or implicitly captured through broader environmental and systemic factors. These findings have significant implications for energy-efficient building design and policymaking, highlighting the need to prioritize climate-appropriate strategies and renewable energy integration to reduce residential energy consumption.

3.2.3. Deployment

As shown in Figure 12, we developed the machine learning models in Python, specifically using the scikit-learn library, called inference models. This model, a mathematical representation of a problem, is trained to make predictions or decisions. The model used in this study, developed using the XGBoost algorithm, consists of a decision tree that provides leaf scores.

Figure 12. Web Interface of RES EUI Estimator

However, the model output, when run in Python, is generally displayed on a terminal; this may not be user-friendly, especially for those without a technical background. Therefore, to make the prediction model accessible and enable users to interact with it intuitively, we developed a Web-based interface using JavaScript for the front end and Flask for the back end.

This Web application allows users to interact with the prediction model conveniently without needing to interact directly with the Python code. Users can enter values for each feature through this Web interface. They can select items from drop-down menus for categorical features based on the training data used during model development. Additionally, this Web interface offers the option to generate random values for each feature and predict the EUI.

3.3. Validation Results and Model Refinement

Validation was conducted by comparing the interior temperature data derived from the EnergyPlus simulations with the data measured using the TR-72nw temperature and humidity logger in the room in April 2024. In addition, direct measurements from the Vantage Pro2 weather station on the rooftop of building S04 were compared with the climate data used in the simulations, sourced from the EPW file. This comparison focused on the outdoor temperature, humidity, and wind speed.

Validation was performed to assess the accuracy of the data points and ensure model reliability. By identifying and addressing discrepancies between the simulation and real-world data, we further customize the model to provide more precise estimates of building energy performance. The comparison results and the measures taken to enhance model precision are as follows (Figure 13):

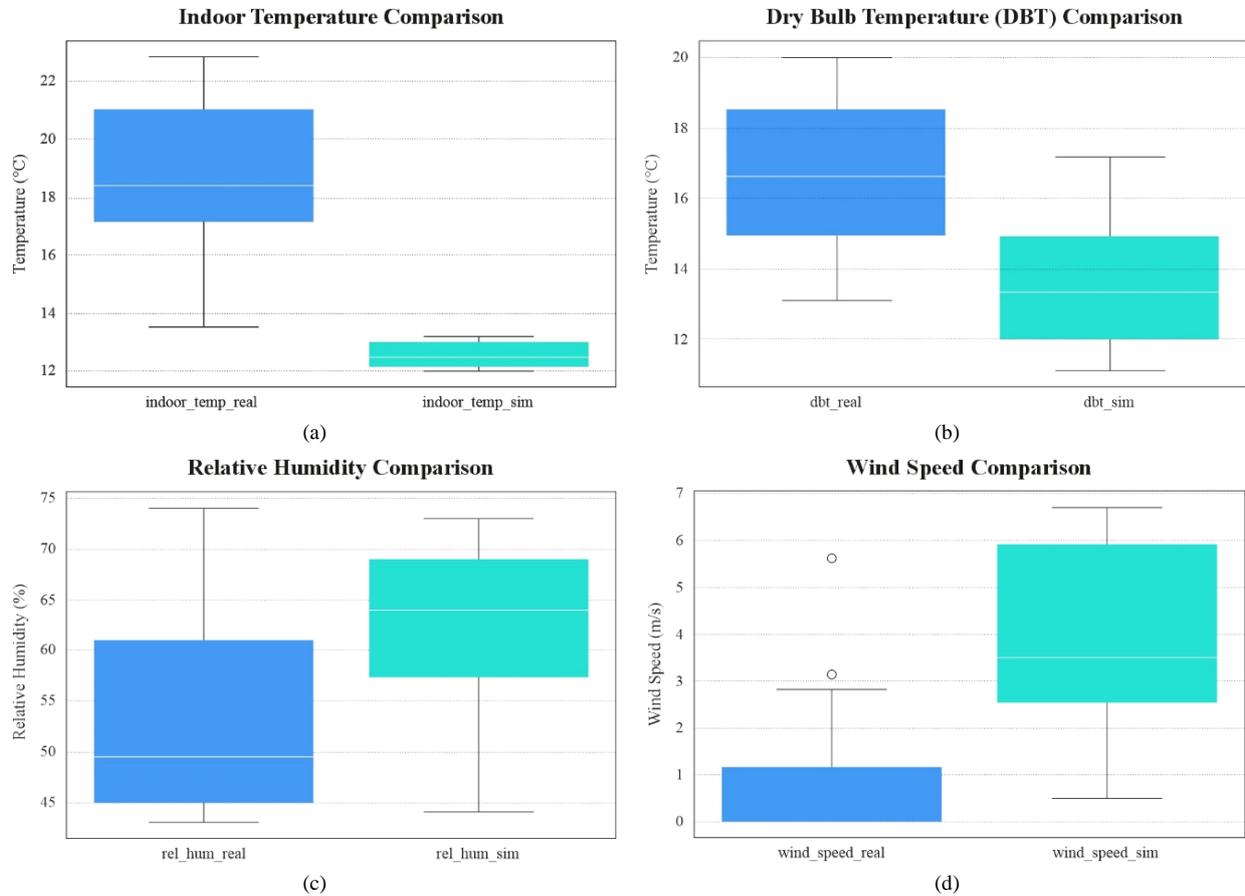


Figure 13. Comparison of Real-World and Simulated Data: (a) Indoor Temperature, (b) Dry Bulb Temperature, (c) Relative Humidity, (d) Wind Speed

The box-plot graphs in Figure 13 indicate notable disparities between the actual and simulated data. In Figure 13a, the indoor temperature distribution demonstrates that the simulated data are more stable, with the median temperature being 13°C to 14°C. The real-world data show more significant fluctuations, with the median temperature varying from 16°C to 18.5°C. In Figure 13b, the simulation's dry bulb temperature (DBT) is consistently 10% to 15% lower than the real-world data. Thus, the simulation tends to generate lower temperature fluctuations.

Figure 13c demonstrates that the simulated data inadequately account for fluctuations in relative humidity, resulting in a discrepancy of up to 20% relative to the actual data. This discrepancy is particularly noticeable during specific periods, such as morning and night. Figure 13d shows that the simulated wind speed is consistently lower and has a narrower range compared with the real-world data, with a discrepancy of up to 30%, particularly during daylight hours. Although the two datasets have comparable overall trends throughout the day, the simulation often fails to represent the actual conditions accurately, indicating limitations in the simulation model.

The correlation analysis in Figure 14 highlights the critical relationships between the real-world and simulated data, providing insights into the model's performance and areas for refinement. The strong positive correlation ($r = 0.97$) between indoor temperature and DBT means the simulation effectively captures the influence of external climatic conditions on indoor environments. This finding is consistent with previous studies emphasizing the role of external temperatures in building energy performance, particularly in regions with limited insulation or low-efficiency HVAC systems. The moderate negative correlation between wind speed and indoor temperature ($r = -0.30$ to $r = -0.40$) reflects the cooling effects of increased ventilation, aligning with research on natural ventilation strategies. However, the weaker correlation suggests that the simulation may not fully capture the variability introduced by wind-driven cooling in buildings with complex ventilation dynamics or nonstandard envelope performance.

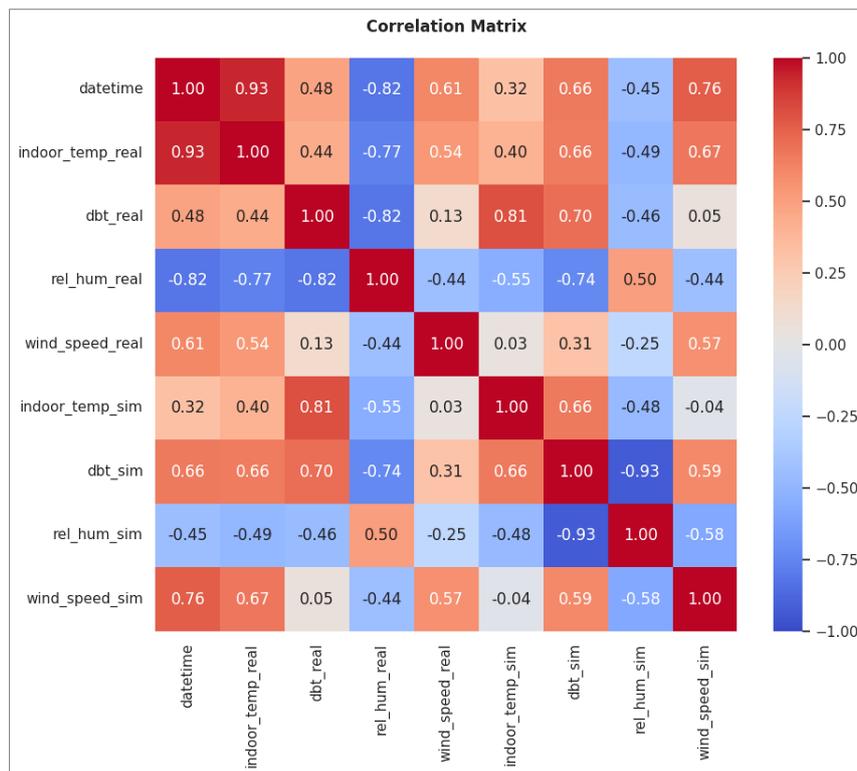


Figure 14. Correlation Matrix between Real-World and Simulated Data

According to the relative humidity correlation between the simulated and real-world data ($r = 0.60$ to $r = 0.70$), although the model captures general trends, it struggles with finer fluctuations. This agrees with prior research highlighting the challenges of modeling humidity due to its dependence on external weather conditions and internal loads, such as occupancy and equipment use. Therefore, the simulation model should be refined to incorporate more detailed representations of internal heat and moisture sources for enhanced accuracy in predicting energy performance.

This study adds to the growing body of NZEB literature by validating simulation results against empirical data, a crucial step in ensuring the reliability of EUI prediction. Comparing such data across diverse climate zones demonstrates the broad applicability of the findings and sheds light on the regional variations influencing energy performance. For example, discrepancies in relative humidity and wind speed indicate the potential need for improvements to better reflect real-world dynamics, particularly in buildings under varying external and internal conditions.

This study focuses on the accuracy of energy consumption and EUI prediction. It does not address factors such as cost implications, socioeconomic influences, or construction conditions. These aspects are beyond the study's objectives, as the primary aim is to integrate advanced machine learning models, such as XGBoost, with parametric architecture to enhance energy performance prediction. Future work can broaden this study by exploring the abovementioned factors, contributing further to the development of NZEB frameworks and sustainable building design practices.

4. Conclusions

We conducted 1,350 simulations using Sefaira to analyze the impacts of different design parameters on building EUI. The simulations included six building types—the existing building and I-, T-, U-, and H-shaped buildings—in eight locations according to the ASHRAE 90.1 climate zones. The simulation results indicate that the building orientation, shape, and climatic conditions are essential factors affecting the EUI. For example, buildings facing 135° , 90° , 135° , and 270° show higher EUI values, whereas the 0° and 180° orientations enable the efficient management of solar light and heat gain.

In addition, façade materials, internal loads, HVAC systems, and PV applications significantly impact the EUI. Facades with triple-pane insulated glass and low-E glass perform the best, whereas metal deck roofs show the worst performance. As for HVAC systems, VRF fan coils are more often efficient than fan coil units and central plants. The use of PV panels with a 20% PV efficiency, a 135° orientation, and a 30° inclination can significantly reduce the EUI, even reaching negative values in some I-shaped buildings, indicating the potential to achieve NZEB status.

This study highlights the role of building orientation and design factors in reducing energy consumption, thus substantially contributing to sustainable building design and optimization. In particular, XGBoost demonstrates exceptional efficacy in EUI forecasting, with an R-squared value of 0.99, an RMSE of 4.57, and an MAE of 1.99.

However, high R-squared values may not accurately represent the variability that can arise in real-world conditions, where the input data may not be as pure or systematic as those generated in a controlled environment, such as that used in this study. A model's reliability and robustness under less controlled conditions can be assessed using a broad range of real-world data, including various building types and materials. The proposed methodology can enhance current architectural design practices by providing comprehensive benefits across a diverse array of building functions and climatic conditions. Architects and planners can create more sustainable and efficient structures by understanding the correlation between local climatic conditions, building orientations, and building design. Aside from contributing to theoretical knowledge, this study has practical implications for the building sector, aiding in the attainment of NZEB standards and reducing the environmental impact of buildings.

In this study, a controlled dataset was used to benchmark the performance of the XGBoost model. Future work will focus on validating the model using noisy real-world datasets. These datasets will include potential challenges, such as measurement errors, missing values, and inconsistent feature distributions, which are often encountered in practical scenarios. To evaluate the robustness of the model to these problems, we will incorporate data augmentation to simulate real-world noise and test the model's performance. In addition, we aim to collaborate with stakeholders in the building energy sector to obtain real-world datasets and assess the model's predictive accuracy and reliability in operational settings. This future direction will ensure that this study's findings are applicable to and impactful in practical energy efficiency applications.

5. Declarations

5.1. Author Contributions

Conceptualization, M.K., M.C.T.M., and J.J.Y.; methodology, M.K. and M.C.T.M.; formal analysis, M.K. and M.C.T.M.; investigation, M.K. and M.C.T.M.; resources, M.K., M.C.T.M., and J.J.Y.; writing—original draft preparation, M.K.; writing—review and editing, M.K. and M.C.T.M.; visualization, M.K.; supervision, J.J.Y.; project administration, J.J.Y.; funding acquisition, J.J.Y. All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

Data presented in this study are available on request from the corresponding author.

5.3. Funding and Acknowledgments

The authors would thank the support of the Basic Science Research Program through the National Research Foundation (NRF) funded by the Korea Ministry of Education (No. 2016R1A6A1A03012812).

5.4. Conflicts of Interest

The authors declare no conflict of interest.

6. References

- [1] Almomani, A., Almeida, R. M. S. F., Vicente, R., & Barreira, E. (2024). Critical Review on the Energy Retrofitting Trends in Residential Buildings of Arab Mashreq and Maghreb Countries. *Buildings*, 14(2), 338. doi:10.3390/buildings14020338.
- [2] Mehedhi, H., AMM, S. A., Mohammad Alamgir, H., & AMM, N. A. (2024). The prospects of zero energy building as an alternative to the conventional building system in Bangladesh (A review). *Journal of Civil Engineering and Environmental Sciences*, 10(2), 039–049. doi:10.17352/2455-488x.000082.
- [3] Noh, Y., Jafarinejad, S., & Anand, P. (2024). A Review on Harnessing Renewable Energy Synergies for Achieving Urban Net-Zero Energy Buildings: Technologies, Performance Evaluation, Policies, Challenges, and Future Direction. *Sustainability (Switzerland)*, 16(8). doi:10.3390/su16083444.
- [4] Xiaoxiang, Q., Junjia, Y., Haron, N. A., Alias, A. H., Law, T. H., & Abu Bakar, N. (2024). Status, Challenges and Future Directions in the Evaluation of Net-Zero Energy Building Retrofits: A Bibliometrics-Based Systematic Review. *Energies*, 17(15), 3826. doi:10.3390/en17153826.
- [5] Hongvityakorn, B., Jaruwatupant, N., Khongphinitbunjong, K., & Aggarangsi, P. (2024). Achieving Nearly Zero-Energy Buildings through Renewable Energy Production-Storage Optimization. *Energies*, 17(19), 4845. doi:10.3390/en17194845.
- [6] Panchal, P., Mehta, V., Thakur, M., & Kumar, B. (2020). Pathway to Net-Zero Energy Buildings. *Plant Archives*, 20 Special (August), 156–159.
- [7] Marszal, A. J., & Heiselberg, P. (2011). Life cycle cost analysis of a multi-storey residential Net Zero Energy Building in Denmark. *Energy*, 36(9), 5600–5609. doi:10.1016/j.energy.2011.07.010.

- [8] Chen, S. Y. (2019). Use of green building information modeling in the assessment of net zero energy building design. *Journal of Environmental Engineering and Landscape Management*, 27(3), 174–186. doi:10.3846/jeelm.2019.10797.
- [9] Becchio, C., Corgnati, S. P., Vio, M., Crespi, G., Prendin, L., Ranieri, M., & Vidotto, D. (2017). Toward NZEB by optimizing HVAC system configuration in different climates. *Energy Procedia*, 140, 115–126. doi:10.1016/j.egypro.2017.11.128.
- [10] Latief, Y., Berawi, M. A., Koesalamwardi, A. B., Sagita, L., & Herzanita, A. (2019). Cost optimum design of a tropical near zero energy house (nZEH). *International Journal of Technology*, 10(2), 376–385. doi:10.14716/ijtech.v10i2.1781.
- [11] Moreno, B., Del Ama Gonzalo, F., Ferrandiz, J. A., Lauret, B., & Hernandez, J. A. (2019). A building energy simulation methodology to validate energy balance and comfort in zero energy buildings. *Journal of Energy Systems*, 3(4), 168–182. doi:10.30521/jes.623285.
- [12] Santos-Herrero, J. M., Lopez-Guede, J. M., & Flores-Abascal, I. (2021). Modeling, simulation and control tools for nZEB: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 142. doi:10.1016/j.rser.2021.110851.
- [13] Mahiwal, S. G., Bhoi, M. K., & Bhatt, N. (2021). Evaluation of energy use intensity (EUI) and energy cost of commercial building in India using BIM technology. *Asian Journal of Civil Engineering*, 22(5), 877–894. doi:10.1007/s42107-021-00352-5.
- [14] Geng, Y., Lin, B., & Zhu, Y. (2020). Comparative study on indoor environmental quality of green office buildings with different levels of energy use intensity. *Building and Environment*, 168(August), 106482. doi:10.1016/j.buildenv.2019.106482.
- [15] Samadi, M., & Fattahi, J. (2021). Energy use intensity disaggregation in institutional buildings – A data analytics approach. *Energy and Buildings*, 235. doi:10.1016/j.enbuild.2021.110730.
- [16] Raimo Simson, Kirsten Engelund Thomsen, Kim Bjarne Wittchen, & Jarek Kurnitski. (2021). NZEB Requirements vs European Benchmarks in Residential Buildings. *The REHVA European HVAC Journal*, 58(2), 40–44.
- [17] Garcia, J. F., & Kranzl, L. (2018). Ambition levels of nearly zero energy buildings (nZEB) definitions: An approach for cross-country comparison. *Buildings*, 8(10), 143. doi:10.3390/buildings8100143.
- [18] Tabrizi, A. (2021). Sustainable Construction, LEED as a Green Rating System and the Importance of Moving to NZEB. *E3S Web of Conferences*, 241. doi:10.1051/e3sconf/202124102001.
- [19] USGBC. (2025). LEED V4.1 Building Design and Construction. U.S. Green Building Council (USGBC), Washington, United States. Available online: https://build.usgbc.org/bd+c_guide (accessed on February 2025).
- [20] Xia, B., Zuo, J., Skitmore, M., Pullen, S., & Chen, Q. (2013). Green Star Points Obtained by Australian Building Projects. *Journal of Architectural Engineering*, 19(4), 302–308. doi:10.1061/(asce)ae.1943-5568.0000121.
- [21] DOE. (2018). Green Building Report 2018. DC. Department of Energy & Environment (DOEE), Washington, United States. Available online: <https://www.slideshare.net/slideshow/2018-green-building-report-finalpptx/254542535#27> (accessed on February 2025).
- [22] Tracy, W. H., Joy, S., Flora, Y., & Audrey, W. (2020). Green Building Rating Systems: Energy Benchmarking Study. *Environment Design Guide*, Civic Exchange, Hong Kong.
- [23] Grainger, C., & Rattenbury, J. (2021). Net Zero New Buildings Evidence and guidance to inform Planning Policy. South West Energy Hub, Bristol, United Kingdom.
- [24] IBEC. (2014). Japan Sustainable Building Consortium, CASBEE for Buildings (New Construction) (1st Ed.). Volume 1, Institute for Building Environment and Energy Conservation (IBEC), Tokyo, Japan.
- [25] Portalatin, M., Roskoski, M., & Shouse, T. (2015). Sustainability How-to Guide Series Green Building Rating Systems. Houston, United States. Available online: <http://cdn.ifma.org/sfcdn/membership-documents/green-rating-systems-htg-final.pdf> (accessed on February 2025).
- [26] Wang, N., Fowler, K. M., & Sullivan, R. S. (2012). Green Building Certification System. PNNL-20966, US Department of Energy, Washington, United States.
- [27] Robar, K. (2018). Comparative analysis study to compare LEED v4 and green globes in Newfoundland and Labrador. Morrison Hershfield Limited, Ottawa, Canada.
- [28] Vierra, S. (2018). Green Building Standards and Certification Systems. Whole Building Design Guide (WBDG), National Institute for Building Sciences, Washington, United States. Available online: https://globalgbc.org/wp-content/uploads/2022/07/034_green-building-standards-and-certification-system.pdf (accessed on February 2025).
- [29] Lee, J., & Shepley, M. (2019). The green standard for energy and environmental design (g-seed) for multi-family housing rating system in Korea: A review of evaluating practices and suggestions for improvement. *Journal of Green Building*, 14(2), 155–176. doi:10.3992/1943-4618.14.2.155.

- [30] Kim, K. H., Chae, C. U., & Cho, D. (2022). Development of an assessment method for energy performance of residential buildings using G-SEED in South Korea. *Journal of Asian Architecture and Building Engineering*, 21(1), 133–144. doi:10.1080/13467581.2020.1838286.
- [31] No, S., & Won, C. (2020). Comparative analysis of energy consumption between green building certified and non-certified buildings in Korea. *Energies*, 13(5), 1049. doi:10.3390/en13051049.
- [32] Nahan, R. T. (2019). *Architect's Guide to Building Performance: Integrating performance simulation in the design process*. The American Institute of Architects, New York, United States.
- [33] Suphavarophas, P., Wongmahasiri, R., Keonil, N., & Bunyarittikit, S. (2024). A Systematic Review of Applications of Generative Design Methods for Energy Efficiency in Buildings. *Buildings*, 14(5), 1311. doi:10.3390/buildings14051311.
- [34] Stevanović, S., Dashti, H., Milošević, M., Al-Yakoob, S., & Stevanović, D. (2024). Comparison of ANN and XGBoost surrogate models trained on small numbers of building energy simulations. *PloS One*, 19(10), e0312573. doi:10.1371/journal.pone.0312573.
- [35] Lu, Y., Wu, W., Geng, X., Liu, Y., Zheng, H., & Hou, M. (2022). Multi-Objective Optimization of Building Environmental Performance: An Integrated Parametric Design Method Based on Machine Learning Approaches. *Energies*, 15(19), 7031. doi:10.3390/en15197031.
- [36] Labib, R. (2022). Integrating Machine Learning with Parametric Modeling Environments to Predict Building Daylighting Performance. *IOP Conference Series: Earth and Environmental Science*, 1085(1), 8DUMMY. doi:10.1088/1755-1315/1085/1/012006.
- [37] Veiga, R. K., Veloso, A. C., Melo, A. P., & Lamberts, R. (2021). Application of machine learning to estimate building energy use intensities. *Energy and Buildings*, 249, 111219. doi:10.1016/j.enbuild.2021.111219.
- [38] Seyedzadeh, S., Pour Rahimian, F., Rastogi, P., & Glesk, I. (2019). Tuning machine learning models for prediction of building energy loads. *Sustainable Cities and Society*, 47. doi:10.1016/j.scs.2019.101484.
- [39] Amasyali, K., & El-Gohary, N. (2021). Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings. *Renewable and Sustainable Energy Reviews*, 142. doi:10.1016/j.rser.2021.110714.
- [40] Solmaz, A. S. (2020). Machine learning based optimization approach for building energy performance. 2020 Building Performance Analysis Conference and SimBuild, 29 September-1 October, Virtual Meeting.
- [41] Guo, H., Duan, D., Yan, J., Ding, K., Xiang, F., & Peng, R. (2022). Machine Learning-Based Method for Detached Energy-Saving Residential Form Generation. *Buildings*, 12(10), 1504. doi:10.3390/buildings12101504.
- [42] Barbaresi, A., Ceccarelli, M., Menichetti, G., Torreggiani, D., Tassinari, P., & Bovo, M. (2022). Application of Machine Learning Models for Fast and Accurate Predictions of Building Energy Need. *Energies*, 15(4), 1266. doi:10.3390/en15041266.
- [43] Kumar, P., Kamalakshi, N., & Karthick, T. (2024). Towards Sustainable Architecture: Machine Learning for Predicting Energy Use in Buildings. *Proceedings 3rd International Conference on Advances in Computing, Communication and Applied Informatics, ACCAI 2024*, 10602461. doi:10.1109/ACCAI61061.2024.10602461.
- [44] Ni, Z., Zhang, C., Karlsson, M., & Gong, S. (2023). Leveraging Deep Learning and Digital Twins to Improve Energy Performance of Buildings. 2023 IEEE 3rd International Conference on Industrial Electronics for Sustainable Energy Systems, IESES 2023, Shanghai, China. doi:10.1109/IESES53571.2023.10253721.
- [45] Gan, H., & Gao, W. (2024). Ensemble machine learning for managing the required thermal energy from the architectural characteristics of residential buildings. *International Journal of Low-Carbon Technologies*, 19, 1222–1230. doi:10.1093/ijlct/ctae064.
- [46] Abdelsattar, M., Ismeil, M. A., Azim Zayed, M. M. A., Abdelmoety, A., & Emad-Eldeen, A. (2024). Assessing Machine Learning Approaches for Photovoltaic Energy Prediction in Sustainable Energy Systems. *IEEE Access*, 12, 107599–107615. doi:10.1109/ACCESS.2024.3437191.
- [47] Barbur, V. A., Montgomery, D. C., & Peck, E. A. (1994). Introduction to Linear Regression Analysis. *The Statistician*, 43(2), 339. doi:10.2307/2348362.
- [48] Chen, T., & Guestrin, C. (2016). XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. doi:10.1145/2939672.2939785.
- [49] Breiman, L. (2001). *Machine Learning*, 45(1), 5–32. doi:10.1023/a:1010933404324.
- [50] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. doi:10.1214/aos/1013203451.
- [51] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199–222. doi:10.1023/b:stco.0000035301.49549.8.