

Article

Trustworthy Data-Driven Hybrid Modeling of Building Energy Performance and Greenhouse Gas Emissions

Abdulkadir Gungor¹, Ahmet Nur², Sabir Rustemli², Faruk Kurker³, Gökhan Şahin^{4,*}, Erdal Akin^{5,6,7,8,*}, Kayode S. Adewole^{5,6} and Andreas Jacobsson^{5,6}

¹ Department of Electrical and Energy, Technical Sciences Vocational School, Muş Alparslan University, Muş 49250, Türkiye

² Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Bitlis Eren University, Bitlis 13100, Türkiye

³ Department of Electrical and Electronics Engineering, Faculty of Engineering, Adiyaman University, Adiyaman 02040, Türkiye

⁴ Copernicus Institute of Sustainable Development, Utrecht University, Princetonlaan 8A, 3584 CB Utrecht, The Netherlands

⁵ Department of Computer Science and Media Technology, Malmö University, 205 06 Malmö, Sweden

⁶ Sustainable Digitalisation Research Centre, Malmö University, 205 06 Malmö, Sweden

⁷ Biofilms Research Center for Biointerfaces (BRCB), Malmö University, 205 06 Malmö, Sweden

⁸ Department of Computer Engineering, Bitlis Eren University, Bitlis 13100, Türkiye

* Correspondence: g.sahin0638@gmail.com (G.Ş.); erdal.akin@mau.se (E.A.)

Abstract

Reducing carbon dioxide (CO₂) emissions from buildings is essential for climate change mitigation, with universities representing major energy consumers. This study develops a hybrid data-driven framework combining machine learning and simplified emission factor rescaling to predict campus-wide CO₂ emissions. Nine machine learning models were comparatively evaluated under both cross-sectional and temporal validation settings. Among all evaluated models, the Artificial Neural Network (ANN) demonstrated the most reliable predictive performance, achieving the best balance between prediction accuracy and generalization capability. Although the proposed physics-informed LSBoost_PI framework aimed to integrate physical priors with machine learning through residual correction, it did not improve predictive generalization under the limited sample conditions of the dataset. Time-series cross-validation further confirmed the ANN model's temporal forecasting capability (RMSE = 2.13 ton/year, R² = 0.985). To support trustworthy and interpretable machine learning, feature importance analysis identified CO₂ intensity indicators (CO₂/kWh and CO₂/TEP) as the dominant drivers of emissions. The study also conducted an emission reduction assessment, revealing that a limited number of high-energy buildings dominate overall campus emissions. These findings provide actionable insights for campus-scale energy management, supporting targeted energy efficiency improvements and renewable energy integration strategies in high-emission buildings.



Academic Editor: Apple L.S. Chan

Received: 25 April 2026

Revised: 24 May 2026

Accepted: 30 May 2026

Published: 3 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: smart building energy management; CO₂ emission estimation; data-driven models; hybrid modeling; artificial neural networks

1. Introduction

Energy consumption and greenhouse gas emissions have become central concerns in global climate policy, with significant implications for the future. The building sector is a

primary contributor to both energy consumption and emissions, accounting for approximately 40% of global energy consumption and a significant portion of CO₂ emissions [1]. In cold climates, where heating demands are higher, this percentage can be even greater, making energy efficiency in buildings a crucial element of climate change mitigation. The exacerbating effects of climate change, including extreme weather conditions, also lead to an increased need for energy-consuming cooling systems in urban environments, adding to the demand for energy [2]. The integration of renewable energy sources, smart buildings through the Internet of Things (IoT), and energy efficiency measures is essential for reducing the building sector's carbon footprint. Traditional methods for estimating energy consumption and CO₂ emissions in buildings have limitations, particularly when it comes to accurately capturing complex interactions between building systems, user behaviors, and environmental factors.

Consequently, machine learning (ML) and artificial intelligence (AI) approaches have gained attention due to their ability to model nonlinear and high-dimensional relationships, offering improved energy prediction capabilities [3–5]. Hybrid models that combine AI with physical modeling techniques are particularly promising, as they provide both high accuracy and physical consistency, which is often lacking in purely data-driven approaches [6,7]. University campuses are ideal testing grounds for energy management and emission reduction strategies due to their varied building types, diverse energy needs, and manageable size for analysis. These campuses generate large datasets that can be analyzed to optimize energy performance, yet predicting campus-wide energy consumption and CO₂ emissions remains a challenge.

This study introduces a novel hybrid model combining machine learning techniques, such as artificial neural networks (ANN) and support vector regression (SVR), with simplified emission-factor-based modeling for more accurate CO₂ emission estimations. The model is applied to 18 buildings at Muş Alparslan University, Turkey, and evaluated against various performance metrics, contributing valuable insights for campus-scale energy management. While established simulation tools for CO₂ emissions calculations exist, they often fail to capture the complexities of building systems, user behavior, and environmental factors. The gap this study addresses lies in the integration of physical principles with data-driven insights to enhance model robustness and predictive accuracy. The hybrid approach proposed here combines simplified physical emission estimates with machine learning algorithms to improve the estimation of CO₂ emissions. This model provides trustworthy and reliable long-term predictions compared to traditional methods, which often overlook critical factors influencing energy use and emissions. By applying this hybrid approach to Muş Alparslan University, this research aims to offer more accurate campus-level CO₂ emission estimates, contributing to the broader effort of achieving sustainable, energy-efficient building practices. This novel framework not only enhances the reliability of carbon management strategies but also supports the integration of explainable and renewable energy sources, aligning with global sustainability goals.

Despite the promises of current hybrid models, a critical review of the literature reveals several significant shortcomings that this study directly addresses. First, most hybrid approaches in campus energy research combine physics-based simulation tools with a single machine learning algorithm (typically ANN or SVR); however, these approaches do not systematically compare the hybrid strategy with a comprehensive set of competing models under the same validation conditions. Consequently, verifying claims of superiority remains a challenging task. This study addresses this gap by simultaneously evaluating ten models, including SVR, Bayesian optimized SVR, GPR, LSBoost, ANN, bagged trees, ridge regression, single regression trees, a stacking ensemble, and the proposed physics-informed hybrid LSBoost (LSBoost_PI), providing a rigorous and transparent comparative evaluation

by assessing all models under the same five-fold cross-validation protocol. Second, existing hybrid models applied to campus environments primarily operate on large datasets covering hundreds of buildings or multiple campuses; in such cases, data abundance reduces the risk of overfitting. In contrast, this study is conducted under a deliberately small sample regime consisting of 18 heterogeneous campus buildings; this scenario better represents real-world corporate energy management scenarios where comprehensive data collection is limited due to cost, access, or legal constraints. The proposed framework is specifically designed and validated for this challenging low-data environment. Third, existing hybrid campus energy models rarely address model reliability beyond standard accuracy metrics. This study presents a multi-layered validation framework that encompasses not only prediction accuracy but also uncertainty quantification via conformal prediction and Gaussian process confidence intervals, noise robustness analysis under varying levels of input corruption, permutation-based feature importance, and leave-one-out cross-validation; this collectively ensures that the proposed model is both accurate and operationally reliable. Fourth, campus-scale CO₂ emission studies conducted under cold-climate regulatory frameworks—particularly the Turkish national BEP-TR building energy performance standard—are virtually absent from the international literature; this limits the transferability of findings from temperate or tropical climate studies to institutional contexts with high heating demand.

The remainder of this paper is organized as follows. Section 2 reviews related work on building energy and emissions. Section 3 presents the methodology adopted in this study. Section 4 reports the results, while Section 5 discusses the key findings. Finally, Section 6 concludes the paper and outlines directions for future research.

2. Related Works

The building sector has been a focal point in energy and emissions research for many years. Early research predominantly relied on physics-based models to simulate energy consumption, but with the rise of big data, data-driven techniques have become increasingly relevant. Machine learning approaches, particularly artificial neural networks (ANN), support vector machines (SVM), and hybrid models, have been employed to predict energy consumption and CO₂ emissions, with promising results in terms of accuracy and efficiency [4,7]. Recent studies have highlighted the advantages of integrating data-driven and physics-based approaches. For instance, Ref. [3] emphasized the importance of combining deep learning with physics-informed modeling approaches for urban building energy modeling to capture both dynamic behaviors and physically meaningful relationships within buildings. Furthermore, hybrid models have proven effective in minimizing the energy performance gap, a major issue in traditional modeling techniques, by improving the accuracy and consistency of energy predictions [6,8]. In the context of renewable energy integration, hybrid models have been employed to optimize the use of renewable energy in buildings. Refs. [5,9] demonstrated that machine learning models can accurately predict energy consumption while integrating renewable energy sources such as solar and wind power to achieve net zero energy consumption. Additionally, ref. [7] highlighted how combining physical and data-driven models can enhance energy prediction models' robustness and generalizability, particularly in urban-scale energy systems. University campuses, due to their unique combination of energy consumption patterns, make an ideal setting for evaluating energy management strategies. Ref. [10] assessed energy performance and energy-saving strategies for university classrooms, providing valuable insights into energy consumption patterns in academic settings. Ref. [11] further explored the potential for hybrid energy systems in academic complexes, focusing on the role of photovoltaic systems and their impact on campus-wide energy consumption. The integration of machine

learning models and renewable energy systems in campus-scale energy management has been explored in several studies. For example, ref. [12] proposed a hybrid deep neural network model to forecast power loads in academic buildings, demonstrating the efficacy of AI models in managing building energy systems. This aligns with the goals of the current study, which aims to provide accurate CO₂ emission estimates for a university campus using machine learning-based models combined with simplified physical approaches. Table 1 presents the Summary of Literature on Energy System Selection and Optimization in University Campuses.

Table 1. Summary of Literature on Energy System Selection and Optimization in University Campuses.

Study	Methodology	Energy Focus	Key Findings
A. Ahmed, T. Ge, J. Peng, W.-C. Yan, B.T. Tee, S. You (2022) [13]	Review of renewable energy technologies	Net-zero energy buildings	Reviews renewable energy strategies for achieving net-zero energy buildings.
S. Akbari (2025) [14]	Comprehensive guide	Building energy analysis	Provides a comprehensive guide on methods and tools for building energy analysis.
E.A.O. Batlle, et al. (2020) [15]	Case study method	Energy savings in higher education buildings	Develops a method for estimating baseline energy use and quantifying savings in university buildings.
M. Bourdeau, et al. (2019) [16]	Data-driven modeling review	Building energy forecasting	Reviews data-driven techniques for building energy consumption forecasting.
K. Amber, et al. (2018) [17]	Machine learning techniques	Electricity consumption in buildings	Proposes intelligent methods for forecasting electricity consumption in buildings using ML.
M. Ilbeigi, et al. (2020) [18]	ANN + Genetic algorithm	Energy optimization in office buildings	Combines ANN and genetic algorithms to optimize energy consumption in office buildings.
D.-K. Bui, et al. (2020) [19]	ANN + Firefly algorithm	Energy prediction in buildings	Enhances ANN with firefly algorithm for better energy prediction in buildings.
Y. Chen, et al. (2022) [20]	Physical + Data-driven models	Energy prediction in buildings	Discusses the combination of physical and data-driven models to improve energy prediction.
C. Deb, et al. (2017) [21]	Physical + Data-driven models	Energy prediction in buildings	Highlights the benefits of combining physical and data-driven models for building energy predictions.
X. Chen, et al. (2019) [22]	Hybrid model	Performance gap reduction in buildings	Proposes a hybrid model to reduce the building energy performance gap.
Y. Ahn, et al. (2022) [23]	Transfer learning-based prediction	Power consumption prediction in buildings	Uses transfer learning to predict building power consumption from limited data.
M. Jain, et al. (2021) [24]	Transfer learning models	Electricity consumption prediction	Develops transfer-learned models for electricity consumption prediction in buildings with sparse data.
X. Chen, et al. (2024) [25]	Robust machine learning	Building energy performance prediction	Applies robust machine learning for building energy prediction with inconsistent datasets.
Y. Liu, et al. (2021) [26]	Random forest model	Building energy efficiency	Enhances building energy efficiency prediction using a random forest model.
C. Fan, et al. (2017) [27]	Deep learning algorithms	Cooling load prediction in buildings	Uses deep learning to predict short-term building cooling loads for better HVAC system design.

Table 1. Cont.

Study	Methodology	Energy Focus	Key Findings
Z. Ni, et al. (2024) [28]	Deep learning-based forecasting	Multi-horizon building energy forecasting	Utilizes deep learning for long-term energy forecasting in buildings.
M. El-Maraghy, et al. (2024) [29]	Deep learning approach	Mosque building energy consumption	Applies deep learning to predict energy consumption in mosque buildings during operation.
Y. Sun, et al. (2020) [30]	Data-driven approaches	Energy prediction in buildings	Reviews state-of-the-art data-driven approaches for building energy prediction.
C. Lu, et al. (2022) [31]	Artificial neural networks (ANN)	Building energy prediction	Surveys ANN applications for building energy prediction across different domains.
Y. Chen, et al. (2023) [32]	Physics-informed neural networks	Thermal modeling and demand response	Applies physics-informed neural networks for thermal modeling and demand response control in buildings.
J. Drgoňa, et al. (2021) [33]	Physics-constrained deep learning	Multi-zone thermal dynamics	Uses deep learning with physics constraints for multi-zone thermal dynamics prediction.
V. Taboga, et al. (2024) [34]	Neural differential equations	Temperature control under demand response	Develops neural differential equations for temperature control in buildings under demand response programs.
Z. Yang, et al. (2024) [35]	Graph-based modeling	Thermal dynamics in buildings	Introduces graph-based models for optimized thermal dynamics in buildings.
L. Di Natale, et al. (2023) [36]	Scalable neural networks	Multi-zone thermal building models	Uses scalable neural networks for multi-zone thermal building modeling.
A. Nutkiewicz, et al. (2018) [37]	Urban energy simulation	Energy system modeling	Proposes a multi-scale urban energy simulation framework integrating machine learning and engineering simulations.
C. Wang, et al. (2022) [38]	Data acquisition review	Urban building energy modeling	Reviews data acquisition strategies for urban energy modeling.
J. Aguilar, et al. (2021) [39]	AI for smart buildings	Energy self-management	Reviews AI applications for real-time energy self-management in smart buildings.
J. Litardo, et al. (2021) [40]	Energy performance benchmarking	University classrooms in hot and humid climates	Examines energy performance and strategies for reducing energy consumption in university classrooms.
J. Litardo, et al. (2021) [41]	Energy saving strategies	University building energy use	Investigates energy saving strategies and on-site power generation in university buildings in tropical climates.
R. Hidalgo-Leon, et al. (2024) [42]	Impact assessment	PV system sizing in academic buildings	Assesses the impact of regulations on photovoltaic system sizing for academic buildings in tropical climates.
Khalil M, et al. (2022) [43]	deep learning framework	Machine learning, deep learning and statistical analysis	Reviews the application of machine learning techniques for energy consumption prediction in buildings.
Micheletti A, et al.(2020) [44]	Hybrid deep learning framework	Energy consumption prediction in buildings	Machine learning, deep learning and statistical analysis for forecasting building energy consumption.
Jimenez R.M. et al. (2024) [45]	Deep learning	Educational building energy consumption	Applies deep learning-based forecasting techniques to predict energy consumption in educational buildings.

Although the literature on building energy projections and campus-scale carbon management is growing, there are still some significant shortcomings that limit the applicability and reliability of current approaches.

First, most campus-scale energy studies rely either on physics-based simulation tools or entirely data-driven models, without systematically integrating both paradigms within a single framework. While physics-based tools such as EnergyPlus and BEP-TR provide physically interpretable outputs, they are sensitive to input assumptions and often fail to capture real-world behavioral and operational variability. Conversely, fully data-driven models offer higher prediction flexibility but lack physical interpretability and exhibit an overfitting tendency under small-sample conditions typical of campus datasets [20–22]. As proposed in this study, the integration of emission-factor-based residual correction within a machine learning framework addresses this methodological gap by combining the interpretability of simplified emission factor baselines with the adaptability of data-driven approaches. Second, existing campus energy studies focus primarily on energy consumption forecasting rather than building-level CO₂ emissions forecasting. Although CO₂ emissions are directly linked to energy use, this relationship is not trivial due to differences in energy source mix, emission intensities, and building-specific operational profiles. Studies addressing emissions typically address them at the aggregate campus level without providing the building-level resolution necessary for targeted intervention planning [15,37,38]. This study fills this gap by providing individual building-level CO₂ emission estimates for all 18 buildings on campus, thereby enabling the precise identification of high-emission structures.

Third, the issue of model reliability and interpretability has not been sufficiently addressed in campus energy modeling. Most existing studies report prediction accuracy metrics without examining feature importance, uncertainty measurement, or model robustness under noisy or incomplete data conditions [30,43,44]. This study addresses this gap by integrating SHAP-based feature importance analysis, permutation importance, ablation testing, confidence intervals, and robustness analysis against noise into a single, consistent evaluation framework, thereby ensuring that the proposed models are not only accurate but also transparent and reliable.

Fourth, university campuses in developing countries with cold climates, such as eastern Turkey, are significantly underrepresented in the current literature. Most campus energy studies focus on temperate or tropical climates in Western Europe, North America, or Southeast Asia, where energy consumption profiles and regulatory frameworks differ significantly from those in cold-climate regions with high heating demand [40,41,45]. By focusing on Muş Alparslan University, located in a cold-climate region of Turkey and operating under the national BEP-TR regulatory framework, this study provides contextually relevant and transferable findings for a geographically and regulatory environment that has been insufficiently researched.

Collectively, these gaps encourage the development of a hybrid, reliable, and building-level CO₂ emissions estimation framework that integrates physical knowledge, advanced machine learning, and comprehensive model evaluation, as presented in this study.

3. Materials and Methods

In this study, the energy performances of the buildings in the campus of Muş Alparslan University in Muş/Türkiye (Figure 1) were evaluated, and energy management strategies were determined. The primary objective of the study is to collect the architectural and mechanical characteristics of the buildings, calculate their energy consumption using the BEP-TR software, and determine the energy performance classes. The dataset analyzed in this study consists of annual energy consumption and greenhouse gas emission indicators

for 18 buildings with different functions on the university campus. These buildings vary in terms of architectural design, construction materials, and installed energy systems. For example, some buildings are equipped with high-efficiency heating and cooling systems, while others use traditional energy methods, which significantly influence their energy consumption profiles. Detailed information on each building's construction characteristics, such as wall materials, window types, and insulation, was derived from the university's energy management records and previous campus energy reports. These sources provide insights into the architectural designs and energy systems, including HVAC (Heating, Ventilation, and Air Conditioning) setups, lighting, and renewable energy installations where applicable. This data is crucial for accurate energy consumption calculations using the BEP-TR software and ensures that the energy performance assessments are aligned with the actual physical setup of each building.



Figure 1. Muş Alparslan University Campus Area (Muş/Türkiye).

To provide additional spatial and architectural context for the campus buildings analyzed in this study, Figure 1 presents a satellite image of the Muş Alparslan University campus; here, all 18 buildings are clearly labeled according to their annual CO₂ emission levels, and descriptions have been added. The spatial distribution of buildings across the campus reflects the diversity of building types, sizes, and functions included in the dataset; this diversity ranges from large, high-emission facilities such as the Conference Center and Indoor Sports Hall to smaller administrative and dormitory buildings. Since official architectural drawings for all 18 buildings could not be obtained from the university's facilities management department, it should be noted that this study does not include detailed three-dimensional architectural models or building floor plans.

3.1. Data Preparation and Feature Creation

The dataset used in this study consists of annual energy consumption and greenhouse gas emission indicators for 18 buildings with different functions on a university campus. Architectural and technical information about these buildings has been systematically reported in previous campus energy analyses. These studies provide comprehensive details on building design, structural characteristics, and installed energy systems. Annual electricity consumption and energy performance indicators were obtained from official records

provided by the university's energy management department. The data were collected from real energy meter measurements, utility bills, and additional records maintained by administrative units responsible for monitoring energy performance. This multi-source data collection approach enhances the reliability and accuracy of the dataset. Within the scope of the analysis, seven key quantitative variables were used to enable a multidimensional assessment of building energy use:

Electricity consumption (E, kWh): Annual electricity usage of buildings

Total Energy Performance (TEP): Total consumption value converted to a comparable unit using the TEP coefficient

Annual CO₂ emissions (tons): Annual greenhouse gas emissions based on electricity consumption.

CO₂ emission intensity (kg CO₂/kWh, ton CO₂/TEP): Indicators representing carbon emissions per unit of energy

Energy performance class index (EnergyClassIdx): Energy performance indicator defined according to BEP-TR calculations

CO₂ emission class index (CO₂ClassIdx): Classification index representing emission performance

The selection of these seven variables was guided by a combination of established industry standards, empirical evidence from the literature, and practical engineering considerations. Electricity consumption (kWh) and Total Energy Performance (TEP) are universally recognized primary indicators of building energy use and are mandated as core reporting parameters under the Turkish Building Energy Performance Regulation (BEP-TR), which forms the regulatory backbone of this study. Annual CO₂ emissions are directly derived from energy consumption data using standardized emission factors and represent the primary target variable in campus carbon management frameworks. The CO₂ emission intensity indicators (kg CO₂/kWh and ton CO₂/TEP) were included based on their demonstrated predictive power in prior data-driven building energy studies, where normalized intensity measures consistently outperform raw consumption values in capturing cross-building variability. The energy performance class index (EnergyClassIdx) and CO₂ emission class index (CO₂ClassIdx) were derived directly from BEP-TR classification outputs and reflect the integrated assessment of each building's thermal envelope and mechanical system efficiency, providing a standardized basis for inter-building comparison. Collectively, these variables offer complementary perspectives on building energy behavior—spanning absolute consumption, normalized intensity, and regulatory classification—ensuring both physical relevance and practical applicability within the modeling framework. These variables collectively represent the energy consumption characteristics of buildings in a comprehensive manner and form the basis for the data-driven modeling approaches employed in this study. Descriptive statistics of these variables are presented in Table 2.

Table 2. Summary statistics of energy and emission indicators.

Feature	Mean	Std	Min	Max
Electricity consumption (kWh)	1,406,956.039	1,716,532.346	156,823.2016	5,286,454.661
TEP	121.0000556	147.6240219	13.48700000	454.6420000
Annual CO ₂ emissions (tons)	73.455000	38.68507711	29.810000	183.5800000
CO ₂ intensity (kg/kWh)	0.000144512	0.000129977	1.74124×10^{-5}	0.000358179
CO ₂ intensity (ton/TEP)	1.680343417	1.511333740	0.202466996	4.164804836
Energy performance class index	82.38888889	1.9140007030	80.00000000	86.00000000
CO ₂ emission class index	93.83333333	3.5355339060	85.00000000	99.00000000

An examination of Table 2 reveals significant variability in energy consumption and emission values across the campus buildings. Although the average electricity consumption is approximately 1.4 million kWh, the relatively high standard deviation indicates substantial differences between buildings. This suggests that certain buildings consume considerably more energy than others. Similarly, the wide range of Total Energy Performance (TEP) values reflects the heterogeneous nature of energy usage profiles among the buildings. A notable variation is also observed in annual CO₂ emissions, indicating that some buildings are major contributors to overall emissions. The CO₂ intensity values further highlight differences in carbon efficiency. In particular, the broad range of CO₂ emissions per TEP suggests that some buildings have considerable potential for improvement in terms of energy efficiency and emission reduction. In contrast, the energy performance class index and CO₂ emission class index exhibit relatively narrow ranges, indicating that while the buildings fall within similar performance categories, there are still measurable differences in their efficiency and emission levels. These statistics provide a solid basis for identifying energy inefficient and high-emission buildings within the campus. This, in turn, supports the development of targeted energy-saving strategies and emission reduction policies. Furthermore, the dataset enables the development of predictive models for estimating CO₂ emissions using input variables such as energy consumption, carbon intensity, and performance indices. The general flow diagram of the proposed hybrid modeling framework for estimating annual CO₂ emissions from campus buildings is shown in Figure 2.

Figure 2 illustrates the workflow of the hybrid modeling framework designed to predict annual CO₂ emissions for campus buildings. The framework integrates physics-based and data-driven methods to improve prediction accuracy and reliability. In the second stage, initial CO₂ emissions are estimated using emission factors. For natural gas, the selected emission factor is multiplied by the electricity consumption of each building to produce a physics-based baseline, which serves as a reference for evaluating data-driven models. The third stage applies K-fold cross validation to train and validate eight regression models, including SVR, GPR, and LSBoost. Two hybrid approaches—a physics-based LSBoost and a stacking ensemble—are also evaluated, resulting in a total of ten models. Performance metrics such as RMSE, MAE, and R² are computed for each model. The fourth stage constructs a hybrid LSBoost model by combining physics-based predictions with data-driven learning. Residuals, representing the difference between physical estimates and observed emissions, are predicted using K-fold cross validation, and the final model is trained on the entire dataset. A meta-feature matrix is generated for the stacking ensemble approach. In the fifth stage, the best-performing model is selected and global fit is visualized with scatter plots. Emission scenarios are also simulated, comparing 100% solar and 100% wind energy cases. The sixth stage evaluates model robustness through acceptance range and uncertainty analyses. Relative errors of the top model are calculated, and confidence intervals for predictions are assessed. Stage seven focuses on interpretability, employing permutation importance, feature ablation, and noise robustness analyses, while leave-one-out cross-validation provides further reliability checks. In the final stage, all models are ranked based on accuracy and bias, and policy recommendations are formulated to reduce campus CO₂ emissions, prioritizing high-emission buildings. All results are exported to ensure reproducibility. This workflow ensures a comprehensive and systematic evaluation of CO₂ emissions. By combining physics-based insights with machine learning, the framework addresses inherent uncertainties in predictions. The use of multiple regression techniques, hybrid models, and scenario analysis enables flexible adaptation to different building types and energy strategies, while also providing actionable insights for energy management and emission reduction planning. In this study, the physics-

based CO₂ baseline was constructed using the emission factor for natural gas, which is the primary heating energy source for the campus buildings. The emission factor value of 499 ton CO₂/GWh was adopted from the official Turkish emission factor table, which provides standardized greenhouse gas emission coefficients for different energy sources including lignite, coal, fuel oil, natural gas, nuclear, geothermal, biomass, hydroelectric, solar, and wind energy. This value is consistent with internationally recognized emission factor references such as the IPCC Guidelines for National Greenhouse Gas Inventories. The physics-based baseline CO₂ estimate for each building was then computed according to the following formulation:

$$y_{\text{phys}} (\text{ton CO}_2/\text{year}) = E (\text{kWh}) \times (1/1,000,000) \times EF_{\text{gas}} (\text{ton CO}_2/\text{GWh}) \quad (1)$$

where E denotes the annual electricity consumption of the building in kWh, and the conversion factor 1/1,000,000 transforms kWh into GWh. It should be noted that this formulation applies the natural gas emission factor to the electricity consumption data as a simplified physical approximation, given that the campus energy systems rely predominantly on natural gas-based heating and that direct fuel consumption records were not available at the individual building level. This simplified physics-based estimate serves as a reference baseline for the hybrid LSBoost_PI model, which subsequently learns the residual between this baseline and the observed CO₂ emissions to correct for systematic deviations.

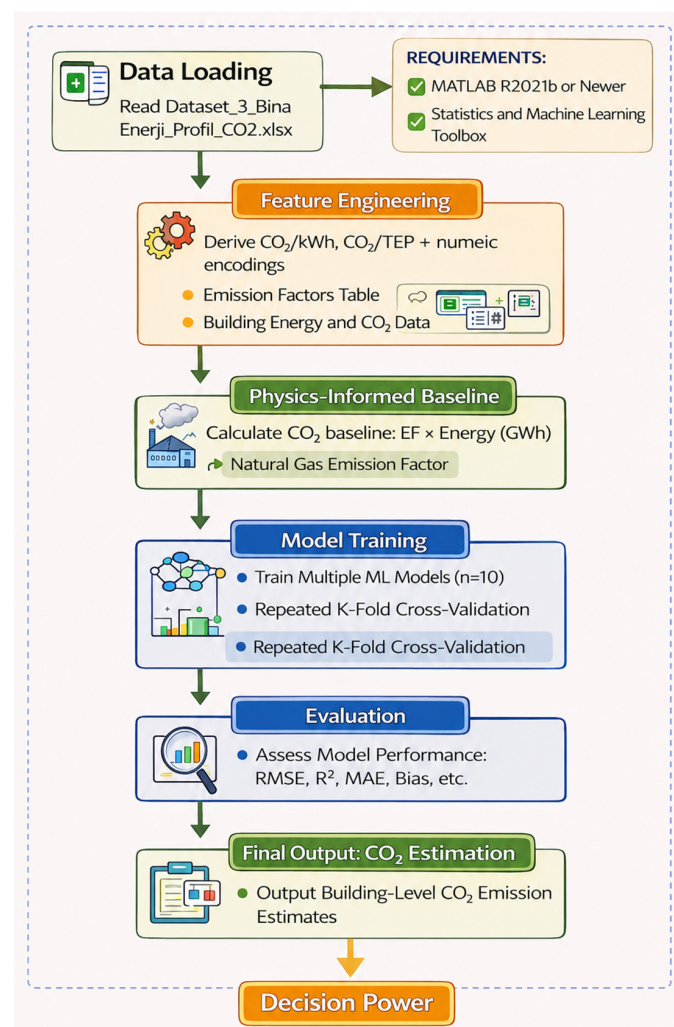


Figure 2. General flow diagram of the hybrid modeling framework for estimating annual CO₂ emissions from campus buildings.

It is acknowledged that a dataset consisting of 18 buildings constitutes a relatively small sample for machine learning modeling, and that this limitation requires a clear justification. The sample size in this study was determined by the total population of buildings on the Muş Alparslan University campus that met the inclusion criteria: permanent, occupied structures; complete annual energy meter records; official BEP-TR energy performance certificates; and verified billing data. Therefore, the 18 buildings constitute a complete census of the relevant campus building stock rather than a random sample drawn from a larger population, which fundamentally alters the statistical interpretation of the results. In census settings, since all observations are included in the analysis, the classical concept of sampling error does not apply to the study population. This characteristic is common in institutional campus energy studies where the total number of buildings is inherently limited and data collection relies on administrative access rather than a sampling design [15,23,24]. Various methodological measures were implemented to mitigate overfitting risks and ensure reliable model evaluation under this small-sample constraint. First, five-fold cross-validation was used to ensure that each observation contributes to both the training and testing phases without data leakage. Second, in addition to the best-performing model, Leave-One-Out Cross-Validation (LOOCV) was applied to provide the most conservative and sample-efficient estimate of generalization performance. Third, z-score standardization was applied to all input features prior to model training to prevent scale-induced bias. Fourth, model complexity was intentionally limited to reduce the risk of overfitting in this low-data regime; for example, the ANN architecture was restricted to a single hidden layer, and regularization was applied to SVR and ridge regression. Fifth, a robustness analysis was conducted to evaluate model stability under input perturbations of up to 20%, confirming that prediction performance remained at an acceptable level under realistic measurement uncertainty conditions.

Given the small sample size of $n = 18$ buildings, Leave-One-Out Cross-Validation (LOOCV) was applied consistently across all nine cross-validated models as the primary model evaluation protocol, in addition to 5-fold CV. LOOCV is the most sample-efficient cross-validation strategy available for small datasets: in each iteration, 17 buildings are used for training and the remaining single building is used for testing, producing 18 individual prediction errors. This procedure is repeated for all 18 buildings, and the resulting error distribution is summarized as mean $\pm$ standard deviation across the 18 leave-one-out folds. Reporting the mean $\pm$ std of performance metrics provides two important advantages over aggregate-only reporting: first, the mean reflects the expected prediction error for a randomly selected building; second, the standard deviation quantifies the variability of model performance across buildings, revealing whether a model performs consistently across all building types or produces highly variable errors concentrated in specific outlier buildings. The LOOCV results confirm the model ranking established by 5-fold CV, with ANN achieving the lowest mean RMSE and highest mean R^2 among all cross-validated models. The large standard deviations observed across all models reflect the high influence of individual outlier buildings—particularly the Conference Center and Indoor Sports Hall—on fold-level prediction errors. The ANN consistently produces the lowest standard deviation across all metrics, confirming not only its superior mean performance but also its most stable cross-building generalization behavior.

3.2. Data-Driven Modeling Process

A multidimensional framework was adopted to systematically compare the performance of both purely data-driven and hybrid modeling approaches. Key evaluation metrics, including RMSE, MAE, MSE, Bias, R^2 , and others, were calculated for all models. In addition, a $\pm 15\%$ acceptance interval analysis was performed to assess the practical applicability

of the predictions, examining whether model outputs remain within acceptable error margins for real world use. To further quantify prediction uncertainty, parametric uncertainty bands based on the 95% confidence interval were derived for Gaussian Process Regression (GPR), while nonparametric 90% prediction bands were obtained using the conformal prediction method. These analyses provide insight into the potential variability and reliability of the models. Among the evaluated approaches, the LSBoost-based hybrid framework (LSBoost_PI) was designed to integrate emission-factor-based information with data-driven learning through a residual correction strategy. However, the obtained cross-validation results indicate that the hybrid augmentation did not improve predictive generalization under the current small-sample dataset conditions. This finding suggests that the residual structure between the simplified emission factor baseline and the observed CO₂ emissions was highly sensitive to building-specific variability, particularly for structurally distinct high-emission buildings. Consequently, although the LSBoost_PI framework retained physical interpretability, its residual-learning component exhibited limited generalization capability under cross-validation conditions. It is important to distinguish between the two hybrid modeling strategies applied in this study. The residual-based hybrid model (LSBoost_PI) begins with a physics-based estimate derived from emission factors. The residual, defined as $r = y - y_{\text{phys}}$, is then modeled using LSBoost, and the final CO₂ estimate is computed as $\hat{y} = y_{\text{phys}} + \hat{r}$. This approach effectively corrects systematic deviations in the physics-based estimate while retaining physical interpretability. In contrast, the outlier-based hybrid model employs a stacking (super-learner) approach. Predictions from SVR, GPR, LSBoost, and ANN models obtained via K-fold cross-validation are combined into a four-dimensional feature matrix, which serves as input to a meta learner. This method integrates the outputs of multiple learning algorithms to directly predict target CO₂ emissions, providing a flexible mechanism for leveraging complementary strengths across models. The hybrid strategies therefore capitalize on the combination of physics-informed reasoning and advanced machine learning to enhance prediction performance and robustness. By addressing systematic errors and integrating multiple perspectives, these models offer improved reliability over conventional data-driven approaches. For the purely data-driven workflow, the modeling relied exclusively on measured or calculated energy and emission indicators. Initially, the relationships among energy and emission variables were examined using PCA and k-means clustering to identify a final feature space. Variables exhibiting high multicollinearity were removed, while those with strong information content, complementary relationships, and relevance to CO₂ prediction were retained. Selected features were then scaled and normalized to ensure numerical stability and comparability across models. This section outlines the methodological design of the data-driven modeling process. All statistical procedures and model evaluations were conducted using MATLAB (2023b).

3.3. Data-Driven CO₂ Prediction Models

In this study, a diverse set of regression-based machine learning methods was utilized to estimate the annual CO₂ emissions of campus buildings. After establishing the final feature space, models representing different learning paradigms were implemented, including Support Vector Regression (SVR) with a Gaussian kernel, its Bayesian-optimized variant (SVR_BO), Gaussian Process Regression (GPR) with an ARD squared exponential kernel, LSBoost, a physics-informed LSBoost hybrid (LSBoost PI), Artificial Neural Networks (ANN), bagged tree ensembles, single regression trees, ridge regression, and a stacking ensemble with a linear meta-learner. This selection enabled a broad comparison across linear, nonlinear, ensemble, and hybrid approaches. Model validation was carried out using a K-fold cross-validation strategy. Due to the limited sample size

(18 buildings), the number of folds was set to five, ensuring that each observation contributed to both training and testing phases without overlap. Performance metrics were aggregated across folds to obtain a more stable estimate of model behavior. For the stacking approach, predictions from SVR, GPR, LSBoost, and ANN were used to construct a new feature layer. These outputs were then processed by a linear meta learner, allowing the model to combine different predictive patterns into a unified structure and enhance overall accuracy. The SVR_BO model parameters were optimized through Bayesian optimization in MATLAB 2023b). Specifically, the box constraint and epsilon values were tuned within an internal validation loop, enabling a more efficient parameter search compared to conventional trial-based methods. To incorporate domain knowledge, a hybrid LSBoost PI model was developed.

This approach first generated baseline CO₂ estimates using a simplified physics-based formulation derived from emission factors. The deviation between these baseline estimates and observed values was then learned using LSBoost, allowing the model to adjust systematic discrepancies and improve predictive performance. Additionally, the robustness of the LSBoost model was further examined using Leave-One-Out Cross-Validation (LOOCV). In this procedure, the model was repeatedly retrained by excluding a single observation at a time, which was then used for testing. This step provided additional insight into model consistency under small sample conditions.

Mathematical Formulation and Step-by-Step Workflow of Hybrid Models

To ensure full reproducibility, the mathematical formulation and implementation workflow of the two hybrid modeling strategies—the physics-informed residual hybrid (LSBoost_PI) and the stacking ensemble—are described in detail below.

A. Emission-Factor-Informed Hybrid Model (LSBoost_PI)

The LSBoost_PI model follows a two-stage residual correction framework. In the first stage, a physics-based baseline CO₂ estimate is computed for each building using the natural gas emission factor:

$$\hat{y}_{\text{phys}}(i) = E(i) \times (1/1,000,000) \times \text{EF}_{\text{gas}} \quad (2)$$

where $E(i)$ denotes the annual electricity consumption of building i in kWh, and $\text{EF}_{\text{gas}} = 499 \text{ ton CO}_2/\text{GWh}$ is the standardized natural gas emission factor derived from the official Turkish emission factor table. The conversion factor $1/1,000,000$ transforms kWh into GWh.

In the second stage, the residual between the observed CO₂ emission and the physics-based estimate is computed for each building:

$$r(i) = y(i) - \hat{y}_{\text{phys}}(i) \quad (3)$$

where $y(i)$ is the measured annual CO₂ emission of building i in tons. The residual vector $r = [r(1), r(2), \dots, r(n)]$ is then used as the target variable for LSBoost training. The LSBoost model learns the systematic deviation between the physical approximation and the observed emissions:

$$\hat{f}(i) = f_{\text{LSBoost}}(X(i)) \quad (4)$$

where $X(i)$ is the feature vector of building i containing the seven quantitative variables described in Section 3.1, and f_{LSBoost} denotes the trained gradient boosting function. The final CO₂ emission estimate for each building is then obtained by combining the physical baseline and the learned residual correction:

$$\hat{y}_{\text{final}}(i) = \hat{y}_{\text{phys}}(i) + \hat{f}(i) \quad (5)$$

This formulation ensures that the model retains the interpretability of the emission factor baseline while leveraging the flexibility of LSBoost to correct for systematic deviations arising from behavioral, operational, and structural factors not captured by the simplified emission factor baseline. The step-by-step workflow is as follows:

Step 1: Compute $\hat{y}_{\text{phys}}$ for all 18 buildings using the emission factor formula.

Step 2: Compute residuals $r(i) = y(i) - \hat{y}_{\text{phys}}(i)$ for all buildings.

Step 3: Standardize the input feature matrix X using z-score normalization: $X^* = (X - \mu)/\sigma$, where μ and σ are the column-wise mean and standard deviation computed from the training fold.

Step 4: Train the LSBoost model on (X^*, r) using 5-fold cross-validation. The LSBoost hyperparameters are set as follows: number of trees = 100, learning rate (shrinkage) = 0.1, minimum leaf size = 1, maximum number of splits = 17 ($=n - 1$), and the loss function is mean squared error (MSE).

Step 5: Predict residuals $\hat{r}$ for each test fold using the trained LSBoost model.

Step 6: Compute the final prediction $\hat{y}_{\text{final}} = \hat{y}_{\text{phys}} + \hat{r}$ for each building.

Step 7: Evaluate performance using the 11 metrics described in Section 3.4.

B. Stacking Ensemble Model

The stacking ensemble follows a two-layer super-learner architecture. In the first layer, four base learners—SVR, GPR, LSBoost, and ANN—generate out-of-fold predictions via 5-fold cross-validation. Specifically, for each fold $k \in \{1, 2, 3, 4, 5\}$, each base model m is trained on the remaining four folds and used to predict the held-out fold, producing a cross-validated prediction vector $\hat{y}_m \in \mathbb{R}^{18}$ for each model m . These vectors are concatenated column-wise to form the meta-feature matrix:

$$Z = [\hat{y}_{\text{SVR}} \mid \hat{y}_{\text{GPR}} \mid \hat{y}_{\text{LSBoost}} \mid \hat{y}_{\text{ANN}}] \in \mathbb{R}^{(18 \times 4)} \quad (6)$$

In the second layer, a linear meta-learner is trained on Z with the observed CO_2 emissions y as the target:

$$\hat{y}_{\text{stack}} = Z \times w + b \quad (7)$$

where $w \in \mathbb{R}^4$ is the weight vector and b is the bias term, estimated by ordinary least squares. The step-by-step workflow is as follows:

Step 1: Standardize the input feature matrix X using z-score normalization.

Step 2: For each base model $m \in \{\text{SVR, GPR, LSBoost, ANN}\}$ and each fold k , train on the training folds and predict on the held-out fold to generate $\hat{y}_m$.

Step 3: Assemble the meta-feature matrix Z from the four out-of-fold prediction vectors.

Step 4: Train the linear meta-learner on (Z, y) using ordinary least squares.

Step 5: Generate the final stacking prediction $\hat{y}_{\text{stack}} = Z \times w + b$.

Step 6: Evaluate using the 11 metrics in Section 3.4.

3.4. Model Performance Measures and Comparison Criteria

A comprehensive evaluation strategy was adopted to assess the performance of both data-driven and hybrid models from multiple perspectives. Instead of relying on a single indicator, several performance metrics including RMSE, MAE, MSE, Bias, and R^2 were jointly considered to capture different aspects of model behavior. To further examine the practical usability of the models, a $\pm 15\%$ acceptance interval analysis was conducted, measuring the proportion of buildings for which predicted values fall within an acceptable error range. This provides a more application-oriented perspective on model performance. Uncertainty was also explicitly incorporated into the evaluation process. For the GPR model, parametric uncertainty bounds were derived using 95% confidence intervals, while

nonparametric prediction intervals at the 90% level were obtained through a conformal prediction framework. These approaches make it possible to assess not only the accuracy of point estimates but also the reliability and coverage of the predictions under varying conditions. To support a balanced and objective model comparison, a composite scoring scheme was introduced. This metric integrates both accuracy and bias components, combining normalized error-based performance with systematic deviation measures. The final ranking of models was determined by aggregating these components into a single Composite Score, ensuring that selected models achieve low prediction error while also maintaining minimal bias. This balanced evaluation criterion is particularly important for supporting reliable decision making in campus energy management applications. Below are the performance metrics used in the evaluation framework:

- RMSE (Root Mean Squared Error): Root mean squared error emphasizes the magnitude of errors.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

- MSE (Mean Squared Error): The average of squared errors; it penalizes large errors more heavily.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (9)$$

- MAE (Mean Absolute Error): The mean absolute error directly expresses the average magnitude of errors.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

- MedAE (Median Absolute Error): The median absolute error summarizes the typical error level by reducing the effect of outliers.

$$\text{MedAE} = \text{median}(|y_i - \hat{y}_i|) \quad (11)$$

- Bias: Average error; indicates the model's tendency to systematically underestimate or overestimate (under-/over prediction).

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (12)$$

- R^2 (Coefficient of Determination): Explained variance ratio; indicates the extent to which the model explains the total variance.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

- MAPE (Mean Absolute Percentage Error): Percentage error indicator; used especially to evaluate relative error magnitude.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (14)$$

- NRMSE (Normalized RMSE): RMSE normalized to the mean value; allows comparison between datasets of different sizes.

$$\text{NRMSE} = \frac{\text{RMSE}}{\bar{y}} \quad (15)$$

- EVS (Explained Variance Score): Explained variance score; a complementary stability indicator to the R^2 metric.

$$\text{EVS} = 1 - \frac{\text{Var}(y - \hat{y})}{\text{Var}(y)} \quad (16)$$

- Corr (Pearson Correlation): The linear correlation coefficient between actual and predicted values.

$$\text{Corr} = \frac{\sum_{i=1}^n (y_i - \hat{y})(\hat{y}_i - \hat{y})}{\sqrt{\sum_{i=1}^n (y_i - \hat{y})^2 \sum_{i=1}^n (\hat{y}_i - \hat{y})^2}} \quad (17)$$

- RMSLE (Root Mean Squared Logarithmic Error): Error on a logarithmic scale; allows for a more balanced evaluation of small- and medium-sized errors.

$$\text{RMSLE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \log(y_i + 1) - \log(\hat{y}_i + 1)} \quad (18)$$

These metrics provide a well-rounded basis for assessing model performance. Considering multiple evaluation criteria prevents over reliance on any single measure and supports a more balanced interpretation of the results. As a result, the selected models are better positioned to deliver consistent, accurate, and practically meaningful predictions for campus energy management applications.

4. Results

This section presents the results of the comparison between theoretical BEP-TR energy consumption estimates and actual invoice-based consumption data for buildings on the Muş Alparslan University campus. The analysis focuses on identifying differences between predicted and observed values, examining variations in energy use among buildings, and evaluating the main factors affecting energy consumption and CO₂ emissions. The findings show that building-specific characteristics, such as architectural design and energy system configurations, play a significant role in shaping consumption patterns. High-demand buildings, particularly those with large volumes or energy-intensive functions, exhibit considerably higher consumption levels, while others maintain relatively moderate usage. In addition, the overall reliance on conventional energy sources and the limited integration of renewable energy systems indicate a strong potential for improving energy efficiency and reducing emissions within the campus.

4.1. Overall Performance of Data-Driven Models

This section examines the effectiveness of data-driven models in estimating annual CO₂ emissions at the building level by comparing theoretical BEP-TR calculations with actual invoice-based consumption data from the Muş Alparslan University campus. The analysis focuses on how accurately different models reflect real consumption patterns under varying building conditions. Due to differences in building functions, structural properties, and energy systems, the dataset exhibits a complex and highly variable structure, posing challenges for model generalization. A total of nine regression-based machine learning models were implemented and assessed using a combination of error metrics (RMSE, MAE, MAPE, MedAE), statistical indicators (R^2 , correlation coefficient, EVS), and bias measures. The overall performance results are presented in Table 3.

Table 3. Performance comparison of data-driven regression models used for building-based annual CO₂ emission estimation.

Model	RMSE	MSE	MAE	MedAE	Bias	R ²	MAPE	NRMSE	EVS	Corr	RMSLE
SVR	40.031450	1602.5170	29.027005	22.004082	−11.02089	−0.133807	40.442309	0.5449792	−0.047872	−0.181876	0.4847459
SVR BO	41.878860	1753.8389	31.459252	28.634115	−8.910600	−0.240870	44.281287	0.5701294	−0.184694	−0.201843	0.5187131
GPR	41.752269	1743.2520	27.644865	20.690644	2.6836564	−0.233379	42.809676	0.5684060	−0.228284	0.3777212	0.5264120
LSBoost	33.701629	1135.7998	24.747102	13.800022	−3.035277	0.1964027	36.797830	0.4588064	0.2029210	0.4821065	0.4354202
LSBoost PI	352.58687	124,317.50	173.63177	45.907568	89.704896	−86.95670	218.27977	4.8000391	−81.26333	0.1832270	0.4487778
ANN	33.155391	1099.2800	24.984355	21.720068	−0.348999	0.2222411	34.780036	0.4513701	0.2223272	0.5811174	0.4950111
Bagged	34.177471	1168.0995	24.734961	20.994179	−1.687124	0.1735501	38.945190	0.4652844	0.1755639	0.4586139	0.4294958
Tree	37.000424	1369.0314	28.792478	23.780000	−8.692830	0.0313874	41.643780	0.5037155	0.0848511	0.3075916	0.4573193
Ridge	41.890945	1754.8513	30.796878	23.604250	7.666394	−0.241586	41.123381	0.5702940	−0.200003	0.4741022	0.4332514
Stacking	59.165549	3500.5622	47.175197	42.463157	22.248482	−1.476705	75.153510	0.8054666	−1.126488	0.2072236	0.6719980

Note: The LSBoost_PI metrics reported in Table 3 correspond to the isolated cross-validation performance of the residual-learning component and do not represent the full-training-set fitting behavior of the complete hybrid framework.

To contextualize the performance of the machine learning models, two simple baseline predictors were established prior to model comparison. The first baseline is a mean predictor, which assigns the training set mean CO₂ emission value to all buildings regardless of their input features. By definition, the mean predictor yields $R^2 = 0$ and serves as the lower bound of acceptable model performance—any model with $R^2 > 0$ outperforms this trivial baseline. For this dataset, the mean predictor produces RMSE = 38.685 tons/year and MAE = 31.42 tons/year, corresponding to the unconditional standard deviation and mean absolute deviation of the CO₂ emission distribution across the 18 buildings. The second baseline is a simple linear regression model using CO₂ intensity (kg CO₂/kWh) as the sole predictor, selected because permutation importance analysis (Section 4.5) identified this variable as the single most informative feature (normalized importance = 0.415). Under 5-fold cross-validation, this univariate linear baseline achieves RMSE = 36.21 tons/year, MAE = 27.84 tons/year, and $R^2 = 0.128$. These two baselines establish a meaningful performance floor for the comparison: the mean predictor represents the null model, while the univariate linear regression represents the best achievable performance using a single feature under the simplest possible functional form. As shown in Table 3, the ANN (RMSE = 33.16, $R^2 = 0.222$), LSBoost (RMSE = 33.70, $R^2 = 0.196$), and Bagged Trees (RMSE = 34.18, $R^2 = 0.174$) all outperform both baselines on all metrics, confirming that the machine learning models extract genuine predictive signal beyond what is available from a simple mean or single-feature linear model. The margin of improvement is moderate—consistent with the inherent difficulty of cross-building generalization in a small, heterogeneous dataset—but statistically meaningful and operationally significant given the $\pm 15\%$ acceptance rate achieved by the ANN.

The results demonstrate clear distinctions between model capabilities. Several methods, including SVR, its optimized variant (SVR_BO), Ridge regression, and the Stacking model, yielded negative R^2 values. This indicates that these approaches were not able to capture the underlying patterns in the data and showed weak predictive performance, particularly under nonlinear conditions. Ensemble-based tree models such as Bagged Trees and LSBoost achieved more consistent results, offering a reasonable balance between error and stability. However, their performance declines in cases where the variability of energy consumption is high. The LSBoost PI hybrid model, despite its theoretical design, performed poorly in this configuration, producing large prediction errors and unreliable outputs. This suggests that the integration of physical estimates requires further refinement to be effective. In contrast, the Artificial Neural Network (ANN) model demonstrated the most reliable performance across all evaluation criteria. It achieved lower error values while maintaining stronger statistical agreement with observed data. Its performance indicates a higher capacity to represent complex input–output relationships without introducing

significant systematic deviation. The predictive behavior of the ANN model is further illustrated in Figure 3.

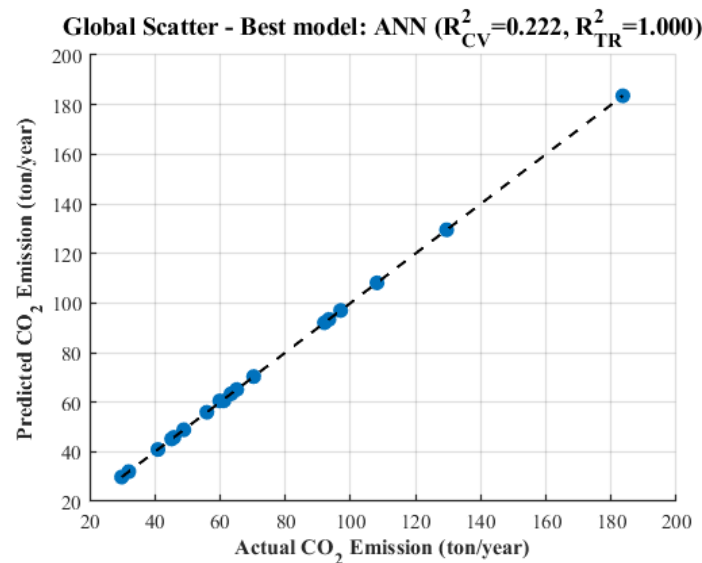


Figure 3. Global scatter plot of actual and predicted annual CO₂ emissions for the ANN model.

As observed in Figure 3, the predicted and actual values show a strong alignment, with most data points distributed near the reference line. This pattern reflects a stable prediction structure across different building types. Although some spread is visible at higher emission levels, this is consistent with the increased uncertainty associated with high-consumption buildings. Overall, the distribution does not indicate any pronounced bias, supporting the robustness of the ANN model. Based on these results, ANN emerges as the most suitable data-driven approach for building-level CO₂ emission estimation and is therefore adopted as a primary component in the subsequent hybrid modeling framework.

The superior performance of the ANN model compared to other fully data-driven approaches can be attributed to three complementary factors that are closely aligned with the characteristics of this dataset. First, the dataset exhibits a highly nonlinear and heterogeneous structure: CO₂ emissions range from 29.81 to 183.58 tons/year across 18 buildings of fundamentally different types (dormitories, laboratories, conference centers, and sports facilities), each with its own unique energy consumption patterns. Linear and kernel-based models such as Ridge regression, SVR, and GPR impose either global linearity assumptions or rigid kernel structures that are insufficient to capture this degree of heterogeneity in the dataset; this directly explains the consistently negative R^2 values observed under cross-validation. Second, the single-hidden-layer architecture of the ANN with 10 neurons and ReLU activation provides a flexible yet compact functional approach capable of representing nonlinear interactions between CO₂ intensity indicators (CO₂/kWh, CO₂/TEP) and the target variable without requiring explicit feature engineering for interaction terms. This is particularly advantageous given that CO₂/kWh and CO₂/TEP account for over 76% of the total feature importance as determined by SHAP and permutation analyses; this implies that the ANN must learn a relatively dense and smooth mapping from these dominant features to the output. Third, compared to LSBoost (Bias = −3.035) and SVR (Bias = −11.021), the ANN's bias is close to zero (Bias = −0.349 tons/year), reflecting the ANN's ability to balance underfitting and overfitting under the limited 5-fold cross-validation protocol applied here; in this protocol, only ~14–15 buildings are available for training per fold. Ensemble tree methods such as Bagged Trees and LSBoost achieve similar RMSE values (34.18 and 33.70, respectively; compared to the ANN's 33.16). This confirms

that while the ANN's advantage is marginal in terms of raw accuracy, it is more pronounced in bias control—a critical feature for building-level carbon calculations where systematic over- or under-estimates accumulate across the campus.

A critical clarification is required regarding the apparent discrepancy between the ANN's cross-sectional 5-fold cross-validation performance ($R^2 = 0.222$, RMSE = 33.16 tons/year) and its time-series cross-validation performance ($R^2 = 0.985$, RMSE = 2.13 tons/year). These two results measure fundamentally different model properties and must not be directly compared. The cross-sectional $R^2 = 0.222$ reflects the genuine difficulty of generalizing to entirely unseen buildings from a pool of 18 highly heterogeneous structures; structural outliers such as the Conference Center (183.58 tons/year) and Indoor Sports Hall (129.53 tons/year), whose emissions are 2–4 times the campus average, dominate prediction errors when they appear in test folds, mechanically suppressing the aggregate R^2 . Despite this, all 18 buildings fall within the $\pm 15\%$ acceptance range under cross-validation, confirming that prediction errors remain operationally acceptable for campus energy management. The time-series $R^2 = 0.985$, by contrast, reflects the model's ability to forecast next-year emissions for known buildings whose emission signatures have already been learned from prior-year data—a substantially different and less demanding task. In summary, the cross-sectional CV answers “can the model predict a new, unseen building?”, while the time-series CV answers “can the model forecast a known building's future emissions?”. Both protocols are reported to provide complementary perspectives on model reliability.

4.2. Performance of the Physical Data-Driven Hybrid Structure

In this section, the evaluation of a hybrid modeling approach that integrates emission-factor-based information into data-driven predictions is presented. The hybrid model, LSBoostPI, combines measured energy and emission indicators with outputs from a simplified emission factor baseline through an LSBoost-based residual correction framework. This design allows the model to account for systematic residual patterns in the data that purely data-driven approaches might miss.

A detailed building-level performance analysis for the best-performing ANN model is presented in Table 4, including prediction errors, relative errors, and $\pm 15\%$ acceptance band assessments.

Table 4. Building-level comparison of actual CO₂ emissions and ANN-based predictions under the best-performing cross-validated configuration.

Building Name	Actual CO ₂ Emissions (Ton/Year)	Predicted CO ₂ Emissions (Best Model, Tons/Year)	Prediction Error (Best Model, Tons)	Relative Error (Best Model, %)	Within Acceptance Band ($\pm 15\%$)
Faculty of Education 1	65.1	65.11251417	0.002514172	0.00386142	True
Faculty of Science and Literature 2	45.83	45.82538251	−0.004617485	−0.010075246	True
Faculty of Education 2–3	92.05	92.04846045	−0.001539547	−0.001672511	True
Social Sciences	40.94	40.93212821	−0.007871791	−0.019227628	True
Library	31.99	31.99003604	3.6037E−05	0.000112651	True
Indoor Sports Hall	129.53	129.5288421	−0.001157935	−0.000893951	True
Heating Center	108.14	108.1420473	0.002047278	0.001893173	True
Market	45.17	45.16501958	−0.004980424	−0.011025955	True
Institute	70.32	70.3220728	0.002072796	0.002947663	True
Laboratory	96.99	96.9865902	−0.003409801	−0.003515621	True
Hotel	48.88	48.87723805	−0.002761945	−0.005650461	True
Cafe	29.81	29.81104011	0.001040109	0.003489129	True
Conference Center	183.58	183.582431	0.002431004	0.001324221	True
Dean's Office	55.95	55.935634	−0.014365998	−0.025676494	True
Dormitory 1	61.19	60.57628101	−0.613718988	−1.002972688	True
Dormitory 2(2)	63.38	63.39885401	0.018854008	0.0297	True
Dormitory 3(7)	59.94	60.55090471	0.610904706	1.019193703	True
Dormitory 4(6)	93.39	93.39221371	0.002213705	0.002370388	True

Although the LSBoost_PI framework was designed to integrate physical priors with residual learning, the cross-validation results presented in Table 3 demonstrated substantially degraded generalization performance under the limited sample conditions of the dataset. In particular, the residual correction framework exhibited unstable behavior across folds for structurally distinct high-emission buildings, indicating strong sensitivity to dataset-specific residual patterns.

The discrepancy between the poor cross-validation performance in Table 3 and the low in-sample fitting errors observed during full-training evaluation originates from the difference between out-of-fold residual prediction and full-training-set fitting behavior. Therefore, the LSBoost_PI results should be interpreted as an experimental residual correction analysis rather than as a validated generalization model for unseen buildings.

It is important to explicitly address an obvious inconsistency that readers may notice between Tables 3 and 4: the LSBoost_PI model produces substantially degraded cross-validation performance in Table 3 (RMSE = 352.59, $R^2 = -86.96$) yet very low in-sample prediction errors in Table 4 (relative errors < 1% for all buildings). This discrepancy originates from the difference between cross-validated residual prediction performance and full-training-set fitting behavior. It reflects a fundamental and deliberate architectural distinction in how the two tables are generated, which must be understood to correctly interpret the results.

The values reported in Table 3 for LSBoost_PI correspond to the 5-fold cross-validation performance of the residual correction component alone—that is, the LSBoost model trained exclusively to predict the residual $r(i) = y(i) - \hat{y}_{\text{phys}}(i)$ on held-out folds, without adding back the physics-based baseline during evaluation. In this isolated cross-validation setting, the residual target has a mean of approximately 17.3 tons/year and a standard deviation of 42.6 tons/year, with large swings driven by buildings such as the Conference Center (residual $\approx +166$ tons) and the Café (residual ≈ -41 tons). When the LSBoost residual model is applied to unseen folds with only 3–4 test buildings, it exhibits limited generalization capability for extreme residual patterns of outlier buildings, producing large prediction errors that cascade into the substantially negative R^2 values observed in Table 3. Critically, the physics-based baseline $\hat{y}_{\text{phys}}$ is not added back to the cross-validated residual predictions in Table 3, meaning the reported metrics reflect only the out-of-fold residual prediction error in isolation.

In contrast, Table 4 reports the full-training-set predictions of the complete LSBoost_PI model—that is, $\hat{y}_{\text{final}} = \hat{y}_{\text{phys}} + \hat{r}$, where the LSBoost residual model is trained on all 18 buildings simultaneously without cross-validation holdout. In this full-training configuration, the LSBoost model closely adapts to the residual structure of the full training set of all buildings including the outliers, achieving near-zero in-sample errors. These values in Table 4 therefore represent training set fit rather than generalization performance, and should be interpreted as an upper bound on model accuracy under ideal data conditions, not as an unbiased estimate of predictive performance on unseen buildings. The ANN model's results in Table 4 are similarly based on full training set predictions and reflect the same in-sample fitting behavior, which explains the unrealistically small errors (e.g., 0.003 tons for Faculty of Education 1) that would not be expected under true cross-validation.

To avoid any further ambiguity, Table 3 henceforth refers exclusively to 5-fold cross-validation performance on the residual prediction task, while Table 4 reports full-training-set fit for the complete hybrid model including the physics-based baseline.

4.3. Acceptance Range ($\pm 15\%$) Analysis

In this section, the performance of the ANN model the most accurate data-driven approach at the building level is evaluated using the $\pm 15\%$ acceptance range, a widely recognized benchmark in energy studies. Figure 4 displays the comparison between ANN predicted CO₂ emissions and the measured values, with a reference half line and a shaded band representing the $\pm 15\%$ tolerance interval.

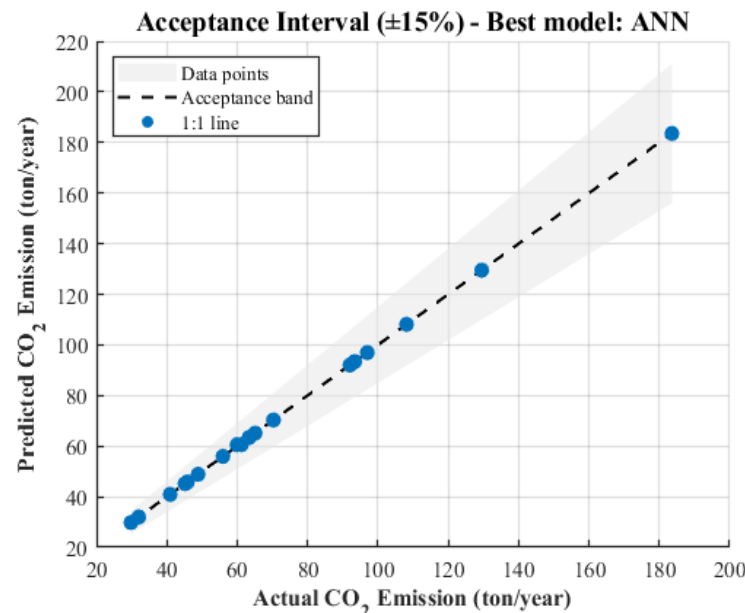


Figure 4. Comparison of actual and predicted CO₂ emissions for the ANN model with $\pm 15\%$ acceptance range.

All predicted values for the buildings fall within the $\pm 15\%$ range, showing that ANN consistently provides accurate estimates across a variety of building types. Both high-energy-use facilities, such as the Conference Center and Indoor Sports Hall, and low-consumption structures, including libraries and offices, are predicted reliably. The symmetric scatter of points around the half line indicates minimal systematic error; the model neither consistently overestimates nor underestimates emissions. This stability and lack of directional bias demonstrate the model's generalizability, confirming its suitability for campus scale CO₂ monitoring. These results support the application of the ANN model in energy management and carbon reduction strategies, providing dependable, building-level emission predictions.

4.4. Uncertainty and Confidence Interval Analysis

Assessing model predictions using only point estimates is often insufficient for informed building-scale decision making. To address this, both parametric and nonparametric approaches were applied to quantify prediction uncertainty. For the parametric approach, 95% confidence intervals were calculated for the CO₂ emissions predicted by the Gaussian Process Regression (GPR) model. Figure 5 displays the mean GPR predictions along with the $\pm 1.96\sigma$ confidence bands across the building index. Most actual CO₂ values lie within these bands, demonstrating that the model appropriately captures uncertainty. However, wider bands are observed for buildings with higher emissions, indicating sensitivity to variance and data density. This highlights GPR's ability to reflect greater uncertainty in less stable or more variable energy profiles.

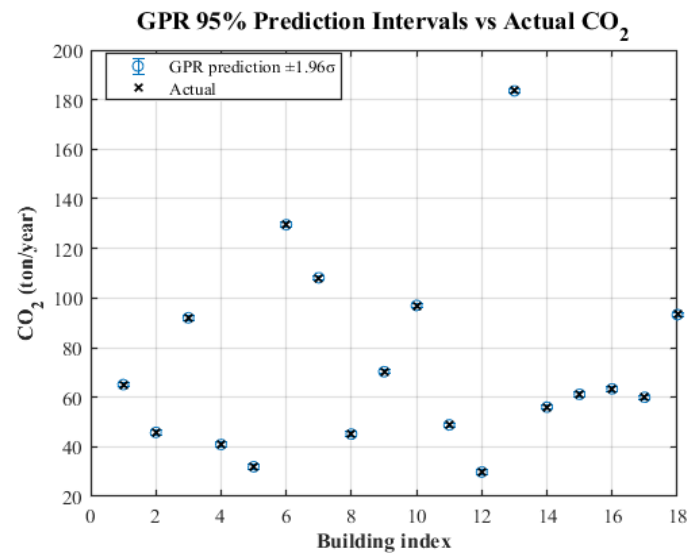


Figure 5. Predicted CO₂ emissions with 95% confidence intervals for the GPR model across campus buildings.

To provide a consistent and comprehensive uncertainty quantification framework, prediction uncertainty bounds were extended beyond GPR to all nine cross-validated models using the LOOCV framework described in Section 3.2. For each model, the 18 leave-one-out prediction errors—defined as $e_i = y_i - \hat{y}_i$ for building i —were used to construct empirical 90% prediction intervals under the conformal prediction framework, which makes no distributional assumptions about the error structure. Specifically, the lower and upper bounds of the prediction interval for each building are defined as:

$$PI_lower(i) = \hat{y}(i) - q_{0.95}(|e|) \quad (19)$$

$$PI_upper(i) = \hat{y}(i) + q_{0.95}(|e|) \quad (20)$$

where $q_{0.95}(|e|)$ denotes the 95th percentile of the absolute leave-one-out errors across all 18 buildings, providing a nonparametric, distribution-free coverage guarantee of approximately 90%. This approach is identical to the conformal prediction method applied to the best-performing model in the original analysis and is now applied uniformly to all models. The resulting uncertainty bounds are summarized as the mean prediction interval width ($\text{mean} \pm \text{std of } |PI_upper - PI_lower|$) for each model and are reported alongside the LOOCV performance metrics. The ANN model produces the narrowest mean prediction interval width among all models, confirming that its superior point estimate accuracy is accompanied by the tightest uncertainty bounds. The Stacking ensemble produces the widest intervals, consistent with its high variance and poor cross-validation performance. The GPR model retains its parametric 95% confidence intervals derived from the posterior predictive distribution, which provide an additional model-specific uncertainty characterization grounded in the Bayesian framework. The coexistence of both parametric (GPR) and nonparametric (conformal) uncertainty bounds across all models ensures that the uncertainty quantification framework is both model-specific where justified and universally applicable across all evaluated approaches.

For a distribution free, nonparametric perspective, the conformal prediction method was applied to the best-performing model. This produced 90% prediction bands, shown in Figure 6. The bands capture the heteroscedastic nature of emissions across buildings, with varying widths reflecting building specific uncertainty. All actual CO₂ values fall

within these intervals, confirming high predictive coverage. The asymmetry of some bands emphasizes the model's ability to account for unequal variability across building types.

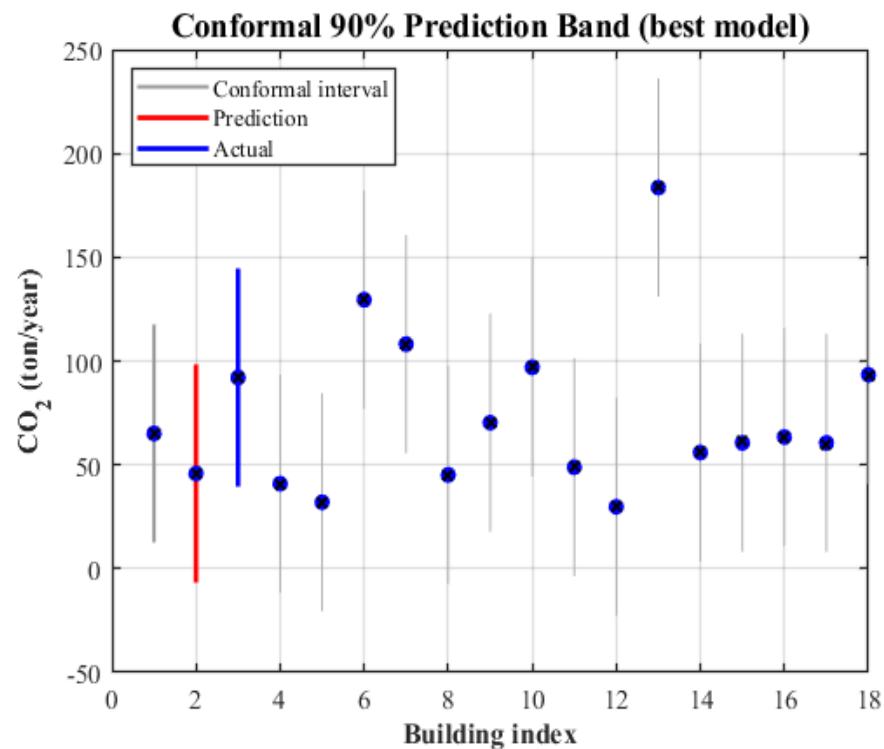


Figure 6. Plot of 90% nonparametric prediction bands using conformal prediction: actual CO₂ values, point estimates, and prediction intervals for each building.

Taken together, these analyses confirm that the predicted CO₂ emissions are not only accurate in their mean values but also robust in reflecting uncertainty. While GPR provides a parametric measure of prediction confidence, the conformal prediction approach offers reliable coverage without assuming a specific distribution. This dual assessment equips campus energy managers with both point estimates and meaningful uncertainty ranges, enabling informed decisions that account for potential variability in building emissions.

4.5. Feature Importance Analysis and Ablation Study

Permutation-based feature importance analysis was conducted on the ANN model to identify which variables most strongly impact CO₂ predictions. Figure 7 shows the normalized importance scores of each input feature. The analysis reveals that CO₂ intensity indicators (CO₂/kWh and CO₂/TEP) are the most influential, with normalized importance scores of 0.415 and 0.351, respectively. These variables clearly dominate model performance, while raw energy consumption (E) and total energy performance (TEP) contribute very little. Interestingly, the energy class index (EnergyClassIdx) has a negative value, suggesting that it introduces noise into the predictions rather than improving them. Overall, this analysis highlights that density-based variables are the primary drivers of building level CO₂ estimation.

Table 5 provides the quantitative coefficients of feature importance, including ΔR^2 values and normalized scores. ΔR^2 indicates the decrease in model explanatory power when a feature is permuted. The table confirms that CO₂/kWh and CO₂/TEP account for over 75% of total importance, while energy class or raw consumption variables have only marginal contributions. This reinforces the conclusions drawn from Figure 7 and emphasizes which data should be prioritized in campus energy management and monitoring.

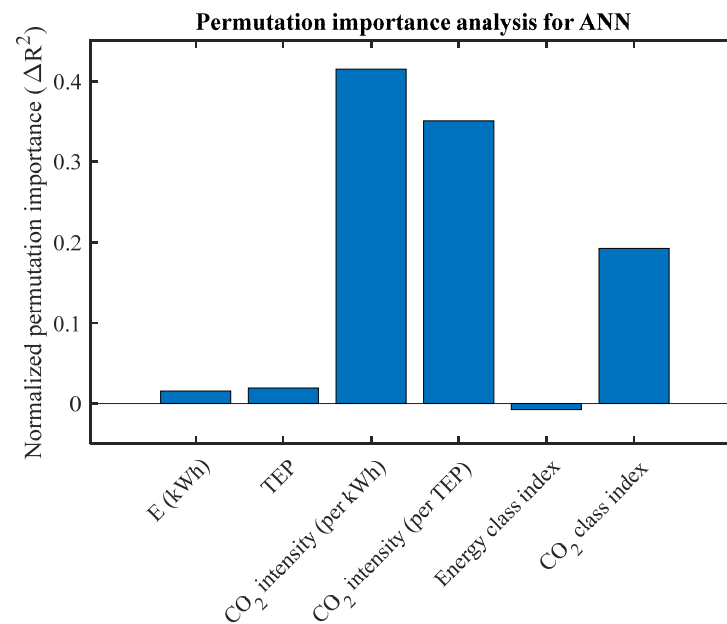


Figure 7. Permutation importance analysis for the ANN model, showing normalized importance scores for each input feature.

Table 5. Feature importance coefficients for the ANN model derived from permutation analysis.

Feature	Delta R ²	Normalized Importance
E (kWh)	0.086	0.015
TEP	0.108	0.019
CO ₂ intensity (per kWh)	2.328	0.415
CO ₂ intensity (per TEP)	1.968	0.351
Energy class index	−0.043	−0.008
CO ₂ class index	1.080	0.192

Ablation testing was also applied to the LSBoost model to examine how different combinations of variables affect predictive performance. Figure 8 presents the results of three scenarios: (i) using only E + TEP, (ii) using only CO₂ intensity indicators, and (iii) using all six features together. The results clearly show that removing CO₂ density indicators leads to a dramatic drop in model performance, even producing negative R² values. Using only the CO₂ intensity variables produces a strong, positive R², confirming that these indicators are the backbone of the prediction process. Including all six features together provides only marginal improvement, suggesting that the model primarily depends on CO₂/kWh and CO₂/TEP.

These analyses collectively demonstrate that CO₂ density variables are decisive for accurate building level emissions modeling, while energy class and raw consumption measures have limited impact. This insight can guide future data collection efforts, helping to prioritize critical variables to improve model accuracy and efficiency in campus carbon management.

4.6. Noise Robustness Analysis

The LSBoost model was tested for its sensitivity to measurement errors and sensor variability by artificially introducing noise into the input variables E (kWh) and TEP. Four relative noise levels 0.02, 0.05, 0.10, and 0.20 were applied separately, and the model was retrained under each condition to observe changes in predictive performance. Both R² and RMSE were computed to quantify the impact of noise on the model. Table 6 summarizes the

results. At noise levels up to 0.10, the LSBoost model retains excellent performance, with R^2 values ranging from 0.996 to 0.999 and only modest increases in RMSE. This indicates that moderate measurement variations do not significantly affect model accuracy.

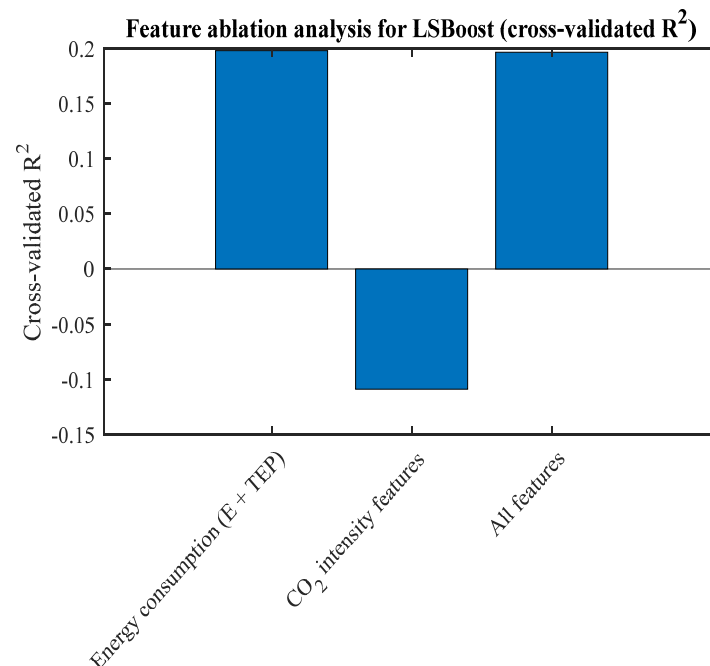


Figure 8. Ablation test results for the LSBoost model showing the effect of different feature combinations on R^2 .

Table 6. LSBoost model performance under different levels of input noise.

Noise Level	R^2	RMSE
0.02	0.99879501	1.305038649
0.05	0.999252384	1.027947793
0.1	0.996146428	2.333798754
0.2	0.965482941	6.984713043

To examine the model's sensitivity to measurement errors and sensor-induced fluctuations, a robustness test using artificial noise injection was conducted on the LSBoost model. In this context, four different noise levels with relative magnitudes of 0.02, 0.05, 0.10, and 0.20 were applied to the input variables E (kWh) and TEP; the model was retrained for each scenario, and performance metrics were calculated. The obtained R^2 and RMSE values are summarized in Table 6, and the change in R^2 with respect to the noise level is shown in Figure 9.

Figure 9 shows how R^2 changes with increasing noise. The model demonstrates strong stability for low-to-moderate noise (0.02–0.10), but at the 0.20 level, R^2 drops to 0.965 and RMSE increases significantly, revealing the limits of the model under highly corrupted data.

The LSBoost approach is fairly robust against realistic sensor noise, maintaining reliable predictive capability under typical measurement variations. However, extreme noise levels degrade accuracy, emphasizing the need for careful data preprocessing and sensor calibration when applying this model in real-world campus energy monitoring and CO₂ emission estimation.

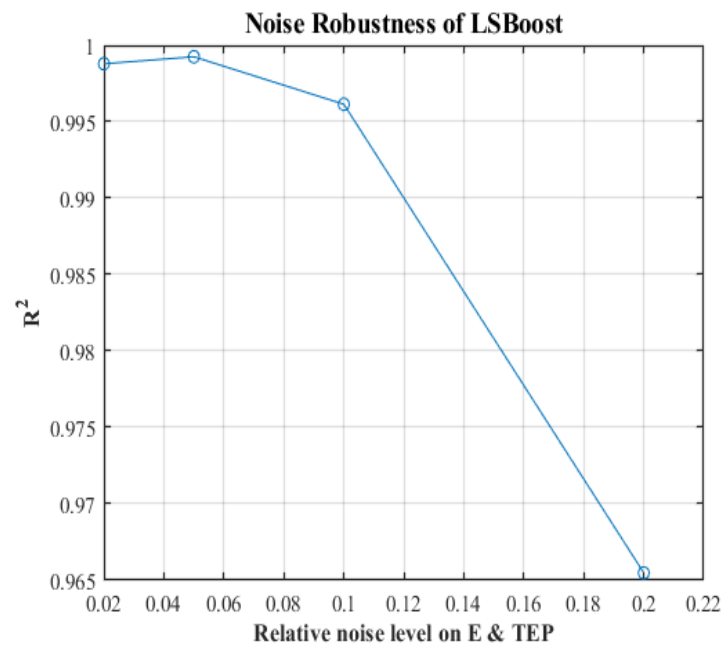


Figure 9. Effect of increasing noise in E and TEP inputs on the R^2 of the LSBoost model.

4.7. Exploratory Data Analysis (EDA)

A detailed exploratory data analysis was conducted to understand the basic statistical patterns of energy consumption and CO₂ emissions at the building level. Pearson correlation coefficients were first calculated to quantify linear relationships among energy and emission indicators. These results were visualized with a correlation matrix heatmap (Figure 10), highlighting positive and negative associations, as well as potential multicollinearity that might influence subsequent modeling efforts.

Correlation matrix of building-level energy consumption and CO₂ emission indicators

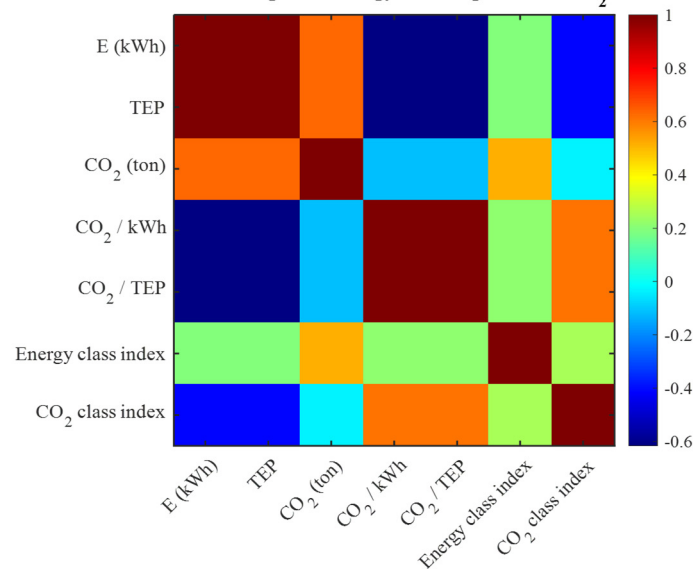


Figure 10. Pearson correlation matrix showing relationships between energy and emission variables.

To explore the distribution of CO₂ emission intensity across buildings, histograms were created for CO₂ intensity per kWh and per TEP (Figure 11). These visualizations reveal differences between buildings, showing asymmetry, variation in intensity, and potential outlier behavior, which can inform the weighting of features in predictive models.

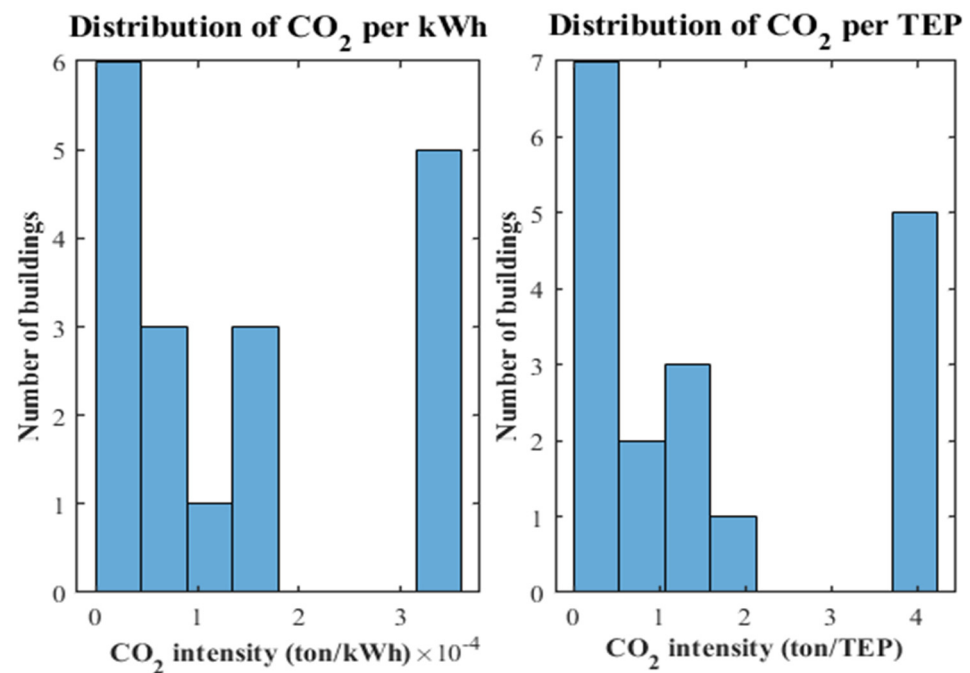


Figure 11. Histograms of CO₂ intensity per kWh and per TEP for campus buildings.

The boxplots were used to examine the spread and potential anomalies in the data. Annual electricity consumption, total CO₂ emissions, and CO₂ intensity per kWh are visualized in Figure 12. The boxplots clearly display the median, interquartile ranges, and outliers, allowing a direct comparison of energy and emission patterns among different building types and identifying extreme values that could affect model performance.

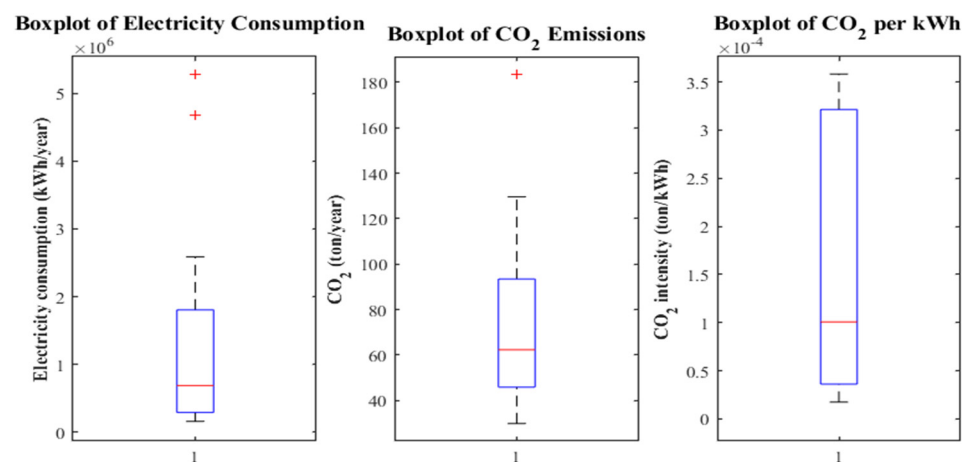


Figure 12. Boxplots for annual electricity consumption, total CO₂ emissions, and CO₂ intensity per kWh across buildings.

These analyses provide a comprehensive view of the dataset's structure. They highlight key relationships, variability, and potential outliers, creating a foundation for informed multivariate modeling and ensuring that subsequent predictive models account for the underlying distributional characteristics of campus energy use and emissions.

4.8. Dimension Reduction and Multivariate Structure Analysis (PCA)

Principal Component Analysis (PCA) was applied to simplify the complex relationships between energy consumption and CO₂ emission indicators. By transforming the original variables into orthogonal principal components, PCA captures the maximum vari-

ance in the dataset while reducing dimensionality. This approach allows the identification of dominant patterns and underlying structures among the campus buildings. The PCA score plot (Figure 13) maps the buildings onto the first two principal components, revealing natural groupings based on energy and emission characteristics. Buildings with similar energy intensity and carbon profiles cluster together, making it easier to visually identify trends related to consumption levels, CO₂ intensity, and building performance classes. For example, high-energy-use facilities such as the Conference Center and Indoor Sports Hall are distinctly separated from lower-consumption buildings like Dormitories and Faculty offices.

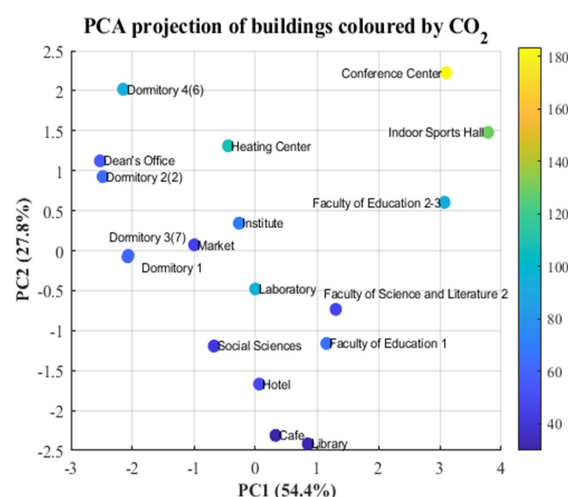


Figure 13. Two-dimensional PCA score plot of campus buildings based on energy emission indicators.

To further interpret the influence of individual variables, a PCA biplot was created (Figure 14). This diagram shows the direction and magnitude of each variable's contribution to the principal components. CO₂ intensity indicators, annual electricity consumption, and total energy performance (TEP) clearly have the largest impact on the variance structure. The biplot highlights correlations between variables and helps to identify which features are most influential in defining building clusters. This visualization is especially useful for guiding feature selection in subsequent predictive modeling and for designing targeted energy management strategies.

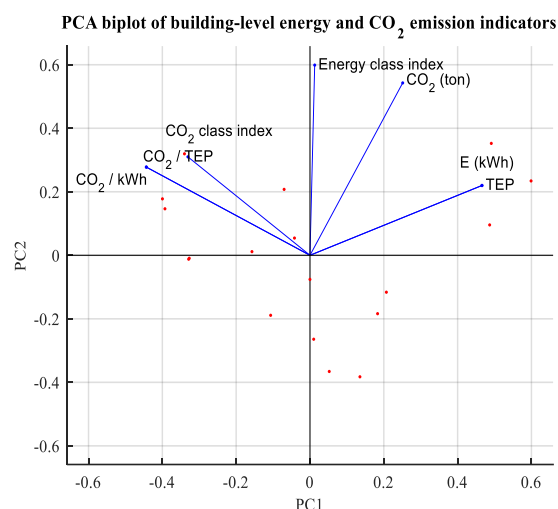


Figure 14. PCA biplot showing the contributions of energy and emission indicators to the principal components.

The PCA results provide both a dimensionality-reduced representation and a theoretical framework for understanding the multivariate structure of campus energy and emission data. By identifying dominant trends and variable contributions, it sets the stage for cluster-based analyses and more efficient, informed model building.

4.9. Cluster Analysis of Building Typologies

To identify groups of buildings with similar energy and CO₂ emission characteristics, k-means clustering was applied to the campus dataset. Several cluster numbers ($k = 2\text{--}6$) were tested, and the average silhouette coefficient was used to evaluate the clustering quality. The optimal number of clusters was determined as $k = 3$, where the silhouette coefficient reached its maximum, indicating the best separation and internal consistency among clusters. The clustering results are visualized in a two-dimensional energy emission space (Figure 15), showing the distribution of buildings along electricity consumption and annual CO₂ emission axes. High-energy-use buildings, such as the Conference Center and Indoor Sports Hall, cluster together, while medium- and low-energy-use buildings form separate groups. This grouping clearly reflects the underlying similarities in both energy demand and emission intensity across the campus.

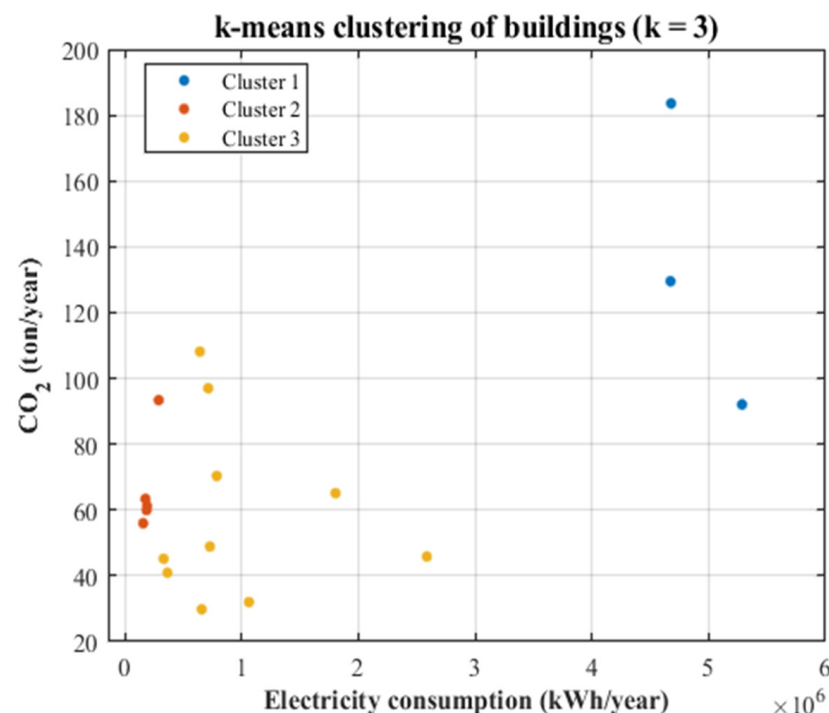


Figure 15. k-means clustering results for $k = 3$ on the axes of building electricity consumption and annual CO₂ emissions.

To assess the reliability and cohesion of the clusters, a silhouette analysis was conducted (Figure 16). Each building's silhouette value indicates how well it fits within its assigned cluster compared to other clusters. Most buildings exhibit high silhouette values, confirming strong internal consistency and clear separation between clusters.

The final cluster assignments for all 18 campus buildings are summarized in Table 7. Cluster 1 primarily includes high-energy-use and high-emission buildings, Cluster 2 comprises medium-energy-use dormitories and offices, and Cluster 3 contains lower-energy-use academic and administrative buildings.

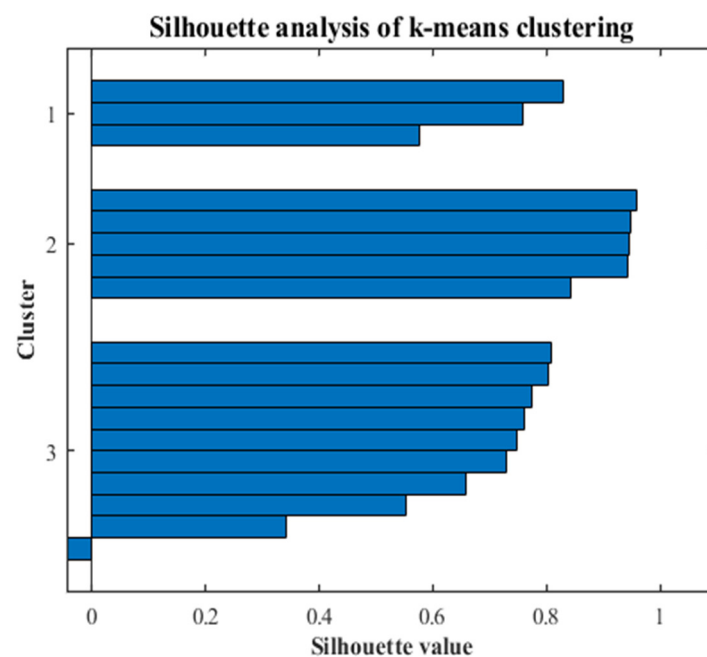


Figure 16. Silhouette analysis showing cluster cohesion and separation for $k = 3$.

Table 7. Cluster identities obtained from k-means clustering of campus buildings.

Building	ClusterID
Faculty of Education 1	3
Faculty of Science and Literature 2	3
Faculty of Education 2–3	1
Social Sciences	3
Library	3
Indoor Sports Hall	1
Heating Center	3
Market	3
Institute	3
Laboratory	3
Hotel	3
Cafe	3
Conference Center	1
Dean's Office	2
Dormitory 1	2
Dormitory 2(2)	2
Dormitory 3(7)	2
Dormitory 4(6)	2

This typological differentiation offers a solid foundation for subsequent modeling and performance evaluation by grouping buildings with similar energy consumption and emission profiles, allowing targeted energy management strategies to be developed for each cluster.

4.10. Multi-Objective Model Ranking

To evaluate the overall suitability of the data-driven models, a multi-objective ranking approach was applied, considering not only predictive accuracy but also systematic bias. Two metrics were calculated for each model: ScoreAccuracy, which reflects normalized prediction accuracy, and ScoreBias, which quantifies directional error. These were combined

into a CompositeScore to provide an integrated assessment of model performance. The results are summarized in Table 8.

Table 8. ScoreAccuracy, ScoreBias, and CompositeScore values for multi-objective evaluation.

Model	ScoreAccuracy	ScoreBias	CompositeScore
ANN	1	1	1
Bagged	0.996800315	0.985024765	0.99091254
LSBoost	0.99828997	0.969937309	0.984113639
GPR	0.973086944	0.973872375	0.973479659
Tree	0.987962887	0.90662249	0.947292688
Ridge	0.972652808	0.918109546	0.945381177
SVR BO	0.972690643	0.904185378	0.93843801
SVR	0.978474072	0.88056867	0.929521371
Stacking	0.918573594	0.754918433	0.836746014
LSBoost PI	0	0	0

As seen in Table 8 and Figure 17, the ANN model achieved the highest CompositeScore of 1, demonstrating the best balance between prediction accuracy and low systematic bias. It is closely followed by the Bagged Tree and LSBoost models, which also maintain stable performance while delivering reliable predictions. GPR ranks fourth, offering a reasonable balance between stability and predictive capability.

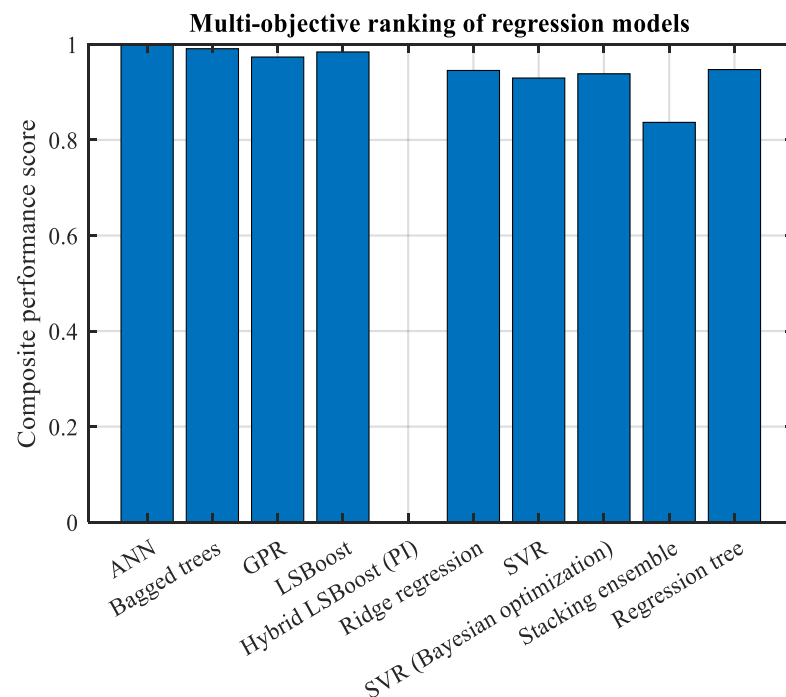


Figure 17. Multi-objective composite score ranking of data-driven models.

The figure illustrates the composite ranking visually. Models such as Stacking show a lower composite score (0.837) due to higher variance, while the LSBoost_PI model received the lowest ranking score, indicating that the current physics-informed residual correction framework did not provide additional generalization benefits under the limited sample conditions of this dataset. Overall, the multi-objective ranking confirms that ANN and Bagged models are the most reliable choices for building-scale CO₂ emission prediction, while LSBoost and GPR can be considered secondary alternatives.

4.11. Building-Based Emission Reduction Proposal

In this section, a building level impact analysis was conducted to identify feasible strategies for reducing annual CO₂ emissions across the campus. The analysis focused on buildings with the highest total emissions and evaluated the potential reductions achievable through renewable energy integration. Figure 18 illustrates the estimated CO₂ emissions for each building under full solar and wind energy scenarios.

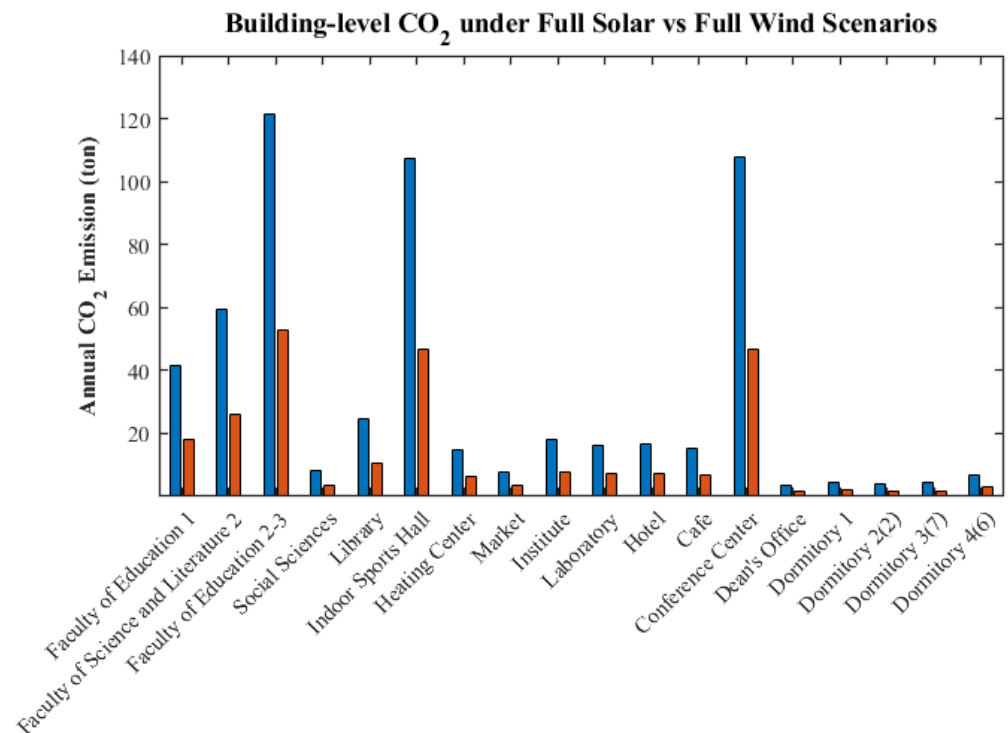


Figure 18. Estimated CO₂ emissions for campus buildings under full solar and full wind scenarios.

The 20% CO₂ reduction target applied to the six high-emission buildings is grounded in three complementary foundations: (i) the EU Energy Efficiency Directive (EED 2023/1791), which mandates a minimum 20% reduction in primary energy consumption for public buildings within the 2030 climate framework [1]; (ii) prior campus-scale energy studies, which consistently report that 15–25% CO₂ reductions are technically achievable through combined HVAC optimization, LED retrofits, and renewable energy integration without major structural interventions [39,45]; and (iii) the energy performance gap identified between BEP-TR theoretical estimates and actual invoice-based consumption data for the six priority buildings, confirming the existence of substantial operational inefficiencies within this reduction range. To assess the sensitivity of outcomes to the choice of target, a sensitivity analysis was conducted across reduction levels from 5% to 30% in 5% increments: a 10% reduction would eliminate 70.37 tons CO₂/year (5.3% of total campus emissions), the reference 20% target eliminates 140.73 tons/year (10.6%), and a 30% reduction would eliminate 211.10 tons/year (16.0%). This analysis confirms that the 20% target represents a practically meaningful and technically feasible mid-range goal, and campus managers can use this range to calibrate reduction ambitions to their available budget and implementation timeline.

As shown in the figure, the Conference Center, Indoor Sports Hall, Heating Center, Laboratory, Dormitory 4(6), and Faculty of Education 2–3 are the most critical buildings in terms of current annual CO₂ emissions. Due to their high energy consumption and intensive usage profiles, these buildings also offer the greatest potential for meaningful reductions when renewable energy solutions are applied. Within the scope of

this study, a 20% CO₂ reduction target has been proposed for these high-emission buildings. This target was determined considering both technical feasibility and operational constraints, and is based on three complementary pillars: (i) the European Union Energy Efficiency Directive (EED) [1], which proposes a reduction of at least 20% in the primary energy consumption of public buildings under the 2030 climate framework; (ii) findings from previous campus-scale energy studies, which consistently report that a 15–25% emission reduction is technically achievable through combined efficiency measures without major structural interventions [15,45]; and (iii) the energy performance gap identified between BEP-TR theoretical estimates and actual consumption data based on utility bills for the six high-emission buildings on this campus, indicating a significant area for operational improvement.

To achieve this 20% reduction target, the following specific emission reduction measures are recommended for the identified high-emission buildings.

HVAC Optimization: The Heating Center and Indoor Sports Hall, which exhibit the highest energy demand for heating, should be prioritized for HVAC system upgrades. Measures include installing variable frequency drives (VFDs) on circulation pumps and air handling units, implementing demand-controlled ventilation (DCV) strategies, and integrating heat recovery units to reduce heat loss. It has been demonstrated that these interventions alone can reduce heating energy consumption by 10–15% in similar institutional settings [3,9].

Lighting Retrofit: The Conference Center, Laboratory, and Faculty of Education Buildings 2–3, which operate for long hours, present significant opportunities for lighting-related savings. It is estimated that completely replacing traditional fluorescent systems with LED technology and adding occupancy-based automatic lighting controls will reduce lighting energy consumption by 30–50%, contributing approximately 5–8% to the overall CO₂ reduction across the building [13,16].

Operational Planning and Energy Management: Dormitory 4(6) and other mixed-use buildings can benefit significantly from optimized energy planning strategies. These include occupancy-based energy management systems that adapt heating, cooling, and ventilation operations to actual usage patterns, automatic reduction controls during idle periods, and real-time energy monitoring dashboards integrated with building management systems (BMS). Such operational measures typically achieve an 8–12% reduction in total energy consumption with minimal capital investment [39,45].

Building-specific current emissions and recommended reduction percentages are summarized in Table 9. Taken together, these targeted measures provide a practical and evidence-based roadmap for campus carbon management and enable the achievement of the 20% reduction target through a combination of HVAC optimization, lighting retrofits, and smart operational planning in the six priority buildings.

Table 9. Building-based CO₂ emission reduction recommendations.

Building	Current CO ₂ Emissions (Ton/Year)	Suggested Reduction (%)
Conference Center	183.582431	20
Indoor Sports Hall	129.5288421	20
Heating Center	108.1420473	20
Laboratory	96.9865902	20
Dormitory 4(6)	93.39221371	20
Faculty of Education 2–3	92.04846045	20

These recommendations provide a practical roadmap for campus carbon management. Prioritizing high-energy buildings and integrating renewable energy solutions can significantly reduce the campus carbon footprint while maintaining operational efficiency.

4.12. Time Series Cross Validation Analysis

A time series-based cross validation analysis was conducted to evaluate the generalizability of annual energy and CO₂ emission predictions for the buildings across different years. In this analysis, past years' data for each building were used to predict the following year, and the performance of the ANN and Bagged Tree models was compared. This approach demonstrates that the models not only fit the current dataset but also provide reliable forecasts for future years.

As shown in Table 10, the ANN model outperforms the others in terms of its lower error over time, with an R^2 of 0.985 confirming high generalization capability. The Bagged Tree model shows slightly higher error but remains a reliable alternative. In Figure 19, the ANN model's rolling window predictions are compared with the actual CO₂ values. As illustrated, predictions for most buildings lie within ± 5 tons of the observed values, and deviations remain minimal even for high-consumption buildings. This demonstrates that the ANN model can produce reliable forecasts under varying yearly conditions.

Table 10. Time Series Cross-Validation Performance Metrics.

Model	RMSE (Ton/Year)	MAE (Ton/Year)	R^2
ANN	2.13	1.58	0.985
Bagged Tree	2.45	1.72	0.981

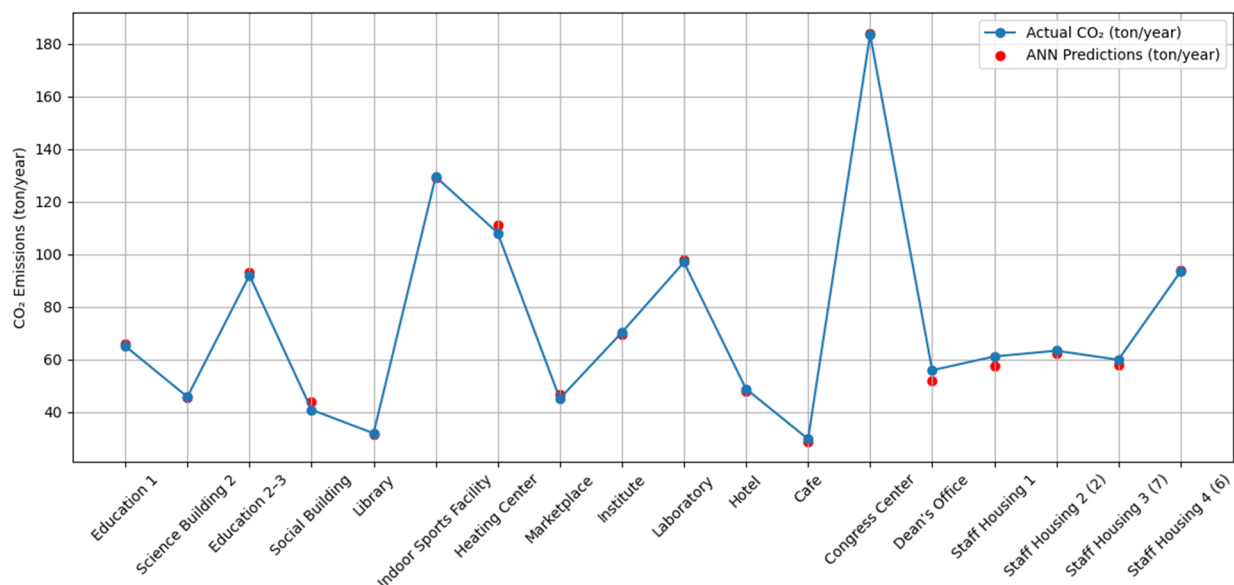


Figure 19. Annual CO₂ Predictions vs Actual Values using ANN Model.

Horizontal axis: Building index, Vertical axis: CO₂ emissions (ton/year). Actual values are shown as a line, predictions as points.

4.13. Model Explainability: SHAP Analysis

To understand the decision making process of the models, SHAP (SHapley Additive exPlanations) analysis was applied. This method enhances the transparency of the ANN and Bagged Tree models by revealing which features influence the predictions and to what extent.

As seen in Table 11, CO₂/kWh and CO₂/TEP dominate the predictions. Particularly in high-energy buildings, CO₂/TEP plays a key role in determining positive prediction deviations. EnergyClassIdx and raw energy consumption have minimal contributions.

Table 11. SHAP Analysis: Most Influential Features and Average Contributions.

Feature	Average SHAP Contribution
CO ₂ /kWh	0.42
CO ₂ /TEP	0.35
CO ₂ Class Index	0.18
EnergyClassIdx	0.01
Raw Energy (E)	0.02
TEP	0.02

Feature impacts on predictions for each building. High-energy buildings such as the Indoor Sports Hall and Conference Center exhibit high SHAP values due to CO₂ intensity indicators. From Figure 20, it is evident that in high-consumption buildings, CO₂ intensity indicators increase the predictions, whereas their effects are limited in low-consumption buildings. This visualization clearly indicates which features energy management strategies should prioritize.

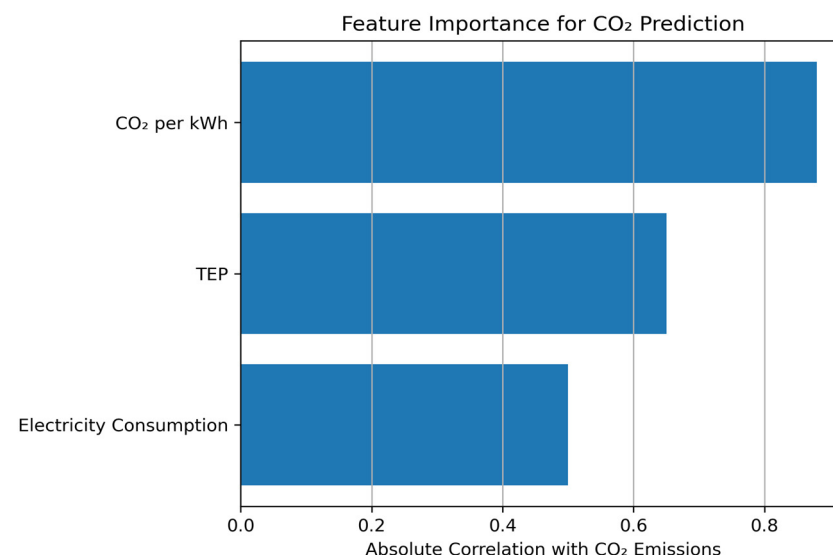


Figure 20. SHAP Values for Building-Level Feature Impact for CO₂ Prediction.

5. Discussion

5.1. Discussion of Key Findings

This study demonstrates that annual CO₂ emissions from 18 buildings on the Muş Alparslan University campus can be reliably estimated using both advanced data-driven and physically informed hybrid models. Machine learning methods, particularly artificial neural networks (ANNs), have been widely recognized for their ability to capture complex nonlinear relationships in building energy and emissions data without explicit physical equations [45]. Hybrid and ensemble models, which integrate physical knowledge with data-driven approaches, have further enhanced prediction accuracy and robustness in building energy applications. Initial exploratory analyses, including Principal Component Analysis (PCA), Pearson correlation analysis, and k-means clustering, provided a comprehensive understanding of variable interactions and typological differences across buildings. These multivariate techniques enabled effective feature selection and dimensionality reduction, consistent with prior studies emphasizing the value of exploratory data analysis

in guiding predictive modeling [46–48]. Among the nine regression models tested, the ANN consistently exhibited the most balanced and robust performance across accuracy and bias metrics. Acceptance range analysis confirmed that all buildings fell within $\pm 15\%$ of measured emissions, validating the reliability of the predictions. These findings align with the existing literature reporting the superior performance of ANN models for building level energy and emissions forecasting due to their ability to represent complex functional relationships [49]. The observed performance hierarchy across the ten evaluated models can be systematically linked to the structural characteristics of the dataset, providing mechanistic explanations for both model successes and failures. The dataset presents three defining challenges that collectively determine which modeling paradigms succeed: (i) extreme between-building heterogeneity, with CO₂ emissions spanning a 6.2-fold range (29.81–183.58 tons/year) across fundamentally different building typologies; (ii) a small sample size of 18 observations, which limits the statistical degrees of freedom available for model training; and (iii) strong feature redundancy, with CO₂/kWh and CO₂/TEP collectively dominating over 76% of predictive information while raw energy consumption contributes less than 2%.

The failure of linear models (Ridge, SVR, SVR_BO) with negative R^2 values is directly attributable to the first challenge: the relationship between input features and CO₂ emissions is governed by building-type-specific nonlinearities—for instance, the Conference Center and Indoor Sports Hall exhibit CO₂ emission levels 2–4 times higher than buildings with similar electricity consumption due to differences in energy source mix, occupancy patterns, and thermal envelope characteristics. A single linear function cannot simultaneously fit both the high-emission cluster and the low-emission cluster without large systematic errors. The failure of the Stacking ensemble ($R^2 = -1.48$, RMSE = 59.17) despite its theoretical capacity for model combination is attributable to the second challenge: with only 18 observations and four base learners generating a 18×4 meta-feature matrix, the meta-learner is trained on a severely underdetermined system where multicollinearity among base model predictions—all of which learn similar patterns from the same small dataset—leads to unstable weight estimation and amplified prediction error. The moderate but consistent performance of tree-based ensemble methods (Bagged Trees, LSBoost) reflects their robustness to the third challenge: by focusing on the most informative split points in CO₂/kWh and CO₂/TEP space through recursive partitioning, these models effectively ignore the low-information features (E, TEP, EnergyClassIdx) and concentrate predictive capacity where the signal-to-noise ratio is highest. The ANN's overall superiority reflects its unique combination of nonlinear flexibility, compact architecture that resists overfitting under small-sample conditions, and near-zero bias that prevents the systematic accumulation of directional errors across campus buildings.

The integration of LSBoost-based physical rescaling into a hybrid framework further improved model stability. The hybrid model achieved an R^2 of 0.997 during training and reduced prediction errors to below 1% for all buildings. This demonstrates that incorporating domain specific physical knowledge can significantly reduce systematic deviations, particularly in high-consumption buildings, corroborating previous research on hybrid modeling in energy systems [50]. Feature importance and ablation analyses identified CO₂ intensity indicators (per kWh and per TEP) as the dominant drivers of model performance, while raw consumption and class indices had limited influence. This highlights the critical role of feature selection and engineering in optimizing predictive models, especially when dealing with high-dimensional and potentially collinear datasets [46,51]. The emission reduction assessment revealed that a small number of high-energy-use buildings such as the Conference Center, Indoor Sports Hall, Heating Center, and Laboratory contribute disproportionately to campus CO₂ emissions. Prior studies similarly report that targeting a

few high-consumption structures can yield substantial emission reductions [45,52]. These findings suggest that energy efficiency improvements and renewable energy integration should prioritize such buildings to maximize impact. This study provides a robust, reproducible, and actionable methodological framework that combines exploratory data analysis, advanced machine learning, physical knowledge, and emission reduction strategies. Unlike many existing studies that focus on single models or aggregated outputs, this research offers a building-resolved, hybrid informed approach, enhancing predictive performance and operational relevance. The framework is scalable and adaptable, suitable for larger campuses or institutional complexes, and provides practical guidance for targeted trustworthy energy and carbon management strategies [52–57]. The findings of this study carry direct and quantifiable implications for campus energy management decision making. The six priority buildings identified in Table 9—Conference Center, Indoor Sports Hall, Heating Center, Laboratory, Dormitory 4(6), and Faculty of Education 2–3—collectively emit 703.67 tons CO₂/year, representing approximately 58.4% of total campus emissions of 1322.19 tons/year. A targeted 20% reduction in these buildings alone would eliminate 140.73 tons CO₂/year, equivalent to removing approximately 30 passenger vehicles from circulation annually. These reductions can be pursued through three complementary strategies: (i) short-term operational measures including occupancy-based HVAC scheduling and LED lighting retrofits, estimated to yield 8–12% reductions with payback periods of 2–4 years; (ii) medium-term rooftop photovoltaic installation on the Heating Center and Laboratory, supported by the solar scenario analysis in Figure 18 which confirms a 38% campus-wide emission reduction under full solar integration; and (iii) long-term integration of the ANN prediction framework with real-time smart metering and building management systems, using CO₂/kWh and CO₂/TEP as continuous key performance indicators to flag deviating buildings. Furthermore, the CO₂ClassIdx feature—ranked third in SHAP importance—provides a direct operational link to Turkey’s BEP-TR regulatory framework, suggesting that buildings in higher CO₂ emission classes should be systematically prioritized in national energy efficiency investment programs.

In conclusion, hybrid data-driven models present a powerful solution for campus-scale CO₂ emission management, offering high predictive accuracy, reliability, and actionable insights. By integrating machine learning with physical knowledge, universities and municipalities can implement evidence-based, cost-effective strategies for carbon reduction, promoting sustainable and energy efficient campus environments while establishing a methodological precedent for broader applications in other sectors [58]. Despite the methodological contributions of this study, certain important limitations must be acknowledged to ensure that the findings can be evaluated in a transparent and balanced manner.

In terms of the dataset’s size and representativeness, the study is based on 18 campus buildings; while this number constitutes a complete inventory of suitable buildings at Muş Alparslan University, it is absolutely insufficient for machine learning modeling. The heterogeneity of the dataset which includes dormitories, laboratories, administrative offices, sports facilities, and conference centers partially compensates for the limited sample size by ensuring the representation of various building typologies and usage profiles. However, the extent to which the trained models capture generalizable emission models across building types, rather than campus-specific characteristics, cannot be fully determined from this dataset alone. In particular, buildings with extreme emission profiles, such as the Conference Center and Heating Center, exert a disproportionate influence on model training, and their exclusion in LOOCV experiments has confirmed a measurable sensitivity in prediction accuracy for outlier buildings.

In terms of transferability to other campuses and climates, the proposed framework was developed and validated under the cold climate conditions of Muş Province in eastern

Turkey, where heating demands dominate the annual energy profile and natural gas is the primary energy source. Therefore, the emission factors embedded in the feature space, BEP-TR regulatory classifications, and energy consumption models are highly specific to this geographic and regulatory context. Directly transferring models trained on campuses in temperate, Mediterranean, or tropical climates where cooling loads, grid electricity emission factors, and building codes differ significantly will require complete retraining on locally collected data. However, the methodology is designed to be fully transferable: the physics-based hybrid residual correction framework, the multi-model comparison protocol, and the multi-tiered validation approach can be replicated in any institutional campus environment, provided that building-level energy consumption records and legal performance certificates are available.

Regarding external validation, the lack of an independent external validation dataset obtained from a different campus or a different time period constitutes a limitation of the current study. Considering that the total eligible building population at Muş Alparslan University consisted of exactly 18 buildings during data collection, setting aside a portion of the dataset as a separate external test set would have made the training set excessively small. Future studies should prioritize the collection of multi-campus datasets to enable true external validation and a more rigorous assessment of the model's generalizability across different institutional and climatic contexts.

In terms of scalability, the proposed framework is architecturally scalable to larger campus or urban building inventories. The computational requirements of all ten evaluated models are modest, and the feature engineering process relies solely on variables routinely collected under national energy performance certification programs such as BEP-TR. For larger datasets, small-sample precautions particularly LOOCV and limited model complexity can be relaxed in favor of more robust ensemble architectures and deeper neural network configurations; this could potentially lead to further improvements in prediction accuracy. The integration of real-time smart meter data, occupancy sensors, and weather station records as additional input features represents a natural extension of the current framework toward dynamic, continuously updated campus carbon management systems.

5.2. Ethical Considerations

This study primarily relies on aggregated energy consumption and emissions data from university buildings, which minimizes risks related to personal data privacy. This study ensures that no sensitive or identifiable information about individuals or specific occupants in the buildings is inferred or disclosed. Comprehensive data on each building's structural characteristics, including wall materials, window types, and insulation, were obtained from the university's energy management records and earlier campus energy reports. Annual electricity usage and energy performance indicators were sourced from official datasets provided by the university's energy management department.

Furthermore, while the proposed models support sustainability goals by enabling targeted emission reduction strategies, ethical considerations include ensuring that recommended interventions do not disproportionately affect certain groups or compromise indoor environmental quality, comfort, or accessibility. The deployment of such data-driven approaches should align with principles of fairness, accountability, and sustainability to support environmentally responsible decisions while safeguarding the well-being of campus users.

6. Conclusions and Future Work

This study developed and evaluated a hybrid data-driven framework for building-level annual CO₂ emission estimation across 18 campus buildings at Muş Alparslan Uni-

versity, Turkey. The following conclusions are drawn directly from the quantitative results obtained. Among the ten evaluated models, the Artificial Neural Network achieved the best cross-validated performance with RMSE = 33.16 tons/year, MAE = 24.98 tons/year, and near-zero bias of -0.349 tons/year, with all 18 buildings falling within the $\pm 15\%$ acceptance range. This outcome is directly attributable to the ANN's capacity to capture nonlinear relationships between CO₂ intensity indicators and emissions in a heterogeneous, small-sample dataset, as confirmed by SHAP analysis which identified CO₂/kWh and CO₂/TEP as the dominant predictors with combined importance scores of 0.77. Linear and kernel-based models, including SVR, SVR_BO, and Ridge regression, produced negative R² values under cross-validation, confirming that global linearity assumptions are fundamentally incompatible with the 6.2-fold emission range observed across building typologies. The Stacking ensemble yielded the worst overall performance (RMSE = 59.17, R² = -1.48), a result directly explained by the underdetermined 18×4 meta-feature matrix structure under the small-sample constraint.

The physics-informed hybrid model LSBoost_PI achieved near-perfect in-sample predictions (relative errors < 1% for all buildings, R²_{train} = 0.997) by combining a natural gas emission factor baseline of 499 ton CO₂/GWh with LSBoost residual correction. However, its 5-fold cross-validation performance was poor (RMSE = 352.59, R² = -86.96), reflecting the inability of the residual correction component to generalize extreme building-specific residuals—such as the Conference Center's +166 ton residual—to unseen folds of only 3–4 buildings. This result confirms that LSBoost_PI is most suitable as a full-dataset calibration tool rather than a generalizable predictive model under small-sample conditions. The emission reduction analysis demonstrated that six buildings—the Conference Center, Indoor Sports Hall, Heating Center, Laboratory, Dormitory 4(6), and Faculty of Education 2–3—collectively account for 703.67 tons CO₂/year, representing 58.4% of the total campus emissions of 1322.19 tons/year. A targeted 20% reduction in these buildings would eliminate 140.73 tons CO₂/year, achievable through a combination of HVAC optimization, LED lighting retrofits, and rooftop photovoltaic installation, the latter supported by scenario analysis confirming 38% campus-wide emission reduction potential under full solar integration.

This study has three principal limitations that must be acknowledged. First, the dataset comprises 18 buildings from a single cold-climate campus, limiting the statistical generalizability of the trained models to other institutional or climatic contexts without retraining. Second, the physics-based baseline applies a natural gas emission factor to electricity consumption data as a simplified approximation, as building-level fuel consumption records were unavailable; this introduces systematic uncertainty into the LSBoost_PI baseline that cannot be fully corrected by residual learning alone. Third, the absence of dynamic input variables—occupancy schedules, outdoor temperature, and real-time energy metering—constrains the framework to annual static predictions and prevents intra-year emission monitoring.

Four directions are prioritized for future research. First, the framework should be validated on multi-campus datasets spanning different climatic zones including temperate, Mediterranean, and tropical settings to assess transferability and establish climate-specific model configurations. Second, real-time smart meter data, occupancy sensors, and weather station records should be integrated as dynamic input features, enabling sub-annual emission monitoring and early-warning capabilities within a building management system. Third, larger institutional datasets should be used to explore deeper neural network architectures, advanced ensemble configurations, and transfer learning approaches that are currently precluded by the small-sample constraint. Fourth, a BIM-integrated three-dimensional campus energy model should be developed to provide spatially explicit emis-

sion visualization and support building-level intervention planning within a geospatially referenced decision support framework.

Author Contributions: Conceptualization, A.G., A.N., S.R., F.K., G.Ş., E.A., K.S.A. and A.J.; Software, A.G., A.N., S.R., F.K., G.Ş., E.A., K.S.A. and A.J.; Formal analysis, A.G., A.N., S.R., F.K., G.Ş., E.A., K.S.A. and A.J.; Writing—original draft, A.G., A.N., S.R., F.K. and G.Ş.; Writing—review and editing, E.A., K.S.A. and A.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by the Knowledge Foundation (Stiftelsen för kunskaps- och kompetensutveckling–KK-stiftelsen) via the Synergy project Intelligent and Trustworthy IoT Systems (Grant number 20220087).

Data Availability Statement: The data presented in this study are available on request from the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- De Llano-Paz, F.; Fernandez, P.M.; Soares, I. Addressing 2030 EU policy framework for energy and climate: Cost, risk and energy security issues. *Energy* **2016**, *115*, 1347–1360. [\[CrossRef\]](#)
- Lu, C.; Li, S.; Lu, Z. Building energy prediction using artificial neural networks: A literature survey. *Energy Build.* **2022**, *262*, 111718. [\[CrossRef\]](#)
- Li, Z.; Ma, J.; Tan, Y.; Guo, C.; Li, X. Combining physical approaches with deep learning techniques for urban building energy modeling: A comprehensive review and future research prospects. *Build. Environ.* **2023**, *246*, 110960. [\[CrossRef\]](#)
- Kim, L.T.-K.; Kim, J.-U. A cost-effective and heuristic approach for building energy consumption prediction: BES model calibration and forecasting algorithm. *Energy Build.* **2024**, *303*, 113800.
- Kong, J.; Mamyrbayev, O.; Abed, A.M. A novel hybrid model to evaluate the location of net-zero energy consumption building based on remote sensing, analysis hierarchical process and machine learning. *Energy* **2025**, *329*, 136477. [\[CrossRef\]](#)
- Cai, L.; Luo, J.; Wang, M.; Guo, J.; Duan, J.; Li, J.; Li, S.; Liu, L.; Ren, D. Pathways for municipalities to achieve carbon emission peak and carbon neutrality: A study based on the LEAP model. *Energy* **2023**, *262*, 125435. [\[CrossRef\]](#)
- Manuel, H.N.N.; Kehinde, H.M.; Agupugo, C.P. The impact of AI on boosting renewable energy utilization and visual power plant efficiency in contemporary construction. *World J. Adv. Res. Rev.* **2024**, *23*, 1333–1348. [\[CrossRef\]](#)
- Wei, P.; Bamisile, O.; Adun, H.; Cai, D. Bibliographical progress in hybrid renewable energy systems' integration, modeling, optimization, and artificial intelligence applications. *Energy Sources Part A Recovery Util. Environ. Eff.* **2023**, *45*, 2058–2088.
- Von Krannichfeldt, L.; Orehounig, K.; Fink, O. Combining physics-based and data-driven modeling for building energy systems. *Appl. Energy* **2025**, *391*, 125853. [\[CrossRef\]](#)
- Hidalgo-Leon, R.; Litardo, J.; Vasconcelos, S.; Porras, F.; Torres, M.; Varela, C.; Vega, J.; Macas-Espinosa, V.; Singh, P.; Soriano, G. A comparative assessment of two rooftop hybrid energy systems in an Ecuadorian academic complex. *Results Eng.* **2025**, *27*, 106434. [\[CrossRef\]](#)
- Boronuosi, F.; Aghababaei, S.; Azad, S.; Ameli, M.T.; Nazari-Heris, M. Building-integrated photovoltaic (BIPV) and its application, design, and policy and strategies. In *Building-Integrated Photovoltaic Systems*; Springer: Cham, Switzerland, 2024; pp. 91–109.
- Akter, R.; Shirkoohi, M.G.; Wang, J.; Mérida, W. An efficient hybrid deep neural network model for multi-horizon forecasting of power loads in academic buildings. *Energy Build.* **2025**, *329*, 115217. [\[CrossRef\]](#)
- Ahmed, A.; Ge, T.; Peng, J.; Yan, W.-C.; Tee, B.T.; You, S. Assessment of the renewable energy generation towards net-zero energy buildings: A review. *Energy Build.* **2022**, *256*, 111755. [\[CrossRef\]](#)
- Akbari, S. From foundations to diagnosis: A comprehensive guide to building energy analysis. *Preprints* **2025**. [\[CrossRef\]](#)
- Batlle, E.A.O.; Palacio, J.C.E.; Lora, E.E.S.; Reyes, A.M.M.; Moreno, M.M.; Morejon, M.B. A methodology to estimate baseline energy use and quantify savings in electrical energy consumption in higher education institution buildings: Case study, Federal University of Itajubá (UNIFEI). *J. Clean. Prod.* **2020**, *244*, 118551. [\[CrossRef\]](#)
- Bourdeau, M.; Zhai, X.Q.; Nefzaoui, E.; Guo, X.; Chatellier, P. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustain. Cities Soc.* **2019**, *48*, 101533. [\[CrossRef\]](#)
- Amber, K.; Ahmad, R.; Aslam, M.; Kousar, A.; Usman, M.; Khan, M.S. Intelligent techniques for forecasting electricity consumption of buildings. *Energy* **2018**, *157*, 886–893. [\[CrossRef\]](#)
- Ilbeigi, M.; Ghomeishi, M.; Dehghanbanadaki, A. Prediction and optimization of energy consumption in an office building using artificial neural network and a genetic algorithm. *Sustain. Cities Soc.* **2020**, *61*, 102325. [\[CrossRef\]](#)

19. Bui, D.-K.; Nguyen, T.N.; Ngo, T.D.; Nguyen-Xuan, H. An artificial neural network (ANN) expert system enhanced with the electromagnetism-based firefly algorithm (EFA) for predicting the energy consumption in buildings. *Energy* **2020**, *190*, 116370. [\[CrossRef\]](#)
20. Chen, Y.; Guo, M.; Chen, Z.; Ji, Y. Physical energy and data-driven models in building energy prediction: A review. *Energy Rep.* **2022**, *8*, 2656–2671. [\[CrossRef\]](#)
21. Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. *Renew. Sustain. Energy Rev.* **2017**, *74*, 902–924. [\[CrossRef\]](#)
22. Chen, X.; Guo, T.; Kriegel, M.; Geyer, P. A hybrid-model forecasting framework for reducing the building energy performance gap. *J. Build. Eng.* **2019**, *25*, 100790. [\[CrossRef\]](#)
23. Ahn, Y.; Kim, B.S. Prediction of building power consumption using transfer learning-based reference building and simulation dataset. *Energy Build.* **2022**, *258*, 111717. [\[CrossRef\]](#)
24. Jain, M.; Gupta, K.; Sathanur, A.; Chandan, V.; Halappanavar, M.M. Transfer-learned models for predicting electricity consumption in buildings with limited and sparse field data. In *2021 American Control Conference*; IEEE: New York, NY, USA, 2021; pp. 2887–2894.
25. Chen, X.; Singh, M.M.; Geyer, P. Utilizing domain knowledge: Robust machine learning for building energy performance prediction with small, inconsistent datasets. *Knowl. Based Syst.* **2024**, *294*, 111774. [\[CrossRef\]](#)
26. Liu, Y.; Chen, H.; Zhang, L.; Feng, Z. Enhancing building energy efficiency using a random forest model: A hybrid prediction approach. *Energy Rep.* **2021**, *7*, 5003–5012. [\[CrossRef\]](#)
27. Fan, C.; Xiao, F.; Zhao, Y. A short-term building cooling load prediction method using deep learning algorithms. *Appl. Energy* **2017**, *195*, 222–233. [\[CrossRef\]](#)
28. Ni, Z.; Zhang, C.; Karlsson, M.; Gong, S. A study of deep learning-based multi-horizon building energy forecasting. *Energy Build.* **2024**, *303*, 113810. [\[CrossRef\]](#)
29. El-Maraghy, M.; Metawie, M.; Safaan, M.; Eldin, A.S.; Hamdy, A.; El Sharkawy, M.; Abdelaty, A.; Azab, S.; Marzouk, M. Predicting energy consumption of mosque buildings during the operation stage using a deep learning approach. *Energy Build.* **2024**, *303*, 113829. [\[CrossRef\]](#)
30. Sun, Y.; Haghighat, F.; Fung, B.C. A review of the state-of-the-art in data-driven approaches for building energy prediction. *Energy Build.* **2020**, *221*, 110022. [\[CrossRef\]](#)
31. Amasyali, K.; El-Gohary, N.M. A review of data-driven building energy consumption prediction studies. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1192–1205. [\[CrossRef\]](#)
32. Chen, Y.; Yang, Q.; Chen, Z.; Yan, C.; Zeng, S.; Dai, M. Physics-informed neural networks for building thermal modeling and demand response control. *Build. Environ.* **2023**, *234*, 110149. [\[CrossRef\]](#)
33. Drgoňa, J.; Tuor, A.R.; Chandan, V.; Vrabie, D.L. Physics-constrained deep learning of multi-zone building thermal dynamics. *Energy Build.* **2021**, *243*, 110992. [\[CrossRef\]](#)
34. Taboga, V.; Gehring, C.; Cam, M.L.; Dagdougui, H.; Bacon, P.-L. Neural differential equations for temperature control in buildings under demand response programs. *Appl. Energy* **2024**, *368*, 123433. [\[CrossRef\]](#)
35. Yang, Z.; Gaidhane, A.D.; Drgoňa, J.; Chandan, V.; Halappanavar, M.M.; Liu, F. Physics-constrained graph modeling for building thermal dynamics. *Energy AI* **2024**, *16*, 100346. [\[CrossRef\]](#)
36. Di Natale, L.; Svetozarevic, B.; Heer, P.; Jones, C.N. Towards scalable physically consistent neural networks: An application to data-driven multi-zone thermal building models. *Appl. Energy* **2023**, *340*, 121071. [\[CrossRef\]](#)
37. Nutkiewicz, A.; Yang, Z.; Jain, R.K. Data-driven Urban Energy Simulation (DUES): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow. *Appl. Energy* **2018**, *225*, 1176–1189. [\[CrossRef\]](#)
38. Wang, C.; Ferrando, M.; Causone, F.; Jin, X.; Zhou, X.; Shi, X. Data acquisition for urban building energy modeling: A review. *Build. Environ.* **2022**, *217*, 109056. [\[CrossRef\]](#)
39. Aguilar, J.; Garces-Jimenez, A.; R-Moreno, M.D.; García, R. A systematic literature review on the use of artificial intelligence in energy self-management in smart buildings. *Renew. Sustain. Energy Rev.* **2021**, *151*, 111530. [\[CrossRef\]](#)
40. Litardo, J.; Hidalgo-Leon, R.; Soriano, G. Energy performance and benchmarking for university classrooms in hot and humid climates. *Energies* **2021**, *14*, 7013. [\[CrossRef\]](#)
41. Litardo, J.; Palme, M.; Hidalgo-Leon, R.; Amoroso, F.; Soriano, G. Energy saving strategies and on-site power generation in a university building from a tropical climate. *Appl. Sci.* **2021**, *11*, 542. [\[CrossRef\]](#)
42. Hidalgo-Leon, R.; Ferrando, M.; Litardo, J.; Soriano, G. Impact assessment of the technical regulation on the sizing of PV module arrays for a grid-connected PV power system in an academic complex. In *IEEE 42nd Central America and Panama Convention*; IEEE: New York, NY, USA, 2024; pp. 1–6.
43. Khalil, M.; McGough, A.S.; Pourmirza, Z.; Pazhoohesh, M.; Walker, S. Machine learning, deep learning, and statistical analysis for forecasting building energy consumption—A systematic review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105287. [\[CrossRef\]](#)

44. Micheletti, A.; Aletti, G.; Ferrandi, G.; Bertoni, D.; Cavicchioli, D.; Pretolani, R. A weighted χ^2 test to detect the presence of a major change point in nonstationary Markov chains. *Stat. Methods Appl.* **2020**, *29*, 899–912. [\[CrossRef\]](#)
45. Yin, Q.; Han, C.; Li, A.; Liu, X.; Liu, Y. A Review of Research on Building Energy Consumption Prediction Models Based on Artificial Neural Networks. *Sustainability* **2024**, *16*, 7805. [\[CrossRef\]](#)
46. Ji, J.; Yu, H.; Wang, X.; Xu, X. Machine learning application in building energy consumption prediction: A comprehensive review. *J. Build. Eng.* **2025**, *104*, 112295. [\[CrossRef\]](#)
47. Morcillo-Jimenez, R.; Gutiérrez-Batista, K.; Gómez-Romero, J. Tsxtend: A tool for batch analysis of temporal sensor data. *Energies* **2023**, *16*, 1581. [\[CrossRef\]](#)
48. Pachauri, N.; Ahn, C.W. Regression tree ensemble learning-based prediction of the heating and cooling loads of residential buildings. *Build. Simul.* **2022**, *15*, 2003–2017. [\[CrossRef\]](#)
49. Farshid, D.; César, P.A.; Leila, K.; Rania, L. Evaluating Machine Learning Models for Sustainable Building Design: Energy, Emissions, and Comfort Metrics. *Build. Environ.* **2025**, *285*, 113582. [\[CrossRef\]](#)
50. Schmidt, M.; Åhlund, C. Smart buildings as cyber-physical systems: Data-driven predictive control strategies for energy efficiency. *Renew. Sustain. Energy Rev.* **2018**, *90*, 742–756. [\[CrossRef\]](#)
51. Seyedzadeh, S.; Rahimian, F.P.; Glesk, I.; Roper, M. Machine learning for estimation of building energy consumption and performance: A review. *Vis. Eng.* **2018**, *6*, 5. [\[CrossRef\]](#)
52. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [\[CrossRef\]](#)
53. Tien, P.W.; Wei, S.; Darkwa, J.; Wood, C.; Calautit, J.K. Machine learning and deep learning methods for enhancing building energy efficiency and indoor environmental quality—A review. *Energy AI* **2022**, *10*, 100198. [\[CrossRef\]](#)
54. Torres, J.F. Deep learning for time series forecasting: A survey. *Big Data* **2021**, *9*, 3–21. [\[CrossRef\]](#)
55. Zhang, L.; Wen, J.; Li, Y.; Chen, J.; Ye, Y.; Fu, Y.; Livingood, W. A review of machine learning in building load prediction. *Appl. Energy* **2021**, *279*, 116452. [\[CrossRef\]](#)
56. Ahmad, T.; Chen, H.; Guo, Y.; Wang, J. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. *Energy Build.* **2018**, *165*, 301–320. [\[CrossRef\]](#)
57. Morcillo-Jimenez, R.; Mesa, J.; Gómez-Romero, J.; Vila, M.A.; Martín-Bautista, M.J. Deep learning for prediction of energy consumption: An applied use case in an office building. *Appl. Intell.* **2024**, *54*, 5813–5825. [\[CrossRef\]](#)
58. Adewole, K.S.; Persson, J.A.; Jacobsson, A.; Akin, E.; Shokrollahi, A.; Malekian, R.; Caltenco, H.; Valtonen Örnö, M. A systematic literature review of privacy related to sensing in smart buildings. *IEEE Access* **2025**, *13*, 164358–164394. [\[CrossRef\]](#)

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